PUBLICATIONS OF THE ASTRONOMICAL SOCIETY OF THE PACIFIC, **114**:1051–1069, 2002 October
© 2002. The Astronomical Society of the Pacific. All rights reserved. Printed in U.S.A.

Review

Deconvolution in Astronomy: A Review

J. L. STARCK AND E. PANTIN

Service d'Astrophysique, SAP/SEDI, CEA-Saclay, F-91191 Gif-sur-Yvette Cedex, France; jstarck@cea.fr, epantin@cea.fr

AND

F. MURTAGH

School of Computer Science, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland; and Observatoire de Strasbourg, Université Louis Pasteur, F-67000 Strasbourg, France; f.murtagh@qub.ac.uk

Received 2002 March 21; accepted 2002 June 28

ABSTRACT. This article reviews different deconvolution methods. The all-pervasive presence of noise is what makes deconvolution particularly difficult. The diversity of resulting algorithms reflects different ways of estimating the true signal under various idealizations of its properties. Different ways of approaching signal recovery are based on different instrumental noise models, whether the astronomical objects are pointlike or extended, and indeed on the computational resources available to the analyst. We present a number of recent results in this survey of signal restoration, including in the areas of superresolution and dithering. In particular, we show that most recent published work has consisted of incorporating some form of multiresolution in the deconvolution process.

1. INTRODUCTION

Deconvolution is a key area in signal and image processing. It is used for objectives in signal and image processing that include the following:

1. deblurring,
2. removal of atmospheric seeing degradation,
3. correction of mirror spherical aberration,
4. image sharpening,
5. mapping detector response characteristics to those of another,
6. image or signal zooming, and
7. optimizing display.

In this article, we focus on one particular but central interface between model and observational data. In observational astronomy, modeling embraces instrument models and also information registration and correlation between different data modalities, including image and catalog. The measure of observing performance that is of greatest interest to us in this context is the instrument degradation function, or point-spread function. How the point-spread function is used to improve image or signal quality lies in deconvolution. We will review a range of important recent results in deconvolution. A central theme for us is how nearly all deconvolution methods, arising from different instrument noise models or from priority given to point-source or extended objects, now incorporate resolution scale into their algorithms. Some other results are very exciting too. A recent result of importance is the potential for superresolution, characterized by a precise algorithmic definition of the "near-black object." A further result of note is dithering as a form of stochastic resonance and not just as a purely ad hoc approach to getting a better signal.

Deconvolution of astronomical images has proven in some cases to be crucial for extracting scientific content. For instance, *IRAS* images can be efficiently reconstructed thanks to a new pyramidal maximum entropy algorithm (Bontekoe, Koper, & Kester 1994). Io volcanism can be studied with a lower resolution of 0.″15, or 570 km on Io (Marchis, Prangé, & Christou 2000). Deconvolved mid-infrared images at 20 μm revealed the inner structure of the active galactic nucleus in NGC 1068, hidden at lower wavelength because of the high extinction (Alloin et al. 2000; see Fig. 1). Research on gravitational lenses is easier and more efficient when applying deconvolution methods (Courbin, Lidman, & Magain 1998). A final example is the high resolution (after deconvolution) of mid-infrared images revealing the intimate structure of young stellar objects (Zavagno, Lagage, & Cabrit 1999). Deconvolution will be even more crucial in the future in order to fully take advantage of increasing numbers of high-quality ground-based telescopes, for which images are strongly limited in resolution by the seeing.

The *Hubble Space Telescope* (*HST*) provided a leading example of the need for deconvolution in the period before the detector system was refurbished. Two proceedings (White & Allen 1990; Hanisch & White 1994) provide useful overviews of this work, and a later reference is Adorf, Hook, & Lucy (1995). While an atmospheric seeing point-spread function (PSF) may be relatively tightly distributed around the mode,

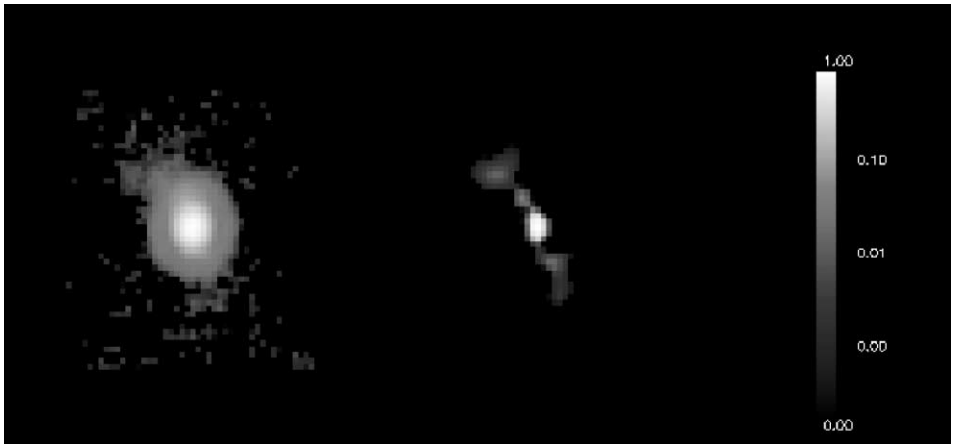

FIG. 1.—Active galactic nucleus of NGC 1068 observed at 20 μm. *Left:* Raw image is highly blurred by telescope diffraction. *Right:* Restored image using the multiscale entropy method reveals the inner structure in the vicinity of the nucleus.

this was not the case for the spherically aberrated *HST* PSF. Whenever the PSF "wings" are extended and irregular, deconvolution offers a straightforward way to mitigate the effects of this and to upgrade the core region of a point source. One usage of deconvolution of continuing importance is in information fusion from different detectors. For example, Faure et al. (2002) deconvolve *HST* images when correlating with ground-based observations. In Radomski et al. (2002), Keck data are deconvolved for study with *HST* data. VLT data are deconvolved in Burud et al. (2002), with other ESO and *HST* data used as well. In planetary work, Coustenis et al. (2001) discuss CFHT data as well as *HST* and other observations.

What emerges very clearly from this small sample—which is in no way atypical—is that a major use of deconvolution is to help in cross-correlating image and signal information.

An observed signal is never in pristine condition, and improving it involves inverting the spoiling conditions, i.e., finding a solution to an inverse equation. Constraints related to the type of signal we are dealing with play an important role in the development of effective and efficient algorithms. The use of constraints to provide for a stable and unique solution is termed regularization. Examples of commonly used constraints include a result image or signal that is nonnegative everywhere, an excellent match to source profiles, necessary statistical properties (Gaussian distribution, no correlation, etc.) for residuals, and absence of specific artifacts (ringing around sources, blockiness, etc.).

Our review opens in § 2 with a formalization of the problem. In § 3, we consider the issue of regularization. In § 4, the CLEAN method, which is central to radio astronomy, is described. Bayesian modeling and inference in deconvolution is reviewed in § 5. In § 6, we introduce wavelet-based methods as used in deconvolution. These methods are based on multiple resolution or scale. In §§ 7 and 8, important issues related to resolution of the output result image are discussed. Section 7 is based on the fact that it is normally not worthwhile to target an output result with better resolution than some limit, for instance, a pixel size. In § 8, we investigate when, where, and how missing information can be inferred to provide superresolution.

2. THE DECONVOLUTION PROBLEM

Noise is the bane of the image analyst's life. Without it we could so much more easily rectify data, compress them, and interpret them. Unfortunately, however, deconvolution becomes a difficult problem due to the presence of noise in high-quality or deep imaging.

Consider an image characterized by its intensity distribution (the "data") I, corresponding to the observation of a "real image" O through an optical system. If the imaging system is linear and shift-invariant, the relation between the data and the image in the same coordinate frame is a convolution:

$$\begin{aligned} I(x, y) &= \int_{x_1=-\infty}^{+\infty} \int_{y_1=-\infty}^{+\infty} P(x - x_1, y - y_1) O(x_1, y_1) dx_1 dy_1 \\ &\quad + N(x, y) \\ &= (P * O)(x, y) + N(x, y), \end{aligned} \tag{1}$$

where P is the PSF of the imaging system and N is additive noise.

In Fourier space, we have

$$\hat{I}(u, v) = \hat{O}(u, v)\hat{P}(u, v) + \hat{N}(u, v). \tag{2}$$

We want to determine $O(x, y)$ knowing I and P. This inverse problem has led to a large amount of work, the main difficulties being the existence of (1) a cutoff frequency of the PSF and (2) the additive noise (see, for example, Cornwell 1989;

Katsaggelos 1993; Bertero & Boccacci 1998; Molina et al. 2001).

A solution can be obtained by computing the Fourier transform of the deconvolved object $\hat{O}$ by a simple division between the image $\hat{I}$ and the PSF $\hat{P}$:

$$\hat{\tilde{O}}(u, v) = \frac{\hat{I}(u, v)}{\hat{P}(u, v)} = \hat{O}(u, v) + \frac{\hat{N}(u, v)}{\hat{P}(u, v)}. \quad (3)$$

This method, sometimes called the *Fourier-quotient method*, is very fast. We need to do only a Fourier transform and an inverse Fourier transform. For frequencies close to the frequency cutoff, the noise term becomes important, and the noise is amplified. Therefore, in the presence of noise, this method cannot be used.

Equation (1) is usually in practice an ill-posed problem. This means that there is no unique and stable solution.

The diversity of algorithms to be looked at in the following sections reflects different ways of recovering a "best" estimate of the source. If one has good prior knowledge, then simple modeling of PSF-convolved sources with a set of variable parameters is often used. In fact, this is often a favored approach, in order to avoid deconvolution, even though its users are unaware of the consequences of its spatially correlated residuals. Lacking specific source information, one then relies on general properties, which have been referred to in § 1. The algorithms described in our review approach these issues in different ways.

With linear regularized methods (§ 3) we use a smoothing/sharpening trade-off. CLEAN assumes our objects are point sources. We discuss the powerful Bayesian methodology in terms of different noise models that can be applicable. Maximum entropy makes a very specific assumption about source structure, but in at least its traditional formulations it was poor at addressing the expected properties of the residuals produced when the estimated source was compared to the observations. Some further work is reviewed that models planetary images or extended objects. So far, all of these methods work, usually iteratively, on the given data.

The story of § 6 is an answer to the question: Where and how do we introduce resolution scale into the methods we review in § 3, 4, and 5, and what are the benefits of doing this?

Some varied directions that deconvolution can take are as follows:

1. Superresolution: object spatial frequency information outside the spatial bandwidth of the image formation system is recovered.
2. Blind deconvolution: the PSF P is unknown.
3. Myopic deconvolution: the PSF P is partially known.
4. Image reconstruction: an image is formed from a series of projections (computed tomography, positron emission tomography [PET], and so on).

We will discuss only the deconvolution and superresolution problems in this paper.

In the deconvolution problem, the PSF is assumed to be known. In practice, we have to construct a PSF from the data or from an optical model of the imaging telescope. In astronomy, the data may contain stars, or one can point toward a reference star in order to reconstruct a PSF. The drawback is the "degradation" of this PSF because of unavoidable noise or spurious instrument signatures in the data. So, when reconstructing a PSF from experimental data, one has to reduce very carefully the images used (background removal, for instance) or otherwise any spurious feature in the PSF would be repeated around each object in the deconvolved image. Another problem arises when the PSF is highly variable with time, as is the case for adaptive optics images. This usually means that the PSF estimated when observing a reference star, after or before the observation of the scientific target, has small differences from a perfect PSF. In this particular case, one has to turn toward myopic deconvolution methods (Christou et al. 1999) in which the PSF is also estimated in the iterative algorithm using a first guess deduced from observations of reference stars.

Another approach consists of constructing a synthetic PSF. Several studies (Buonanno et al. 1983; Moffat 1969; Djorgovski 1983; Molina et al. 1992) have suggested a radially symmetric approximation to the PSF:

$$P(r) \propto \left(1 + \frac{r^2}{R^2}\right)^{-\beta}. \quad (4)$$

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

3. LINEAR REGULARIZED METHODS

It is easy to verify that the minimization of $\| I(x, y) - P(x, y) * O(x, y)\|^2$, where the asterisk means convolution, leads to the least-squares solution:

$$\hat{\tilde{O}}(u, v) = \frac{\hat{P}^*(u, v)\hat{I}(u, v)}{|\hat{P}(u, v)|^2}, \quad (5)$$

which is defined only if $\hat{P}(u, v)$ (the Fourier transform of the PSF) is different from zero. A tilde indicates an estimate. The problem is generally ill-posed and we need to introduce *regularization* in order to find a unique and stable solution.

Tikhonov regularization (Tikhonov et al. 1987) consists of minimizing the term

$$J_T(O) = \| I(x, y) - (P * O)(x, y) \| + \lambda \| H * O \|, \quad (6)$$

where H corresponds to a high-pass filter. This criterion contains two terms. The first, $\|I(x, y) - P(x, y) * O(x, y)\|^2$, expresses fidelity to the data $I(x, y)$, and the second,

$\lambda \| H * O\|^2$, expresses smoothness of the restored image; λ is the regularization parameter and represents the trade-off between fidelity to the data and the smoothness of the restored image. The solution is obtained directly in Fourier space:

$$\hat{\tilde{O}}(u, v) = \frac{\hat{P}^*(u, v)\hat{I}(u, v)}{| \hat{P}(u, v)|^2 + \lambda | \hat{H}(u, v)|^2}. \quad (7)$$

Finding the optimal value λ necessitates use of numerical techniques such as cross-validation (Golub, Heath, & Wahba 1979; Galatsanos & Katsaggelos 1992). This method works well, but computationally it is relatively lengthy and produces smoothed images. This second point can be a real problem when we seek compact structures such as is the case in astronomical imaging.

This regularization method can be generalized, and we write

$$\hat{\tilde{O}}(u, v) = \hat{W}(u, v)\frac{\hat{I}(u, v)}{\hat{P}(u, v)}, \quad (8)$$

which leads directly to Wiener filtering when the W filter depends on both signal and noise behavior (see eq. [16] below, which introduces Wiener filtering in a Bayesian framework); W must satisfy the following conditions (Bertero & Boccacci 1998). We give here the window definition in one dimension:

1. $| \hat{W}(\nu) | \leq 1$, for any $\nu > 0$.
2. $\lim_{\nu \to 0} \hat{W}(\nu) = 1$, for any ν such that $\hat{P}(\nu) \neq 0$.
3. $\hat{W}(\nu)/\hat{P}(\nu)$ bounded for any $\nu > 0$.

Any function satisfying these three conditions defines a regularized linear solution. The most commonly used windows are Gaussian, Hamming, Hanning, and Blackman (Bertero & Boccacci 1998). The function can also be derived directly from the PSF (Pijpers 1999). Linear regularized methods have the advantage of being very attractive from a computation point of view. Furthermore, the noise in the solution can easily be derived from the noise in the data and the window function. For example, if the noise in the data is Gaussian with a standard deviation σ_d, the noise in the solution is $\sigma_s^2 = \sigma_d^2 \sum W_k^2$. But this noise estimation does not take into account errors relative to inaccurate knowledge of the PSF, which limits its interest in practice.

Linear regularized methods present also a number of severe drawbacks:

1. Creation of Gibbs oscillations in the neighborhood of the discontinuities contained in the data. The visual quality is therefore degraded.

2. No a priori information can be used. For example, negative values can exist in the solution, while in most cases we know that the solution must be positive.

3. Since the window function is a low-pass filter, the resolution is degraded. There is trade-off between the resolution we want to achieve and the noise level in the solution. Other methods such as wavelet-based methods do not have such a constraint.

4. CLEAN

The CLEAN method (Högbom 1974) is a mainstream one in radio astronomy. This approach assumes that the object is only composed of point sources. It tries to decompose the image (called the dirty map) into a set of δ-functions. This is done iteratively by finding the point with the largest absolute brightness and subtracting the PSF (dirty beam) scaled with the product of the loop gain and the intensity at that point. The resulting residual map is then used to repeat the process. The process is stopped when some prespecified limit is reached. The convolution of the δ-functions with an ideal PSF (clean beam) plus the residual equals the restored image (clean map). This solution is only possible if the image does not contain large-scale structures.

In the work of Champagnat, Goussard, & Idier (1996) and Kaaresen (1997), the restoration of an object composed of peaks, called *sparse spike trains*, has been treated in a rigorous way.

5. BAYESIAN METHODOLOGY

5.1. Definition

The Bayesian approach consists of constructing the conditional probability density relationship

$$p(O \mid I) = \frac{p(I \mid O)p(O)}{p(I)}, \quad (9)$$

where $p(I)$ is the probability of our image data and $p(O)$ is the probability of the real image, over all possible image realizations. The Bayes solution is found by maximizing the right part of the equation. The maximum likelihood (ML) solution maximizes only the density $p(I \mid O)$ over O:

$$\mathrm{ML}(O) = \max_O p(I \mid O). \quad (10)$$

The maximum a posteriori (MAP) solution maximizes over O the product $p(I \mid O)p(O)$ of the ML and a prior:

$$\mathrm{MAP}(O) = \max_O p(I \mid O)p(O). \quad (11)$$

The term $p(I)$ is considered as a constant value that has no effect in the maximization process and is ignored. The ML

solution is equivalent to the MAP solution assuming a uniform probability density for $p(O)$.

5.2. Maximum Likelihood with Gaussian Noise

The probability $p(I \mid O)$ is

$$p(I \mid O) = \frac{1}{\sqrt{2}\pi\sigma_N} \exp - \frac{(I - P * O)^2}{2\sigma_N^2}, \tag{12}$$

and, assuming that $p(O)$ is a constant, maximizing $p(O \mid I)$ is equivalent to minimizing

$$J(O) = \frac{\| I - P * O\|^2}{2\sigma_n^2}. \tag{13}$$

We obtain the least-squares solution using equation (5). This solution is not regularized. A regularization can be derived by minimizing equation (13) using an iterative algorithm such as the steepest descent minimization method. A typical iteration is

$$O^{n+1} = O^n + \gamma P^* * (I - P * O^n), \tag{14}$$

where $P^*(x, y) = P(-x, -y)$, P^* is the transpose of the PSF, and $O^{(n)}$ is the current estimate of the desired "real image." This method is usually called the Landweber method (Landweber 1951), but sometimes also the *successive approximations* or Jacobi method (Bertero & Boccacci 1998). The number of iterations plays an important role in these iterative methods. Indeed, the number of iterations can be considered as a regularization parameter. When the number of iterations increases, the iterates first approach the unknown object and then potentially go away from it (Bertero & Boccacci 1998). Furthermore, some constraints can be incorporated easily in the basic iterative scheme. Commonly used constraints are the positivity (i.e., the object must be positive), the support constraint (i.e., the object belongs to a given spatial domain), or the band-limited constraint (i.e., the Fourier transform of the object belongs to a given frequency domain). More generally, the constrained Landweber method is written as

$$O^{n+1} = \mathcal{P}_C[O^n + \alpha(I - P * O^n)], \tag{15}$$

where $\mathcal{P}_C$ is the projection operator that enforces our set of constraints on O^n.

5.3. Gaussian Bayes Model

If the object and the noise are assumed to follow Gaussian distributions with zero mean and variance, respectively, equal to σ_O and σ_N, then a Bayes solution leads to the Wiener filter:

$$\hat{O}(u, v) = \frac{\hat{P}^*(u, v)\hat{I}(u, v)}{|\hat{P}(u, v)|^2 + [\sigma_N^2(u, v)] / [\sigma_O^2(u, v)]}. \tag{16}$$

Wiener filtering has serious drawbacks (artifact creation such as ringing effects) and needs spectral noise estimation. Its advantage is that it is very fast.

5.4. Maximum Likelihood with Poisson Noise

The probability $p(I \mid O)$ is

$$p(I \mid O) = \prod_{x,y} \frac{[(P * O)(x, y)]^{I(x,y)} \exp\{-(P * O)(x, y)\}}{I(x, y)!}. \tag{17}$$

The maximum can be computed by taking the derivative of the logarithm:

$$\frac{\partial \ln p(I \mid O)(x, y)}{\partial O(x, y)} = 0. \tag{18}$$

Assuming the PSF is normalized to unity, and using Picard iteration (Issacson & Keller 1966), we get

$$O^{n+1}(x, y) = \left[\frac{I(x, y)}{(P * O^n)(x, y)} * P^*(x, y)\right] O^n(x, y), \tag{19}$$

which is the Richardson-Lucy algorithm (Richardson 1972; Lucy 1974; Shepp & Vardi 1982), also sometimes called the *expectation maximization* (EM) method (Dempster, Laird, & Rubin 1977). This method is commonly used in astronomy. Flux is preserved and the solution is always positive. Constraints can also be added by using the following iterative scheme:

$$O^{n+1} = \mathcal{P}_C\left\{O^n\left[\frac{I}{(P * O^n)} * P^*\right]\right\}. \tag{20}$$

5.5. Poisson Bayes Model

We formulate the object probability density function (PDF)

as

$$p(O) = \prod_{x,y} \frac{M(x, y)^{O(x,y)} \exp\{-M(x, y)\}}{O(x, y)!}. \quad (21)$$

The MAP solution is

$$O(x, y) = M(x, y) \exp\left\{\left[\frac{I(x, y)}{(P * O)(x, y)} - 1\right] * P^*(x, y)\right\}, \quad (22)$$

and choosing $M = O^n$ and using Picard iteration leads to

$$O^{n+1}(x, y) = O^n(x, y) \exp\left\{\left[\frac{I(x, y)}{(P * O^n)(x, y)} - 1\right] * P^*(x, y)\right\}. \quad (23)$$

5.6. Maximum Entropy Method

In the absence of any information on the solution O except its positivity, a possible course of action is to derive the probability of O from its entropy, which is defined from information theory. Then if we know the entropy H of the solution, we derive its probability as

$$p(O) = \exp[-\alpha H(O)]. \quad (24)$$

The most commonly used entropy functions are

1. Burg (1978):[1] $H_b(O) = -\sum_x \sum_y \ln[O(x, y)]$;
2. Frieden (1978): $H_f(O) = -\sum_x \sum_y O(x, y) \ln[O(x, y)]$;
3. Gull & Skilling (1991):

$$H_g(O) = \sum_x \sum_y O(x, y) - M(x, y) - O(x, y) \ln[O(x, y)|M(x, y)].$$

The last definition of the entropy has the advantage of having a zero maximum when O equals the model M, usually taken as a flat image.

5.7. Other Regularization Models

In this section, we discuss approaches to deconvolving images of extended objects.

Molina et al. (2001) present an excellent review of taking the spatial context of image restoration into account. Some appropriate prior is used for this. One such regularization constraint is

$$\| CI\|^2 = \sum_x \sum_y I(x, y) - \frac{1}{4}[I(x, y + 1) + I(x, y - 1) + I(x + 1, y) + I(x - 1, y)]. \quad (25)$$

Similar to the discussion above in § 5.2, this is equivalent to defining the prior

$$p(O) \propto \exp\left\{-\frac{\alpha}{2} \| CI\|^2\right\}. \quad (26)$$

Given the form of equation (25), such regularization can be viewed as setting a constraint on the Laplacian of the restoration. In statistics this model is a simultaneous autoregressive (SAR) model (Ripley 1981).

Alternative prior models can be defined, related to the SAR model of equation (25). In

$$p(O) \propto \exp\left\{-\alpha \sum_x \sum_y [I(x, y) - I(x, y + 1)]^2 + [I(x, y) - I(x + 1, y)]^2\right\}, \quad (27)$$

constraints are set on first derivatives.

Blanc-Féraud & Barlaud (1996) and Charbonnier et al. (1997) consider the following prior:

$$p(O) \propto \exp\left\{-\alpha \sum_x \sum_y \phi(\| \nabla I \| (x, y))\right\} \quad (28)$$

$$\propto \exp\left\{-\alpha \sum_x \sum_y \left[\phi(I(x, y) - I(x, y + 1))^2 + \phi(I(x, y) - I(x + 1, y))^2\right]\right\}. \quad (29)$$

The function ϕ, called a *potential function*, is an edge-preserving function. The term $\alpha \sum_x \sum_y \phi(\| \nabla I \| (x, y))$ can also be interpreted as the Gibbs energy of a Markov random field.

The ARTUR method (Charbonnier et al. 1997), which has been used for helioseismic inversion (Corbard et al. 1999), uses the function $\phi(t) = \log(1 + t^2)$. Anisotropic diffusion (Perona & Malik 1990; Alvarez, Lions, & Morel 1992) uses similar functions, but in this case the solution is computed using *partial differential equations*.

The function $\phi(t) = t$ leads to the *total variation* method

[1] "Multichannel Maximum Entropy Spectral Analysis," paper presented at the Annual Meeting of the International Society of Exploratory Geophysics.

(Rudin, Osher, & Fatemi 1992; Acar & Vogel 1994); the constraints are on first derivatives, and the model is a special case of a conditional autoregressive (CAR) model. Molina et al. (2001) discuss the applicability of CAR models to image restoration involving galaxies. They argue that such models are particularly appropriate for the modeling of luminosity exponential and $r^{1/4}$ laws.

The priors reviewed above can be extended to more complex models. In Molina et al. (1996, 2000), a compound Gauss Markov random field (CGMRF) model is used, one of the main properties of which is to target the preservation and improvement of line processes. Another prior again was used in Molina & Cortijo (1992) for the case of planetary images.

6. WAVELET-BASED DECONVOLUTION

6.1. Introduction

The regularized methods presented in the previous sections give rise to a range of limitations: Fourier-based methods such as Wiener or Tikhonov methods lead to a band-limited solution, which is generally not optimal for astronomical image restoration, especially when the data contain point sources. The CLEAN method cannot correctly restore extended sources. The maximum entropy method (MEM) cannot recover simultaneously both compact and extended sources. MEM regularization presents several drawbacks, which are discussed in Starck et al. (2001b). The main problems are (1) results depend on the background level; (2) the proposed entropy functions give poor results for negative structures, i.e., structures under the background level (such as absorption bands in a spectrum); and (3) spatial correlation in the images is not taken into account. Iterative regularized methods such as Richardson-Lucy or the Landweber method do not prevent noise amplification during the iterations. Finally, if Markov random field based methods can be very useful for images with edges such as planetary images, they are ill-adapted for other cases, insofar as the majority of astronomical images contain objects that are relatively diffuse and do not have a "border."

6.2. Toward Multiresolution

The Fourier domain diagonalizes the convolution operator, and we can identify and reduce the noise that is amplified during the inversion. When the signal can be modeled as stationary and Gaussian, the Wiener filter is optimal. But when the signal presents spatially localized features such as singularities or edges, these features cannot be well represented with Fourier basis functions, which extend over the entire spatial domain. Other basis functions, such as wavelets, are better suited to represent a large class of signals.

The wavelet transform, its conceptual links with Fourier and Gabor transforms, its indirect links with Karhunen-Loève and other transforms, and its generalization to multiresolution transforms, are all dealt with at length in Starck, Murtagh, & Bijaoui (1998a), Starck & Murtagh (2002), and many articles in the mainstream astronomy literature. Perhaps among the most important properties of the wavelet transform are the following:

1. A resolution scale decomposition of the data is provided, using high-pass, bandpass or detail, and low-pass or smooth coefficients.
2. The transformed data are more compact than the original. Indeed, the noise is uniformly distributed over all coefficients while the signal of interest is concentrated in a few coefficients. Therefore, the signal-to-noise ratio of these coefficients is high, which opens the way toward data filtering and denoising.
3. Of even greater relevance for data denoising, noise models defined for the original data often carry over well into wavelet transform space.

The last point is of tremendous importance in the physical sciences: as a result of the instrument or sensor used, we generally know the noise model we are dealing with. Direct definition of this noise model's parameters from the observed data is not at all easy. Determining the noise parameters in wavelet space is a far more effective procedure.

In this short briefing on the capital reasons explaining the importance of wavelet and multiresolution transforms, we note that wavelet transforms differ in the wavelet function used and in a few different schemes for arranging the high- and low-frequency information used to define our data in wavelet space. Such schemes include the graphical "frequency domain tiling" used below in §§ 6.3 and 6.4, which provide powerful summaries of transform properties. We have generally espoused so-called redundant transforms (i.e., each wavelet resolution scale has exactly the same number of pixels as the original data) whenever "pattern recognition" uses of the wavelet scale are uppermost in our minds, as opposed to compression.

A final point to note: the manner in which the wavelet transform is incorporated into more traditional deconvolution approaches varies quite a bit. In CLEAN, the scale-based decomposition is used to help us focus in on the solution. In § 6.4 below, the question of how noise is propagated into multiresolution transform space is uppermost in our minds. In § 6.5, we "siphon off" part of the restored data at each iteration of an iterative deconvolution, clean it by denoising, and feed it back into the following iteration.

We will return below to look at particular wavelet transform algorithms. In the remainder of this section, we will review various approaches that have links or analogies with multiresolution approaches.

The concept of multiresolution was first introduced for deconvolution by Wakker & Schwarz (1988) when they proposed the multiresolution CLEAN algorithm for interferometric image deconvolution. During the last 10 years, many developments have taken place in order to improve the existing methods (CLEAN, Landweber, Lucy, MEM, and so on), and these results have led to the use of different levels of resolution.

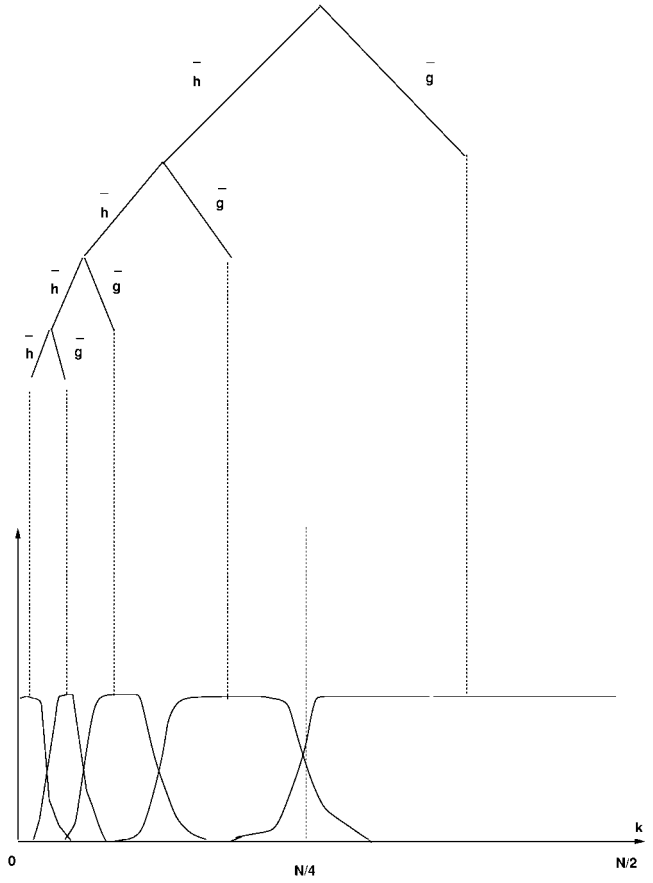

FIG. 2.—Frequency domain tiling by the one-dimensional wavelet transform. The filter pair, h and $\bar{g}$, are, respectively, low pass and bandpass. The tree shows the order of application of these filters. The tiling shows where these filters have an effect in frequency space.

The Lucy algorithm was modified (Lucy 1994) in order to take into account a priori information about stars in the field where both position and brightness are known. This is done by using a two-channel restoration algorithm, one channel representing the contribution relative to the stars, and the second to the background. A smoothness constraint is added on the background channel. This method, called PLUCY, was then refined first (and called CPLUCY) for considering subpixel positions (Hook 1999), and a second time (and called GIRA; Pirzkal, Hook, & Lucy 2000) for modifying the smoothness constraint.

A similar approach has been followed by Magain, Courbin, & Sohy (1998), but more in the spirit of the CLEAN algorithm. Again, the data are modeled as a set of point sources on top of spatially varying background, leading to a two-channel algorithm.

MEM has also been modified by several authors (Weir 1992; Bontekoe et al. 1994; Pantin & Starck 1996; Núñez & Llacer 1998; Starck et al. 2001b). First, Weir proposed the *multichannel MEM*, in which an object is modeled as the sum of objects at different levels of resolution. The method was then improved by Bontekoe et al. (1994) with the *pyramid MEM*. In particular, many regularization parameters were fixed by the introduction of the dyadic pyramid. The link between pyramid MEM and wavelets was underlined in Pantin & Starck (1996) and Starck et al. (2001b), and it was shown that all the regularization parameters can be derived from the noise modeling. Wavelets were also used in Núñez & Llacer (1998) in order to create a segmentation of the image, each region being then restored with a different smoothness constraint, depending on the resolution level where the region was found. This last method, however, has the drawback of requiring user interaction for deriving the segmentation threshold in the wavelet space.

The *Pixon* method (Dixon et al. 1996; Puetter & Yahil 1999) is relatively different from the previously described methods. This time, an object is modeled as the sum of pseudoimages smoothed locally by a function with position-dependent scale, called the Pixon shape function. The set of pseudoimages defines a dictionary, and the image is supposed to contain only features included in this dictionary. But the main problem lies in the fact that features that cannot be detected directly in the data or in the data after a few Lucy iterations will not be modeled with the Pixon functions, and they will be strongly regularized as background. The result is that the faintest objects are overregularized while strong objects are well restored. This is striking in the example shown in Figure 8.

Wavelets offer a mathematical framework for the multiresolution processing. Furthermore, they furnish an ideal way to include noise modeling in the deconvolution methods. Since the noise is the main problem in deconvolution, wavelets are very well adapted to the regularization task.

6.3. The Wavelet Transform

We begin with the wavelet transform used in the overwhelming majority of practical applications, for the simple reason that its performance in compression is well proven. It is used in the JPEG2000 standard, for instance. Of course, for compression, it is a nonredundant transform. The schema used in the wavelet transform output, and the associated frequency domain tiling, will be familiar to anyone who has studied wavelets in the nonastronomy (and compression) context. In § 6.4, we will contrast the bi-orthogonal wavelet transform with the recently developed innovative use of the wavelet-vaguelette approach to noise filtering.

We denote $\mathcal{W}$ the (bi-) orthogonal wavelet transform and w the wavelet transform of a signal s, $w = \mathcal{W}s$; w is composed of a set of wavelet bands w_j and a coarse version c_J of s, $w = \{w_1, \ldots, w_J, c_J\}$, where J is the number of scales used in the wavelet transform. Roughly speaking, the Fourier transform of a given wavelet band w_j is localized in a frequency band with support $[1/2^{j+1}, 1/2^j]$, and the Fourier transform of the smoothed array c_J is localized in the frequency band with

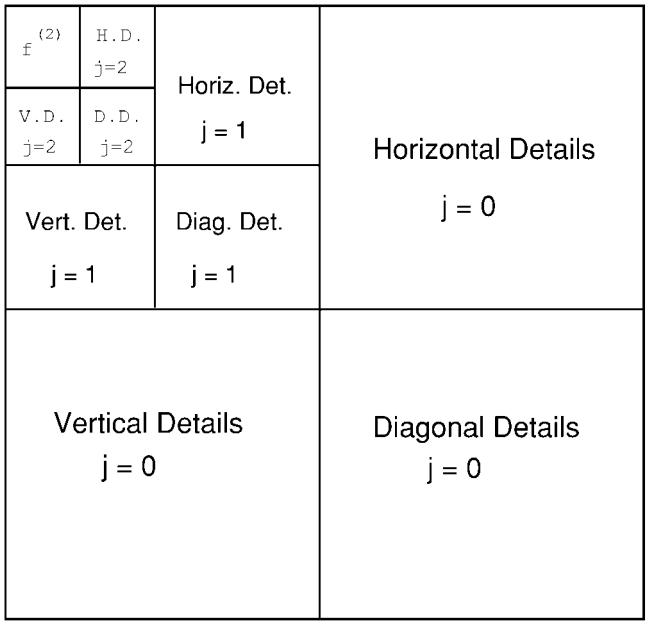

FIG. 3.—Orthogonal wavelet transform representation of an image.

support $[0, 1/2^J]$. Thus, the algorithm outputs $J + 1$ subband arrays. The indexing is such that, here, $j = 1$ corresponds to the finest scale (high frequencies). Coefficients $c_{j,l}$ and $w_{j,l}$ are obtained by means of the filters h and g, $c_{j+1,l} = \sum_k h(k - 2l)c_{j,k}$ and $w_{j+1,l} = \sum_k g(k - 2l)c_{j,k}$, where the filters h and g are derived from the analysis wavelet function ψ; h and g can be interpreted as, respectively, a low- and a high-pass filter. Another important point in this algorithm is the critical sampling. Indeed, the number of pixels in the transformed data w is equal to the number of pixels N in the original signal. This is possible because of the decimation performed at each resolution level. The signal c_j at resolution level j (with N_j pixels and $c_0 = s$) is decomposed into two bands w_{j+1} and c_{j+1}, both of them containing $N_j/2$ pixels. Finally, the signal s can be reconstructed from its wavelet coefficients: $s = \mathcal{W}^{-1}w$, using the inverse wavelet transform.

Figure 2 shows the frequency domain tiling by the one-dimensional wavelet transform. The first convolution step by the two filters h and g separates the frequency band into two parts, the high frequencies and the low frequencies. The first scale w_1 of the wavelet transform corresponds to the high frequencies. The same process is then repeated several times on the low-frequency band, which is separated into two parts at each step. We get the wavelet scales $w_2, \ldots, w_J$ and c_J. The wavelet transform can be seen as a set of passband filters that has the properties of being reversible (the original data can be reconstructed from the wavelet coefficients) and nonredundant.

The two-dimensional algorithm is based on separate variables leading to prioritizing of horizontal, vertical, and diagonal directions. The detail signal is obtained from three wavelets: the vertical wavelet, the horizontal wavelet, and the diagonal wavelet. This leads to three wavelet subimages at each resolution level. This transform is nonredundant, which means that the number of pixels in the wavelet-transformed data is the same as in the input data.

Figure 3 shows the spatial representation of a two-dimensional wavelet transform. The first step decomposes the image into three wavelet coefficient bands (i.e., horizontal band, vertical band, and diagonal band) and the smoothed array. The same process is then repeated on the smoothed array. Figure 4 shows the wavelet transform of a galaxy (NGC 2997) using four resolution levels.

Other discrete wavelet transforms exist. The à trous wavelet transform is very well suited for astronomical data and has been discussed in many papers and books (Starck et al. 1998a). By using the à trous wavelet transform algorithm, an image I can be defined as the sum of its J wavelet scales and the last smooth array: $I(x, y) = c_J(x, y) + \sum_{j=1}^{J} w_{j,x,y}$, where the first term on the right is the last smoothed array and w denotes a wavelet scale. This algorithm is quite different from the previous one. It is redundant; i.e., the number of pixels in the transformed data is larger than in the input data (each wavelet scale has the same size as the original image, hence the redundancy factor is $J + 1$); and it is isotropic (there are no favored orientations). Both properties are useful for the purpose of astronomical image restoration.

Figure 5 shows the à trous transform of the galaxy NGC 2997. Three wavelet scales are shown (*upper left*, *upper right*, *lower left*) and the final smoothed plane (*lower right*). The original image is given exactly by the sum of these four images.

Since the wavelet transform is a linear transform, the noise behavior in wavelet space can be well understood and correctly modeled. This point is fundamental since one of the main problems of deconvolution is the presence of noise in the data.

6.4. Wavelet-Vaguelette Decomposition

The wavelet-vaguelette decomposition, proposed by Donoho (1995), consists of first applying an inverse filtering:

$$F = P^{-1} * I + P^{-1} * N = O + Z, \tag{30}$$

where P^{-1} is the inverse filter $[\hat{P}^{-1}(u, v) = 1/\hat{P}(u, v)]$. The noise $Z = P^{-1} * N$ is not white but remains Gaussian. It is amplified when the deconvolution problem is unstable. Then a wavelet transform is applied to F, the wavelet coefficients are soft- or hard-thresholded (Donoho 1993), and the inverse wavelet transform furnishes the solution.

The method has been refined by adapting the wavelet basis to the frequency response of the inverse of P (Kalifa 1999; Kalifa, Mallat, & Rougé 2000). This *mirror wavelet basis* has a time-frequency tiling structure different from conventional wavelets, and it isolates the frequency where $\hat{P}$ is close to zero, because a singularity in $\hat{P}^{-1}(u_s, v_s)$ influences the noise vari-

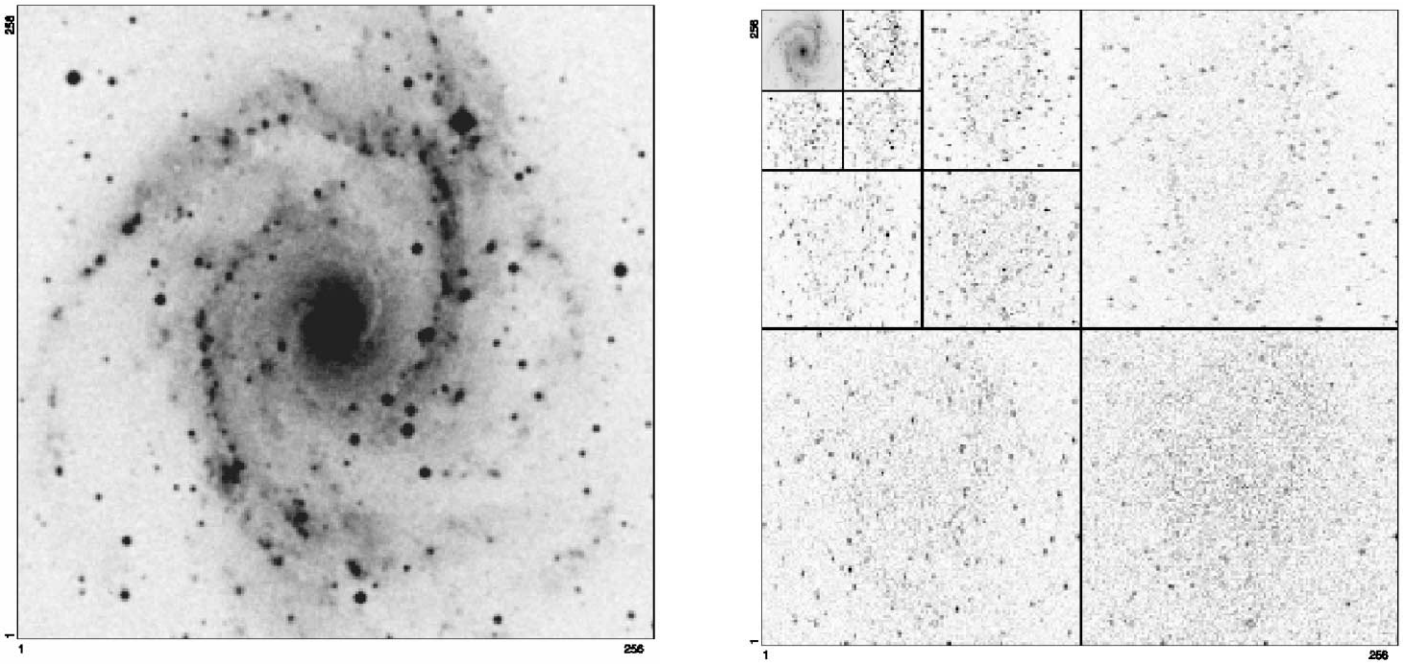

FIG. 4.—Galaxy NGC 2997 and its bi-orthogonal wavelet transform.

ance in the wavelet scale corresponding to the frequency band that includes (u_s, v_s). Figures 6 and 7 show the decomposition of the Fourier space, respectively, in one dimension and two dimensions.

Because it may be not possible to isolate all singularities, Neelamani (1999) and Neelamani, Choi, & Baraniuk (2001) advocated a hybrid approach, proposing to still use the Fourier domain so as to restrict excessive noise amplification. Regularization in the Fourier domain is carried out with the window function W_λ:

$$\hat{W}_\lambda(u, v) = \frac{|\hat{P}(u, v)|^2}{|\hat{P}(u, v)|^2 + \lambda T(u, v)}, \tag{31}$$

where $T(u, v) = \sigma^2/\hat{S}(u, v)$, S being the power spectral density of the observed signal:

$$F = W_\lambda * P^{-1} * I + W_\lambda * P^{-1} * N. \tag{32}$$

The regularization parameter λ controls the amount of Fourier-domain shrinkage and should be relatively small (<1; Neelamani et al. 2001). The estimate F still contains some noise, and a wavelet transform is performed to remove the remaining noise. The optimal λ is determined using a given cost function. See Neelamani et al. (2001) for more details.

This approach is fast and competitive compared to linear methods, and the wavelet thresholding removes the Gibbs oscillations. It presents, however, several drawbacks:

1. The regularization parameter is not so easy to find in practice (Neelamani et al. 2001) and requires some computation time, which limits the usefulness of the method.
2. The positivity a priori is not used.
3. The power spectrum of the observed signal is generally not known.
4. It is not trivial to consider non-Gaussian noise.

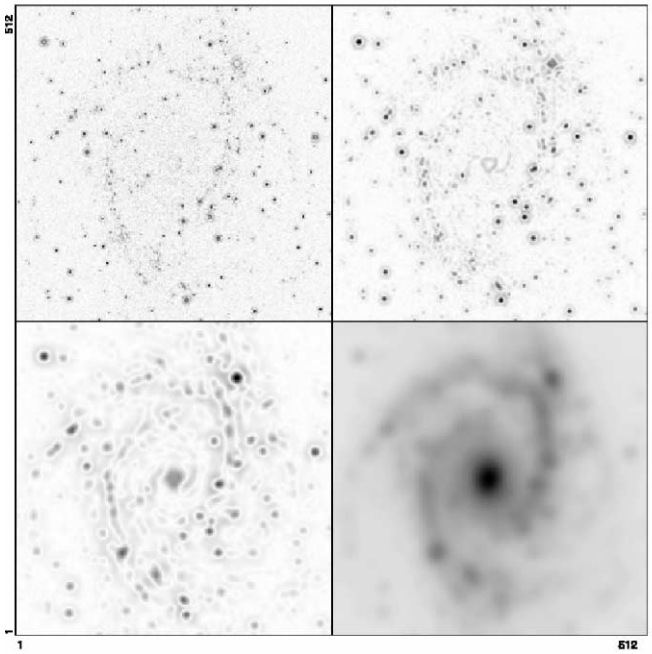

FIG. 5.—Wavelet transform of NGC 2997 by the à trous algorithm.

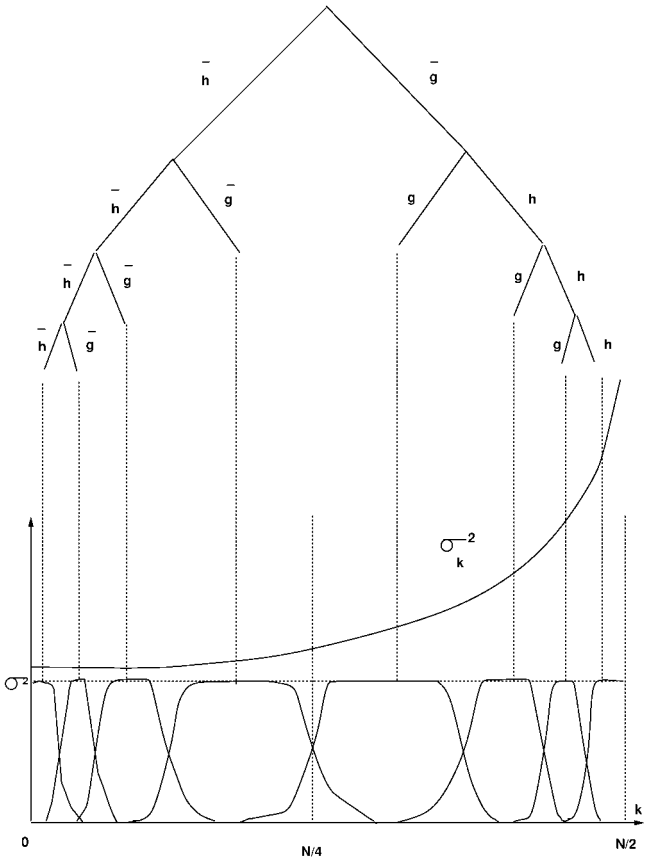

FIG. 6.—Frequency domain tiling using a wavelet packet decomposition with a mirror basis. The variance of the noise has a hyperbolic growth. See text for discussion of the implications of the noise variation.

The second point is important for astronomical images. It is well known that the positivity constraint has a strong influence on the solution quality (Kempen & van Vliet 2000). We will see in the following that it is straightforward to modify the standard iterative methods in such a way that they benefit from the capacity of wavelets to separate the signal from the noise.

6.5. Regularization from the Multiresolution Support

6.5.1. Noise Suppression Based on the Wavelet Transform

We have noted how, in using an iterative deconvolution algorithm such as van Cittert or Richardson-Lucy, we define $R^{(n)}(x, y)$, the residual at iteration n:

$$R^n(x, y) = I(x, y) - (P * O^n)(x, y). \tag{33}$$

By using the à trous wavelet transform algorithm, R^n can be defined as the sum of its J wavelet scales and the last smooth array:

$$R^n(x, y) = c_J(x, y) + \sum_{j=1}^{J} w_{j, x, y}, \tag{34}$$

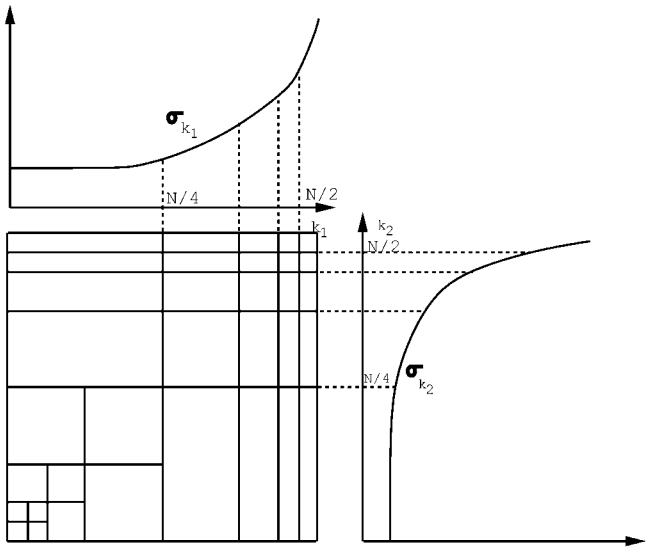

FIG. 7.—Mirror wavelet basis in two-dimensional space. See text for discussion of the implications of the noise variation.

where the first term on the right is the last smoothed array and w denotes a wavelet scale.

The wavelet coefficients provide a mechanism to extract only the significant structures from the residuals at each iteration. Normally, a large part of these residuals are statistically nonsignificant. The significant residual (Murtagh & Starck 1994; Starck & Murtagh 1994) is then

$$\bar{R}^n(x, y) = c_{J, x, y} + \sum_{j=1}^{J} M(j, x, y) w_{j, x, y}, \tag{35}$$

where $M(j, x, y)$ is the multiresolution support and is defined by

$$M(j, x, y) = \begin{cases} 1 & \text{if } w_{j, x, y} \text{ is significant,} \\ 0 & \text{if } w_{j, x, y} \text{ is nonsignificant.} \end{cases} \tag{36}$$

This describes in a logical or Boolean way if the data contain information at a given scale j and at a given position (x, y). Assuming that the noise follows a given distribution, $w_j(x, y)$ is significant if the probability that the wavelet coefficient is due to noise is small $[P(|\, W > w_{j, x, y}\, |) < \epsilon]$. In the case of Gaussian noise, $w_{j, x, y}$ is significant if $w_{j, x, y} > k\sigma_j$, where σ_j is the noise standard deviation at scale j and k is a constant generally taken between 3 and 5. Different noise models are discussed in Starck et al. (1998a). If a priori information is available, such as star positions, it can easily be introduced into the multiresolution support.

An alternative approach was outlined in Murtagh, Starck, & Bijaoui (1995) and Starck, Bijaoui, & Murtagh (1995): the support was initialized to zero and built up at each iteration of the restoration algorithm. Thus, in equation (35) above, $M(j, x, y)$ was additionally indexed by n, the iteration number.

In this case, the support was specified in terms of significant pixels at each scale, j; in addition, pixels could become significant as the iterations proceeded but could not be made nonsignificant.

6.5.2. Regularization of Van Cittert's Algorithm

Van Cittert's iteration (van Cittert 1931) is

$$O^{n+1}(x, y) = O^n(x, y) + \alpha R^n(x, y), \quad (37)$$

with $R^n(x, y) = I^n(x, y) - (P * O^n)(x, y)$. Regularization using significant structures leads to

$$O^{n+1}(x, y) = O^n(x, y) + \alpha \bar{R}^n(x, y). \quad (38)$$

The basic idea of this regularization method consists of detecting, at each scale, structures of a given size in the residual $R^n(x, y)$ and putting them in the restored image $O^n(x, y)$. The process finishes when no more structures are detected. Then, we have separated the image $I(x, y)$ into two images $\tilde{O}(x, y)$ and $R(x, y)$; $\tilde{O}$ is the restored image, which ought not to contain any noise, and $R(x, y)$ is the final residual, which ought not to contain any structure; R is our estimate of the noise $N(x, y)$.

6.5.3. Regularization of the One-Step Gradient Method

The one-step gradient iteration is

$$O^{n+1}(x, y) = O^n(x, y) + P^*(x, y) * R^n(x, y), \quad (39)$$

with $R^n(x, y) = I(x, y) - (P * O^n)(x, y)$. Regularization by significant structures leads to

$$O^{n+1}(x, y) = O^n(x, y) + P^*(x, y) * \bar{R}^n(x, y). \quad (40)$$

6.5.4. Regularization of the Richardson-Lucy Algorithm

From equation (1), we have $I^n(x, y) = (P * O^n)(x, y)$. Then $R^n(x, y) = I(x, y) - I^n(x, y)$, and hence $I(x, y) = I^n(x, y) + R^n(x, y)$.

The Richardson-Lucy equation is

$$O^{n+1}(x, y) = O^n(x, y)\left[\frac{I^n(x, y) + R^n(x, y)}{I^n(x, y)} * P^*(x, y)\right],$$

and regularization leads to

$$O^{n+1}(x, y) = O^n(x, y)\left[\frac{I^n(x, y) + \bar{R}^n(x, y)}{I^n(x, y)} * P^*(x, y)\right].$$

6.5.5. Convergence

The standard deviation of the residual decreases until no more significant structures are found. Convergence can be estimated from the residual. The algorithm stops when a user-specified threshold is reached:

$$(\sigma_{R^{n-1}} - \sigma_{R^n})/\sigma_{R^n} < \epsilon. \quad (41)$$

6.5.6. Examples

A simulated *Hubble Space Telescope* Wide Field Camera image of a distant cluster of galaxies is shown in Figure 8*a*. The image used was one of a number described in Caulet & Freudling (1993) and Freudling & Caulet (1993). The simulated data are shown in Figure 8*b*. Four deconvolution methods were tested: Richardson-Lucy, Pixon, wavelet-vaguelette, and wavelet-Lucy. Deconvolved images are presented, respectively, in Figures 8*c*, 8*d*, 8*e*, and 8*f*. The Richardson-Lucy method amplifies the noise, which implies that the faintest objects disappear in the deconvolved image. The Pixon method introduces regularization, and the noise is under control, while objects where "Pixons" have been detected are relatively well protected from the regularization effect. Since the "Pixon" features are detected from noisy partially deconvolved data, the faintest objects are not in the Pixon map and are strongly regularized. The wavelet-vaguelette method is very fast and produces relatively high quality results when compared to Pixon or Richardson-Lucy, but the wavelet-Lucy method seems clearly the best of the four methods. There are fewer spurious objects than in the wavelet-vaguelette method, it is stable for any kind of PSF, and any kind of noise modeling can be considered.

6.6. Wavelet CLEAN

The CLEAN solution is only available if the image does not contain large-scale structures. Wakker & Schwarz (1988) introduced the concept of multiresolution CLEAN (MRC) in order to alleviate the difficulties occurring in CLEAN for extended sources. The MRC approach consists of building two intermediate images, the first one (called the smooth map) by smoothing the data to a lower resolution with a Gaussian function, and the second one (called the difference map) by subtracting the smoothed image from the original data. Both of these images are then processed separately. By using a standard CLEAN algorithm on them, the smoothed clean map and difference clean map are obtained. The recombination of these two maps gives the clean map at the full resolution. This algorithm may be viewed as an artificial recipe, but it has been shown (Starck et al. 1994, 1998a; Starck & Bijaoui 1994) that it is linked to multiresolution analysis. Wavelet analysis leads to a generalization of MRC from a set of scales. The wavelet CLEAN (WCLEAN) method consists of the following steps:

1. Apply the wavelet transform to the image: we get W_I.

2. Apply the wavelet transform to the PSF: we get W_P.
3. Apply the wavelet transform to the CLEAN beam: we get W_C.
4. For each scale j of the wavelet transform, apply the CLEAN algorithm using the wavelet scale j of both W_I and W_P.
5. Apply an iterative reconstruction algorithm using W_C.

More details can be found in Starck et al. (1994, 1998a).

6.7. Multiscale Entropy

6.7.1. Introduction

In Starck, Murtagh, & Gastaud (1998b), Starck et al. (2001b), and Starck & Murtagh (1999), the benchmark properties for a good "physical" definition of entropy were discussed, and it was proposed to consider that the entropy of a signal is the sum of the information at each scale of its wavelet transform (Starck et al. 1998b), and the information of a wavelet coefficient is related to the probability of it being due to noise. Denoting h the information relative to a single wavelet coefficient, we have $H(X) = \sum_{j=1}^{J} \sum_{k=1}^{N_j} h(w_{j,k})$, with $h(w_{j,k}) = -\ln p(w_{j,k})$, where J is the number of scales and N_j is the number of samples (pixels, time-, or wavelength-interval values) in band (scale) j. For Gaussian noise, the information is proportional to the energy of the wavelet coefficients. The larger the value of a normalized wavelet coefficient, then the lower will be its probability of being noise and the higher will be the information furnished by this wavelet coefficient. Since H is corrupted by the noise, it can be decomposed into two components, one (H_s) corresponding to the noncorrupted part and the other (H_n) to the corrupted part (Starck et al. 1998b): $H(X) = H_s(X) + H_n(X)$; H_s is called the signal information and H_n the noise information. For each wavelet coefficient $w_{j,k}$, we have to estimate the proportions h_n and h_s of h [with $h(w_{j,k}) = h_n(w_{j,k}) + h_s(w_{j,k})$] that should be assigned to H_n and H_s. Hence, signal information and noise information are defined by

$$H_s(X) = \sum_{j=1}^{J} \sum_{k=1}^{N_j} h_s(w_{j,k}),$$

$$H_n(X) = \sum_{j=1}^{J} \sum_{k=1}^{N_j} h_n(w_{j,k}). \quad (42)$$

If a wavelet coefficient is small, its value can be due to noise, and the information h relative to this single-wavelet coefficient should be assigned to H_n. More details can be found in Starck et al. (2001b).

Following the Bayesian scheme, the functional to minimize is

$$J(O) = \sum_{k=1}^{N} \frac{[I_k - (P * O)_k]^2}{2\sigma_I^2} + \alpha \sum_{j=1}^{J} \sum_{k=1}^{N_j} \frac{w_{j,k}^2}{2\sigma_j^2}, \quad (43)$$

where σ_j is the noise at scale j, N_j is the number of pixels at the scale j, σ_I is the noise standard deviation in the data, and J is the number of scales.

Rather than minimizing the amount of information in the solution, we may prefer to minimize the amount of information that can be due to the noise. The function is now

$$J(O) = \sum_{k=1}^{N} \frac{[I_k - (P * O)_k]^2}{2\sigma_I^2} + \alpha H_n(O), \quad (44)$$

and for Gaussian noise, H_n has been defined by

$$H_n(X) = \sum_{j=1}^{J} \sum_{k=1}^{N_j} \frac{1}{\sigma_j^2} \int_0^{|w_{j,k}|} u \,\mathrm{erf}\left(\frac{|w_{j,k}| - u}{\sqrt{2}\sigma_j}\right) du. \quad (45)$$

Minimizing H_n can be seen as a kind of adaptive soft-thresholding in the wavelet terminology. Finally, equation (44) can be generalized by (Starck et al. 2001b)

$$J(O) = H_s(I - P * O) + \alpha H_n(O); \quad (46)$$

i.e., we want to minimize the minimum of information due to the signal in the residual and the minimum of information due to the noise in the solution.

6.7.2. Example: β Pictoris Image Deconvolution

The β Pictoris image (Pantin & Starck 1996) was obtained by integrating 5 hr on-source using a mid-infrared camera, TIMMI, placed on the 3.6 ESO telescope (La Silla, Chile). The raw image has a peak signal-to-noise ratio of 80. It is strongly blurred by a combination of seeing, diffraction (0.″7 on a 3 m class telescope), and additive Gaussian noise. The initial disk shape in the original image has been lost after the convolution with the PSF (see Fig. 9*a*). Thus, we need to deconvolve such an image to get the best information on this object, i.e., the exact profile and thickness of the disk, and subsequently to compare the results to models of thermal dust emission.

After filtering (see Fig. 9*b*), the disk appears clearly. For detection of faint structures (the disk here), one can calculate that the application of such a filtering method to this image provides a gain of observing time of a factor of around 60. The deconvolved image (Fig. 9*c*) shows that the disk is extended at 10 μm and asymmetrical. The multiscale entropy method is more effective for regularizing than other standard methods and leads to good reconstruction of the faintest structures of the dust disk.

6.8. Summary of Scale-based Deconvolution

As already mentioned, our objective in § 6 has been to take the major categories of deconvolution methods—linear inversion, CLEAN, Bayesian optimization, and maximum entropy (respectively, §§ 3, 4, and 5)—and seek where and how resolution scale information could be incorporated. A key requirement for us is that the basic method should stay the same

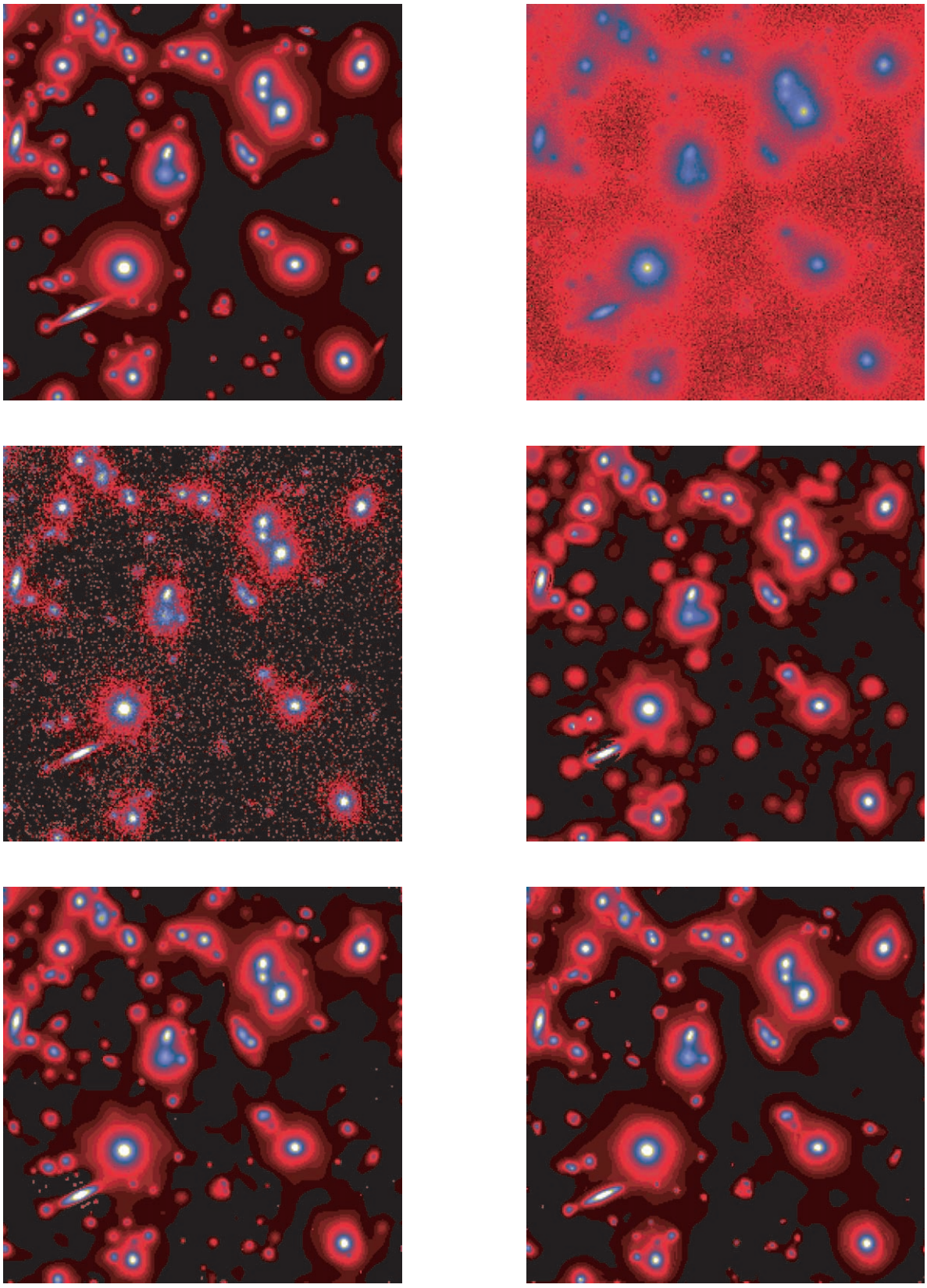

FIG. 8.—Simulated *Hubble Space Telescope* Wide Field Camera image of a distant cluster of galaxies. (*a*) Original, unaberrated, and noise-free. (*b*) Input, aberrated, noise added. (*c*) Restoration, Richardson-Lucy. (*d*) Restoration, Pixon method. (*e*) Restoration, wavelet-vaguelette. (*f*) Restoration, wavelet-Lucy.

in all cases (e.g., well-behaved convergence properties should remain well behaved, new artifacts or other degradation must not be introduced, and so on). It follows that the essential properties of the methods described earlier (e.g., appropriateness of particular noise models, use or otherwise of particular a priori assumptions, etc.) hold in the multiple-resolution setting also. We have cited examples of further reading on these enhanced methods.

7. DECONVOLUTION AND RESOLUTION

In many cases, there is no sense in trying to deconvolve an image at the resolution of the pixel (especially when the PSF is very large). The idea to limit the resolution is relatively old, because it is already this concept that is used in the CLEAN algorithm (Högbom 1974). Indeed, the clean beam fixes the resolution in the final solution. This principle was also developed by Lannes & Roques (1987) in a different form. This concept was reinvented, first by Gull & Skilling (1991), who called the clean beam the *intrinsic correlation function* (ICF), and more recently by Magain et al. (1998) and Pijpers (1999).

The ICF is usually a Gaussian, but in some cases it may be useful to take another function. For example, if we want to compare two images I_1 and I_2, which are obtained with two wavelengths or with two different instruments, their PSFs P_1 and P_2 will certainly be different. The classic approach would be to deconvolve I_1 with P_2 and I_2 with P_1, so we are sure that both are at the same resolution. Unfortunately, however, we lose some resolution in doing this. Deconvolving both images is generally not possible because we can never be sure that both solutions O_1 and O_2 will have the same resolution.

A solution would be to deconvolve only the image that has the worse resolution (say, I_1) and to limit the deconvolution to the second image resolution (I_2). Then, we just have to take P_2 for the ICF. The deconvolution problem is to find $\tilde{O}$ (hidden solution) such that

$$I_1 = P_1 * P_2 * \tilde{O}, \tag{47}$$

and our real solution O_1 at the same resolution as I_2 is obtained by convolving $\tilde{O}$ with P_2; O_1 and I_2 can then be compared.

Introducing an ICF G in the deconvolution equation leads to just considering a new PSF P', which is the convolution of P and G. The deconvolution is carried out using P', and the solution must be reconvolved with G at the end. In this way, the solution has a constrained resolution, but aliasing may occur during the iterative process, and it is not sure that the artifacts will disappear after the reconvolution with G. Magain et al. (1998) proposed an innovative alternative to this problem by assuming that the PSF can be considered as the convolution product of two terms, the ICF G and an unknown S: $P = G * S$. Using S instead of P in the deconvolution process and a sufficiently large FWHM value for G implies that the Shannon sampling theorem (Shannon 1948) is never violated. But the problem is now to calculate S, knowing P and G, which is again a deconvolution problem. Unfortunately, this delicate point was not discussed in the original paper. Propagation of the error on the S estimation in the final solution has also until now not been investigated, even if this issue seems to be quite important.

8. SUPERRESOLUTION

8.1. Definition

Superresolution consists of recovering object spatial frequency information outside the spatial bandwidth of the image formation system. In other terms, frequency components where $\hat{P}(\nu) = 0$ have to be recovered. It has been demonstrated (Donoho et al. 1992) that this is possible under certain conditions. The observed object must be *nearly black*, i.e., nearly zero in all but a small fraction of samples. Denoting n as the number of samples, m as the number of nonzero values in the Fourier transform of the PSF, and $\epsilon = m/n$ as the incompleteness ratio, it has been shown that an image (Donoho et al. 1992):

1. Must admit superresolution if the object is $\frac{1}{2}\epsilon$-black.
2. Might admit superresolution if the object is ϵ-black. In this case, it depends on the noise level and the spacing of nonzero elements in the object. Well-spaced elements favor the possibility of superresolution.
3. Cannot admit superresolution if the object is not ϵ-black.

Near-blackness is both necessary and sufficient for superresolution. Astronomical images often give rise to such data sets, where the real information (stars and galaxies) is contained in very few pixels. If the $\frac{1}{2}\epsilon$-blackness of the object is not verified, a solution is to limit the Fourier domain Ω of the restored object. Several methods have been proposed in different contexts for achieving superresolution.

8.2. Gerchberg-Saxon-Papoulis Method

The Gerchberg-Saxon-Papoulis (Gerchberg 1974) method is iterative and uses the a priori information on the object, which is its positivity and its support in the spatial domain. It was developed for interferometric image reconstruction, where we want to recover the object O from some of its *visibilities*, i.e., some of its frequency components. Hence, the object is supposed to be known in a given Fourier domain Ω, and we need to recover the object outside this domain. The problem can also be seen as a deconvolution problem $I = P * O$, where

$$P(u, v) = \begin{cases} 1 & \text{if } (u, v) \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{48}$$

We denote $\mathcal{P}_{C_s}$ and $\mathcal{P}_{C_f}$ the projection operators in the spatial

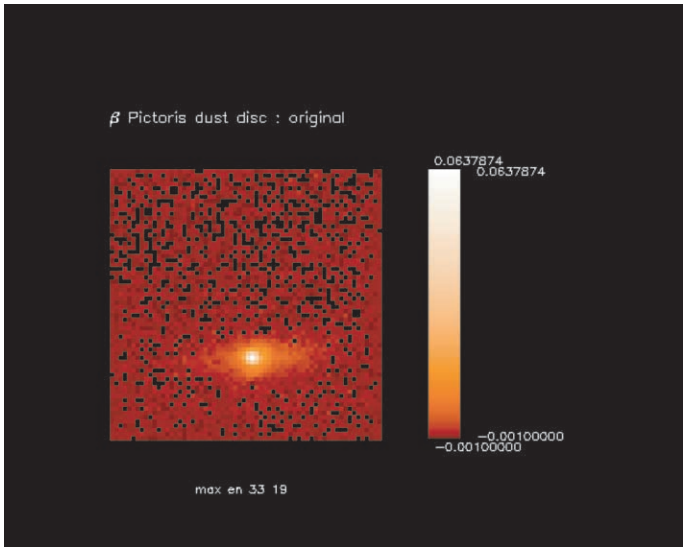

Fig. 9*a*

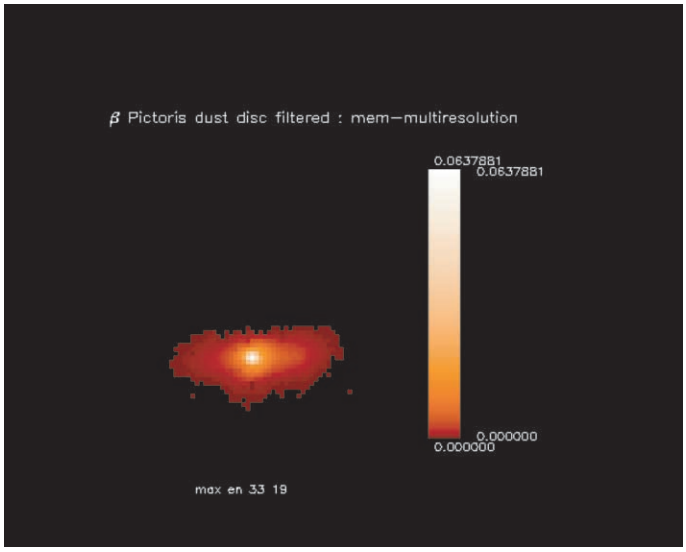

Fig. 9*b*

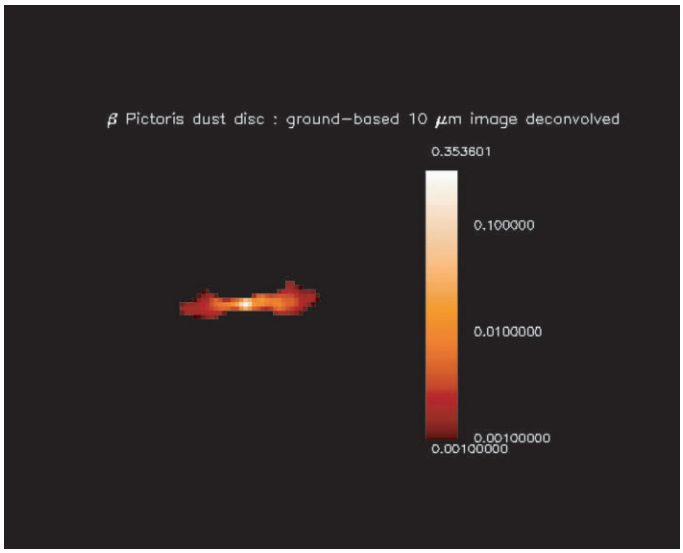

Fig. 9*c*

FIG. 9.—(*a*) β Pictoris raw data; (*b*) filtered image; (*c*) deconvolved image.

and the Fourier domain:

$$\mathcal{P}_{C_s}(X(x, y)) = \begin{cases} X(x, y) & \text{if } (x, y) \in \mathcal{D}, \\ 0 & \text{otherwise;} \end{cases} \tag{49}$$

$$\mathcal{P}_{C_f}(\hat{X}(u, v)) = \begin{cases} \hat{I}(u, v) = \hat{O}(u, v) & \text{if } (u, v) \in \Omega, \\ \hat{X}(u, v) & \text{otherwise.} \end{cases}$$

The projection operator $\mathcal{P}_{C_s}$ replaces by zero all pixel values that are not in the spatial support defined by $\mathcal{D}$, and $\mathcal{P}_{C_f}$ replaces all frequencies in the Fourier domain Ω by the frequencies of the object O. The Gerchberg algorithm is as follows:

1. Compute $\tilde{O}^0$ = inverse Fourier transform of $\hat{I}$, and set $i = 0$.
2. Compute $X_1 = \mathcal{P}_{C_s}(\tilde{O}^i)$.
3. Compute $\hat{X}_1$ = Fourier transform of X_1.
4. Compute $\hat{X}_2 = \mathcal{P}_{C_f}(\hat{X}_1)$.
5. Compute X_2 = inverse Fourier transform of $\hat{X}_2$.
6. Compute $\tilde{O}^{i+1} = \mathcal{P}_{C_s}(\hat{X}_2)$.
7. Set $X_1 = \tilde{O}^{i+1}$, $i = i + 1$, and go to 2.

The algorithm consists just of forcing iteratively the solution to be zero outside the spatial domain $\mathcal{D}$ and equal to the observed visibilities inside the Fourier domain Ω. It has been

shown that this algorithm can be derived from the Landweber method (Bertero & Boccacci 1998), and therefore its convergence and regularization properties are the same as for the Landweber method. It is straightforward to introduce the positivity constraint by replacing $\mathcal{P}_{C_s}$ by $\mathcal{P}^+_{C_s}$:

$$\mathcal{P}^+_{C_s}(X(x, y)) = \begin{cases} \max(X(x, y), 0) & \text{if } (x, y) \in \mathcal{D}, \\ 0 & \text{otherwise.} \end{cases} \quad (50)$$

The Gerchberg method can be generalized (Bertero & Boccacci 1998) using the Landweber iteration:

$$O^{n+1} = \mathcal{P}^+_{C_s}[O^n + \alpha(P^* * L - P^* * P * O^n)], \quad (51)$$

where $L = \mathcal{P}^+_{C_s}(I)$.

8.3. Deconvolution with Interpolation

The MAP Poisson algorithm, combined with an interpolation, can be used to achieve superresolution (Hunt 1994):

$$O^{n+1} = O^n \exp\left\{\left(\frac{I}{(P * O^n)_\downarrow} - 1\right)_\uparrow * P^*\right\}, \quad (52)$$

where up-arrow and down-arrow notation describes, respectively, the oversampling and downsampling operators. The PSF P must be sampled on the same grid as the object.

8.4. Undersampled PSF

When considering sampling, we should remember that detector pixels are "bins" for capturing flux and hence that there is inherent area integration. At all times, therefore, signal detection implies signal convolution. Passing from one spatial resolution to another (which is implicit, for example, when cross-correlating signals from different detectors) necessarily involves deconvolution, however this is actually achieved in practice.

We now look at the case where observations are made with an undersampled PSF. When the observation is repeated several times with a small shift between two measurements, we can reconstruct a deconvolved image on a smaller grid. We denote $D(i, j, k)$ the kth observation ($k = 1 \ldots, n$), $\Delta_{i,k}$, $\Delta_{j,k}$ the shift in both directions relative to the first frame, $\mathcal{L}_\uparrow$ the operator that co-adds all the frame on a smaller grid, and $\mathcal{L}^{-1}_\downarrow$ the operator that estimates D from $\mathcal{L}_\uparrow D$ using shifting and downsampling operations. The $\Delta_{i,k}$, $\Delta_{j,k}$ shifts are generally derived from the observations using correlation methods or a PSF fitting (if a star is in the field) but can also be the jitter information if the data are obtained from space. Note also that $\mathcal{L}^{-1}_\downarrow \mathcal{L}_\uparrow D \neq D$. The PSF P can generally be derived on a finer grid using a set of observations of a star or using an optical modeling of the instrument. The deconvolution iteration becomes

$$O^{n+1} = O^n + \alpha P^*[\mathcal{L}_\uparrow(D - \mathcal{L}^{-1}_\downarrow(P * O^n))] \quad (53)$$

and the positivity and spatial constraints can also be used:

$$O^{n+1} = \mathcal{P}^+_{C_s}\{O^n + \alpha P^*[\mathcal{L}_\uparrow(D - \mathcal{L}^{-1}_\downarrow(P * O^n))]\}. \quad (54)$$

The co-addition operator $\mathcal{L}_\uparrow$ can be implemented in different ways. All frames can first be interpolated to the finer grid size, shifted using an interpolation function, and then co-added.

"Dithering" or "jitter" have been terms applied to purposeful use of offsets in imaging (Hook & Fruchter 2000). An ad-hoc method called "drizzling" was developed by Hook & Fruchter (2000) and implemented in IRAF, based on mapping pixels to a finer grid and assuming knowledge of geometric distortion.

Dithering may be described as a way of recovering Nyquist sampling. Moving the source image to get new samples is easy to understand, particularly with the effective PSF concept, where the effective PSF combines the native optical PSF and the areal integral within the detector. In recent work, Gammaitoni et al. (1998) consider dithering as a special case of "stochastic resonance," which in general seeks to amplify weak signals by the assistance of small quantities of noise.

Lauer (1999) ignores geometric distortion and instead addresses the problem of aliasing resulting from combining undersampled images. A linear combination of Fourier transforms of the offset images is used, which mitigates aliasing artifacts in direct space.

8.5. Multiscale Support Constraint

The constraint operator $\mathcal{P}^+_{C_s}$ may not always be easy to determine, especially when the observed object is extended. Furthermore, if the object is very extended, the support will be very large and the support constraint may have very small influence on the final result. For this reason, it may be convenient to replace the support constraint by the multiresolution support constraint. The advantages are the following:

1. It can be automatically calculated from the noise modeling in the wavelet space.
2. Extended objects are generally also smooth. This means that the support on the small scales will be small, and therefore this introduces a smoothness constraint on extended objects and no constraint on pointlike objects.

9. CONCLUSIONS

We conclude with a short look at how multiscale methods used in deconvolution are evolving and maturing.

We have seen that the recent improvement in deconvolution methods has led to use of multiscale approaches. These can be summarized as follows:

1. Linear inverse filtering leading to wavelet-vaguelette decomposition.
2. CLEAN leading to wavelet-CLEAN.
3. Fixed-step gradient, Lucy, and van Cittert leading to regularization by the multiresolution support.
4. MEM leading to the multiscale entropy method.

The multiscale entropy method (Starck et al. 2001b), which generalized the wavelet-regularized iterative methods, allows us to separate the deconvolution problem into two separate problems: noise control from one side and solution smoothness control on the other side. The advantage of this approach is that noise control is better carried out in the image domain, while smoothness control can only be carried out in the object domain.

The success of wavelets is due to the fact that wavelet bases represent well a large class of signals. Other multiscale methods, such as the ridgelet or the curvelet transform (Candès & Donoho 1999, 2000; Donoho & Duncan 2000; Starck, Candès, & Donoho 2001a) will certainly play a role in the future.

An old view of astronomy in practice is that of the observational specialist, aided by one or more data-analysis specialists. This view is changing fast with the coming of astronomy "collaboratories" supported by middleware and Web services. Our review of deconvolution is addressed to the teams that are leading the evolution toward use of the Grid and "virtual observatory."

The virtual observatory in astronomy is premised on the fact that all usable astronomy data are digital. High-performance information cross-correlation and fusion, and long-term availability of information, are required. The term "virtual" in this context means the use of reduced or processed on-line data.

A second and closely associated development is that of the Grid. The computational Grid is to provide an algorithmic and processing infrastructure for the virtual science of the future. The data Grid is to allow ready access to information from our tera- and petabyte data stores. The information Grid is to provide active and dynamic retrieval of information, not just pointers to where information might or might not exist.

The evolution of the way we do science, driven by these themes, is inextricably linked to the problem areas and often recently developed algorithmic solutions surveyed in this article.

As just one location for further information on innovative developments in this broad area, see "Computational and Information Infrastructure in the Astronomical DataGrid" on the iAstro Web site.[2] This is a 4 year (from late 2001) European collaborative project.

We are grateful to an anonymous referee for various comments on an earlier version of this article.

[2] http://www.iAstro.org.

REFERENCES

Acar, R., & Vogel, C. 1994, Physica D, 10, 1217
Adorf, H., Hook, R., & Lucy, L. 1995, Int. J. Imaging Syst. Tech., 6, 339
Alloin, D., Pantin, E., Lagage, P. O., & Granato, G. L. 2000, A&A, 363, 926
Alvarez, L., Lions, P.-L., & Morel, J.-M. 1992, SIAM J. Numer. Anal., 29, 845
Bertero, M., & Boccacci, P. 1998, Introduction to Inverse Problems in Imaging (London: Inst. Phys.)
Blanc-Féraud, L., & Barlaud, M. 1996, Vistas Astron., 40, 531
Bontekoe, T., Koper, E., & Kester, D. 1994, A&A, 284, 1037
Buonanno, R., Buscema, G., Corsi, C., Ferraro, I., & Iannicola, G. 1983, A&A, 126, 278
Burg, J. 1978, in Modern Spectral Analysis, ed. D. G. Childers (New York: IEEE Press), 34
Burud, I., et al. 2002, A&A, 383, 71
Candès, E. & Donoho, D. 1999, Philos. Trans. R. Soc. London A, 357, 2495
———. 2000, Proc. SPIE, 4119, 1
Caulet, A., & Freudling, W. 1993, ST-ECF Newsl., 20, 5
Champagnat, F., Goussard, Y., & Idier, J. 1996, IEEE Trans. Image Process., 44, 2988
Charbonnier, P., Blanc-Féraud, L., Aubert, G., & Barlaud, M. 1997, IEEE Trans. Image Process., 6, 298
Christou, J. C., Bonnacini, D., Ageorges, N., & Marchis, F. 1999, Messenger, 97, 14
Corbard, T., Blanc-Féraud, L., Berthomieu, G., & Provost, J. 1999, A&A, 344, 696
Cornwell, T. 1989, in Diffraction-Limited Imaging with Very Large Telescopes, ed. D. Alloin & J. Mariotti (Dordrecht: Kluwer), 273
Courbin, F., Lidman, C., & Magain, P. 1998, A&A, 330, 57
Coustenis, A., et al. 2001, Icarus, 154, 501
Dempster, A., Laird, N., & Rubin, D. 1977, J. Royal Stat. Soc. B, 39, 1
Dixon, D., et al. 1996, A&AS, 120, 683
Djorgovski, S. 1983, J. Astrophys. Astron., 4, 271
Donoho, D. 1993, in Proc. Symp. Applied Mathematics 47, Different Perspectives on Wavelets, ed. I. Daubechies (Providence: Am. Math. Soc.), 173
———. 1995, Appl. Comput. Harmonic Anal., 2, 101
Donoho, D., & Duncan, M. 2000, Proc. SPIE, 4056, 12
Donoho, D., Johnson, I., Hoch, J., & Stern, A. 1992, J. Royal Stat. Soc. B, 54, 41
Faure, C., Courbin, F., Kneib, J. P., Alloin, D., Bolzonella, M., & Burud, I. 2002, A&A, 386, 69
Freudling, W., & Caulet, A. 1993, in Proc. 5th ESO/ST-ECF Data Analysis Workshop, ed. P. Grosbøl (Garching: ESO), 63
Frieden, B. 1978, Image Enhancement and Restoration (Berlin: Springer)
Galatsanos, N., & Katsaggelos, A. 1992, IEEE Trans. Image Process., 1, 322
Gammaitoni, L., Hänggi, P., Jung, P., & Marchesoni, F. 1998, Rev. Mod. Phys., 70, 223
Gerchberg, R. 1974, Opt. Acta, 21, 709

Golub, G., Heath, M., & Wahba, G. 1979, Technometrics, 21, 215
Gull, S., & Skilling, J. 1991, MEMSYS5 Quantified Maximum Entropy User's Manual (Suffolk: Maximum Entropy Data Consultants)
Hanisch, R. J. & White, R. L., eds. 1994, Proc. STScI Workshop, The Restoration of HST Images and Spectra II (Baltimore: STScI)
Högbom, J. 1974, A&AS, 15, 417
Hook, R. 1999, ST-ECF Newsl., 26, 3
Hook, R., & Fruchter, A. 2000, in ASP Conf. Ser. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 521
Hunt, B. 1994, Int. J. Mod. Phys. C, 5, 151
Issacson, E., & Keller, H. 1966, Analysis of Numerical Methods (New York: Wiley)
Kaaresen, K. 1997, IEEE Trans. Image Process., 45, 1173
Kalifa, J. 1999, Ph.D. thesis, Ecole Polytechnique
Kalifa, J., Mallat, S., & Rougé, B. 2000, IEEE Trans. Image Process., submitted
Katsaggelos, A. 1993, Digital Image Processing (Berlin: Springer)
Kempen, G., & van Vliet, L. 2000, J. Opt. Soc. Am. A, 17, 425
Landweber, L. 1951, Am. J. Math., 73, 615
Lannes, A., & Roques, S. 1987, J. Opt. Soc. Am., 4, 189
Lauer, T. 1999, PASP, 111, 227
Lucy, L. 1974, AJ, 79, 745
———. 1994, in The Restoration of HST Images and Spectra II, ed. R. J. Hanisch & R. L. White (Boston: STScI), 79
Magain, P., Courbin, F., & Sohy, S. 1998, ApJ, 494, 472
Marchis, R., Prangé, R., & Christou, J. 2000, Icarus, 148, 384
Moffat, A. 1969, A&A, 3, 455
Molina, R., & Cortijo, F. 1992, in Proc. International Conference on Pattern Recognition, ed. E. S. Gelsema & E. Backer (Vol. 3; Los Alamitos: IEEE IEEE Comp. Soc.), 147
Molina, R., Katsaggelos, A., Mateos, J., & Abad, J. 1996, Vistas Astron., 40, 539
Molina, R., Katsaggelos, A., Mateos, J., Hermoso, A., & Segall, A. 2000, Pattern Recognition, 33, 555
Molina, R., Núñez, J., Cortijo, F., & Mateos, J. 2001, IEEE Signal Process. Magazine, 18, 11
Molina, R., Ripley, B., Molina, A., Moreno, F., & Ortiz, J. 1992, AJ, 104, 1662
Murtagh, F., & Starck, J. 1994, ST-ECF Newsl., 21, 19
Murtagh, F., Starck, J., & Bijaoui, A. 1995, A&AS, 112, 179
Neelamani, R. 1999, M.S. thesis, Rice Univ.
Neelamani, R., Choi, H., & Baraniuk, R. G. 2001, IEEE Trans. Image Process., submitted
Núñez, J., & Llacer, J. 1998, A&AS, 131, 167
Pantin, E., & Starck, J. 1996, A&AS, 118, 575
Perona, P. & Malik, J. 1990, IEEE Trans. Pattern Analysis and Machine Intelligence, 12, 629
Pijpers, F. P. 1999, MNRAS, 307, 659
Pirzkal, N., Hook, R., & Lucy, L. 2000, in ASP Conf. Ser. 216, Astronomical Data Analysis Software and Systems IX, ed. N. Manset, C. Veillet, & D. Crabtree (San Francisco: ASP), 655
Puetter, R., & Yahil, A. 1999, in ASP Conf. Ser. 172, Astronomical Data Analysis Software and Systems VIII, ed. D. M. Mehringer, R. L. Plante, & D. A. Roberts (San Francisco: ASP), 307
Radomski, J. T., Piña, R. K., Packham, C., Telesco, C. M., & Tadhunter, C. N. 2002, ApJ, 566, 675
Richardson, W. 1972, J. Opt. Soc. Am., 62, 55
Ripley, B. 1981, Spatial Statistics (New York: Wiley)
Rudin, L., Osher, S., & Fatemi, E. 1992, Physica D, 60, 259
Shannon, C. 1948, Bell Syst. Tech. J., 27, 379
Shepp, L., & Vardi, Y. 1982, IEEE Trans. Medical Imaging, 2, 113
Starck, J., & Bijaoui, A. 1994, Signal Process., 35, 195
Starck, J., Bijaoui, A., Lopez, B., & Perrier, C. 1994, A&A, 283, 349
Starck, J., Bijaoui, A., & Murtagh, F. 1995, CVGIP: Graphical Models & Image Processing, 57, 420
Starck, J., Candès, E., & Donoho, D. 2001a, IEEE Trans. Image Process., 11, 670
Starck, J., & Murtagh, F. 1994, A&A, 288, 342
———. 1999, Signal Process., 76, 147
———. 2002, Astronomical Image and Data Analysis (Berlin: Springer), in press
Starck, J., Murtagh, F., & Bijaoui, A. 1998a, Image Processing and Data Analysis: The Multiscale Approach (Cambridge: Cambridge Univ. Press)
Starck, J., Murtagh, F., & Gastaud, R. 1998b, IEEE Trans. on Circuits and Systems II, 45, 1118
Starck, J., Murtagh, F., Querre, P., & Bonnarel, F. 2001b, A&A, 368, 730
Tikhonov, A., Goncharski, A., Stepanov, V., & Kochikov, I. 1987, Sov. Phys.–Doklady, 32, 456
van Cittert, P. 1931, Z. Phys., 69, 298
Wakker, B., & Schwarz, U. 1988, A&A, 200, 312
Weir, N. 1992, in ASP Conf. Ser. 25, Astronomical Data Analysis, Software, and Systems I, ed. D. Worral, C. Biemesderfer, & J. Barnes (San Francisco: ASP), 186
White, R. L., & Allen, R. J., eds. 1990, Proc. STScI Workshop, The Restoration of HST Images and Spectra (Baltimore: STScI)
Zavagno, A., Lagage, P. O., & Cabrit, S. 1999, A&A, 344, 499