Sparse priors in unions of bases for radiointerferometic image reconstruction

Mary, D.*, Bourguignon, S.**, Theys, C.*, Lanteri, H.*

Abstract

The image reconstruction problem posed by radiointerferometric data is a Fourier synthesis or a deconvolution problem. This paper proposes to handle this problem as a sparse approximation problem. The sparsity assumption is expressed in a union of bases. This allows firstly to account for the morphological diversity of astronomical images in a flexible way, and secondly to use fast transforms without which iterative reconstruction algorithms cannot be implemented on large images. Sparse solutions are obtained using greedy and global optimization algorithms, for which numerical comparisons are proposed. A particular attention is paid to the norms of the actual dictionary that is used in the reconstruction, which depends not only on the elaborated union of bases, but also on the noise covariance matrix and on the Fourier sampling. This analysis provides an interpretation of the regularisation and stopping parameters of the reconstruction algorithms in terms of false alarm rate. Numerical results show that this approach proposes relatively efficient though simple solutions to a recurrent problem of image reconstruction in Astronomy, which is the ability of the reconstruction algorithms to synthesize in the same time extended and local structures.

Keywords: image reconstruction, Fourier synthesis, image deconvolution, sparse priors, approximation theory, radiointerferometry, denoising

1. Signal model for interferometic image reconstruction

Imaging in Astronomy is not only possible with classical monopupil telescopes. The interferometric technique combines the light beams coming from several telescopes or antennas pointed toward the same celestial scene to create interference fringes. The light of two interfering beams contains information on the two-dimensional Fourier spectrum of the observed scene. The transfer function in the Fourier plane is given by the autocorrelation function of the 2-aperture pupil. As the Earth rotates, the position of the antennas or telescopes w.r.t. the sky changes, yielding new interference fringes that in turn produce new samples. This results in a collection of samples which provide a noisy, sparsely sampled Fourier spectrum of the observed celestial scene. This paper deals with methods aimed at recovering a two-dimensional image from such data.

A word on the notations before turning to the mathematical formulation of the signal model: vectors will be denoted by underlined lowercase letters, matrices by upper case letters. The ith entry of a vector $\underline{v}$ is denoted by v_i, and M_{ij} denotes the entry of the matrix M on its ith row and jth column. The symbols ^{t}, * and † denote respectively transpose, complex conjugate and complex conjugate transpose.

In this framework, a model of interferometric data acquisition can be written in its simplest form as the following underdetermined system (see e.g. [1, 2] for details and main assumptions of this model):

$$\underline{\widetilde{y}} = \widetilde{F}\underline{x} + \underline{\widetilde{n}}, \tag{1}$$

where $\underline{x} \in \mathbb{R}^{+N}$ is the image to be reconstructed, $\widetilde{F}$ is the $\frac{M}{2} \times N$ complex Fourier transform matrix restricted to the set of probed frequencies (which are assumed to belong to the regularly spaced frequency grid produced by the Discrete

*UMR 6525 CNRS H. FIZEAU, UNS, OCA, Campus Valrose, F-06108 Nice Cedex 2

**UMR 6202 Cassiopée, CNRS, Observatoire de la Côte d'Azur, Boulevard de l'Observatoire, F-06304 Nice Cedex 4.

Email addresses: David.Mary@unice.fr (Mary, D.), Sebastien.Bourguignon@oca.eu (Bourguignon, S.), Celine.Theys@unice.fr (Theys, C.), Henri.Lanteri@unice.fr (Lanteri, H.)

Fourier Transform, allowing to use the FFT algorithm), $\widetilde{\underline{y}} \in \mathbb{C}^{\frac{M}{2}}$ are the $\frac{M}{2}$ available Fourier samples, and $\widetilde{\underline{n}} \in \mathbb{C}^{\frac{M}{2}}$ is the noise. The complex data vector $\widetilde{\underline{y}}$ actually contains M degrees of freedom in its real and imaginary parts $\widetilde{\underline{y}}_R$ and $\widetilde{\underline{y}}_I$. In order to deal with real values we will write the data as $\underline{y} = [\widetilde{\underline{y}}_R^t \; \widetilde{\underline{y}}_I^t]^t$, and the model equation (1) becomes

$$\underline{y} = F\underline{x} + \underline{n}, \tag{2}$$

where the $M \times N$ real matrix F contains the cosine and sine parts of $\widetilde{F}$, and $\underline{y} \in \mathbb{R}^M$. Because the antennas are different one from another and may suffer from correlated perturbations (e.g. temperature variations), the noise affecting different Fourier samples may not be the same in power, and may be correlated. In this paper the noise $\underline{n} \in \mathbb{R}^M$ is assumed to be Gaussian with covariance matrix C: $\underline{y} \sim \mathcal{N}(\underline{0}, C)$. Since $N > M$ there is indeed an infinity of solutions in general to eq. (2), even in the absence of noise. The problem is therefore two-folded: the first difficulty comes from the missing data, and the second difficulty from the noise perturbations.

2. Sparse priors and reconstruction algorithms

2.1. Introduction and relation to previous works

One way to counteract these difficulties is to introduce some prior knowledge. A sparse[1] (synthesis) prior assumes that $\underline{x}$ is composed from a few (i.e. less than M, and much less than N here) salient features. This assumption is widely used in data modeling for denoising, compression, pattern recognition or inpainting applications for instance, because natural signals and images tend to be sparse in appropriate spaces [3]. The sparsity is expressed via representation spaces (orthonormal transforms corresponding to bases, or more generally redundant dictionaries), whose atoms (the columns of the dictionary) correspond to geometrical features that are likely to compose the unknown image. A large variety of such representations has been elaborated in the image processing literature, e.g., impulsions basis indeed (corresponding to point-like structures), Discrete Cosine Transform (DCT, 2-D plane waves), wavelets (localized patterns in time and frequency), curvelets (elongated and curved patterns), shapelets, etc... The choice of a representation space is made with respect to a class of images (astronomical images in our case). This choice also crucially depends on the existence of fast operators, without which iterative algorithms cannot be implemented in reasonable time, especially in radiointerferometry where the number of Fourier samples can be of the order of hundreds of thousands, and N much higher.

Let us note here that the now famous and numerous results of Compressed Sensing issued from [4, 5] heavily rely on the sparsity assumption and on approximation algorithms such as those used below. However, real astronomical images are not strictly sparse in general. For this reason, we will not intend here to verify the theoretical results of CS related to the partial Fourier matrix, in terms for example of unicity or of perfect recovery of a sparse support.

In order to be able to model complex images with various features, one way is to concatenate a set of K representation bases or sub-dictionaries $\{B_i\}_{i=1\dots K}$ into a large dictionary D of $T > N$ vectors [6, 7]. Of course, redundant dictionaries can be obtained in many other ways; we will here focus on union of bases for which a fast transform is associated to each basis, so that the analysis and the synthesis of images using the whole dictionary can be performed with fast transforms. The coefficients vectors $\underline{u}_1, \dots \underline{u}_K$ associated to each basis can be collected in a vector $\underline{u} = [\underline{u}_1^t \dots \underline{u}_K^t]^t$. The sparsity assumption of $\underline{x}$ can then be formalized as

$$\underline{x} \approx D\underline{u} \;\text{ with }\; D = [B_1 \; B_2 \dots B_K], \;\text{ and }\; \underline{u} = [\underline{u}_1^t \dots \underline{u}_K^t]^t \in \mathbb{R}^T \;\text{ is sparse.} \tag{3}$$

The model (2) becomes

$$\underline{y} = FD\underline{u} + \underline{n}, \quad \text{with } \underline{u} \text{ sparse.} \tag{4}$$

In this setting the interferometric image reconstruction problem can be seen as a sparse approximation problem: select a few coefficients $\hat{u}_i$ that best explain (in the sense of some fidelity criterion) the data $\underline{y}$.
Sparsity was already used in [2] for radiointerferometric image reconstruction with wavelets dictionaries. In contrast to these works, the present paper intends to illustrate the advantages that redundant dictionaries using union of bases

[1] A vector is sparse if many of its entries are 0.

may present for astronomical images. The emphasis is also on statistical considerations related to the reconstruction algorithms which, athough quite standard, must be adapted to correctly account for the effects of the noise structure and of the Fourier sampling.

Comprehensive surveys of the literature about image reconstruction methods for radiointerferometric data can be found in [8] (including methods from model fitting to non parametric deconvolution) and in [1], which emphasizes the problem of reconstructing both extended and point sources. On this last topic, the literature is rather sparse: in [9] and [10], the point-like sources are written in a parametric manner based on amplitudes and peaks positions. The extended morphology is accounted for in [9] with a Tikhonov regularisation, and in [10] with a Gaussian correlation. In [1], a global criterion is minimized which is composed of the l_1 norm of the point-like component plus the l_2 norm of the gradient of the extended component, and a positivity constraint is imposed. As described above, the approach followed here is different.

Two approaches are mainly used in practice to solve the approximation problems posed by (4). The first approach is based on the explicit minimization of a cost function including a data fidelity term and a regularization (or penalization) term. Within this approach, the data fidelity term and the regularization functions inducing sparse solution vectors $\underline{u}$ may find interesting interpretations in terms of Bayesian Maximum A Posteriori (MAP) estimation. The second approach decreases a cost function including only the data fidelity term, but in a greedy manner. In this second approach, sparsity is achieved by stopping the reconstruction after having iteratively estimated only the most active components of $\underline{u}$.

2.2. Sparse priors in the Maximum A Posteriori (MAP) framework

The MAP framework puts a probabilistic setting on $\underline{u}$. The MAP estimate maximizes on $\underline{u}$ the probability of $\underline{u}$ given the data. By the Bayes theorem, this probability is proportional to the product of the likelihood $\mathcal{L}(\underline{y};\underline{u})$ by the prior density $p(\underline{u})$:

$$\begin{aligned}
\widehat{\underline{u}}^{\mathrm{MAP}} &= \arg\max_{\underline{u}} p(\underline{u}|\underline{y}), \text{ with } p(\underline{u}|\underline{y}) \propto \mathcal{L}(\underline{y};\underline{u})p(\underline{u}) \\
&= \arg\min_{\underline{u}} \underbrace{-\log \mathcal{L}(\underline{y};\underline{u})}_{J\mathrm{data}_{(\underline{u},\underline{y})}} \underbrace{-\log p(\underline{u})}_{J\mathrm{prior}_{(\underline{u})}} \\
&= \arg\min_{\underline{u}} J^{\mathrm{data}}(\underline{u},\underline{y}) + J^{\mathrm{prior}}(\underline{u}).
\end{aligned}$$

The data attachment term $J^{\mathrm{data}}(\underline{u},\underline{y})$ is the neg-loglikelihood and is related density of the noise in eq. (2) by

$$J^{\mathrm{data}}(\underline{u},\underline{y}) = -\log \exp^{-\frac{1}{2}(\underline{y}-FD\underline{u})^t C^{-1}(\underline{y}-FD\underline{u})} = \frac{1}{2}\|\underline{y}-FD\underline{u}\|_C^2 = \frac{1}{2}\|\underbrace{C^{-\frac{1}{2}}\underline{y}}_{\underline{z}} - \underbrace{C^{-\frac{1}{2}}FD}_{D_\nu}\underline{u}\|^2.$$

The data attachment term therefore relates the Euclidean norm of the error between the whitened data $\underline{z} = C^{-\frac{1}{2}}\underline{y}$, and the data synthesized by means of $\underline{u}$ and of the "transformed dictionary" $D_\nu = C^{-\frac{1}{2}}FD$. The model with whitened data can consequently be rewritten as

$$\underbrace{C^{-\frac{1}{2}}\underline{y}}_{\underline{z}} = \underbrace{C^{-\frac{1}{2}}FD}_{D_\nu}\underline{u} + \underbrace{C^{-\frac{1}{2}}\underline{n}}_{\underline{\epsilon}}, \text{ with } \underline{\epsilon} = C^{-\frac{1}{2}}\underline{n} \sim \mathcal{N}(\underline{0}, I). \tag{5}$$

Since the likelihood term suggests to work with the whitened data, the image reconstruction approach will the following:

- Choose K sparsifying bases $\{B_i\}_{i=1\ldots K}$ and set $D = [B_1\ B_2 \ldots B_K]$.
- Find a sparse $\underline{\hat{u}} = [\underline{\hat{u}}_1^t \ldots \underline{\hat{u}}_K^t]^t$ (either by means of a sparsity inducing prior, or greedily) that well approximates (in the sense of the l_2 norm) the whitened data $\underline{z}$.

- Synthesize the reconstructed image as

$$\underline{\hat{x}} = B_1\underline{\hat{u}}_1 + \ldots + \underline{\hat{u}}_K B_K = D\underline{\hat{u}}. \tag{6}$$

Note that eq. (6) naturally provides the separated components $\{B_i\underline{\hat{u}}_i\}$, as previously noticed in [11].
In the MAP framework, the sparse $\underline{\hat{u}}$ is found by solving

$$\arg\min_{\underline{u}} J = \arg\min_{\underline{u}} \left(J^{\text{data}} + J^{\text{prior}} \right) = \arg\min_{\underline{u}} \left(\frac{1}{2}\|\underline{z} - D_v\underline{u}\|^2 + J^{\text{prior}} \right), \tag{7}$$

where J^{prior} is chosen as a sparsity inducing function. There are many such functions. The l^0 "norm"[2] of $\underline{u}$ best expresses the sparsity as it corresponds to the number of nonzeros coefficients u_i. This penalization function corresponds to hard thresholding when the dictionary is an orthonormal basis. For a redundant dictionary, it leads to a generally untractable combinatorial optimization. The l^1 norm ($= \sum_i |u_i|$) is often seen as a convex approximation to l^0. As in this case the problem (7) involves then two convex terms, it is convex, and can be solved using well known convex optimization techniques. The $l1$ norm corresponds to a soft threshloding function for orthonormal bases, and to a Laplace prior for $\underline{u}$ in the MAP framework. The l^p "norms", with $0 < p < 1$ ($= \sum_i |u_i|^p$) are also sometimes used. A lower p induces a more sparse solution, but the resulting J^{prior} is not convex. Such penalizations lead to non analytical thresholding functions in general, and correspond to generalized Gaussian priors on $\underline{u}$ [12].

In the following we will focus on an l_1 penalization, which from eq. (7) entails finding

$$\underline{\hat{u}} = \arg\min_{\underline{u}} J = \arg\min_{\underline{u}} \left(\frac{1}{2}\|\underline{z} - D_v\underline{u}\|^2 + \lambda\|\underline{u}\|_1 \right), \tag{8}$$

where $\lambda \in \mathbb{R}^+$. This problem has the general form of the Basis Pursuit Denoising (BPDN) of [6]. An important issue is the choice of the parameter λ, which balances the respective weights of data fidelity and sparsity. As pointed out in [13], an interesting interpretation of λ can be drawn from the first order Kuhn-Tucker conditions [14, 15] for problem (8). If $\underline{\hat{u}}$ denotes the minimizer of (8), then there exists a real vector $\underline{h}$ such that

$$D_v^t(D_v\underline{\hat{u}} - \underline{z}) + \lambda\underline{h} = \underline{0}, \tag{9}$$

with $h_i = \text{sign}(\hat{u}_i)$ if $\hat{u}_i \neq 0$ (NSC1), and $|h_i| \leq 1$ if $\hat{u}_i = 0$ (NSC2). NSC2 implies that if $\hat{u}_i = 0$ the ith component of $\underline{h}$ satisfies

$$|h_i| \leq 1 \text{ and } h_i = \left(\frac{D_v^t(\underline{z} - D_v\underline{\hat{u}})}{\lambda}\right)_i \Leftrightarrow |(D_v^t(\underline{z} - D_v\underline{\hat{u}}))_i| \leq \lambda. \tag{10}$$

This means that the final component $\hat{u}_i$ will be 0 (i.e. not detected) if the ith component of the residual $(\underline{z} - D_v\underline{\hat{u}})$ retroprojected in the transform coefficients space by D_v^t, has a magnitude less than λ. The parameter λ can consequently be seen as a detection threshold for $\hat{u}_i$. Let us assume that the residual $(\underline{z} - D_v\underline{\hat{u}})$ contains only the noise of model (5). Then the ith component of the retroprojected residual will be Gaussian with variance $(D_v^tD_v)_{ii} = \|\underline{D_v}_i\|^2$. If one wants the false alarm rate to be equal for all components, for example to ρ times the standard deviation, the threshold λ should be component dependent and taken as $\lambda_i = \rho\|\underline{D_v}_i\|$. The problem of choosing a priori the ρ that will yield the best (in the sense of some distance measure) reconstructed image is not easy because the optimum value depends on the considered image. As in denoising however, values of ρ around 3 or 4 yield good results (see the simulations). Note that the norms of the transform dictionary atoms (i.e. the columns of $C^{-\frac{1}{2}}FD$) depend on the noise correlation, on the sampling function in the Fourier plane and on the dictionary. Consequently the atoms $\{\underline{D_v}_i\}$ generally have very different norms. These norms must be computed beforehand. As a practical consideration, it is important to note that the transformed dictionnary cannot be normalized once for all before the reconstruction : in practice, D_v is not stored because this would require a huge matrix. Instead, fast transforms are used. To compute the atoms' norms, it is necessary to compute, beforehand, each atom using fast transforms.

The solution of (8) can be found by several algorithms, among which the Iterative Soft Thresholding Algorithm [16]. This algorithm can be written in our case:

[2]The l^p norm of $\underline{u}$ is $\|\underline{u}\|_p = (\sum_i |u_i|^p)^{\frac{1}{p}}$ for $p \neq 0$. By convention $\|\underline{u}\|_0 = \text{supp}(\underline{u})$.

- *Computation of the norms* of the dictionary atoms $\{\|\underline{D_{\nu}}_i\|\}$.
- *Initialisation*: $k = 0$. Choose $\underline{\hat{u}}^0 = \underline{0}$ and the value of ρ.
- While the conditions NSC1 and NSC2 are not satisfied :
  - *Gradient step*: $\underline{\hat{u}}'^k = \underline{\hat{u}}^k + \beta D_{\nu}^t(\underline{z} - D_{\nu}\underline{\hat{u}}^k)$, where β is a sufficiently small stepsize.
  - *Soft thresholding*: $\hat{u}_i^{k+1} = 0$ if $\hat{u}_i'^k \leq \lambda_i$, or $\hat{u}_i^{k+1} = \hat{u}_i'^k - \lambda_i \ \mathrm{sgn}(\hat{u}_i'^k)$ otherwise , with $\lambda_i = \rho\|\underline{D_{\nu}}_i\|$.
  - *Criterion stop*: If the conditions NSC1 and NSC2 are satisfied up to a numerical tolerance, set $\underline{\hat{u}} = \underline{\hat{u}}^{k+1}$.
- *Image synthesis*: Compute $\underline{\hat{x}} = D\underline{\hat{u}}$.

Many accelerated versions of this algorithms exist, e.g. FISTA, GPSR, SPARSA, GPAS [17].
A back projection may be performed in order to improve the estimation of the nonzero synthesis coefficients $\{\hat{u}_i\}$, which are biased [15], and may be reestimated by minimizing $\|\underline{y} - \sum_{i:\ \hat{u}_i \neq 0} \hat{u}_i \underline{D_{\nu}}_i\|^2$. The improvement is not systematic, however, because the reestimation may give inappropriate weights to the selected components to enforce agreement with the data. In the simulations below, we show reestimated amplitudes when they improve the reconstruction.

2.3. Greedy approach via the Matching Pursuit (MP)

The second approach to find a sparse estimate $\hat{u}_i$ is to minimize $J^{\mathrm{data}} = \frac{1}{2}\|\underline{z} - D_{\nu}\underline{u}\|^2$ greedily. This yields sparse solutions by construction. The MP algorithm [18] performs such a minimization, by iteratively removing the contribution of one atom at a time : the atom which is the most correlated with the data residual $\underline{r}$. As above, a special attention must be paid to the fact that the atoms of the transformed dictionary have different norms. This can be handled as follows : let us assume that at iteration k the whitened data residual $\underline{r}^k$ is composed only from noise plus one signal component $\underline{D_{\nu}}_s$ with weight u_s:

$$\underline{r}^k = \underline{D_{\nu}}_s u_s + \underline{\epsilon}. \tag{11}$$

Finding the maximum likelihood estimate of the signal component leads to minimize the quantity $\frac{1}{2}\|\underline{r}^k - \underline{D_{\nu}}_i u_i\|^2$ in both the component index i and the weight u_i. For a given component $\underline{D_{\nu}}_i$, it is easily seen that the optimal weight is

$$\hat{u}_i = \arg\min_{u_i} \frac{1}{2}\|\underline{r}^k - \underline{D_{\nu}}_i u_i\|^2 = \frac{< \underline{D_{\nu}}_i, \underline{r}^k >}{\|\underline{D_{\nu}}_i\|^2}. \tag{12}$$

For this $\hat{u}_i$, the optimal index is then

$$m = \arg\min_i \frac{1}{2}\|\underline{z} - \hat{u}_i \underline{D_{\nu}}_i\|^2 = \arg\max_i \frac{| < \underline{D_{\nu}}, \underline{r}^k_i > |}{\|\underline{D_{\nu}}_i\|}. \tag{13}$$

The MP can be interpreted as an iterative maximum likelihood algorithm, where the model (11) is assumed at each iteration, until $\underline{r}^k$ reaches the noise level. If in the end the residual $\underline{r}^k$ contains only $\underline{\epsilon}$ (i.e. unit-variance uncorrelated Gaussian noise), the normalized correlations $\{\frac{<\underline{r}^k, \underline{D_{\nu}}_i>}{\|\underline{D_{\nu}}_i\|}\}$ are also unit-variance and Gaussian. Comparing these quantities to a threshold μ therefore yields an equal detection threshold for all components. Note that $\mu_i = \mu \ \|\underline{D_{\nu}}_i\|$ plays the roles of the detection threshold for the correlation $< \underline{r}^k, \underline{D_{\nu}}_i >$, in a way very similar to the λ_i of BPDN (this similarity was noticed in [19]). This analysis, which shows that the different atoms norms affect the rule of selection of the components and their weights, leads to the following MP reconstruction algorithm:

- *Computation of the norms* of the dictionary atoms $\{\|\underline{D_{\nu}}_i\|\}$.
- *Initialisation*: $k = 0$, $\underline{\hat{u}}^0 = \underline{0}$ and $\underline{r}^0 = \underline{z} = C^{-\frac{1}{2}}\underline{y}$. Choose a threshold μ.
- While some correlations $\{\frac{|<\underline{r}^k, \underline{D_{\nu}}_i>|}{\|\underline{D_{\nu}}_i\|}\}$ are above μ:

- *Best match*: Find $m = \arg\max_i \frac{|<\underline{r}^m, \underline{D}_{v_i}>|}{\|\underline{D}_{v_i}\|}$, with $\frac{|<\underline{r}^m, \underline{D}_{v_m}>|}{\|\underline{D}_{v_m}\|} > \mu$.
- *Update*: $\underline{r}^{k+1} = \underline{r}^k - \frac{<\underline{r}^k, \underline{D}_{v_m}>}{\|\underline{D}_{v_m}\|^2}\underline{D}_{v_m}$ and $\hat{u}_m^{k+1} = \hat{u}_m^k + \frac{<\underline{r}^k, \underline{D}_{v_m}>}{\|\underline{D}_{v_m}\|^2}$.
- *Criterion stop*: Set $\underline{\hat{u}} = \underline{\hat{u}}^{k+1}$ if all $\{\frac{|<\underline{r}^{k+1}, \underline{D}_{v_i}>|}{\|\underline{D}_{v_i}\|}\}$ are below μ.

- *Image synthesis*: Compute $\underline{\hat{x}} = D\underline{\hat{u}}$.

As for BPDN a back projection may be performed before the synthesis of the final image.

2.4. A note on the CLEAN algorithm

A reference in radiointerferometric image reconstruction remains the CLEAN deconvolution algorithm. CLEAN was proposed by [20], and was related to earlier methods [21, 22, 23] by [24] . See also [25] for its impact on Astronomy and beyond.

CLEAN resembles the MP above, but works in the image space. In order to work with real images made from incomplete Fourier samples, each sample $\widetilde{y}_i$ corresponding to a frequency (u_i, v_i) in the (u, v) plane can be mirrored to correspond to the frequency $(-u_i, -v_i)$ by setting the corresponding sample to $\widetilde{y}_i^*$. Since the resulting Fourier data possess the Hermitian symmetry, the corresponding image is real. One can gather all the resulting data in a complex vector of $2M$ entries $\underline{\widetilde{y}}_{2M} = [\underline{\widetilde{y}}^t \; \underline{\widetilde{y}}^\dagger]^t$, and the acquisition model (1) can be written as

$$\underline{\widetilde{y}}_{2M} = \widetilde{F}_{2M}\underline{x} + \underline{\widetilde{n}}_{2M}, \tag{14}$$

where $\widetilde{F}_{2M}$ is the $2M \times N$ complex Fourier transform restricted to the symmetric frequencies, and $\widetilde{n}_{2M}$ is the noise vector mirrored in the same way as $\underline{\widetilde{y}}$. The "dirty image" is $\widetilde{F}_{2M}^\dagger \underline{\widetilde{y}}_{2M} = \widetilde{F}_{2M}^\dagger \widetilde{F}_{2M}\underline{x} + \widetilde{F}_{2M}^\dagger \underline{\widetilde{n}}_{2M}$. Ignoring the noise, the CLEAN algorithm sees the dirty image as a linear combination of the columns of the circulant convolution matrix $\widetilde{F}_{2M}^\dagger \widetilde{F}_{2M}$, which acts in the image space as a dictionary of shifted PSFs ("dirty beams"). CLEAN iteratively selects the location of the maxima in the residual dirty map and subtracts from it (a fraction γ of) the contribution of the shifted PSF at the selected location. By selecting the most correlated PSF to the dirty map, CLEAN decreases greedily the norm $\|\widetilde{F}_{2M}^\dagger \underline{\widetilde{y}}_{2M} - \widetilde{F}_{2M}^\dagger \widetilde{F}_{2M}\underline{x}\|^2$, which is equivalent to $\frac{1}{2}\|\underline{\widetilde{y}}_{2M} - \widetilde{F}_{2M}\underline{x}\|^2$ because $\widetilde{F}_{2M}\widetilde{F}_{2M}^\dagger = I_{2M}$. Thus, from the point of view of the Fourier space, CLEAN aims to fit the available Fourier samples by means of a transformed dictionary of impulsions $\widetilde{F}_{2M} I_N$. CLEAN stops when the maximum of the residual map falls below a threshold κ. Note that a loop gain similar to that of CLEAN can indeed be introduced in the MP algorithm.

The difference between CLEAN and the MP working with a transform dictionary of impulsions $D_v = FI_N$ is first that CLEAN consider only positive correlations in the image space : this naturally imposes a positivity constraint. On the other hand, CLEAN in its original form above is limited to a dictionary of impulsions and tends consequently to yield spiky images even in case of diffuse sources - unless the result is further reconvolved, but then the resolution is uniformly decreased[3].

3. Numerical results

The reconstructed image $\underline{\hat{x}}$ can be evaluated qualitatively by comparing visually $\underline{\hat{x}}$ to $\underline{x}$. For a quantitative measure, we use here the Signal to Noise Ratio (SNR), defined as $\mathrm{SNR}(\underline{x}, \underline{\hat{x}}) = 10\log_{10}\frac{\sum_i x_i^2}{\sum_i (x_i - \hat{x}_i)^2}$, in dB. The higher SNR, the better reconstruction.

The first results regard a comparison between the global optimization BPDN for finding a sparse solution, and the greedy (local) approaches of CLEAN and MP. The image is chosen as a field of 100 randomly placed point sources, for which CLEAN works best. The dictionary used for MP and BPDN is indeed $D = I_N$. The test image has $N = 4096$ pixels, and the number of real constraints is $M = 851$. The (u, v) plane is sampled randomly according to a Gaussian

[3] Note that in [20] a matrix W can be applied to $\underline{\widetilde{y}}_{2M}$ to weight the Fourier samples.

density. The SNR for the data, SNR($\underline{y}, \underline{y} + \underline{n}$) equals 15.6 dB. The top row of figures in Fig. (1) shows respectively the original image $\underline{x}$, the sampling function and the dirty map (SNR($\underline{x}, \widetilde{F}^{\dagger}_{2M}\widetilde{\underline{y}}_{2M}$) = 1.18 dB). The bottom row of figures shows respectively the SNR for MP, CLEAN (both with a loop gain of $\gamma = 0.01$) and BPDN, as a function of their respective thresholding parameters μ, κ and ρ. For ISTA, convergence is declared when NSC1 is strictly satisfied, and NSC2 is satisfied up to a numerical tolerance of 10^{-14}. At best, CLEAN yields better SNR than MP, because of the positivity constraint, but less than BDPN. For the images of best SNR (Fig. 2), the number of nonzero pixels is 460 for MP, 358 for CLEAN and 123 for BDPN. Additional components (the correct support is 100) can be interpreted as false detections. In this sense, the global optimisation of eq. (8) is more robust as greedy approaches. Note also that the SNR of the reconstructed image can be higher than the SNR in the data, showing that the denoising capability of these algorithms may remain (MP and BPDN are often used in denoising applications), despite the additional underdetermination of the reconstruction problem.

Figures (3) and (4) present simulations of an interferometric system on two images chosen for their morphologic complexity. The transfer function (i.e. the autocorrelation function of the pupils, top row of (3), middle) yields $M = 6327$, and $N = 512^2 = 262144$. There is no noise here because we wish to focus on the effect of the missing Fourier samples on the reconstruction. The MP, using here a dictionary of impulsions and of sym6-wavelets, and CLEAN were stopped close to their maximum SNR, which for MP corresponded to $\mu \approx 2$ and for CLEAN to $\kappa \approx \max(\widetilde{F}^{\dagger}_{2M}\widetilde{\underline{y}}_{2M})/100$. Both algorithms used $\gamma = 0.1$.

Figure (3) shows that CLEAN yields spiky or clotted images, even when the result is reconvolved with a synthetic PSF of size comparable to the central part (main lobe) of the original PSF (SNR=4.4 dB). The best SNR for the CLEAN result is obtained by reconvolving the image by a large kernel ($21\times$ 21 pixels^2) in order to smooth enough the image (SNR=7.9 dB), but then all sharp features are lost in this smoothing. Of course, the optimum size of the convolution kernel to be used by CLEAN is not known in practice as it is image dependent. In contrast, the reconstructed image obtained with a MP using a union of bases (impulsions and wavelets) includes both spiky and extended structures (SNR=8.4 dB, bottom row, left). This image is only reconvolved by the central part of the PSF (of size 9×9 pixels^2) and no further improvement is obtained by further smoothing because diffuse structures have already been restituted. Some artifacts appear however, because the positivity of the reconstructed image is not imposed by the algorithm. Note that from eq. (6), the reconstructed image on the lower left corner is the *exact sum* of the last two images : the diffuse sources (bottom row, middle) and point-like sources (bottom row, right).

Figure (4) shows another such comparison on a less extended object. The aliasing (which is somewhat paradoxically more visible than on the previous example, because the object only partially overlaps) is fairly removed by both algorithms. Note however that the bright zone of the original image at position (310,270), which creates strong aliases in the dirty map, and not fully removed around (50,270) and (450,270) for the MP. Comparing the results of CLEAN and of MP with impulsions and wavelets, the same conclusions as above can be drawn: for CLEAN (SNR=15.7 dB at best), the whole image is either homogeneously spiky, or homogeneously smooth when reconvolved with a kernel sufficently large to recover diffuse sources. On the other hand, a dictionary with complementary features allows to restitute both structures in the same time (SNR=15.6 dB). Finally, the last image (bottom row, right) shows that although the positivity constraint is not imposed by MP, a simple thresholding of the MP image above zero brings some qualitative improvement (though the SNR of 15.7 dB is not increased much).

4. Conclusions

This paper showed that the image reconstruction problems that arises in interferometry can be treated as a sparse approximation problem. The union of bases used to express the sparsity allows to use fast transforms and to account naturally for the morphological diversity of astronomical images. Sparse solutions were obtained using greedy and global optimization algorithms, for which a numerical comparison was proposed. A particular attention was paid to the norms of the atoms of the actual dictionary that is used in the reconstruction, which depends not only on the chosen union of bases, but also on the noise covariance matrix and on the Fourier sampling. We showed that these norms must be accounted for in both BPDN and MP, and gave an interpretation of the regularisation parameters in terms of a false alarm rate. Numerical results showed that the proposed approach allows to deal with a recurrent problem of image reconstruction in Astronomy, which is the ability of the reconstruction algorithm to reconstruct in the same time extended and local structures. Further improvements are expected by imposing the positivity constraint in the

MP and BPDN algorithms, and by refining the dictionary to account for the morphological specifities of particular classes of astronomical images.

References

[1] J. Giovannelli, A. Coulais, Positive deconvolution for superimposed extended source and point sources, Astronomy and Astrophysics 439 (2005) 401–412.

[2] Y. Wiaux, L. Jacques, G. Puy, A. M. M. Scaife, P. Vandergheynst, Compressed sensing imaging techniques for radio interferometry, MNRAS 395 (2009) 1733–1742.

[3] S. Mallat, A wavelet tour of signal processing : the sparse way, Academic Press, 2008.

[4] D. L. Donoho, Compressed sensing, IEEE Transactions on Information Theory 52 (2006) 1289–1306.

[5] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory 52 (2006) 489–509.

[6] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM Journal of Scientific Computing 20 (1998) 33–61.

[7] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Transactions on Information Theory 49 (2003) 3320–3325.

[8] J. Starck, E. Pantin, F. Murtagh, PASP 114 (2002) 1051.

[9] P. Magain, F. Courbin, S. Sohy, Astrophysical Journal 494 (1998).

[10] N. Pirzkal, R. Hook, L. Lucy, In ASP Conference Series, Astronomical Data Analysis, Software and Systems IX, Paris, ed. N. Manset, C. Veuillet, D. Crabtree 216 (2000).

[11] J. Bobin, J. luc Starck, J. M. Fadili, Y. Moudden, D. L. Donoho, Morphological component analysis: An adaptive thresholding strategy, IEEE Transactions on Image Processing 16 (2007) 2675–2681.

[12] P. Moulin, J. Liu, Analysis of multiresolution image denoising schemes using generalized-gaussian and complexity priors, IEEE Transactions on Information Theory 45 (1998) 909–919.

[13] S. Bourguignon, D. Mary, E. Slezak, Sparsity-based denoising of hyperspectral astrophysical data with colored noise. Application to the MUSE instrument, in: IEEE Whispers Conference Conference Series.

[14] S. Alliney, S. Ruzinski, An algorithm for the minimization of mixed l1 and l2 norms with application to bayesian estimation, IEEE Transactions on Signal Processing 42 (1994) 618–627.

[15] J. Fuchs, On sparse representations in arbitrary redundant bases, IEEE Transactions Information Theory 50(6) (2004) 1341–1344.

[16] I. Daubechies, M. Defrise, C. De Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Communications on Pure and Applied Math. 57 (2004).

[17] Conference on Advanced Inverse Problems, Vienna, 2009, http://www.ricam.oeaw.ac.at/conferences/aip2009/minisymposia/.

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

[19] V. Solo, A modified CLEAN algorithm does l1-denoising, in: Proceedings ICASSP, 2008.

[20] J. A. Högbom, Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines, AAPS 15 (1974) 417–426.

[21] R. Southwell, Stress-calculation in frameworks by the method of systematic relaxation of constraints, Proceedings of the Royal Academical Society 151 (1935) 56–95.

[22] G. Temple, The general theory of relaxation methods applied to linear systems, Proceedings of the Royal Academical Society 169 (1939) 476–500.

[23] G. Forsythe, W. Wasow, Finite difference methods for partial differential equations, J. Wiley New York, 1960.

[24] U. Schwarz, Mathematical-statistical description of the iterative beam removing technique (method CLEAN), Astronomy and Astrophysics 65 (1978) 417–426.

[25] T. Cornwell, Hogbom's CLEAN algorithm. Impact on astronomy and beyond, Astronomy and Astrophysics, Special Issue 500 (2009) 65–66.

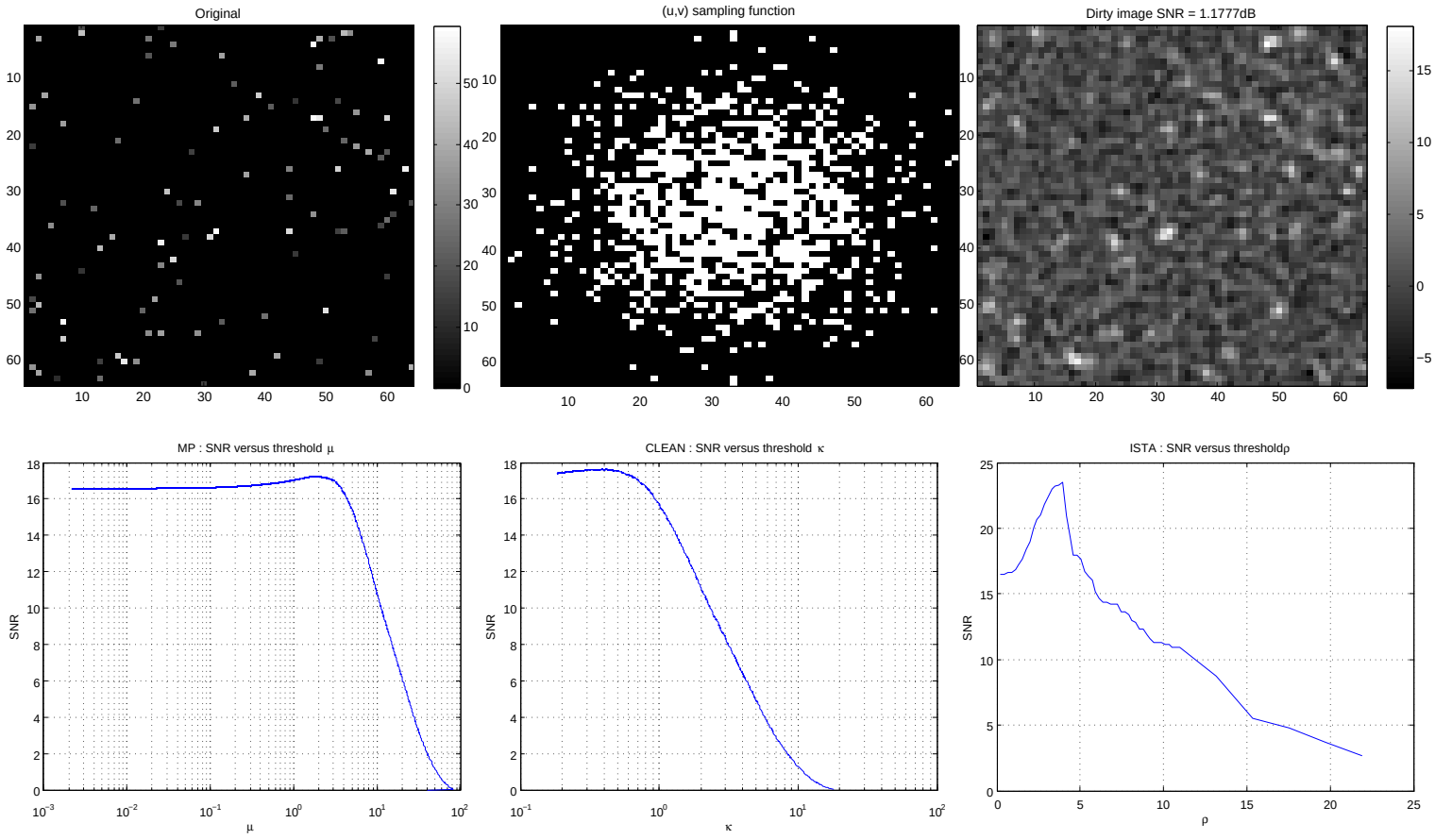

Figure 1: Top row, from left ro right : Original image (100 point sources), sampling function in the Fourier plane, dirty image. Bottom row : SNR as a function of the threshold for MP with impulsions basis (left), for CLEAN (middle) and for BDPN (right). The best SNR are respectively 17.25 dB (MP), 17.64 dB (CLEAN) and 23.49 dB (BDPN) and the corresponding images and errors are shown in Fig. (2).

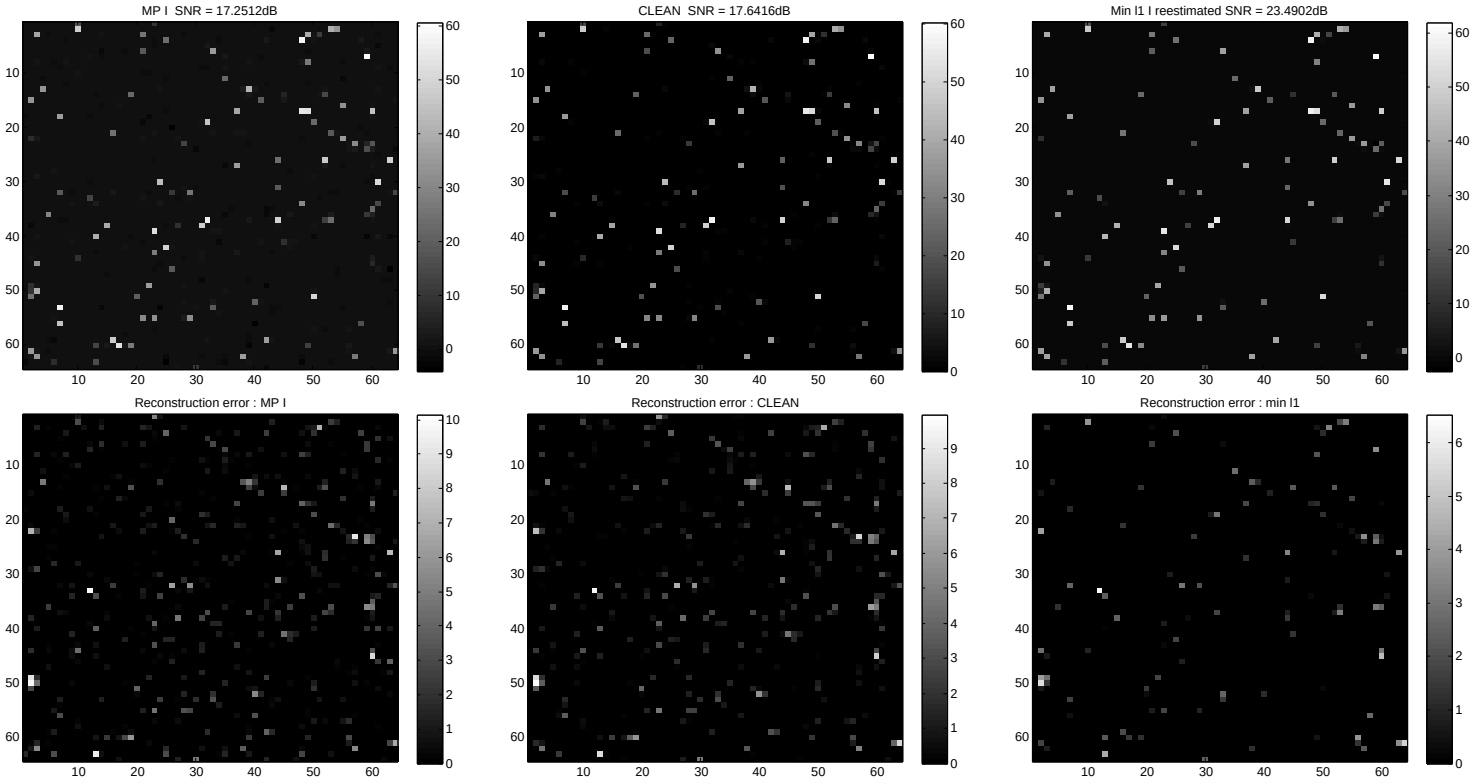

Figure 2: Top row, from left ro right : Best reconstructed images using MP with impulsions basis (left), for CLEAN (middle) and for BDPN (right). Bottom row : corresponding absolute values of the difference images (original minus reconstructed) for MP with impulsions basis (left), CLEAN (middle) and BDPN (right). The number of nonzeros components for the reconstructed images are 460 for MP, 358 for CLEAN and 123 for BDPN.

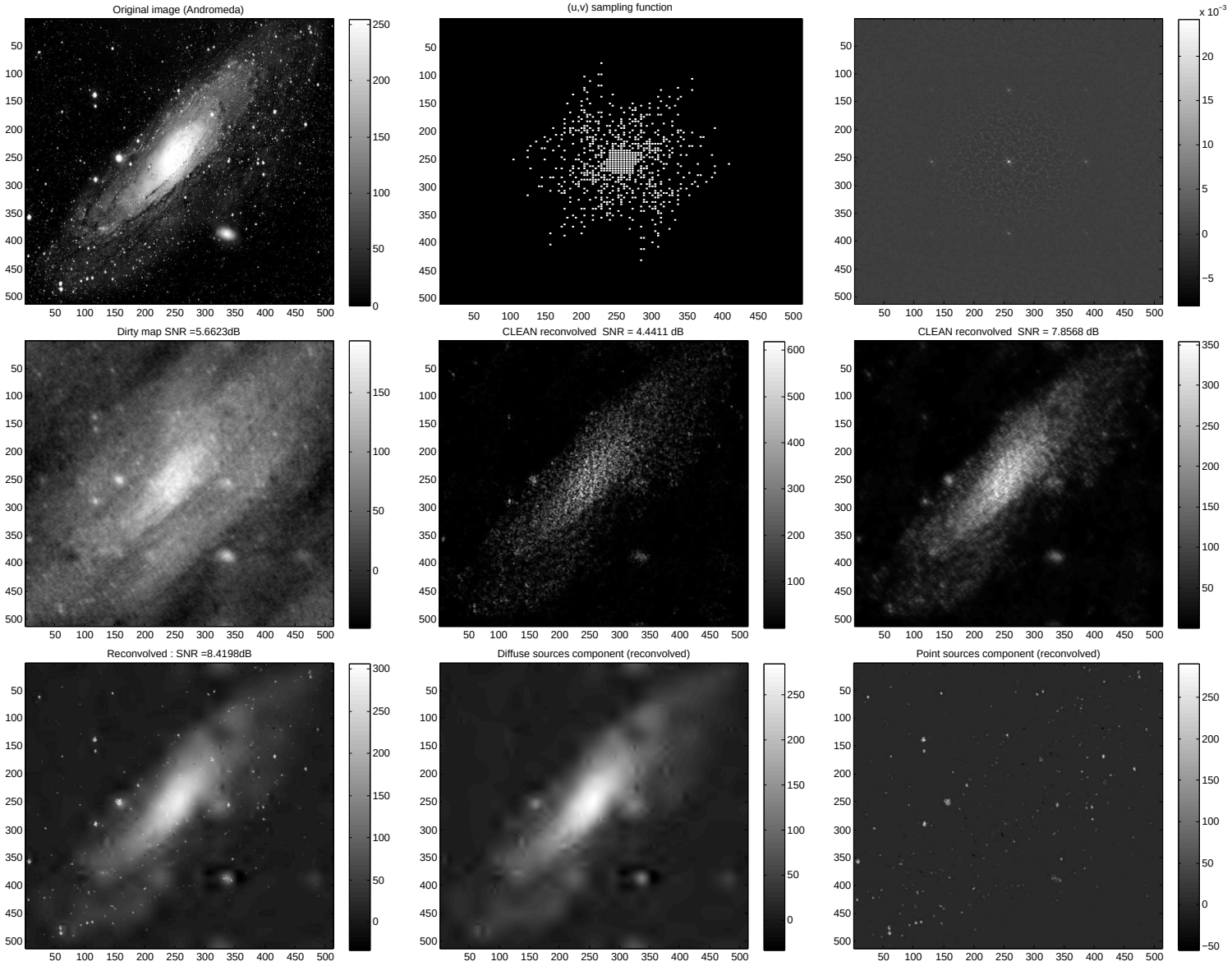

Figure 3: Top row, from left ro right : Original image, (u, v) sampling and corresponding PSF. Middle row : Dirty image (left), CLEAN reconstruction after reconvolving by the central lobe of the PSF (a beam of 9×9 pixels), and after reconvolving by a larger beam (21×21). Bottom row, left: image reconstructed using the MP with a dictionary of impulsions and of sym6-wavelets. This image is the sum of diffuse sources (middle) and compact sources (right).

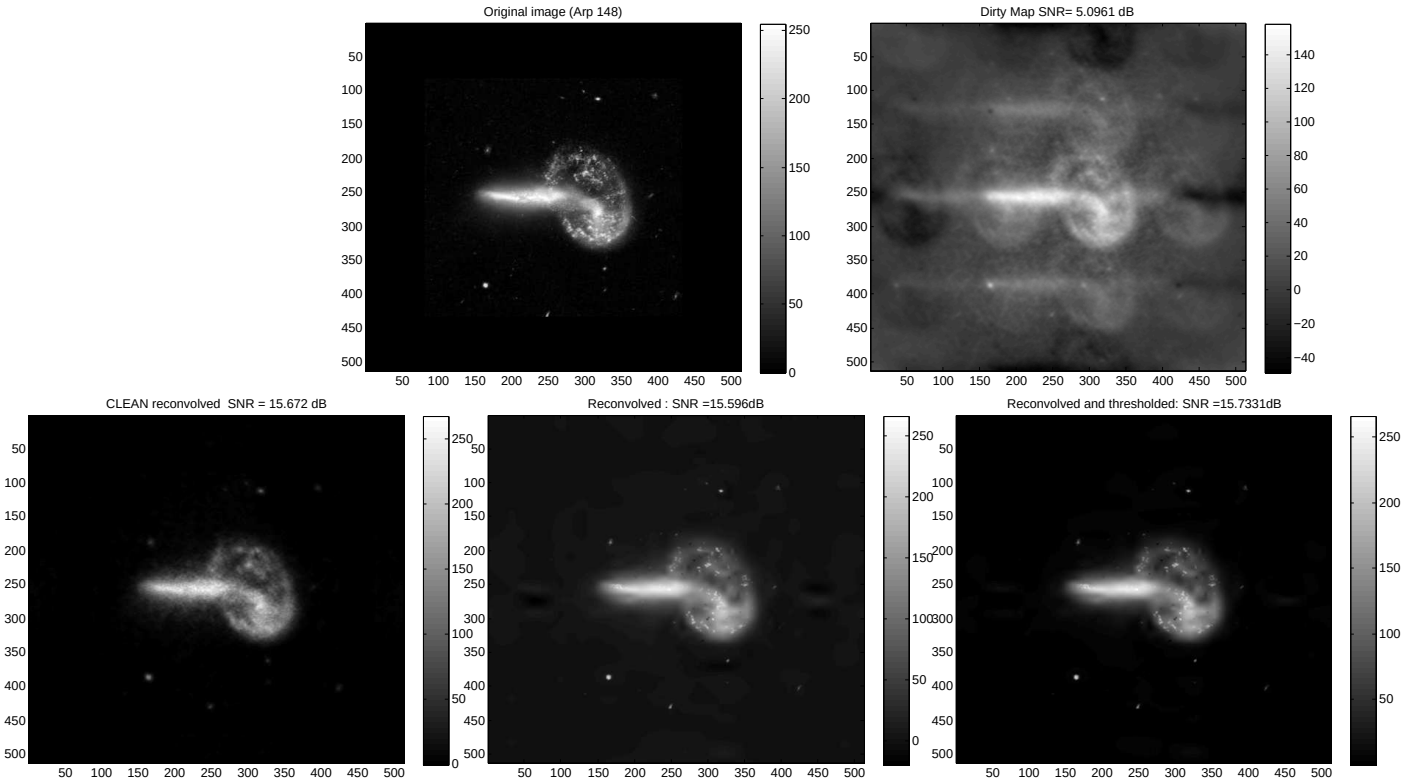

Figure 4: Top row, left : Original image. Right : dirty map showing a clear aliasing, corresponding to the important, regularly spaced sidelobes visible the PSF. Bottom row, left : best result obtained by CLEAN after reconvolving by a beam of size 13×13 pixels). Middle : left image reconstructed using the MP with a dictionary of impulsions and of sym6-wavelets, and reconvolved by the central lobe of the PSF. Right : Same as the middle image, but with all negative values set to 0.