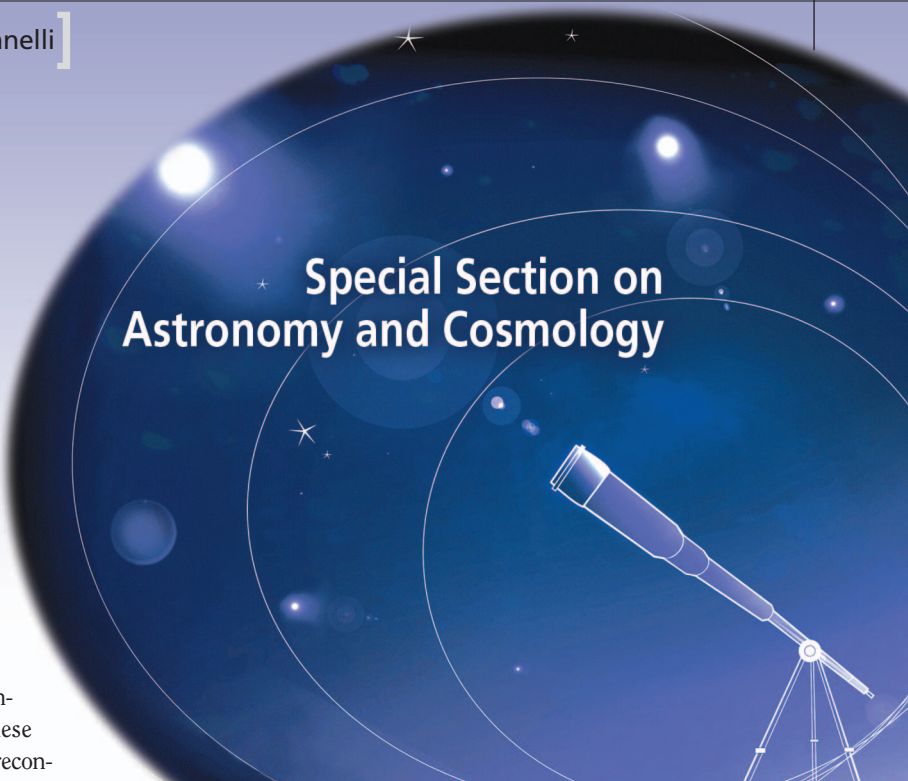


Since the first multitelescope optical interferometer [1], considerable technological improvements have been achieved. Optical (visible/infrared) interferometers are now widely open to the astronomical community and provide the means to obtain unique information from observed objects at very high angular resolution (submilliarcsecond). There are numerous astrophysical applications, such as stellar surfaces, environment of premain sequence or evolved stars, and central regions of active galaxies. See [2]–[4] for comprehensive reviews about optical interferometry and recent astrophysical results. As interferometers do not directly provide images, reconstruction methods are needed to fully exploit these instruments. This article aims at reviewing image reconstruction algorithms in astronomical interferometry using a general framework to formally describe and compare the different methods. The challenging issues in image reconstruction from interferometric data are introduced in the general framework of inverse problem approach. This framework is then used to describe existing image reconstruction algorithms in radio interferometry and the new methods specifically developed for optical interferometry.

Multitelescope interferometers provide sparse measurements of the Fourier transform of the brightness distribution of the observed objects (cf. the section “Interferometric Data”). Hence the first problem in image reconstruction from interferometric data is to cope with voids in the sampled spatial frequencies. This can be tackled in the framework of inverse problem approach (cf. the section “Imaging from Sparse Fourier Data”). At optical wavelengths, addi-

Special Section on Astronomy and Cosmology



© PHOTODISC

tional problems arise due to the missing of part of Fourier phase information, and to the nonlinearity of the direct model. These issues had led to the development of specific algorithms which can also be formally described in the same general framework (cf. the section “Image Reconstruction from Nonlinear Data”).

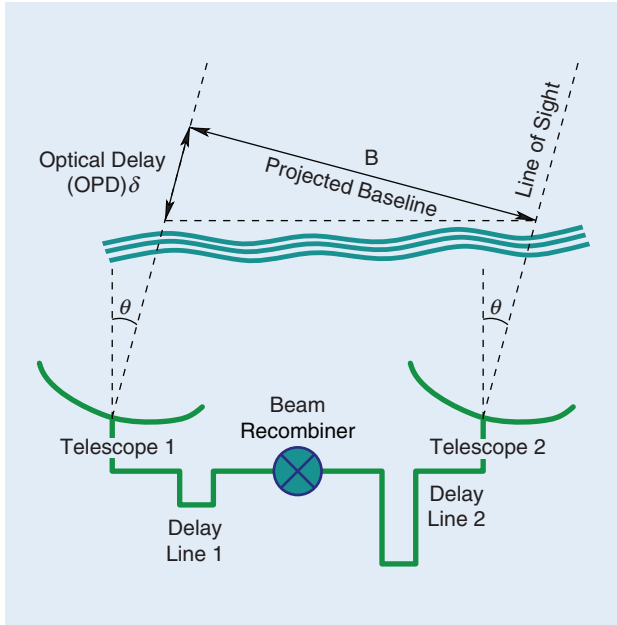
INTERFEROMETRIC DATA

The instantaneous output of an optical interferometer is the so-called complex visibility $V_{j_1, j_2}(t)$ of the fringes given by the interferences of the monochromatic light from the j_1 th and the j_2 th telescopes at instant t [3]

$$V_{j_1, j_2}(t) = g_{j_1}(t)^* g_{j_2}(t) \hat{I}(v_{j_1, j_2}(t)), \quad (1)$$

Image Reconstruction in Optical Interferometry

Using a general framework to formally
describe and compare different methods



[FIG1] Geometrical layout of an interferometer where B is the projected baseline, θ is the view angle, and δ is the geometrical optical path difference that is compensated by the delay lines.

where $\hat{I}(\nu)$ is the Fourier transform of $I(\theta)$, the brightness distribution of the observed object in angular direction θ , $g_j(t)$ is the complex amplitude throughput for the light from the j th telescope and $\nu_{j_1, j_2}(t)$ is the spatial frequency sampled by the pair of telescopes (j_1, j_2) (see Figure 1)

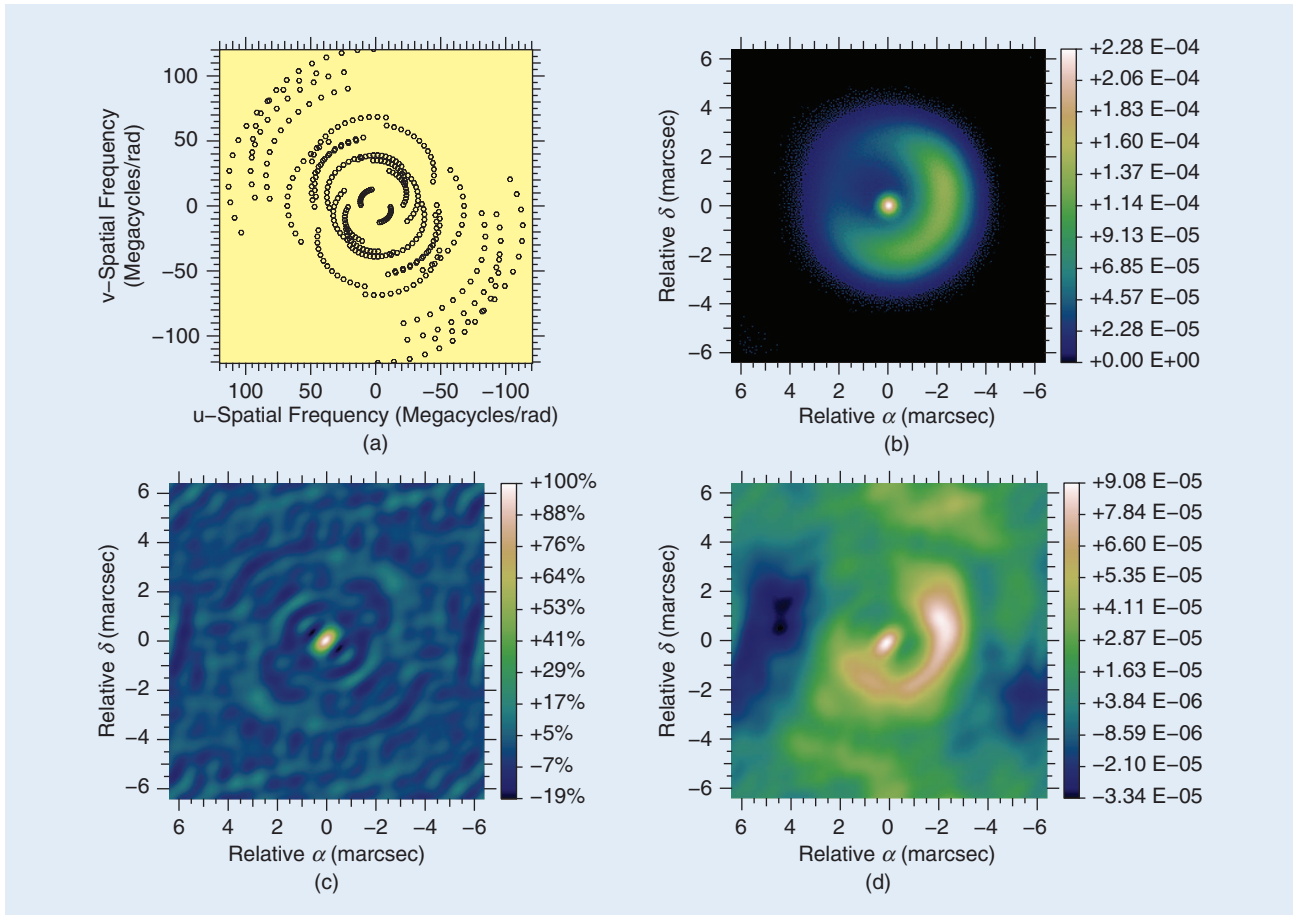
$$\nu_{j_1, j_2}(t) = \frac{r_{j_2}(t) - r_{j_1}(t)}{\lambda} \quad (2)$$

with λ the wavelength and $r_j(t)$ the projected position of the j th telescope on a plane perpendicular to the line of sight. These equations assume that the diameters of the telescopes are much smaller than their projected separation and that the object is an incoherent light source. An interferometer therefore provides sparse measurements of the Fourier transform of the brightness distribution of the observed object. Figure 2(a) shows an example of the sampling of spatial frequencies by an interferometer.

In practice, the complex visibility is measured during a finite exposure duration

$$V_{j_1, j_2, m}^{\text{data}} = \langle V_{j_1, j_2}(t) \rangle_m + V_{j_1, j_2, m}^{\text{err}} \quad (3)$$

where $\langle \rangle_m$ denotes averaging during the m th exposure and $V_{j_1, j_2, m}^{\text{err}}$ stands for the errors due to noise and modeling



[FIG2] (a) (u, v) coverage, (b) observed object, (c) dirty beam, and (d) dirty image. Object model and (u, v) coverage are from the "2004 Beauty Contest" [16].

approximations. The exposure duration is short enough to consider the projected baseline $r_{j_2}(t) - r_{j_1}(t)$ as constant, thus

$$\langle V_{j_1,j_2}(t) \rangle_m \simeq G_{j_1,j_2,m} \hat{I}(\nu_{j_1,j_2,m}) \quad (4)$$

with $\nu_{j_1,j_2,m} = \langle \nu_{j_1,j_2}(t) \rangle_m \simeq \nu_{j_1,j_2}(t_m)$, $t_m = \langle t \rangle_m$ the mean exposure time and $G_{j_1,j_2,m} = \langle g_{j_1}(t)^* g_{j_2}(t) \rangle_m$ the effective optical transfer function (OTF). The fast variations of the instantaneous OTF are mainly due to the random optical path differences (OPD) caused by the atmospheric turbulence. In long baseline interferometry, two telescopes are separated by more than the outer scale of the turbulence, hence their OPDs are independent. Furthermore, the exposure duration is much longer than the evolution time of the turbulence (a few tens of milliseconds) and averaging can be approximated by expectation

$$\langle g_{j_1}(t)^* g_{j_2}(t) \rangle_m \simeq E\{g_{j_1}(t)^*\}_m E\{g_{j_2}(t)\}_m$$

with $E\{\}_m$ the expectation during the m th exposure. During this exposure, the phase of $g_j(t)$ is $\phi_j(t) = \phi_{j,m} + (2\pi/\lambda) \delta_j(t)$ with $\phi_{j,m} = \langle \phi_j(t) \rangle_m$ the static phase aberration and $\delta_j(t) \sim \mathcal{N}(0, \sigma_\delta^2)$ the OPD which is a zero-mean Gaussian variable with the same standard deviation for all telescopes [5]. For a given telescope, the amplitude and phase of the complex throughput can be assumed independent, hence

$$E\{g_j(t)\}_m \simeq E\{|g_j(t)|\}_m E\{e^{i\phi_j(t)}\}_m \simeq g_{j,m} e^{-\frac{1}{2}\sigma_\phi^2}$$

with $g_{j,m} = |g_j(t_m)| \exp(-i\phi_{j,m})$ and $\sigma_\phi^2 = (2\pi/\lambda)^2 \sigma_\delta^2$ the variance of the phase during an exposure. The OTF is finally

$$G_{j_1,j_2,m} = \langle g_{j_1}(t)^* g_{j_2}(t) \rangle_m \simeq g_{j_1,m}^* g_{j_2,m} e^{-\sigma_\phi^2}. \quad (5)$$

At long wavelengths (radio), the phase variation during each exposure is small, hence $G_{j_1,j_2,m} \simeq g_{j_1,m}^* g_{j_2,m} \neq 0$. If some means to calibrate the $g_{j,m}$ s are available, then image reconstruction amounts to deconvolution (cf. the section “Imaging from Sparse Fourier Data”); otherwise, self-calibration has been developed to jointly estimate the OTF and the brightness distribution of the object given the measured complex visibilities.

At short wavelengths (optical), the phase variance exceeds a few squared radians and $G_{j_1,j_2,m} \simeq 0$, hence the object's complex visibility cannot be directly measured. A first solution would be to compensate for the OPD errors in real time using fast delay lines. This solution, however, requires a bright reference source in the vicinity of the observed object and dedicated instrumentation [6] that is currently in development and not yet available. An alternative solution consists in integrating nonlinear estimators that are insensitive to telescope-wise phase errors. This requires high acquisition rates (about 1,000 Hz in the near infrared) and involves special data processing but otherwise no special instrumentation.

To overcome loss in visibility transmission due to fast varying OPD errors, current optical interferometers integrate the power spectrum (for $j_1 \neq j_2$)

$$S_{j_1,j_2,m} = \langle |V_{j_1,j_2}(t)|^2 \rangle_m \simeq \rho_{j_1,m} \rho_{j_2,m} |\hat{I}(\nu_{j_1,j_2,m})|^2 \quad (6)$$

with $\rho_{j,m} = \langle |g_j(t)|^2 \rangle_m$ the mean squared modulus of the complex throughput of the j th telescope during the m th exposure. By construction, the $\rho_{j,m}$ s are insensitive to the phase errors and so is the power spectrum. Unlike that of the complex visibility, the transfer function $\rho_{j_1,m} \rho_{j_2,m}$ of the power spectrum is not negligible. This transfer function can be estimated by simultaneous photometric calibration and, to compensate for remaining static effects, from the power spectrum of a reference source (a so-called calibrator). Hence the object power spectrum $|\hat{I}(\nu_{j_1,j_2,m})|^2$ can be measured by $S_{j_1,j_2,m}$ in spite of phase errors due to the turbulence.

To obtain Fourier phase information (which is not provided by the power spectrum), the bispectrum of the complex visibilities is measured

$$\begin{aligned} B_{j_1,j_2,j_3,m} &= \langle V_{j_1,j_2}(t) V_{j_2,j_3}(t) V_{j_3,j_1}(t) \rangle_m \\ &\simeq \rho_{j_1,m} \rho_{j_2,m} \rho_{j_3,m} \\ &\quad \times \hat{I}(\nu_{j_1,j_2,m}) \hat{I}(\nu_{j_2,j_3,m}) \hat{I}(\nu_{j_3,j_1,m}), \end{aligned} \quad (7)$$

where j_1 , j_2 and j_3 denote three different telescopes. As for the power spectrum, the transfer function $\rho_{j_1,m} \rho_{j_2,m} \rho_{j_3,m}$ of the bispectrum can be calibrated. Since this transfer function is real, it has no effect on the phase of the bispectrum (the so-called phase closure) that is equal to that of the object

$$\begin{aligned} \beta_{j_1,j_2,j_3,m} &\equiv \arg(B_{j_1,j_2,j_3,m}) \\ &= \arg(\hat{I}(\nu_{j_1,j_2,m}) \hat{I}(\nu_{j_2,j_3,m}) \hat{I}(\nu_{j_3,j_1,m})). \end{aligned} \quad (8)$$

However, some phase information is missing. Indeed, from all the interferences between T telescopes (in a nonredundant configuration), $T(T-1)/2$ different spatial frequencies are sampled but the phase closure only yields $(T-1)(T-2)/2$ linearly independent phase estimates [3]. The deficiency of phase information is most critical for a small number of telescopes. Whatever the number of telescopes is, at least the information of absolute position of the observed object is lost.

In practice, obtaining the power spectrum and the bispectrum involves measuring the instantaneous complex visibilities (that is, for a very short integration time compared to the evolution of the turbulence) and averaging their power spectrum and bispectrum over the effective exposure time. Being nonlinear functions of noisy variables, these quantities are biased but the biases are easy to remove [7], [8]. To simplify the description of the algorithms, we will consider that the debiased and calibrated power spectrum and bispectrum are available as input data for image reconstruction, thus

$$S_{j_1,j_2,m}^{\text{data}} = |\hat{I}(\nu_{j_1,j_2,m})|^2 + S_{j_1,j_2,m}^{\text{err}}, \quad (9)$$

$$\begin{aligned} B_{j_1,j_2,j_3,m}^{\text{data}} &= \hat{I}(\nu_{j_1,j_2,m}) \hat{I}(\nu_{j_2,j_3,m}) \hat{I}(\nu_{j_3,j_1,m}) \\ &\quad + B_{j_1,j_2,j_3,m}^{\text{err}} \end{aligned} \quad (10)$$

where $S_{j_1,j_2,m}^{\text{err}}$ and $B_{j_1,j_2,j_3,m}^{\text{err}}$ are zero-mean terms that account for noise and model errors. Instead of the complex bispectrum data, we may consider the phase closure data

$$\beta_{j_1, j_2, j_3, m}^{\text{data}} = \arccos(\varphi(\nu_{j_1, j_2, m}) + \varphi(\nu_{j_2, j_3, m}) + \varphi(\nu_{j_3, j_1, m}) + \beta_{j_1, j_2, j_3, m}^{\text{err}}), \quad (11)$$

where $\varphi(\nu) = \arg(\hat{I}(\nu))$ is the Fourier phase of the object brightness distribution, $\arccos(\cdot)$ wraps its argument in the range $(-\pi, +\pi]$ and $\beta_{j_1, j_2, j_3, m}^{\text{err}}$ denotes the errors.

IMAGING FROM SPARSE FOURIER DATA

We consider here the simplest problem of image reconstruction given sparse Fourier coefficients (the complex visibilities) and first assuming that the OTF has been calibrated.

DATA AND IMAGE MODELS

To simplify the notation, we introduce the data vector $y \in \mathbb{C}^L$ that collates all the measurements $y_\ell = V_{j_1, j_2, m}^{\text{data}}$ with $\ell \sim (j_1, j_2, m)$ to denote a one-to-one mapping between index ℓ and triplet (j_1, j_2, m) . Long baseline interferometers provide data for a limited set $\mathcal{L} = \{\nu_k\}_{k=1, \dots, K}$ of observed spatial frequencies. For each ν_k , there is a nonempty set \mathcal{B}_k of telescope pairs and exposures such that

$$(j_1, j_2, m) \in \mathcal{B}_k \Leftrightarrow r_{j_2, m} - r_{j_1, m} = \lambda \nu_k$$

or equivalently

$$\mathcal{B}_k \stackrel{\text{def}}{=} \{(j_1, j_2, m) \in \mathcal{A}^2 \times \mathbb{E}; r_{j_2, m} - r_{j_1, m} = \lambda \nu_k\} \quad (12)$$

with \mathcal{A} and \mathcal{E} the sets of apertures (telescopes or antennae) and exposure indexes, and $r_{j, m} = \langle r_j(t) \rangle_m$ the mean position of the j th telescope during the m th exposure. Introducing \mathcal{B}_k and the set \mathcal{L} of observed frequencies is a simple way to account for all possible cases (with or without redundancies, multiple data sets, or observations from different interferometers). Note that, if every spatial frequency is only observed once, then $L = K$ and we can use $\ell = k$.

The image is a parametrized representation of the object brightness distribution. A very general description is given by a linear expansion

$$I(\theta) = \sum_{n=1}^N x_n b_n(\theta) \xrightarrow{\text{F.T.}} \hat{I}(\nu) = \sum_{n=1}^N x_n \hat{b}_n(\nu), \quad (13)$$

where $\{b_n(\theta)\}_{n=1, \dots, N}$ are basis functions and $x \in \mathbb{R}^N$ are the image parameters, for instance, the values of the image pixels, or wavelet coefficients. Given a grid of angular directions $\mathcal{G} = \{\theta_n\}_{n=1, \dots, N}$ and taking $b_n(\theta) = b(\theta - \theta_n)$, a grid model is obtained

$$I(\theta) = \sum_{n=1}^N x_n b(\theta - \theta_n) \xrightarrow{\text{F.T.}} \hat{I}(\nu) = \hat{b}(\nu) \sum_{n=1}^N x_n e^{-i 2 \pi \theta_n \cdot \nu}. \quad (14)$$

Using an equispaced grid, the usual pixelized image representation is obtained with pixel shape $b(\theta)$. The function $b(\theta)$ can also be used as a building block for image reconstruction [9]. Alternatively, $b(\theta)$ may be seen as the neat beam that sets the effective resolution of the image [10].

The size of the synthesized field of view and the image resolution must be chosen according to the extension of the observed object and to the resolution of the interferometer; see e.g., [10]. To avoid biases and rough approximations caused by the particular image model, the grid spacing $\Delta\theta$ should be well beyond the limit imposed by the longest baseline

$$\Delta\theta \ll \frac{\lambda}{2 B_{\max}}, \quad (15)$$

where $B_{\max} = \max_{j_1, j_2, t} |r_{j_1}(t) - r_{j_2}(t)|$ is the maximum projected separation between interfering telescopes. Oversampling by a factor of at least two is usually used and the pixel size is given by $\Delta\theta \approx \lambda/(4 B_{\max})$. To avoid aliasing and image truncation, the field of view must be chosen large enough and without forgetting that the reciprocal of the width of the field of view also sets the sampling step of the spatial frequencies.

The model of the complex visibility at the observed spatial frequencies is

$$V_k(x) = \hat{I}(\nu_k) = \sum_{n=1}^N T_{k,n} x_n, \quad (16)$$

where the coefficients of the matrix $T \in \mathbb{C}^{K \times N}$ are $T_{k,n} = \hat{b}_n(\nu_k)$ or $T_{k,n} = \hat{b}(\nu_k) e^{-i 2 \pi \theta_n \cdot \nu_k}$ depending which model of (13) or (14) is used. The matrix T performs the Fourier transform of nonequispaced data, which is a very costly operation. This problem is not specific to interferometry, similar needs in crystallography, tomography, and biomedical imaging have led to the development of fast algorithms to approximate this operation [11]. For instance

$$T \approx R \cdot F \cdot S, \quad (17)$$

where $F \in \mathbb{C}^{N \times N}$ is the fast Fourier transform (FFT) operator, $R \in \mathbb{C}^{K \times N}$ is a linear operator to interpolate the discrete Fourier transform of the image $\hat{x} = F \cdot x$ at the observed spatial frequencies, and S is diagonal and compensates the field of view apodization (or spectral smoothing) caused by R .

In radio astronomy a different technique called regridding [12], [13] is generally used, which consists of interpolating the data (not the model) onto the grid of discrete frequencies. The advantage is that, when there is a large number of measurements, the number of data points is reduced, which speeds up further computations. There are, however, a number of drawbacks to the regridding technique. First, it is not possible to apply the technique to nonlinear estimators such as the power spectrum and the bispectrum. Second, owing to the structure of the regridding operator, the regridded data are correlated even if the original data are not. These correlations are usually ignored in further processing and the pseudodata are assumed to be independent, which results in a poor approximation of the real noise statistics. This can be a critical issue with low signal to noise data [14].

Putting all together, the direct model of the data is affine

$$y = A \cdot x + e \quad (18)$$

with e the error vector ($e_\ell = V_{j_1, j_2, m}^{\text{err}}$), $A = G \cdot T$ the linear model operator and $G \in \mathbb{C}^{L \times K}$ the OTF operator given by

$$G_{\ell k} = \begin{cases} g_{j_1, m}^* g_{j_2, m} & \text{if } \ell \sim (j_1, j_2, m) \in \mathcal{B}_k \\ 0 & \text{else.} \end{cases} \quad (19)$$

Applying the pseudoinverse $\mathbf{T}^+ = \mathbf{S}^{-1} \cdot \mathbf{F}^{-1} \cdot \mathbf{R}^+$ of \mathbf{T} to the data yields the so-called dirty image [see Figure 2])

$$\hat{\mathbf{y}} = \mathbf{T}^+ \cdot \mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \hat{\mathbf{e}}, \quad (20)$$

where $\hat{\mathbf{e}} = \mathbf{T}^+ \cdot \mathbf{e}$ and $\mathbf{H} = \mathbf{T}^+ \cdot \mathbf{G} \cdot \mathbf{T}$. Apart from the apodization, \mathbf{H} essentially performs the convolution of the image by the dirty beam (see Figure 2). From (18) and (20), image reconstruction from interferometric data can be equivalently seen as a problem of interpolating missing Fourier coefficients or as a problem of deconvolution of the dirty map by the dirty beam [15].

INVERSE PROBLEM APPROACH

Since many Fourier frequencies are not measured, fitting the data alone does not uniquely define the sought image. Such an ill-posed problem can be solved by an inverse problem approach [17] by imposing a priori constraints to select a unique image among all those that are consistent with the data. The requirements for the priors are that they must help to smoothly interpolate voids in the (u, v) coverage while avoiding high frequencies beyond the diffraction limit. Without loss of generality, we assume that these constraints are monitored by a penalty function $f_{\text{prior}}(\mathbf{x})$ that measures the agreement of the image with the priors: the lower $f_{\text{prior}}(\mathbf{x})$, the better the agreement. In inverse problem framework, $f_{\text{prior}}(\mathbf{x})$ is termed as the regularization. Then the parameters \mathbf{x}^+ of the image which best matches the priors while fitting the data are obtained by solving a constrained optimization problem

$$\mathbf{x}^+ = \arg \min_{\mathbf{x}} f_{\text{prior}}(\mathbf{x}), \quad \text{subject to: } \mathbf{A} \cdot \mathbf{x} = \mathbf{y}. \quad (21)$$

Other strict constraints may apply. For instance, assuming the image brightness distribution must be positive and normalized, the feasible set is

$$\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{R}^N; \mathbf{x} \geq 0, \sum_n x_n = 1 \right\}, \quad (22)$$

where $\mathbf{x} \geq 0$ means $\forall n, x_n \geq 0$. Besides, due to noise and model approximations, there is some expected discrepancy between the model and actual data. As for the priors, the distance of the model to the data can be measured by a penalty function $f_{\text{data}}(\mathbf{x})$. We then require that, to be consistent with the data, an image must be such that $f_{\text{data}}(\mathbf{x}) \leq \eta_{\text{data}}$ where η_{data} is set according to the level of errors

$$\mathbf{x}^+ = \arg \min_{\mathbf{x} \in \mathcal{X}} f_{\text{prior}}(\mathbf{x}), \quad \text{subject to: } f_{\text{data}}(\mathbf{x}) \leq \eta_{\text{data}}. \quad (23)$$

The Lagrangian of this constrained optimization problem can be written as

$$\mathcal{L}(\mathbf{x}; \ell) = f_{\text{prior}}(\mathbf{x}) + \ell (f_{\text{data}}(\mathbf{x}) - \eta_{\text{data}}), \quad (24)$$

where ℓ is the Lagrange multiplier associated to the inequality constraint $f_{\text{data}}(\mathbf{x}) \leq \eta_{\text{data}}$. If the constraint is active then $\ell > 0$

and $f_{\text{data}}(\mathbf{x}) = \eta_{\text{data}}$ [18]. Conversely, the constraint being inactive would imply that $\ell = 0$, which would mean that the data are useless, which is hopefully not the case. Dropping the constant η_{data} , which does not depend on \mathbf{x} , the solution is obtained by solving either of the following problems:

$$\begin{aligned} \mathbf{x}^+ &= \arg \min_{\mathbf{x} \in \mathcal{X}} \{f_{\text{prior}}(\mathbf{x}) + \ell f_{\text{data}}(\mathbf{x})\} \\ &= \arg \min_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x}; \mu), \end{aligned}$$

where

$$f(\mathbf{x}; \mu) = f_{\text{data}}(\mathbf{x}) + \mu f_{\text{prior}}(\mathbf{x}) \quad (25)$$

is the penalty function and $\mu = 1/\ell > 0$ has to be tuned to match the constraint $f_{\text{data}}(\mathbf{x}) = \eta_{\text{data}}$. Hence, we can equivalently consider that we are solving the problem of maximizing the agreement of the model with the data subject to the constraint that the priors be below a preset level

$$\mathbf{x}^+ = \arg \min_{\mathbf{x} \in \mathcal{X}} f_{\text{data}}(\mathbf{x}), \quad \text{subject to: } f_{\text{prior}}(\mathbf{x}) \leq \eta_{\text{prior}}. \quad (26)$$

For convex penalties and providing that the Lagrange multipliers (μ and ℓ) and the thresholds (η_{data} and η_{prior}) are set consistently, the image restoration is achieved by solving either of the problems in (23) and (26) or by minimizing the penalty function in (25). However, choosing which of these particular problems to solve can be a deciding issue for the efficiency of the method. For instance, if $f_{\text{data}}(\mathbf{x})$ and $f_{\text{prior}}(\mathbf{x})$ are both smooth functions, direct minimization of $f(\mathbf{x}; \mu)$ in (25) can be done by using general-purpose optimization algorithms [18] but requires to know the value of the Lagrange multiplier. If the penalty functions are not smooth or if one wants to have the Lagrange multiplier automatically tuned given η_{data} or η_{prior} , specific algorithms must be devised. As we will see, specifying the image reconstruction as a constrained optimization problem provides a very general framework suitable to describe most existing methods; it however hides important algorithmic details about the strategy to search the solution. In what remains of this section, we first derive expressions of the data penalty terms and, then, the various regularizations that have been considered for image reconstruction in interferometry.

DISTANCE TO THE DATA

The ℓ_2 norm is a simple means to measure the consistency of the model image with the data

$$f_{\text{data}}(\mathbf{x}) = \| \mathbf{y} - \mathbf{A} \cdot \mathbf{x} \|_2^2. \quad (27)$$

However, to account for correlations and for the inhomogeneous quality of the measurements, the distance to the data has to be defined according to the statistics of the errors $\mathbf{e} = \mathbf{y} - \mathbf{A} \cdot \mathbf{x}$ given the image model. Assuming Gaussian statistics, this leads to

$$f_{\text{data}}(\mathbf{x}) = (\mathbf{y} - \mathbf{A} \cdot \mathbf{x})^T \cdot \mathbf{W}_{\text{err}} \cdot (\mathbf{y} - \mathbf{A} \cdot \mathbf{x}), \quad (28)$$

where the weighting matrix $\mathbf{W}_{\text{err}} = \mathbf{C}_{\text{err}}^{-1}$ is the inverse of the covariance matrix of the errors. There is a slight issue because we are dealing with complex values. Since complex numbers are just pairs of reals, complex valued vectors (such as \mathbf{y} , \mathbf{e} and $\mathbf{A} \cdot \mathbf{x}$) can be flattened into ordinary real vectors (with doubled size) to use standard linear algebra notation and define the covariance matrix as $\mathbf{C}_{\text{err}} = \langle \mathbf{e} \cdot \mathbf{e}^T \rangle$. This is what is assumed in (28).

There are some possible simplifications. For instance, the complex visibilities are measured independently, hence the weighting matrix \mathbf{W}_{err} is block diagonal with 2×2 blocks. Furthermore, if the real and imaginary parts of a given measured complex visibility are uncorrelated and have the same variance, then f_{data} takes a simple form

$$f_{\text{data}}(\mathbf{x}) = \sum_{\ell} w_{\ell} |\mathbf{y}_{\ell} - (\mathbf{A} \cdot \mathbf{x})_{\ell}|^2, \quad (29)$$

where the weights are given by

$$w_{\ell} = \text{Var}(\text{Re}(\mathbf{y}_{\ell}))^{-1} = \text{Var}(\text{Im}(\mathbf{y}_{\ell}))^{-1}. \quad (30)$$

This expression of $f_{\text{data}}(\mathbf{x})$, popularized by Goodman [19], is very commonly used in radio interferometry.

Real data may, however, have different statistics. For instance, OI-FITS, which is the standard file exchange format for optical interferometric data, assumes that the amplitude and the phase of complex data (complex visibility or triple product) are independent [20]. The thick lines in Figure 3 display the isocontours of the corresponding log-likelihood that forms a nonconvex valley in the complex plane. Assuming Goodman statistics would yield circular isocontours in this figure and is obviously a bad approximation of the true criterion in that case. To improve on the Goodman model while avoiding nonconvex criteria, Meimon et al. [14] have proposed quadratic convex approximations of the true log-likelihood (see Figure 3) and have shown that their so-called local

approximation yields the best results, notably when dealing with low signal-noise data. For a complex datum $\mathbf{y}_{\ell} = \rho_{\ell} \exp(i \varphi_{\ell})$, their local quadratic approximation writes

$$f_{\text{data}}(\mathbf{x}) = \sum_{\ell} \left\{ \frac{\text{Re}(\mathbf{e}_{\ell} \mathbf{e}_{\ell}^* e^{-i \varphi_{\ell}})^2}{\sigma_{\parallel, \ell}^2} + \frac{\text{Im}(\mathbf{e}_{\ell} \mathbf{e}_{\ell}^* e^{-i \varphi_{\ell}})^2}{\sigma_{\perp, \ell}^2} \right\}, \quad (31)$$

where $\mathbf{e} = \mathbf{y} - \mathbf{A} \cdot \mathbf{x}$ denotes the complex residuals and the variances along and perpendicular to the complex datum vector are

$$\sigma_{\parallel, \ell}^2 = \text{Var}(\rho_{\ell}) \quad (32)$$

$$\sigma_{\perp, \ell}^2 = \rho_{\ell}^2 \text{Var}(\varphi_{\ell}). \quad (33)$$

The Goodman model is retrieved when $\rho_{\ell}^2 \text{Var}(\varphi_{\ell}) = \text{Var}(\rho_{\ell})$.

MAXIMUM ENTROPY METHODS

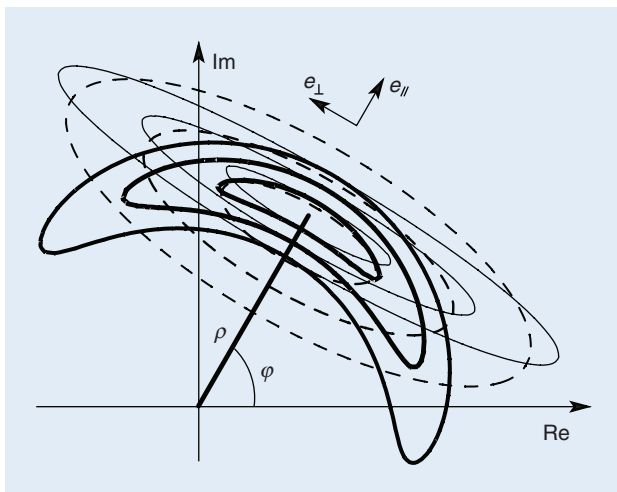
Maximum entropy methods (MEMs) are based on the 1950s work of Jaynes on information theory; the underlying idea is to obtain the least informative image that is consistent with the data [21]. This amounts to minimizing a criterion like the one in (25) with $f_{\text{prior}}(\mathbf{x}) = -S(\mathbf{x})$, where the entropy $S(\mathbf{x})$ measures the informational contents of the image \mathbf{x} . In this framework, $f_{\text{prior}}(\mathbf{x})$ is sometimes called negentropy. Among all the expressions considered for the negentropy of an image, one of the most popular is [22]

$$f_{\text{prior}}(\mathbf{x}) = \sum_j [x_j \log(x_j/\bar{x}_j) - x_j + \bar{x}_j] \quad (34)$$

with \bar{x} the default image; that is, the one that would be recovered in the absence of any data. Back to information theory, this expression is similar to the Kullback-Leibler divergence between \mathbf{x} and \bar{x} (with additional terms that cancel for normalized distributions). The default image \bar{x} can be taken as being a flat image, an image previously restored, or an image of the same object at a lower resolution. Narayan and Nityananda [23] reviewed MEMs for radio-interferometry imaging and compared the other forms of the negentropy that have been proposed. They argued that only nonquadratic priors can interpolate missing Fourier data and noted that such penalties also forbid negative pixel values. The fact that there is no need to explicitly impose positivity is sometimes put forward by the proponents of these methods.

MEM penalties are usually separable, which means that they do not depend on the ordering of the pixels. To explicitly enforce some correlation between close pixels in the sought image \mathbf{x} (hence, some smoothness), the prior can be chosen to depend on \mathbf{x} . For instance: $\bar{x} = \mathbf{P} \cdot \mathbf{x}$, where \mathbf{P} is some averaging or smoothing linear operator. This type of floating prior has been used to loosely enforce constraints such as radial symmetry [24]. Alternatively, an intrinsic correlation function (ICF) can be explicitly introduced by a convolution kernel to impose the correlation structure of the image [25].

Minimizing the joint criterion in (25) with entropy regularization has a number of issues as the problem is highly nonlinear and as the number of unknowns is very large (as many as there are pixels). Various methods have been proposed, but the most effective algorithm [26] seeks for the solution by a nonlinear optimization in a local subspace of search directions with the Lagrange multiplier μ tuned on the fly to match the constraint that $f_{\text{data}}(\mathbf{x}) = \eta_{\text{data}}$.



[FIG3] Convex quadratic approximations of complex data. Thick lines are isocontours of the log-likelihood f_{data} (at 1, 2, and 3 rms levels) for a complex datum with independent amplitude and phase. Dashed lines are isocontours for the global quadratic approximation. Thin lines are isocontours for the local quadratic approximation.

OTHER PRIOR PENALTIES

Bayesian arguments can be invoked to define other types of regularization. For instance, assuming that the pixels have a Gaussian distribution leads to quadratic penalties such as

$$f_{\text{prior}}(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T \cdot \mathbf{C}_{\text{prior}}^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}}) \quad (35)$$

with $\mathbf{C}_{\text{prior}}$ the prior covariance and $\bar{\mathbf{x}}$ the prior solution. Tikhonov's regularization [27], $f_{\text{prior}}(\mathbf{x}) = \|\mathbf{x}\|_2^2$, is the simplest of these penalties. By Parseval's theorem, this regularization favors zeroes for unmeasured frequencies, it is therefore not recommended for image reconstruction in interferometry. Yet, this does not rule out all quadratic priors. For instance, compactness is achieved by a very simple quadratic penalty

$$f_{\text{prior}}(\mathbf{x}) = \sum_n w_n^{\text{prior}} x_n^2, \quad (36)$$

where the weights are increasing with the distance to the center of the image thus favoring structures concentrated within this part of the image. Under strict normalization constraint and in the absence of any data, the default image given by this prior is $\bar{x}_n \propto 1/w_n^{\text{prior}}$ where the factor comes from the normal-

ization requirement [28]. Although simple, this regularizer, coupled with the positivity constraint, can be very effective as shown by Figure 4(b). Indeed, smooth Fourier interpolation follows from the compactness of the brightness distribution which is imposed by $f_{\text{prior}}(\mathbf{x})$ in (36) and by the positivity as it plays the role of a floating support.

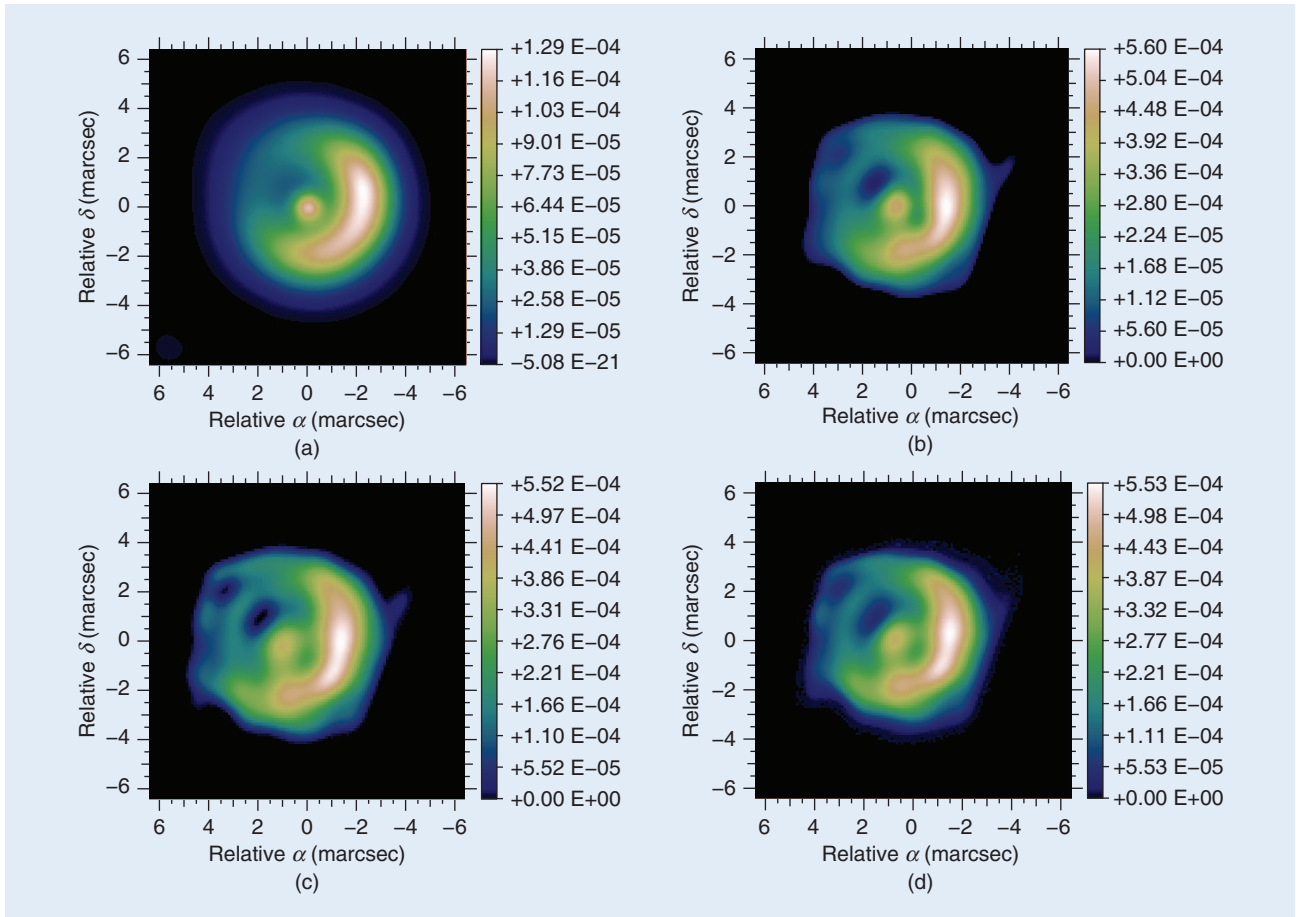
Other prior penalties commonly used in image restoration methods can be useful for interferometry. For instance, edge-preserving smoothness is achieved by

$$f_{\text{prior}}(\mathbf{x}) = \sum_{n_1, n_2} \sqrt{\epsilon^2 + |\nabla \mathbf{x}|_{n_1, n_2}^2}, \quad (37)$$

where $\epsilon > 0$ is a chosen threshold and $|\nabla \mathbf{x}|^2$ is the squared magnitude of the spatial gradient of the image

$$|\nabla \mathbf{x}|_{n_1, n_2}^2 = (x_{n_1+1, n_2} - x_{n_1, n_2})^2 + (x_{n_1, n_2+1} - x_{n_1, n_2})^2.$$

The penalization in (37) behaves as a quadratic (respectively, linear) function where the magnitude of the spatial gradient is small (respectively, large) compared to ϵ . Thus reduction of small local variations without penalizing too much strong sharp features is achieved by this regularization. In the limit $\epsilon \rightarrow 0$,



[FIG4] Image reconstruction with various types of regularization. (a) Original object smoothed to the resolution of the interferometer (FWHM ~ 15 marcsec); (b) reconstruction with a quadratic regularization given by (36) and that imposes a compact field of view; (c) reconstruction with edge-preserving regularization as in (37); and (d) reconstruction with maximum entropy regularization as in (34). All reconstructions by algorithm MiRA and from the power spectrum and the phase closures.

edge-preserving smoothness behaves like total variation [29], which has proved successful in imposing sparsity.

CLEAN METHOD

Favoring images with a limited number of significant pixels is a way to avoid the degeneracies of image reconstruction from sparse Fourier coefficients. This could be formally done by searching for the least ℓ_0 -norm image consistent with the data; hence using $f_{\text{prior}}(x) = \|x\|_0$. However, due to the number of parameters, minimizing the resulting mixed criterion is a combinatorial problem that is too difficult to solve directly. The CLEAN algorithm [30], [31] implements a matching pursuit strategy to attempt to find this kind of solution. The method proceeds iteratively as follows. Given the data in the form of a dirty image, the location of the brightest point source that best explains the data is searched. The model image is then updated by a fraction of the intensity of this component, and this fraction times the dirty beam is subtracted from the dirty image. The procedure is repeated for the new residual dirty image that is searched for evidence of another point-like source. When the level of the residuals becomes smaller than a given threshold set from the noise level, the image is convolved with the clean beam [usually a Gaussian-shaped point spread function (PSF)] to set the resolution according to the extension of the (u, v) coverage. Once most point sources have been removed, the residual dirty image is essentially due to the remaining extended sources that may be smooth enough to be insensitive to the convolution by the dirty beam. Hence, adding the residual dirty image to the clean image produces a final image consisting in compact sources (convolved by the clean beam) plus smooth extended components. Although designed for point sources, CLEAN works rather well for extended sources and remains one of the preferred methods in radio interferometry.

It has been demonstrated that the matching pursuit part of CLEAN is equivalent to an iterative deconvolution with early stopping [32] and that it is an approximate algorithm for obtaining the image of minimum total flux consistent with the observations [33]. Hence, under the nonnegativity constraint, this would, at best, yield the least ℓ_1 -norm image consistent with the data. This objective is supported by recent results in *Compressive Sensing* [34] showing that, in most practical cases, regularization by the ℓ_1 -norm of x enforces the sparsity of the solution. However, the matching pursuit strategy implemented by CLEAN is slow, it has some instabilities and it is known to be suboptimal [33].

OTHER METHODS

This section briefly reviews other image reconstruction methods applied in astronomical interferometry.

1) *Multiresolution*: These methods aim at reconstructing images with different scales. They basically rely on recursive decomposition of the image in low and high frequencies. Multiresolution CLEAN [35] first reconstructs an image of the broad emission and then iteratively updates this map at full resolution as in the original CLEAN algorithm. This approach has been generalized by using a wavelet expansion to describe the image, which could be formally expressed in terms of (13), and achieved multiresolution deconvolution by a matching

pursuit algorithm applied to the wavelet coefficients and such that the solution satisfies positivity and support constraints [36]. The multiscale CLEAN algorithm [37] explicitly describes the image as a sum of components with different scales and makes use of a weighted matching pursuit algorithm to search for the scale and position of each image update. The main advantages of multiscale CLEAN are its ability to leave very few structures in the final residuals and to correctly estimate the total flux of the observed object. This method is widely used in radio astronomy and is part of standard data processing packages [38]. In the context of MEM, the multichannel maximum entropy image reconstruction method [39] introduces a multiscale structure in the image by means of different intrinsic correlation functions [25]. The reconstructed image is then the sum of several extended sources with different levels of correlation. This approach was extended by using a pyramidal image decomposition [40] or wavelet expansions [41], [42].

2) *WIPE method*: The WIPE method [10] is a regularized fit of the interferometric data under positivity and support constraints. The model image is given by (13) using an equally spaced grid and the effective resolution is explicitly set by the basis function $b(\theta)$, the so-called neat beam, with an additional penalty to avoid super resolution. The image parameters are the ones that minimize

$$f_{\text{WIPE}}(x) = \sum_{\ell} w_{\ell} |\hat{b}_{\ell} y_{\ell} - (A \cdot x)_{\ell}|^2 + \sum_{k, |v_k| > \nu_{\text{eff}}} |(F \cdot x)_k|^2$$

with y the calibrated complex visibility data, \hat{b}_{ℓ} the Fourier transform of the neat beam at the spatial frequency of datum y_{ℓ} , $\nu_{\text{eff}} \geq \sup_{\nu \in \mathcal{L}} |\nu|$ an effective cutoff frequency, F the Fourier transform operator, A the model matrix given by (18) accounting for the subsampled Fourier transform and the neat beam, and

$$w_{\ell} = 1 / \left(\sigma_{\ell}^2 \sum_{\ell'} \sigma_{\ell'}^{-2} \right),$$

where $\sigma_{\ell}^2 = |\hat{b}_{\ell}|^2 \text{Var}(y_{\ell})$ assumes the Goodman approximation. In the criterion minimized by WIPE, one can identify the distance of the model to the data and a regularization term. There is no hyper-parameter to tune the level of this latter term. The optimization is done by a conjugate gradient search with a stopping criterion derived from the analysis of the conditioning of the problem. This analysis is built up during the iterations.

3) *Bimodel method*: The case of an image model explicitly mixing extended source and point sources has also been addressed [43], [44] and more recently [15]. The latter have considered an image $x = x_e + x_p$ made of two maps: x_e for extended structures and x_p for point-like components. The maps x_e and x_p are respectively regularized by imposing smoothness and sparsity. With additional positivity and, optionally, support constraints, it turns out that the two kinds of regularization can be implemented by quadratic penalties. Their method amounts to minimize

$$f_{\text{mix}}(\mathbf{x}_e, \mathbf{x}_p) = \|\mathbf{y} - \mathbf{A} \cdot (\mathbf{x}_e + \mathbf{x}_p)\|^2 + \lambda_s \mathbf{c}^T \cdot \mathbf{x}_p + \epsilon_s \|\mathbf{x}\|_p^2 + \lambda_c \|\mathbf{x}_e\|_{C_{\text{prior}}}^2 + \epsilon_m (\mathbf{c}^T \cdot \mathbf{x}_e)^2$$

with $\|\mathbf{x}_e\|_{C_{\text{prior}}}^2$ a local finite difference norm similar to (35), and \mathbf{c} a vector with all components set to one, hence, $\mathbf{c}^T \cdot \mathbf{x} = \sum \mathbf{x}_n$. There are four tuning parameters for the regularization terms: $\lambda_s \geq 0$ and $\epsilon_s > 0$ control the sparsity of \mathbf{x}_p , $\lambda_c > 0$ controls the level of smoothness in the extended map \mathbf{x}_e , and $\epsilon_m > 0$ (or $\epsilon_m \geq 0$ if there is a support constraint) insures strict convexity of the regularization with respect to \mathbf{x}_e . Circulant approximations are used to implement a very fast minimization of $f_{\text{mix}}(\mathbf{x}_e, \mathbf{x}_p)$ under the constraints that $\mathbf{x}_e \geq 0$ and $\mathbf{x}_p \geq 0$ [15].

SELF-CALIBRATION

When the OTF cannot be calibrated (e.g., there is no reference or the OTF is significantly varying due to the turbulence), the problem is not only to derive the image parameters \mathbf{x} , but also the unknown complex throughputs \mathbf{g} . As there is no correlation between the throughputs and the observed object, the inverse approach leads to solve

$$(\mathbf{x}, \mathbf{g})^+ = \underset{\mathbf{x} \in \mathcal{X}, \mathbf{g}}{\operatorname{argmin}} \{f_{\text{data}}(\mathbf{x}, \mathbf{g}) + \mu_{\text{img}} f_{\text{prior}}^{\text{img}}(\mathbf{x}) + \mu_{\text{gain}} f_{\text{prior}}^{\text{gain}}(\mathbf{g})\} \quad (38)$$

with $\mu_{\text{img}} f_{\text{prior}}^{\text{img}}(\mathbf{x})$ and $\mu_{\text{gain}} f_{\text{prior}}^{\text{gain}}(\mathbf{g})$ the regularization terms for the image parameters and for the complex throughputs. The latter can be derived from prior statistics about the turbulence [5]. In principle, global optimization should be required to minimize the nonconvex criterion in (38). Fortunately, a simpler strategy based on alternate minimization with respect to \mathbf{x} only and then with respect to \mathbf{g} only has proved effective to solve this problem. This method has been called self-calibration because it uses the current estimate of the sought image as a reference source to calibrate the throughputs. The algorithm begins with an initial image $\mathbf{x}^{[0]}$ and repeats until convergence the following steps (starting with $n = 1$ and incrementing n after each iteration):

1) *Self-calibration step.* Given the image $\mathbf{x}^{[n-1]}$, find the best complex throughputs $\mathbf{g}^{[n]}$ by solving

$$\mathbf{g}^{[n]} = \underset{\mathbf{g}}{\operatorname{argmin}} \{f_{\text{data}}(\mathbf{x}^{[n-1]}, \mathbf{g}) + \mu_{\text{gain}} f_{\text{prior}}^{\text{gain}}(\mathbf{g})\}.$$

2) *Image reconstruction step.* Apply image reconstruction algorithm to recover a new image estimate given the data and the complex throughputs

$$\mathbf{x}^{[n]} = \underset{\mathbf{x}}{\operatorname{argmin}} \{f_{\text{data}}(\mathbf{x}, \mathbf{g}^{[n]}) + \mu_{\text{img}} f_{\text{prior}}^{\text{img}}(\mathbf{x})\}.$$

Note that any image reconstruction algorithm described previously can be used in the second step of the method. The criterion in (38) being nonconvex, the solution should depend on the initialization. Yet, this does not appear to be an issue in practice even if simple local optimization methods are used to solve the self-calibration step (such as the one recently proposed by [45]).

Self-calibration was initially proposed by Readhead and Wilkinson [46] to derive missing Fourier phase information

from phase closure data, and the technique was later improved by Cotton [47]. Schwab [48] was the first to solve the problem by explicitly minimizing a nonlinear criterion similar to $f_{\text{data}}(\mathbf{x}, \mathbf{g})$ in (29). Schwab's approach was further improved by Cornwell and Wilkinson [49] who introduced priors for the complex gains, that is, the term $\mu_{\text{gain}} f_{\text{prior}}^{\text{gain}}(\mathbf{g})$ in the global penalty. However, for most authors, no priors about the throughputs are assumed, hence $\mu_{\text{gain}} = 0$.

Self-calibration is a particular case of the blind, or myopic, deconvolution methods [50] that have been developed to improve the quality of blurred images when the PSF is unknown. Indeed, when the PSF can be completely described by phase aberrations in the pupil plane, blind deconvolution amounts to solving the same problem as self-calibration [51].

IMAGE RECONSTRUCTION FROM NONLINEAR DATA

At optical wavelengths, the complex visibilities (whether they are calibrated or not) are not directly measurable, the available data (cf. the section "Interferometric Data") are the power spectrum, the bispectrum, and/or the phase closure. Image reconstruction algorithms can be designed following the same inverse problem approach as before. In particular, the regularization can be implemented by the same f_{prior} penalties as in the section "Imaging From Sparse Fourier Data." However, the direct model of the data is now nonlinear and specific expressions to implement f_{data} have to be derived. The nonlinearity has also some incidence on the optimization strategy.

DATA PENALTY

The power spectrum, the bispectrum, and the phase closure data have non-Gaussian statistics: the power spectrum is a positive quantity, the phase closure is wrapped in $(-\pi, +\pi]$, etc. Most algorithms, however, make use of quadratic penalties with respect to the measurements that implies Gaussian statistics in a Bayesian framework. Another assumption generally made is the independence of the measurements, which leads to separable penalties. Under such approximations, the penalty with respect to the power spectrum data writes

$$f_{\text{data}}^{\text{ps}}(\mathbf{x}) = \sum_{m, j_1 < j_2} \frac{(S_{j_1, j_2, m}^{\text{data}} - S_{j_1, j_2, m}^{\text{model}}(\mathbf{x}))^2}{\text{Var}(S_{j_1, j_2, m}^{\text{data}})}, \quad (39)$$

with $S_{j_1, j_2, m}^{\text{model}}(\mathbf{x}) = |\hat{I}(\nu_{j_1, j_2, m})|^2$ the model of the power spectrum. For the penalty with respect to the bispectrum data, there is the additional difficulty to deal with complex data. The Goodman approximation [19] yields

$$f_{\text{data}}^{\text{bisp}}(\mathbf{x}) = \sum_{m, j_1 < j_2 < j_3} w_{j_1, j_2, j_3, m}^{\text{bisp}} |B_{j_1, j_2, j_3, m}^{\text{data}} - B_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x})|^2 \quad (40)$$

with $B_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x}) = \hat{I}(\nu_{j_1, j_2, m}) \hat{I}(\nu_{j_2, j_3, m}) \hat{I}(\nu_{j_3, j_1, m})$ the model of the bispectrum and weights derived from the variance of the bispectrum data. An expression similar to that in (31) can be derived for bispectrum data with independent modulus and phase errors. To account for phase wrapping, Haniff [52] proposed to define the penalty with respect to the phase closure data as

$$f_{\text{data}}^{\text{cl}}(\mathbf{x}) = \sum_{m, j_1 < j_2 < j_3} \frac{\arccos^2(\beta_{j_1, j_2, j_3, m}^{\text{data}} - \beta_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x}))}{\text{Var}(\beta_{j_1, j_2, j_3, m}^{\text{data}})} \quad (41)$$

with $\beta_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x}) = \varphi(v_{j_1, j_2, m}) + \varphi(v_{j_2, j_3, m}) + \varphi(v_{j_3, j_1, m})$ the model of the phase closure. This penalty is, however, not continuously differentiable with respect to \mathbf{x} , which can prevent the convergence of optimization algorithms. This problem can be avoided by using the complex phasors [53]

$$f_{\text{data}}^{\text{cl}}(\mathbf{x}) = \sum_{m, j_1 < j_2 < j_3} \frac{|e^{i\beta_{j_1, j_2, j_3, m}^{\text{data}}} - e^{i\beta_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x})}|^2}{\text{Var}(\beta_{j_1, j_2, j_3, m}^{\text{data}})}, \quad (42)$$

which is approximately equal to the penalty in (41) in the limit of small phase closure errors.

Depending on which set of data is available, and assuming that the different types of data have statistically independent errors, the total penalty with respect to the data is simply a sum of some of the penalties given by (39)–(42). For instance, to fit the power spectrum and the phase closure data

$$f_{\text{data}}(\mathbf{x}) = f_{\text{data}}^{\text{ps}}(\mathbf{x}) + f_{\text{data}}^{\text{cl}}(\mathbf{x}). \quad (43)$$

IMAGE RECONSTRUCTION ALGORITHMS

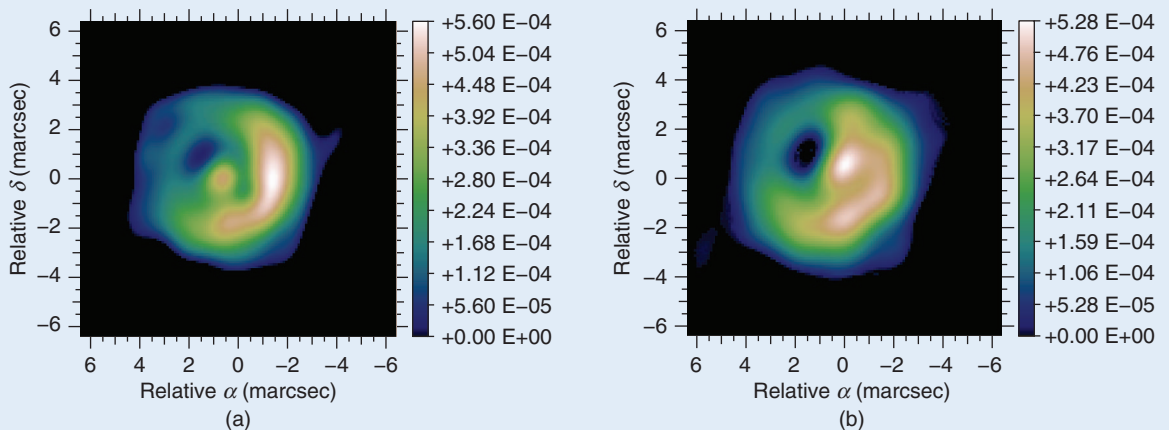
We describe here the image reconstruction methods used with some success on realistic optical interferometric data in astronomy and that can be considered as ready to process real data. In addition to cope with sparse Fourier data, these methods were specifically designed to tackle the nonlinear direct model, to account for the particular statistics of the data [14], and to handle the new data format [20]. These image reconstruction methods can all be formally described in terms of a criterion to optimize, perhaps under some strict constraints, and an optimization strategy. Some of these algorithms have clearly inherited from methods previously developed: BSMEM [54], the building block method [9], and WISARD [55] are related to MEM, CLEAN, and self-calibration, respectively.

1) *The BSMEM algorithm* [54], [56] makes use of a MEM to regularize the problem of image restoration from the measured bispectrum (hence its name). The improved BSMEM version [56] uses the Gull and Skilling entropy, see (34), and a likelihood term with respect to the complex bispectrum, which assumes independent Gaussian noise statistics for the amplitude and phase of the measured bispectrum. The optimization engine is MEMSYS, which implements the strategy proposed by Skilling and Bryan [26] and automatically finds the most likely value for the hyper-parameter μ . The default image is either a Gaussian, a uniform disk, or a Dirac centered in the field of view. Because it makes no attempt to directly convert the data into complex visibilities, a strength of BSMEM is that it can handle any type of data sparsity (such as missing closures). Thus, in principle, BSMEM could be used to restore images when Fourier phase data are completely missing (see Figure 5).

2) *The building block method* [9] is similar to the CLEAN method but designed for reconstructing images from bispectrum data obtained by means of speckle or long baseline interferometry. The method proceeds iteratively to reduce a cost function $f_{\text{data}}^{\text{bisp}}$ equal to that in (40) with weights set to a constant or to an expression motivated by Wiener filtering. The minimization of the penalty is achieved by a matching pursuit algorithm which imposes sparsity of the solution. The image is given by the building block model in (13) and (14) and, at the n th iteration, the new image $I^{[n]}(\theta)$ is obtained by adding a new building block at location $\theta^{[n]}$ with a weight $\alpha^{[n]}$ to the previous image, so as to maintain the normalization

$$I^{[n]}(\theta) = (1 - \alpha^{[n]}) I^{[n-1]}(\theta) + \alpha^{[n]} b(\theta - \theta^{[n]}).$$

The weight and location of the new building block is derived by minimizing the criterion $f_{\text{data}}^{\text{bisp}}$ with respect to these parameters. Strict positivity and support constraint can be trivially enforced by limiting the possible values for $\alpha^{[n]}$ and $\theta^{[n]}$. To improve the convergence, the method allows to add/remove more than one block at a time. To avoid super resolution artifacts, the final



[FIG5] Image reconstruction with (a) phase closure and (b) without any Fourier phase information.

image is convolved with a smoothing function with size set according to the spatial resolution of the instrument.

3) *The Markov Chain Imager (MACIM) algorithm* [57], aims at maximizing the posterior probability

$$\Pr(\mathbf{x}|\mathbf{y}) \propto \exp\left(-\frac{1}{2}f_{\text{data}}(\mathbf{x}) - \frac{\mu}{2}f_{\text{prior}}(\mathbf{x})\right).$$

MACIM implements MEM regularization and a specific regularizer that favors large regions of dark space in-between bright regions. For this latter regularization, $f_{\text{prior}}(\mathbf{x})$ is the sum of all pixels with zero flux on either side of their boundaries. MACIM attempts to maximize $\Pr(\mathbf{x}|\mathbf{y})$ by a simulated annealing algorithm with the Metropolis sampler. Although maximizing $\Pr(\mathbf{x}|\mathbf{y})$ is the same as minimizing $f_{\text{data}}(\mathbf{x}) + \mu f_{\text{prior}}(\mathbf{x})$, the use of normalized probabilities is required by the Metropolis sampler to accept or reject the image samples. In principle, simulated annealing is able to solve the global optimization problem of maximizing $\Pr(\mathbf{x}|\mathbf{y})$ but the convergence of this kind of Monte-Carlo method for such a large problem is very slow and critically depends on the parameters that define the temperature reduction law. A strict Bayesian approach can also be exploited to derive, in a statistical sense, the values of the hyper-parameters (such as μ) and some a posteriori information such as the significance level of the image.

4) *The multitelescope image reconstruction (MiRA) algorithm* [53] defines the sought image as the minimum of the penalty function in (25). Minimization is done by a limited variable memory method (based on BFGS updates) with bound constraints for the positivity [58]. Since this method does not implement global optimization, the image restored by MiRA depends on the initial image. MiRA is written in a modular way: any type of data can be taken into account by providing a function that computes the corresponding penalty and its gradient. For the moment, MiRA handles complex visibility, power spectrum, and closure-phase data via penalty terms given by (31), (39), and (42). Also, many different regularizers are built into MiRA (such as negentropy, quadratic or edge-preserving smoothness, compactness, and total variation) and provisions are made to implement custom priors. MiRA can cope with any missing data, in particular, it can be used to restore an image given only the power spectrum (i.e., without any Fourier phase information) with at least a 180° orientation ambiguity. An example of reconstruction with no phase data is shown in Figure 5. In the case of nonsparse (u, v) coverage, the problem of image reconstruction from the modulus of its Fourier transform has been addressed by Fienup [59] by means of an algorithm based on projections onto convex sets (POCS).

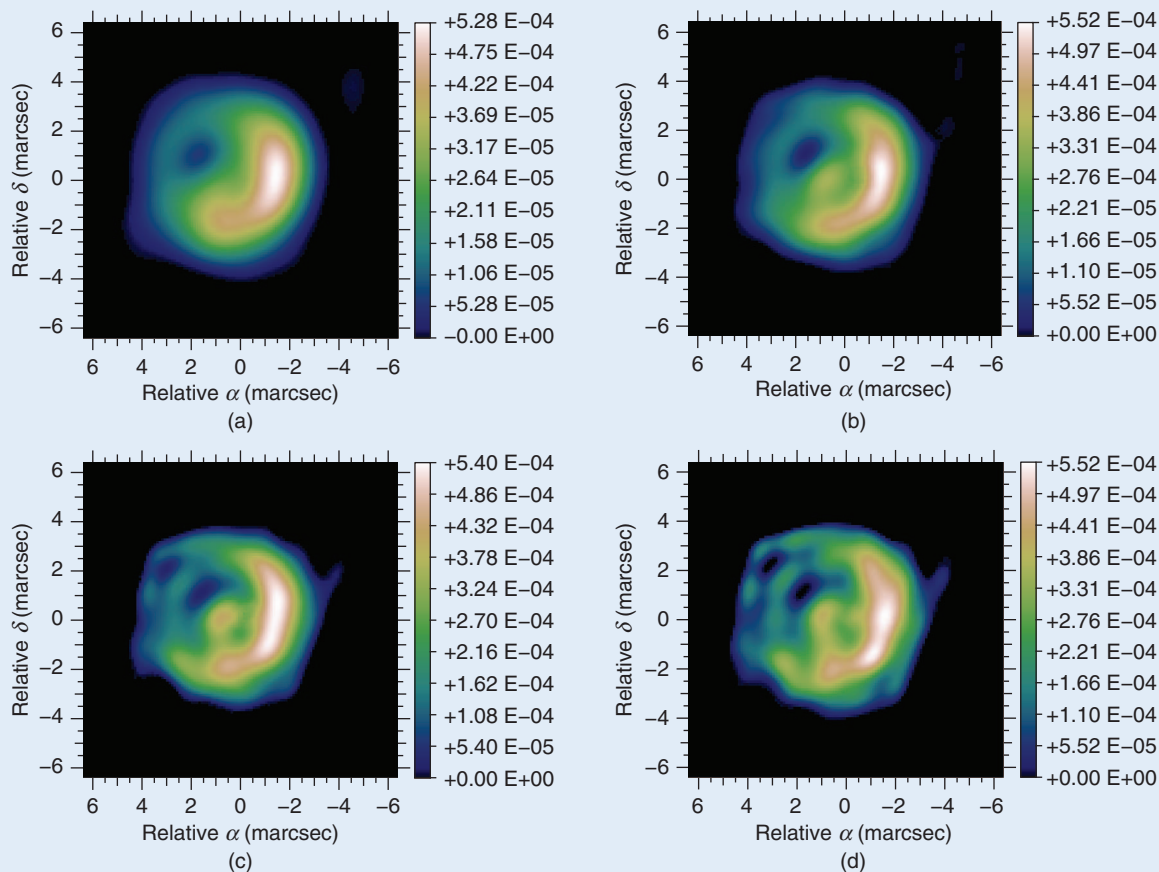
5) *The WISARD algorithm* [55] recovers an image from power spectrum and phase closure data. It exploits a self-calibration approach (cf. the section “Self-Calibration” for a more detailed look at this approach) to recover missing Fourier phases. Given a current estimate of the image and the phase closure data, WISARD first derives missing Fourier phase information in such a way as to minimize the number of unknowns. Then, the synthesized Fourier phases are com-

bined with the square root of the measured power spectrum to generate pseudocomplex visibility data that are fitted by the image restoration step. This step is performed by using the chosen regularization and a penalty with respect to the pseudocomplex visibility data. However, to account for a more realistic approximation of the distribution of complex visibility errors, WISARD makes use of a quadratic penalty that is different from the usual Goodman approximation [14]. Taken separately, the image restoration step is a convex problem with a unique solution, the self-calibration step is not strictly convex but (like in original self-calibration method) does not seem to pose insurmountable problems. Nevertheless, the global problem is multimodal and, at least in difficult cases, the final solution depends on the initial guess. There are many possible regularizers built into WISARD, such as the one in (36) and the edge-preserving smoothness prior in (37).

MiRA and WISARD have been developed in parallel and share some common features. They use the same optimization engine [58] and means to impose positivity and normalization [28]. They, however, differ in the way missing data is taken into account: WISARD takes a self-calibration approach to explicitly solve for missing Fourier phase information; while MiRA implicitly accounts for any lack of information through the direct model of the data [28].

All of these algorithms have been compared on simulated data during “Interferometric Beauty Contests” [16], [60], [61]. The results of the contest were very encouraging. Although quite different algorithms, BSMEM, the building block method, MiRA, and WISARD give good image reconstructions where the main features of the objects of interest can be identified in spite of the sparse (u, v) coverage, the lack of some Fourier phase information and the nonlinearities of the measurements. BSMEM and MiRA appear to be the most successful algorithms (they respectively won the first two and last contests).

With their tuning parameters and, for some of them, the requirement to start with an initial image, these algorithms still need some expertise to be used successfully. But this is quite manageable. For instance, the tuning of the regularization level can be derived from Bayesian considerations but can also almost be done by visual inspection of the restored image. From Figure 6, one can see the effects of under-regularization (which yields more artifacts) and over-regularization (which yields over simplification of the image). In that case, a good regularization level is probably between $\mu = 10^5$ and $\mu = 10^4$ and any choice in this range would give a good image. Figure 4 shows image reconstructions from one of the data sets of the “2004 Beauty Contest” [16] and with different types of regularization. These synthesized images do not greatly differ and are all quite acceptable approximations of the reality (compare, for instance, with the dirty image in Figure 2). Hence, provided that the level of the priors is correctly set, the particular choice of a given regularizer can be seen as a refinement that can be done after some reconstruction attempts with a prior that is simpler to tune. At least, the qualitative type of prior is what really matters, not the specific expression of the penalty imposing the prior.



[FIG6] Image reconstruction under various regularization levels. Algorithm is MiRA with edge-preserving regularization given in (37) with (a) $\epsilon = 10^{-4}$ and $\mu = 10^6$, (b) $\mu = 10^5$, (c) $\mu = 10^4$, and (d) $\mu = 3 \times 10^3$.

DISCUSSION

The main issues in image reconstruction from interferometric data are the sparsity of the measurements (which sample the Fourier transform of the object brightness distribution) and the lack of part of the Fourier phase information. The inverse problem approach appears to be suitable to describe the most important existing algorithms in this context. Indeed, the image reconstruction methods can be stated as the minimization of a mixed criterion under some strict constraints such as positivity and normalization. Two different types of terms appear into this criterion: likelihood terms that enforce consistency of the model image with the data, and regularization terms that maintain the image close to the priors required to lever the degeneracies of the image reconstruction problem. Hence, the differences between the various algorithms lie in the kind of measurements considered, in the approximations for the direct model and for the statistics of the errors and in the prior imposed by the regularization. For nonconvex criteria that occur when the OTF is unknown or when nonlinear estimators are measured to overcome turbulence effects, the initial solution and the optimization strategy are also key components of the algorithms. Although global optimization is required to solve such multimodal problems, most existing algorithms are successful whereas they only

implement local optimization. These algorithms are not fully automated black boxes: at least some tuning parameters and the type of regularization are left to the user choice. Available methods are however now ready for image reconstruction from real data. Nevertheless, a general understanding of the mechanisms involved in image restoration algorithms is mandatory to correctly use these methods and to analyze possible artifacts in the synthesized images. From a technical point of view, future developments of these algorithms will certainly focus on global optimization and unsupervised reconstruction. However, to fully exploit the existing instruments, the most worthwhile tracks to investigate are multispectral imaging and accounting for additional data such as a low-resolution image of the observed object to overcome the lack of short baselines.

AUTHORS

Eric Thiébaud (thiebaut@obs.univ-lyon1.fr) graduated from the École Normale Supérieure in 1987 and received the Ph.D. degree in astrophysics from the Université Pierre & Marie Curie (Paris VII, France), in 1994. Since 1995, he has been an astronomer at the Centre de Recherche Astrophysique de Lyon. His main interests are in the fields of signal processing and image reconstruction. He has made various contributions

to blind deconvolution, optical interferometry, and optimal detection with applications in astronomy, biomedical imaging, and digital holography. He is a Member of the IEEE.

Jean-François Giovannelli (Giova@IMS-Bordeaux.fr) graduated from the École Nationale Supérieure de l'Électronique et de ses Applications in 1990. He received the Ph.D. degree in 1995 and the Habilitation à Diriger des Recherches degree in physics (signal processing) in 2005. From 1997 to 2008, he was an assistant professor with the Université Paris-Sud, and a researcher with the Laboratoire des Signaux et Systèmes. He is currently a professor with the Université de Bordeaux and a researcher with the Laboratoire d'Intégration du Matériau au Système, Équipe Signal-Image. He is interested in regularization and Bayesian methods for inverse problems in signal and image processing. His application fields essentially concern astronomical, medical, proteomics, and geophysical imaging.

REFERENCES

- [1] A. Labeyrie, "Interference fringes obtained on VEGA with two optical telescopes," *Astrophys. J. Lett.*, vol. 196, no. 2, pp. L71–L75, 1975.
- [2] A. Quirrenbach, "Optical interferometry," *Annu. Rev. Astron. Astrophys.*, vol. 39, no. 1, pp. 353–401, 2001.
- [3] J. Monnier, "Optical interferometry in astronomy," *Rep. Prog. Phys.*, vol. 66, no. 5, pp. 789–857, 2003.
- [4] G. Perrin, "VLTI science highlights," in *Proc. Astrophys. Space Science: Science with the VLT in the ELT Era*, 2009, pp. 81–87.
- [5] F. Roddier, "The effects of atmospheric turbulence in optical astronomy," *Prog. Opt.*, vol. 19, pp. 281–376, 1981.
- [6] F. Delplancke, F. Derie, F. Paresce, A. Glindemann, F. Lévy, S. Lévêque, and S. Ménardi, "Prima for the VLTI—Science," *Astrophys. Space Sci.*, vol. 286, no. 1, pp. 99–104, 2003.
- [7] J. Dainty and A. Greenaway, "Estimation of spatial power spectra in speckle interferometry," *J. Opt. Soc. Amer.*, vol. 69, no. 5, pp. 786–790, 1979.
- [8] B. Wirtitzer, "Bispectral analysis at low light levels and astronomical speckle masking," *J. Opt. Soc. Amer. A*, vol. 2, no. 1, pp. 14–21, 1985.
- [9] K.-H. Hofmann and G. Weigelt, "Iterative image reconstruction from the bispectrum," *Astron. Astrophys.*, vol. 278, no. 1, pp. 328–339, 1993.
- [10] A. Lannes, E. Anterrieu, and P. Maréchal, "Clean and wipe," *Astron. Astrophys. Suppl.*, vol. 123, pp. 183–198, 1997.
- [11] D. Potts, G. Steidl, and M. Tasche, "Fast Fourier transforms for nonequispaced data: A tutorial," in *Modern Sampling Theory: Mathematics and Applications*, J. Benedetto and P. Ferreira, Eds., Cambridge, MA: Birkhäuser, 2001, pp. 249.
- [12] A. Thompson and R. Bracewell, "Interpolation and Fourier transformation of fringe visibilities," *Astron. J.*, vol. 79, no. 1, pp. 11–24, 1974.
- [13] R. Sramek and F. Schwab, "Imaging," in *Synthesis Imaging in Radio Astronomy*, R. Perley, F. Schwab, and A. Bridle, Eds., vol. 6, pp. 117–138, 1989.
- [14] S. Meimon, L. Mugnier, and G. Le Besnerais, "Convex approximation to the likelihood criterion for aperture synthesis imaging," *J. Opt. Soc. Amer. A*, vol. 22, no. 11, pp. 2348–2356, 2005.
- [15] J.-F. Giovannelli and A. Coulaes, "Positive deconvolution for superimposed extended source and point sources," *Astron. Astrophys.*, vol. 439, pp. 401–412, 2005.
- [16] P. Lawson, W. Cotton, C. Hummel, J. Monnier, M. Zhao, J. Young, H. Thorsteinsson, S. Meimon, L. Mugnier, G. Le Besnerais, E. Thiébaud, and P. Tuthill, "The 2004 optical/IR interferometry imaging beauty contest," *Bull. Amer. Astron. Soc.*, vol. 36, pp. 1605–1618, 2004.
- [17] A. Tarantola, *Inverse Problem Theory and Methods for Model Parameter Estimation*. Philadelphia, PA: SIAM, 2005.
- [18] J. Nocedal and S. Wright, *Numerical Optimization*. New York: Springer-Verlag, 2006.
- [19] J. Goodman, *Statistical Optics*. New York: Wiley, 1985.
- [20] T. Pauls, J. Young, W. Cotton, and J. Monnier, "A data exchange standard for optical (visible/IR) interferometry," *Publ. Astron. Soc. Pac.*, vol. 117, pp. 1255–1262, 2005.
- [21] J. Ables, "Maximum entropy spectral analysis," *Astron. Astrophys. Suppl.*, vol. 15, pp. 383–393, 1974.
- [22] S. Gull and J. Skilling, "The maximum entropy method," in *Measurement and Processing for Indirect Imaging*, J. Roberts, Ed., Cambridge, U.K.: Cambridge Univ. Press, 1984, p. 267.
- [23] R. Narayan and R. Nityananda, "Maximum entropy image restoration in astronomy," *Annu. Rev. Astron. Astrophys.*, vol. 24, pp. 127–170, 1986.
- [24] K. Horne, "Images of accretion discs. I. The eclipse mapping method," *Month. Notices Roy. Astron. Soc.*, vol. 213, pp. 129–141, 1985.
- [25] S. Gull, "Developments in maximum entropy data analysis," in *Maximum Entropy and Bayesian Methods*, J. Skilling, Ed., Dordrecht, The Netherlands: Kluwer, 1989, pp. 53–71.
- [26] J. Skilling and R. Bryan, "Maximum entropy image reconstruction: General algorithm," *Month. Notices Roy. Astron. Soc.*, vol. 211, no. 1, pp. 111–124, 1984.
- [27] A. Tikhonov and V. Arsenin, "Solution of ill-posed problems," in *Scripta Series in Mathematics*. Washington, D.C.: Winston & Sons, 1977.
- [28] G. Le Besnerais, S. Lacour, L. Mugnier, E. Thiébaud, G. Perrin, and S. Meimon, "Advanced imaging methods for long-baseline optical interferometry," *IEEE J. Select. Topics Signal Processing*, vol. 2, no. 5, pp. 767–780, 2008.
- [29] D. Strong and T. Chan, "Edge-preserving and scale-dependent properties of total variation regularization," *Inverse Problems*, vol. 19, pp. S165–S187, 2003.
- [30] E. Fomalont, "Earth-rotation aperture synthesis," *Proc. IEEE* (Special Issue on Radio and Radar Astronomy), vol. 61, no. 9, pp. 1211–1218, 1973.
- [31] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astron. Astrophys. Suppl.*, vol. 15, pp. 417–426, 1974.
- [32] U. Schwarz, "Mathematical-statistical description of the iterative beam removing technique (method CLEAN)," *Astron. Astrophys.*, vol. 65, pp. 345–356, 1978.
- [33] K. Marsh and J. Richardson, "The objective function implicit in the CLEAN algorithm," *Astron. Astrophys.*, vol. 182, pp. 174–178, 1987.
- [34] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.
- [35] B. Wakker and U. Schwarz, "The multi-resolution CLEAN and its application to the short-spacing problem in interferometry," *Astron. Astrophys.*, vol. 200, pp. 312–322, 1988.
- [36] J.-L. Starck, A. Bijaoui, B. Lopez, and C. Perrier, "Image reconstruction by the wavelet transform applied to aperture synthesis," *Astron. Astrophys.*, vol. 283, pp. 349–360, 1994.
- [37] T. Cornwell, "Multiscale CLEAN deconvolution of radio synthesis images," *IEEE J. Select. Topics Signal Processing*, vol. 2, no. 5, pp. 793–801, 2008.
- [38] J. Rich, W. de Blok, T. Cornwell, E. Brinks, F. Walter, I. Bagetakos, and R. Kenicutt, "Multi-scale CLEAN: A comparison of its performance against classical CLEAN on galaxies using THINGS," *Astron. J.*, vol. 136, no. 6, pp. 2897–2920, 2008.
- [39] N. Weir, "A multi-channel method of maximum entropy image restoration," in *Proc. Astron. Soc. Pacific Conf.: Astronomical Data Analysis Software and Systems I*, 1992, vol. 25, pp. 186–190.
- [40] T. Bontekoe, E. Koper, and D. Kester, "Pyramid maximum entropy images of IRAS survey data," *Astron. Astrophys.*, vol. 284, pp. 1037–1053, 1994.
- [41] E. Pantin and J.-L. Starck, "Deconvolution of astronomical images using the multiscale maximum entropy method," *Astron. Astrophys. Suppl.*, vol. 118, pp. 575–585, 1996.
- [42] J.-L. Starck, F. Murtagh, P. Querre, and F. Bonnarel, "Entropy and astronomical data analysis: Perspectives from multiresolution analysis," *Astron. Astrophys.*, vol. 368, pp. 730–746, 2001.
- [43] P. Magain, F. Courbin, and S. Sohy, "Deconvolution with correct sampling," *Astrophys. J.*, vol. 494, pp. 472–477, Feb. 1998.
- [44] N. Pirzkal, R. Hook, and L. Lucy, "GIRA—Two channel photometric restoration," in *Proc. Astron. Soc. Pacific Conf.: Astronomical Data Analysis Software and Systems IX*, 2000, vol. 216, pp. 655–658.
- [45] S. Lacour, E. Thiébaud, and G. Perrin, "High dynamic range imaging with a single-mode pupil remapping system: A self-calibration algorithm for highly redundant interferometric arrays," *Month. Notices Roy. Astron. Soc.*, vol. 374, no. 3, pp. 832–846, 2007.
- [46] A. Readhead and P. Wilkinson, "The mapping of compact radio sources from VLBI data," *Astrophys. J.*, vol. 223, no. 1, pp. 25–36, July 1978.
- [47] W. Cotton, "A method of mapping compact structure in radio sources using VLBI observations," *Astron. J.*, vol. 84, no. 8, pp. 1122–1128, Aug. 1979.
- [48] F. Schwab, "Adaptive calibration of radio interferometer data," in *Proc. SPIE*, 1980, vol. 231, pp. 18–25.
- [49] T. Cornwell and P. Wilkinson, "A new method for making maps with unstable radio interferometers," *Month. Notices Roy. Astron. Soc.*, vol. 196, pp. 1067–1086, 1981.
- [50] P. Campisi and K. Egiazarian, Eds., *Blind Image Deconvolution: Theory and Applications*. Boca Raton, FL: CRC, 2007.
- [51] T. Schulz, "Multiframe blind deconvolution of astronomical images," *J. Opt. Soc. Amer. A*, vol. 10, no. 5, pp. 1064–1073, 1993.
- [52] C. Haniff, "Least-squares Fourier phase estimation from the modulo 2π bispectrum phase," *J. Opt. Soc. Amer. A*, vol. 8, no. 1, pp. 134–140, 1991.
- [53] E. Thiébaud, "MiRA: An effective imaging algorithm for optical interferometry," in *Proc. SPIE: Astronomical Telescopes and Instrumentation*, 2008, vol. 7013, pp. 701311–701312.
- [54] D. Buscher, "Direct maximum-entropy image reconstruction from the bispectrum," in *Proc. IAU Symp. 158: Very High Angular Resolution Imaging*, 1994, pp. 91–93.
- [55] S. Meimon, L. Mugnier, and G. Le Besnerais, "Reconstruction method for weak-phase optical interferometry," *Opt. Lett.*, vol. 30, no. 14, pp. 1809–1811, 2005.
- [56] F. Baron and J. Young, "Image reconstruction at Cambridge University," in *Proc. SPIE: Astronomical Telescopes and Instrumentation*, 2008, vol. 7013, p. 144.
- [57] M. Ireland, J. Monnier, and N. Thureau, "Monte-Carlo imaging for optical interferometry," in *Proc. SPIE: Advances in Stellar Interferometry*, vol. 6268, 2008, pp. 62681T–62681T8.
- [58] E. Thiébaud, "Optimization issues in blind deconvolution algorithms," in *Proc. SPIE: Astronomical Data Analysis II*, vol. 4847, 2002, pp. 174–184.
- [59] J. Fienup, "Reconstruction of an object from the modulus of its Fourier transform," *Opt. Lett.*, vol. 3, no. 1, pp. 27–29, 1978.
- [60] P. Lawson, W. Cotton, C. Hummel, F. Baron, J. Young, S. Kraus, K.-H. Hofmann, G. Weigelt, M. Ireland, J. Monnier, E. Thiébaud, S. Rengaswamy, and O. Chesneau, "The 2006 interferometry image beauty contest," in *Proc. SPIE: Advances in Stellar Interferometry*, 2006, vol. 6268, pp. 62681U–62681U12.
- [61] W. Cotton, J. Monnier, F. Baron, K.-H. Hofmann, S. Kraus, G. Weigelt, S. Rengaswamy, E. Thiébaud, P. Lawson, W. Jaffe, C. Hummel, T. Pauls, Henrique, P. Tuthill, and J. Young, "2008 imaging beauty contest," in *Proc. SPIE: Astronomical Telescopes and Instrumentation*, 2008, vol. 7013, pp. 70131N–70131N14.