Image reconstruction with optical interferometers

E. Thiébaut

Centre de Recherche Astrophysique de Lyon

Abstract

This lecture is an introduction to image synthesis with emphasis set on the specific problems encountered in optical (visible/IR) interferometry. Image synthesis is first presented in the context of complex visibility data obtained with a known transfer function. This simple case is used to introduce the concept of regularization which is needed to lever degeneracies (due to missing spatial frequencies) of the image reconstruction problem. When the complex amplitude transmissions are unknown but vary slowly with respect to the exposure time, the self calibration method can recover both the amplitude transmissions and the image. When the complex amplitude transmissions vary too quickly, as it is generally the case in optical interferometry, the data consist in non-linear measurements such as the powerspectrum and the bispectrum which are insensitive to the fast varying aberrations set by the turbulence at the cost of some loss of Fourier phase information. We show how the inverse problem approach presented for the simpler cases can be extended to obtain effective image restoration algorithms.

Key words: aperture synthesis, imaging, inverse problem.

The first multi-telescope optical interferometer (Labeyrie, 1975) had two telescopes and could only measure the squared visibility (the *powerspectrum*) of some bright stars. With up to 6 telescopes, current optical interferometers (Monnier, 2003; Quirrenbach, 2009) provide enough data to consider image reconstruction at sub-milliarcsecond resolution of the observed objects which can be fainter and more complex in structure. There are an increasing number of astrophysical applications: stellar surfaces, environment of pre-main sequence or evolved stars, central regions of active galaxies, *etc.* Quirrenbach (2001), Monnier (2003), and Perrin (2009) give comprehensive reviews of optical interferometry and summarize some astrophysical results. For examples of image reconstructions of stellar surfaces and nearby star environments, see Lacour et al. (2008), Le Bouquin et al. (2009), and Lacour et al. (2009).

Multi-telescope interferometers do not directly provide images but sparse

Email address: thiebaut@obs.univ-lyon1.fr (E. Thiébaut)
URL: http://www-obs.univ-lyon1.fr/labo/perso/eric.thiebaut/ (E. Thiébaut)

measurements of the complex visibility of the observed objects (the Fourier transform of their brightness distribution; *cf.* Section 1). Hence reconstruction methods are needed to fully exploit these instruments. When the complex visibilities can be reliably measured, algorithms similar to those developed for radio-interferometry are able to cope with the missing information; that is, the voids in the sampled spatial frequencies. These methods can be derived in the general framework of inverse problem approach (*cf.* Section 2). At optical wavelengths, additional problems arise due to the additional loss of part of Fourier phase information. These issues had led to the development of specific algorithms for image reconstruction from the powerspectrum and the phase closure. These new algorithms can also be formally described in the same general framework (*cf.* Section 3).

1. Principle of optical interferometry

The instantaneous output of an optical interferometer is the so-called *complex visibilities* V of the interference fringes between two telescopes and is directly related to the Fourier transform $\hat{I}(\boldsymbol{\nu})$ of the object brightness distribution[1] $I(\boldsymbol{\theta})$. At instant t, the complex visibility $V_{j_1,j_2}(t)$ of the fringes given by the j_1-th and j_2-th telescopes writes:

$$V_{j_1,j_2}(t) = g_{j_1}(t)^\star \, g_{j_2}(t) \, \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2}(t)\right) \tag{1}$$

where the *gain* $g_j(t)$ is the complex amplitude throughput of the j-th telescope and $\boldsymbol{\nu}_{j_1,j_2}(t)$ is the spatial frequency measured by the pair of telescopes (j_1, j_2). Assuming that the diameters of the telescopes are much smaller than their projected separation (on a plane perpendicular to the line of sight), the measured spatial frequency writes:

$$\boldsymbol{\nu}_{j_1,j_2}(t) = \frac{\boldsymbol{r}_{j_2}(t) - \boldsymbol{r}_{j_1}(t)}{\lambda} \tag{2}$$

where $\boldsymbol{r}_j(t)$ is the projected position of j-th telescope and λ is the wavelength. Usually, there is a small number of recombined telescopes, hence interferometry yields a very sparse coverage of the spatial frequencies. Figure 2 shows a typical case of u-v coverage, so-called after the usual notation (u, v) for the coordinates of the spatial frequencies. Observe how the Earth rotation results in approximately elliptical tracks whereas changes in wavelength yield radial extension of the u-v coverage.

The transmission gains are dimensionless complex numbers:

$$g_j(t) = \tau_j(t) \, \mathrm{e}^{\mathrm{i}\,\phi_j(t)} \tag{3}$$

with amplitude $\tau_j(t) = |g_j(t)|$ and phase $\phi_j(t)$. Thanks to photometric calibration (requiring that the instantaneous flux transmitted by each telescope

[1] It is assumed here that the object brightness distribution remains the same during the observations.

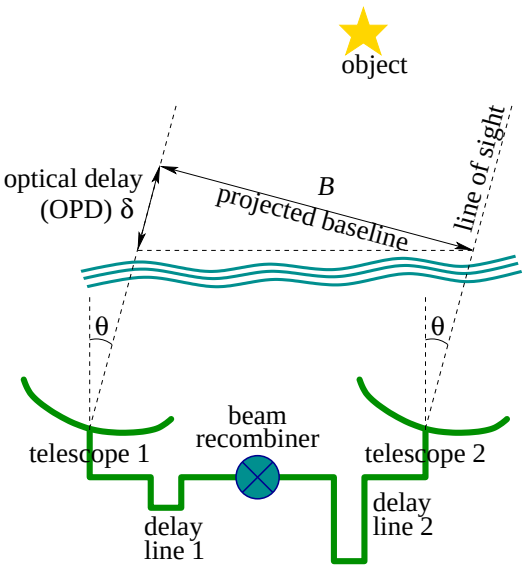

Figure 1: Geometrical layout of an interferometer. B is the projected baseline, θ is the view angle and δ is the geometrical optical path difference which is compensated by the delay lines.

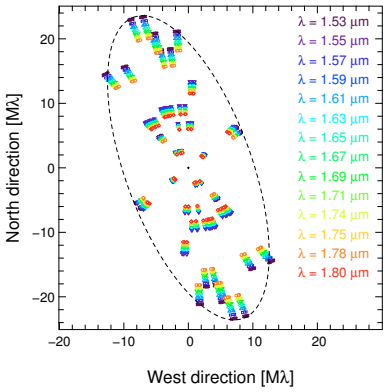

Figure 2: u-v coverage at different wavelengths in the H-band for observations of Arcturus in May 2006 with IOTA and achieved with a maximum baseline of 37.7 m (Lacour et al., 2008).

be measured simultaneously), the amplitude of the effective gain can be made almost equal to one: $\tau_j(t) = |g_j(t)| \simeq 1$. Besides, the phase of the gain is due to an optical path difference (OPD) $\delta_j(t)$ which is variable and different for each telescope:

$$\phi_j(t) = 2\,\pi\,\delta_j(t)/\lambda\,. \tag{4}$$

The OPD is caused by tracking errors in the delay lines and by the atmospheric turbulence. Hence, it may be quickly variable and is mostly unknown.

Image synthesis from interferometric data has to cope with two main issues. First, the sparseness of the u-v coverage implies that the measured complex visibilities alone are not sufficient to uniquely define an image. In effect, any brightness distribution of which the Fourier transform is compatible with data at sampled spatial frequencies and which takes any value elsewhere is a legitimate image. In other words, the image restoration problem is ill-posed and additional constraints are required to select a unique image among all those which are in agreement with the data. We deal with this problem in Section 2. Second, the complex gains may be poorly calibrated or unknown and may vary quickly during a single exposure. This is a very important issue at short (visible/IR) wavelengths because of the atmospherics turbulence. Depending on how fast the complex gains change, two different approaches can be used to overcome this problem. For gains approximately constant during an exposure, self-calibration methods (*cf.* Section 3.2) guess the complex transmissions from the data and the current image. Then, *standard* image restoration methods (*cf.* Section 2) can be used to build a new image given the complex gains and the data. For fast varying throughput, instead of the complex visibility, other quantities such as the powerspectrum and the phase closure (bispectrum) which are insensitive to OPD errors are integrated during each exposure. This is to the cost of missing part of the Fourier phase information (*cf.* Section 3.3). Restoring an image from the powerspectrum and the bispectrum (or closure phase) require specific algorithms described in Section 3.5.

2. Image synthesis

We introduce here means to properly solve the problem of synthesizing an image from sparse measurements of its Fourier transform. For the moment, we assume that the complex gains are known and that the exposure time is short enough to freeze any temporal evolution. The cases when these assumptions do no apply is considered later in Section 3.

2.1. Model of the data

From the assumptions, the expected complex visibilities at the sampled frequencies write:

$$V_{j_1,j_2,m} = \langle V_{j_1,j_2}(t)\rangle_m \simeq V_{j_1,j_2}(t_m) \simeq g^\star_{j_1,m}\, g_{j_2,m}\, \hat{I}\,(\boldsymbol{\nu}_{j_1,j_2,m}) \tag{5}$$

where $\langle\,\rangle_m$ denotes averaging during m-th exposure at mean time $t_m = \langle t\rangle_m$ and:

$$g_{j,m} = \langle g_j(t)\rangle_m \simeq g_j(t_m)\,, \tag{6}$$

$$\boldsymbol{r}_{j,m} = \langle \boldsymbol{r}_j(t)\rangle_m \simeq \boldsymbol{r}_j(t_m)\,, \tag{7}$$

$$\boldsymbol{\nu}_{j_1,j_2,m} = \langle \boldsymbol{\nu}_{j_1,j_2}(t)\rangle_m \simeq \left(\boldsymbol{r}_{j_2,m} - \boldsymbol{r}_{j_1,m}\right)/\lambda\,. \tag{8}$$

Accounting for the noise and approximations in our description, the measured complex visibilities are:

$$V^{\mathsf{data}}_{j_1,j_2,m} = g^{\star}_{j_1,m}\, g_{j_2,m}\, \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right) + e_{j_1,j_2,m} \tag{9}$$

where $\boldsymbol{e}$ denotes the errors, that is, all discrepancies between the data and the model $g^{\star}_{j_1,m}\, g_{j_2,m}\, \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)$ in particular noise and modeling errors.

Looking closely at Eq. (9), one can note that image restoration given the data is very similar to a (multi-frame) deconvolution problem where the terms $g^{\star}_{j_1,m}\, g_{j_2,m}$ would play the role of the optical transfer function (OTF) at spatial frequency $\boldsymbol{\nu}_{j_1,j_2,m}$. That is why some authors use the terms *image deconvolution* instead of *image synthesis*.

2.2. Image model and its Fourier transform

Now that we have a model linking the data and the object of interest, we need a parametrized model of the *image*, that is, the object brightness distribution. However it is not question here to use a restricted parametric model such as a limb darkening law but rather to discretize the brightness distribution of the object (as the *pixels* of a numerical image are a discretized version of the observed scene). For instance, a general linear expansion yields:

$$I(\boldsymbol{\theta}) = \sum_{n=1}^{N} x_n\, b_n(\boldsymbol{\theta}) \quad \overset{\text{F.T.}}{\longrightarrow} \quad \hat{I}(\boldsymbol{\nu}) = \sum_{n=1}^{N} x_n\, \hat{b}_n(\boldsymbol{\nu})\,, \tag{10}$$

where $\{b_n(\boldsymbol{\theta}); n = 1,\ldots,N\}$ are basis functions and $\boldsymbol{x} \in \mathbb{R}^N$ are the image parameters. Given a grid of angular directions $\mathcal{G} = \{\boldsymbol{\theta}_n; n = 1,\ldots,N\}$, a *grid model* is obtained by using a single *building block function* $b(\Delta\boldsymbol{\theta})$ shifted at the grid nodes:

$$I(\boldsymbol{\theta}) = \sum_{n=1}^{N} x_n\, b(\boldsymbol{\theta} - \boldsymbol{\theta}_n) \quad \overset{\text{F.T.}}{\longrightarrow} \quad \hat{I}(\boldsymbol{\nu}) = \hat{b}(\boldsymbol{\nu}) \sum_{n=1}^{N} x_n\, \mathrm{e}^{-2\,\mathrm{i}\,\pi\,\boldsymbol{\theta}_n\cdot\boldsymbol{\nu}}\,. \tag{11}$$

Using an equispaced grid, the usual pixelized image representation is obtained with the pixel shape given by the building block function $b(\Delta\boldsymbol{\theta})$. The model in Eq. (11) is however more general. For instance, the basis function can be used to set the effective resolution of the synthesized image whereas the grid spacing is purposely much smaller to avoid grid biases (Lannes et al., 1997).

Whatever is the choice of the parametrization, for the set of sampled spatial frequencies $\mathcal{L} = \{\boldsymbol{\nu}_k; k = 1, \ldots, K\}$, the model of the complex visibility is linear:

$$V_k(\boldsymbol{x}) = \hat{I}(\boldsymbol{\nu}_k) = \sum_{n=1}^{N} A_{k,n}\, x_n \,, \tag{12}$$

where the coefficients of the matrix $\mathbf{A} \in \mathbb{C}^{K \times N}$ are:

$$A_{k,n} = \hat{b}_n(\boldsymbol{\nu}_k) \,, \tag{13}$$

for the general linear expansion in Eq. (10); and

$$A_{k,n} = \hat{b}(\boldsymbol{\nu}_k)\, \mathrm{e}^{-2\,\mathrm{i}\,\pi\,\boldsymbol{\theta}_n \cdot \boldsymbol{\nu}_k} \,, \tag{14}$$

for the *building block model* in Eq. (11). The data equation (9) now writes:

$$\begin{aligned} V^{\mathsf{data}}_{j_1,j_2,m} &= g^{\star}_{j_1,m}\, g_{j_2,m}\, V_k(\boldsymbol{x}) + e_{j_1,j_2,m} \\ &= g^{\star}_{j_1,m}\, g_{j_2,m}\, (\mathbf{A} \cdot \boldsymbol{x})_k + e_{j_1,j_2,m} \end{aligned}$$

for any triplet (j_1, j_2, m) that belongs to the set $\mathcal{B}_k$ of telescope indices and exposure index which sample the k-th spatial frequency, that is:

$$(j_1, j_2, m) \in \mathcal{B}_k \quad \Longleftrightarrow \quad \frac{\boldsymbol{r}_{j_2,m} - \boldsymbol{r}_{j_1,m}}{\lambda} = \boldsymbol{\nu}_k$$

or equivalently:

$$\mathcal{B}_k \stackrel{\text{def}}{=} \left\{ (j_1, j_2, m) \in \mathcal{A}^2 \times \mathcal{E};\ \frac{\boldsymbol{r}_{j_2,m} - \boldsymbol{r}_{j_1,m}}{\lambda} = \boldsymbol{\nu}_k \right\} \tag{15}$$

where $\mathcal{A}$ and $\mathcal{E}$ are respectively the lists of indices of apertures (telescopes or antennae) and exposures. Introducing $\mathcal{B}_k$ and the list of observed frequencies $\mathcal{L} = \{\boldsymbol{\nu}_k; k = 1, \ldots, K\}$ is a simple means to account for all possible cases (*e.g.* with or without redundancies, multiple data sets, observations from different interferometers *etc.*).

Finally, the data equation (9) can be written in a very general way as:

$$\boldsymbol{d} = \boldsymbol{m}(\boldsymbol{x}; \boldsymbol{g}) + \boldsymbol{e} \tag{16}$$

where $\boldsymbol{d} = \{V^{\mathsf{data}}_{j_1,j_2,m}; \forall (j_1, j_2, m)\}$ is the set of measured complex visibilities, $\boldsymbol{m}(\boldsymbol{x}; \boldsymbol{g})$ is the model which depends on the parameters $\boldsymbol{x}$ and on $\boldsymbol{g} = \{g_{j,m}; \forall (j, m)\}$ which characterizes the instrumental response, and $\boldsymbol{e} = \{e_{j_1,j_2,m}; \forall (j_1, j_2, m)\}$ accounts for errors due to noise and approximations. The model of the complex visibilities in Eq. (12) is linear with respect to the parameters and so is the model of the data, hence Eq. (16) simplifies to an affine equation:

$$\boldsymbol{d} = \mathbf{G} \cdot \mathbf{A} \cdot \boldsymbol{x} + \boldsymbol{e} = \mathbf{M} \cdot \boldsymbol{x} + \boldsymbol{e} \tag{17}$$

where $\mathbf{M} = \mathbf{G} \cdot \mathbf{A}$ and $\mathbf{G}$ is a diagonal matrix which accounts for the amplitude transmission gains:

$$G_{j_1,j_2,m,k} = \begin{cases} g^\star_{j_1,m}\, g_{j_2,m} & \text{if } (j_1, j_2, m) \in \mathcal{B}_k \\ 0 & \text{else.} \end{cases} \tag{18}$$

Compared to more conventional imaging techniques, $\mathbf{G}$ and $\mathbf{A}$ are the analogous for aperture synthesis of the optical transfer function (OTF) and of the Fourier transform respectively.

2.3. Fourier interpolation and regridding

We will see in what follows that image restoration is achieved by iterative algorithms which involve many multiplications by the model matrix $\mathbf{A}$ or its conjugate transpose. The matrix $\mathbf{A}$ has $K \times N$ complex coefficients, where K is the number of different spatial frequencies that have been sampled during the observations and N is the number of image parameters (*e.g.* the pixels). Accounting for the fact that the parameters $\boldsymbol{x}$ are real, matrix multiplication by $\mathbf{A}$ involves $3\,K\,N$ operations. This may be prohibitive in terms of memory requirements and computational burden. For instance, a 256×256 image and 1000 measurements would require ~ 1 gigabyte of memory to store the matrix and ~ 200 millions of floating point operations to apply it.

Under certain restrictions, multiplication by $\mathbf{A}$ can be approximated by a discrete Fourier transform (DFT) which can be computed in a much faster way (and with negligible memory needs) thanks to the fast Fourier transform (FFT). This however imposes that the image and its spectrum be periodical discretized functions and is restricted to equally spaced sampling of the direct and conjugate space with sampling steps such that:

$$\Delta\nu_d = \frac{1}{N_d\,\Delta\theta_d} = \frac{1}{\Omega_d}\,, \tag{19}$$

where N_d is the number of pixels along d-th dimension, $\Delta\nu_d$ and $\Delta\theta_d$ are the spatial frequency and angular direction steps along this dimension, and Ω_d is the width of the field of view (FOV) in this dimension. The counterpart of Eq. (19) is:

$$\Delta\theta_d = \frac{1}{N_d\,\Delta\nu_d} = \frac{1}{2\,\nu_d^{\mathsf{cut}}}\,, \tag{20}$$

where $\nu_d^{\mathsf{cut}} = \max|\nu_d|$ is the Nyquist frequency along d-th dimension which is the maximum frequency in the model without aliasing. Note that, to avoid anisotropic effects, the number of elements and the pixel size are usually chosen to be the same along the two spatial dimensions. We therefore drop the index d in what follows.

Using the FFT imposes that the discrete Fourier transform of the model image be sampled on an equally spaced grid of so-called *frequels*. It is unlikely that observed spatial frequencies do coincide with anyone of the frequels. The complex visibility at observed spatial frequencies must therefore be interpolated

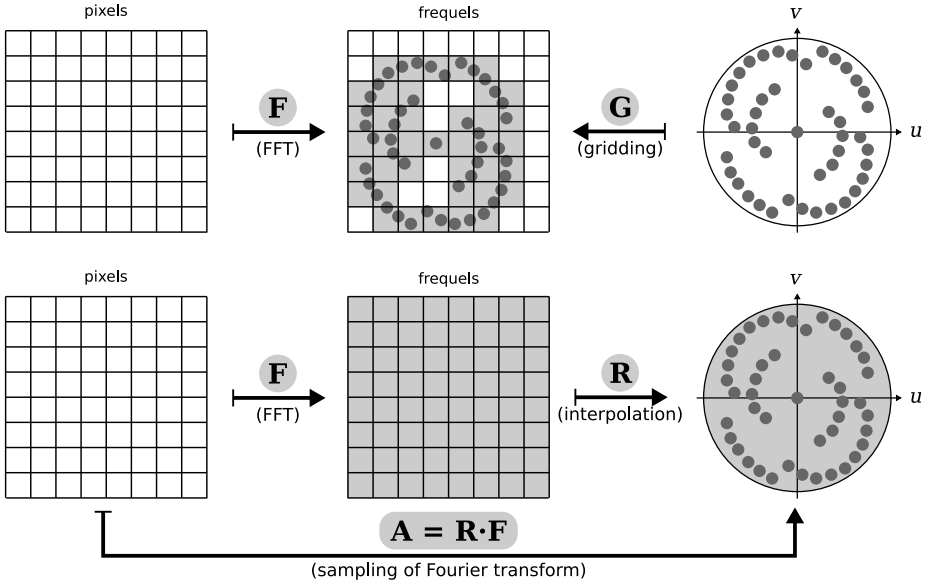

Figure 3: Regridding of the data (top) and Fourier interpolation (bottom).

from the FFT of the image. A first possibility is to use the nearest discrete frequency. This however leads to a very coarse approximation (especially for the lowest frequencies) unless a very large field of view (FOV) is used. Interpolation from the few (usually 4 or 16) nearest frequels is cheap to compute[2] and gives a substantially more accurate approximation. Denoting by $\mathbf{R}$ the interpolation operator in the spatial frequency plane, the model matrix $\mathbf{A}$ is approximated by:

$$\mathbf{A} \simeq \mathbf{R} \cdot \mathbf{F}\,, \tag{21}$$

where $\mathbf{F}$ is the discrete Fourier transform operator (computed by FFT). Note that Fourier interpolation by the operator $\mathbf{R}$ is equivalent to convolving the Fourier spectrum with the interpolation kernel (a square box for nearest neighbor interpolation, a bi-linear spline for bi-linear interpolation, *etc.*). Hence $\mathbf{R}$ does some *Fourier smoothing* which implies that the model image is implicitly multiplied by a tapering function equals to the inverse Fourier transform of the interpolation kernel. This set a limit on the effective field of view of the synthesized image. If the basis function $b(\Delta\boldsymbol{\theta})$ is not a Dirac distribution (delta function), then multiplication by its Fourier transform, see Eq. (14), can be accounted for by $\mathbf{R}$.

If P is the number of neighbor frequels considered in the interpolation of a sampled spatial frequency, $\mathbf{R}$ can have has few as $K \times P$ non zero coefficients

[2]To interpolate a given spatial frequency, the number of operations scales as the number of nearest frequels considered.

and necessitates $2\,(2\,P-1)\,K \sim 4\,K\,P$ operations to be applied. The memory requirements of the FFT are very modest (though it depends on the particular implementation) and its computational cost scales as $\eta\,N\,\log N$ with, *e.g.*, $\eta \sim 5$ for a real-complex transform by FFTW (Frigo and Johnson, 2005). The approximation $\mathbf{R}\cdot\mathbf{F}$ of the model matrix is therefore much more inexpensive to store and apply than the exact transform by $\mathbf{A}$.

In radio astronomy a different technique called *regridding* (Thompson and Bracewell, 1974; Sramek and Schwab, 1989) is generally used and which consists in interpolating the data onto the grid of frequels. The advantage is that, when there are a great number of measurements, the number of pseudo-data is reduced which may speed up further computations. There are however a number of drawbacks to the regridding technique. (i) Strictly speaking, the pseudo-data are correlated even if the original data are not. These correlations are usually ignored in further processing and the pseudo-data are assumed to be independent which results in a poor approximation of the real likelihood. (ii) It is not easy to resample non-linear estimators such as the phase closure and powerspectrum at least because these quantities are not smooth (the phase and partial derivatives of the powerspectrum are undefined at spatial frequencies where the complex visibility is zero, and the phase is further wrapped). Figure 3 schematically compares regridding (that is, interpolation of the data) to Fourier interpolation (that is, interpolation of the model).

Similar needs in crystallography, tomography and bio-medical imaging have lead to a number of significant progresses in the development of fast algorithms to approximate the Fourier transform of non-equispaced data (Potts et al., 2001). These algorithms may be worth considering for interferometric imaging as they are not only fast but also achieve good numerical precision and would give better approximations for $\mathbf{A}$ than simple Fourier interpolation, or regridding.

2.4. Choosing the resolution and field of view of the synthetic image

Another consequence of using FFT's to compute the model of the complex visibility is that the choice of the number of pixels and of the spatial sampling dictates the resolution of the model in the u-v plane, see Eq. (19) or Eq. (20). To avoid that the model imposes too much spectral smoothness (with exact Fourier transform) or field of view aliasing (with the discrete Fourier transform), the size of the field of view must be chosen large enough to properly sample the u-v plane. Also, to not be too much biased by the particular image model, *e.g.*, by the pixel shape $b(\Delta\boldsymbol{\theta})$, the spatial resolution of the model should be well beyond the limit imposed by the longest baseline:

$$\Delta\theta \ll \frac{\lambda}{2\,B_{\mathsf{max}}} \tag{22}$$

where $B_{\mathsf{max}} = \max_{j_1,j_2,t} |\boldsymbol{r}_{j_1}(t) - \boldsymbol{r}_{j_2}(t)|$ is the maximum projected separation between interfering telescopes. As a rule of thumb, a super resolution by, at

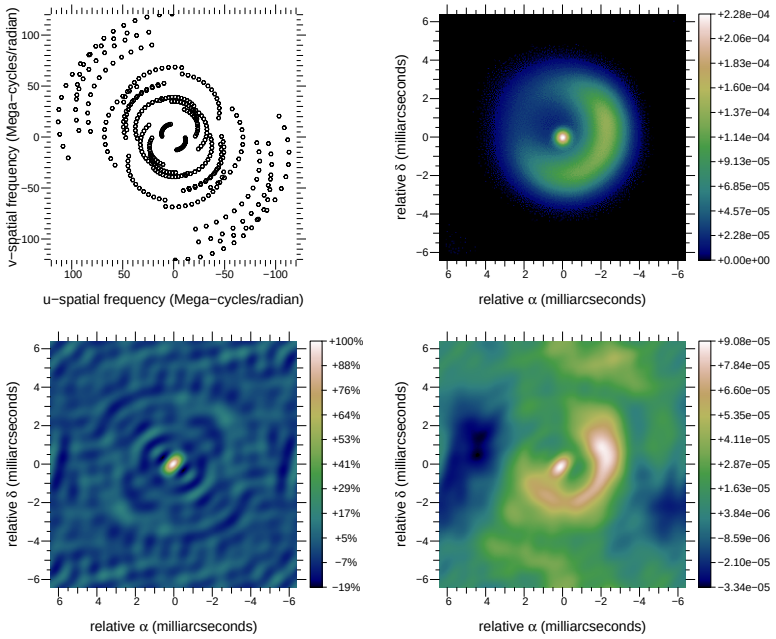

Figure 4: Top left: u-v coverage. Top right: observed object. Bottom left: dirty beam. Bottom right: dirty image. Object model and u-v coverage are from the *2004' Beauty Contest* (Lawson et al., 2004).

least, a factor of 2 is usually used and the pixel size can be given by:

$$\Delta\theta \lesssim \frac{\lambda}{4\,B_{\mathsf{max}}}\,. \tag{23}$$

2.5. Maximum likelihood yields "dirty" images

Now that we are equipped with a model of the measured complex visibilities related to our image parameters, we can attempt to derive practical means to recover the image from the data. Owing to measurement noise and model approximations, we expect some discrepancy between our model and the actual data and a criterion is needed to select the *best* image given the data. For instance, we can seek for the model parameters that maximize the probability of having observed the data. This leads to the so-called *maximum likelihood* (ML) solution:

$$\boldsymbol{x}_{\mathsf{ML}} = \arg\max_{\boldsymbol{x}} \Pr(\boldsymbol{d}|\boldsymbol{x};\boldsymbol{g}) \tag{24}$$

where the mathematical notation "$\ldots = \arg\max_{\boldsymbol{x}} \ldots$" means that the left hand side is equal to the particular value of $\boldsymbol{x}$ that maximizes the right hand side expression, the *data vector* $\boldsymbol{d} = \{V^{\mathsf{data}}_{j_1,j_2,m}; \forall(j_1,j_2,m)\}$ collects the measured complex visibilities, $\boldsymbol{g} = \{g_{j,m}; \forall(j,m)\}$ characterizes the instrumental response, and $\boldsymbol{x}$ are the image parameters. Probabilities are numerically unpractical,

it is common practice to use the opposite of their logarithm instead. Hence, introducing the maximum likelihood penalty:

$$f_{\mathsf{ML}}(\boldsymbol{x}) \stackrel{\text{def}}{=} -c_1 \, \log \Pr(\boldsymbol{d}|\boldsymbol{x};\boldsymbol{g}) + c_0 \,, \tag{25}$$

where $c_1 > 0$ and c_0 are arbitrary constants chosen to simplify the expression of f_{ML}, the maximum likelihood solution becomes:

$$\boldsymbol{x}_{\mathsf{ML}} = \arg\min_{\boldsymbol{x}} f_{\mathsf{ML}}(\boldsymbol{x}) \,. \tag{26}$$

Assuming that errors $\boldsymbol{e}$, see Eq. (16), have a centered Gaussian distribution yields:

$$f_{\mathsf{ML}}(\boldsymbol{x}) = [\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x};\boldsymbol{g})]^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot [\boldsymbol{d} - \boldsymbol{m}(\boldsymbol{x};\boldsymbol{g})] \tag{27}$$

where $\boldsymbol{m}(\boldsymbol{x};\boldsymbol{g})$ is the parametrized model of the data and the weighting matrix $\mathbf{W}_{\mathsf{err}} = \mathbf{C}_{\mathsf{err}}^{-1}$ is the inverse of the covariance matrix of the errors $\mathbf{C}_{\mathsf{err}} = \langle \boldsymbol{e} \cdot \boldsymbol{e}^{\mathrm{T}} \rangle$. There is an additional issue to take care of because we are dealing with complex data. In fact, since complex numbers are just pairs of reals (the real and imaginary parts of the complex), complex valued vectors (such as $\boldsymbol{d}$, $\boldsymbol{m}$ and $\boldsymbol{e}$) can be *flattened* into ordinary real vectors (with twice the number of complex elements) to use ordinary linear algebra notation for the dot product and transpose operation. This is what is assumed in Eq. (27).

From Eq. (17), the model $\boldsymbol{m}(\boldsymbol{x};\boldsymbol{g}) = \mathbf{M}(\boldsymbol{g}) \cdot \boldsymbol{x}$ is linear and the maximum likelihood penalty writes:

$$f_{\mathsf{ML}}(\boldsymbol{x}) = [\boldsymbol{d} - \mathbf{M}(\boldsymbol{g}) \cdot \boldsymbol{x}]^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot [\boldsymbol{d} - \mathbf{M}(\boldsymbol{g}) \cdot \boldsymbol{x}] \,. \tag{28}$$

Then, the minimum of $f_{\mathsf{ML}}(\boldsymbol{x})$ with respect to the parameters $\boldsymbol{x}$ solves the so-called *normal equations*:

$$\mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \mathbf{M} \cdot \boldsymbol{x}_{\mathsf{ML}} = \mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \boldsymbol{d} \,, \tag{29}$$

obtained by stated that the gradient of $f_{\mathsf{ML}}(\boldsymbol{x})$ with respect to $\boldsymbol{x}$ must be null at the solution $\boldsymbol{x}_{\mathsf{ML}}$ since it is a stationary point. Due to voids in the u-v coverage, the matrix $\mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \mathbf{M}$ is singular and there is no unique solution to the normal equations. However, since $\mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \mathbf{M}$ is positive (semi-definite), all the solutions of the normal equations are local minima of $f_{\mathsf{ML}}(\boldsymbol{x})$. This means that, as we already have intuitively guessed, the data alone are not sufficient to recover an image without ambiguity. Additional constraints or assumptions are needed to overcome this problem. The simplest solution is to use the pseudo-inverse of $\mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \mathbf{M}$ and leads to:

$$\boldsymbol{x}_{\mathsf{dirty}} = [\mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \mathbf{M}]^{\dagger} \cdot \mathbf{M}^{\mathrm{T}} \cdot \mathbf{W}_{\mathsf{err}} \cdot \boldsymbol{d} \tag{30}$$

which is called the *dirty map* or *dirty image* for reasons that we clarify later. By using the pseudo-inverse, $\boldsymbol{x}_{\mathsf{dirty}}$ is, in fact, the solution of the modified optimization problem:

$$\boldsymbol{x}_{\mathsf{dirty}} = \lim_{\mu \to 0^+} \arg\min_{\boldsymbol{x}} \left\{ f_{\mathsf{ML}}(\boldsymbol{x}) + \mu \, \|\boldsymbol{x}\|_2^2 \right\} \tag{31}$$

where $\|\boldsymbol{x}\|_2 = \sqrt{\boldsymbol{x}^{\mathrm{T}} \cdot \boldsymbol{x}}$ is the Euclidean (ℓ_2) norm of $\boldsymbol{x}$. Hence the dirty map is the minimum norm image that maximizes the likelihood of the data.

Computation and interpretation of the dirty image is very easy in the following simplified case. First, we assume white Gaussian errors, hence the weighting matrix is proportional to the identity, $\mathbf{W}_{\mathsf{err}} \propto \mathbf{I}$. Second, we consider *calibrated data* (all the gains are equal to one) with observed spatial frequencies being a subset of the frequel grid, hence the model matrix becomes $\mathbf{M} = \mathbf{A} = \mathbf{S} \cdot \mathbf{F}$ where $\mathbf{F}$ is the FFT operator and $\mathbf{S}$ is a sub-sampling operator which selects measured spatial frequencies among all the frequels (its coefficients are only 0's or 1's). Then, the dirty map is simply obtained by inverse Fourier transforming the complex visibilities with values set to zero at missing spatial frequencies:

$$\boldsymbol{x}_{\mathsf{dirty}} = \mathbf{F}^{-1} \cdot \mathbf{S}^{\mathrm{T}} \cdot \boldsymbol{d}\,. \tag{32}$$

which is straightforward to obtain as $\mathbf{F}$ is invertible and $\mathbf{S}^{\dagger} = \mathbf{S}^{\mathrm{T}}$. Figure 4 shows that the resulting image has a lot of artifacts hence its name of *dirty map*. Replacing the data vector $\boldsymbol{d}$ by its expression in Eq. (17) and taking the expectation of Eq. (32) yields:

$$\langle \boldsymbol{x}_{\mathsf{dirty}} \rangle = \mathbf{F}^{-1} \cdot \mathbf{S}^{\mathrm{T}} \cdot \mathbf{S} \cdot \mathbf{F} \cdot \boldsymbol{x}_{\mathsf{true}} \tag{33}$$

where $\boldsymbol{x}_{\mathsf{true}}$ is the *true* image. $\mathbf{S}^{\mathrm{T}} \cdot \mathbf{S}$ being a diagonal matrix, $\langle \boldsymbol{x}_{\mathsf{dirty}} \rangle$ is just the true brightness distribution convolved by an equivalent point spread function (PSF), the so-called *dirty beam*:

$$\boldsymbol{h}_{\mathsf{dirty}} = \mathbf{F}^{-1} \cdot \mathrm{diag}(\mathbf{S}^{\mathrm{T}} \cdot \mathbf{S})\,, \tag{34}$$

where $\mathrm{diag}(\mathbf{A})$ is the vector built from the diagonal of matrix $\mathbf{A}$. The 2-D OTF $\hat{\boldsymbol{h}}_{\mathsf{dirty}} = \mathrm{diag}(\mathbf{S}^{\mathrm{T}} \cdot \mathbf{S})$ is equals to 0 or 1 depending whether the corresponding frequel has been sampled by the data or not.

Clearly (*cf.* Fig. 4), the dirty image is not a satisfactorily solution: we need better means to select the best image among all those compatible with the data. This is the purpose of the next section.

2.6. Maximum a posteriori

Coming back to probabilities, additional constraints for the sought image can be set by assigning a high or a low probability to images depending whether they match the priors or not. In order to account for such a prior probability (which does not depend on the data) and for the data, we are led to select the most probable image given the data and the instrumental response which is the so-called *maximum a posteriori* (MAP) solution:

$$\boldsymbol{x}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{x}} \Pr(\boldsymbol{x}|\boldsymbol{d};\boldsymbol{g})\,. \tag{35}$$

By Bayes' theorem and assuming the instrumental response is known, the joint probability of the data and the parameters expands as:

$$\Pr(\boldsymbol{x},\boldsymbol{d};\boldsymbol{g}) = \Pr(\boldsymbol{d};\boldsymbol{g})\ \Pr(\boldsymbol{x}|\boldsymbol{d};\boldsymbol{g}) = \Pr(\boldsymbol{x})\ \Pr(\boldsymbol{d}|\boldsymbol{x};\boldsymbol{g}) \tag{36}$$

where it has been accounted for the fact that the image parameters $\boldsymbol{x}$ do not depend *a priori* on the instrumental transmissions $\boldsymbol{g}$. Moreover, the probability of the data does not depend on the sought parameters, hence:

$$\boldsymbol{x}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{x}} \Pr(\boldsymbol{d}|\boldsymbol{x};\boldsymbol{g})\ \Pr(\boldsymbol{x}) \tag{37}$$

and taking the negative log-probabilities:

$$\boldsymbol{x}_{\mathsf{MAP}} = \arg\min_{\boldsymbol{x}} \left\{ f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) + f_{\mathsf{prior}}(\boldsymbol{x}) \right\} \tag{38}$$

where

$$f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) = -c_1\ \log \Pr(\boldsymbol{d}|\boldsymbol{x};\boldsymbol{g}) + c_0 = f_{\mathsf{ML}}(\boldsymbol{x})\,, \tag{39}$$

$$f_{\mathsf{prior}}(\boldsymbol{x}) = -c_1\ \log \Pr(\boldsymbol{x}) + c_0'\,. \tag{40}$$

The posterior penalty is the sum of two terms: a *likelihood* term $f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) = f_{\mathsf{ML}}(\boldsymbol{x})$ which enforces agreement of the model with the data, and a *regularization* term $f_{\mathsf{prior}}(\boldsymbol{x})$ which accounts for prior constraints. Compared to the maximum likelihood case (*cf.* Section 2.5), the additional regularization penalty helps to select, among all image model compatible with the data, the one which most fit our priors.

2.7. Pragmatic Bayesian approach

The previous reasoning is based on strict Bayesian approach but, in practice, while the statistics of the data is (approximately) known, the exact a priori statistics is often unknown[3]. At least, the priors must have some mathematical properties: (i) they must lever degeneracies due to the ill-posedness of the problem in order to lead to a unique solution; (ii) they must avoid noise amplification for ill-conditioned inverse problems[4]. For instance, the regularization can impose sufficient smoothness to avoid noise amplification, or compactness which helps filling voids in u-v coverage. In interferometry, the problem is mostly ill-posed because of the sparseness of the u-v coverage and other missing data. These considerations give qualitative hints to the kind of required priors, yet we do not know to what level they must be imposed.

In words, we want to match the priors (*e.g.* the restored image must be compact and/or smooth) as much as possible, that is, minimize the regularization penalty:

$$\min_{\boldsymbol{x}} f_{\mathsf{prior}}(\boldsymbol{x}) \tag{41}$$

whereas we want that the model remains compatible with the data:

$$f_{\mathsf{data}}(\boldsymbol{x}) \leq \eta_{\mathsf{data}} \tag{42}$$

[3] Despite theoretical arguments that could be invoked, see for instance the debate in radio-astronomy to justify the definition of the entropy of an image

[4] The problem is said to be *ill-conditioned* when a small change in the data yields an arbitrarily large change in the solution obtained by direct inversion, that is, without regularization.

where η_{data} is set according to the noise level, and, possibly, accounting for strict constraints such as positivity and normalization:

$$\boldsymbol{x} \geq 0 \quad \text{and} \quad \sum\nolimits_n x_n = 1\,, \tag{43}$$

where $\boldsymbol{x} \geq 0$ means: $\forall n,\ x_n \geq 0$. Putting all these requirements together, we are lead to solve the constrained optimization problem:

$$\boldsymbol{x}^+ = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathcal{X}} f_{\mathsf{prior}}(\boldsymbol{x}) \quad \text{s.t.} \quad f_{\mathsf{data}}(\boldsymbol{x}) \leq \eta_{\mathsf{data}}\,, \tag{44}$$

where $\mathcal{X} = \{\boldsymbol{x} \in \mathbb{R}^N; \boldsymbol{x} \geq 0, \sum_n x_n = 1\}$ is the set of feasible solutions according to Eq. (43). The Lagrangian of this constrained optimization problem writes:

$$\mathcal{L}(\boldsymbol{x}; \ell) = f_{\mathsf{prior}}(\boldsymbol{x}) + \ell \left(f_{\mathsf{data}}(\boldsymbol{x}) - \eta_{\mathsf{data}} \right) \tag{45}$$

where ℓ is the Lagrange multiplier associated to the constraint $f_{\mathsf{data}}(\boldsymbol{x}) \leq \eta_{\mathsf{data}}$. If the constraint is *active*[5], then $\ell > 0$ and such that $f_{\mathsf{data}}(\boldsymbol{x}) = \eta_{\mathsf{data}}$ (Nocedal and Wright, 2006; Bonnans et al., 2006). Taking $\mu = 1/\ell$ and dropping the constant η_{data} (which does not depend on $\boldsymbol{x}$), the solution is obtained by solving one of the following equivalent optimization problems:

$$\begin{aligned}
\boldsymbol{x}^+ &= \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathcal{X}} \{ f_{\mathsf{prior}}(\boldsymbol{x}) + \ell\, f_{\mathsf{data}}(\boldsymbol{x}) \} \\
&= \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathcal{X}} \{ f_{\mathsf{data}}(\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) \} \\
&= \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}; \mu)\,,
\end{aligned}$$

where

$$f(\boldsymbol{x}; \mu) = f_{\mathsf{data}}(\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) \tag{46}$$

is the penalty function and $\mu > 0$ has to be tuned to match the constraint $f_{\mathsf{data}}(\boldsymbol{x}) = \eta_{\mathsf{data}}$. These developments also show that we can equivalently consider that we are solving the problem of maximizing the likelihood of the data, that is, minimize $f_{\mathsf{data}}(\boldsymbol{x})$, subject to the constraint that the prior be below a preset level η_{prior}:

$$\boldsymbol{x}^+ = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathcal{X}} f_{\mathsf{data}}(\boldsymbol{x}) \quad \text{s.t.} \quad f_{\mathsf{prior}}(\boldsymbol{x}) \leq \eta_{\mathsf{prior}}\,. \tag{47}$$

Providing the Lagrange multipliers (μ and ℓ) and the thresholds (η_{data} and η_{prior}) are set consistently, the image restoration is equivalently achieved by solving either of the problems in Eq. (44), Eq. (47) or by minimizing the penalty function in Eq. (46). However choosing which of these particular problems to solve can be a deciding issue for the efficiency of the method. For instance, if

[5] Conversely, the constraint being *inactive* would imply that $\ell = 0$ which would mean that the data are useless, which is hopefully not the case...

$f_{\mathsf{data}}(\boldsymbol{x})$ and $f_{\mathsf{prior}}(\boldsymbol{x})$ are both *smooth functions*, direct minimization of $f(\boldsymbol{x};\mu)$ in Eq. (46) can be done by using general purpose optimization algorithms 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}, more specific algorithms must be devised. As we will see in the following, specifying the image synthesis as a constrained optimization problem provides a very general framework suitable to describe most existing algorithms and their properties but it may hide important algorithmic details. We can already anticipate that image synthesis algorithms will differ by the following ingredients:

- the type of data taken into account (*e.g.* complex visibilities, uncalibrated data, powerspectrum and phase closure);
- the model approximations (*e.g.* exact Fourier transform, Fourier interpolation, regridding);
- the definition of the feasible domain $\mathcal{X}$, that is, the strict constraints such as positivity and normalization;
- the prior type and level;
- the numerical algorithm to solve the problem.

All these ingredients, but the last one, completely define the constrained optimization problem. Hence, for a (strictly) convex penalty $f(\boldsymbol{x})$ and a convex feasible domain $\mathcal{X}$, they are sufficient to specify the solution sought by the image synthesis algorithm. In that case, the numerical optimization algorithm has only an incidence on the computational time required to obtain the solution. However, for non-convex criteria, as it is the case when the gains are unknown (*cf.* Section 3.2) or when dealing with powerspectrum and bispectrum data (*cf.* Section 3.3), the numerical optimization method used to solve the problem has a more important incidence as different methods applied to the same (non convex) problem may provide different solutions.

Going back to the dirty image case (*cf.* Section 2.5), we see that the degeneracy due to incomplete u-v coverage was solved by using the pseudo-inverse which is identical to compute the solution of a regularization problem in the limit of a negligible (but still non null) value of μ. The term $\mu \|\boldsymbol{x}\|^2$ is the simplest quadratic regularization that could be used, incidentally Tikhonov (Tikhonov and Arsenin, 1977) was the first to propose this kind of regularization to solve inverse problems.

2.8. Likelihood criterion

Assuming complex visibility data follow a (multi-variable) Gaussian distribution, then Eq. (27) can be used to define the likelihood criterion f_{data}. However, there are some possible simplifications. For instance, if the complex visibility

measurements are independent Gaussian variables and if the real and imaginary parts have the same variance, then the likelihood penalty takes a simple form:

$$f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) = \sum_k \sum_{(j_1,j_2,m)\in\mathcal{B}_k} w_{j_1,j_2,m} \left| V^{\mathsf{data}}_{j_1,j_2,m} - g^\star_{j_1,m}\, g_{j_2,m}\, V_k(\boldsymbol{x}) \right|^2 , \tag{48}$$

where the weights are given by:

$$w_{j_1,j_2,m} = \mathrm{Var}\left(\mathrm{Re}\left(V^{\mathsf{data}}_{j_1,j_2,m}\right)\right)^{-1} = \mathrm{Var}\left(\mathrm{Im}\left(V^{\mathsf{data}}_{j_1,j_2,m}\right)\right)^{-1} . \tag{49}$$

This expression is very commonly used in interferometry and was popularized by Goodman (1985). Using the same *simplified* notation as in Sect. 2.5, the likelihood writes:

$$f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) = \sum_k w_k \left| d_k - m_k(\boldsymbol{x};\boldsymbol{g}) \right|^2 , \tag{50}$$

where d_k and m_k are the complex visibility data and model in k-th spatial frequency.

Real data may however have a different statistics. For instance, the OI-FITS file exchange format for optical interferometric data assumes that the amplitude and the phase of complex data (complex visibility or triple product) are independent (Pauls et al., 2005). The thick lines in Fig. 5 display the isocontours of the corresponding likelihood which forms a non-convex valley in the complex plane. Assuming Goodman statistics would yields circular isocontours in this figure and is obviously a bad approximation of the *true* criterion in that case. To improve on Goodman model while avoiding non-convex criteria, Meimon et al. (2005a) have proposed quadratic convex approximations of the true likelihood (see Fig. 5) and have shown that their so-called *local approximation* yields the best results, notably when dealing with low signal to noise data. For a complex datum $d_k = \rho_k \exp(\mathrm{i}\,\varphi_k)$, the local quadratic approximation writes:

$$f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) = \sum_k \left\{ \frac{\mathrm{Re}\left(e_k\, \mathrm{e}^{-\mathrm{i}\,\varphi_k}\right)^2}{\sigma^2_{/\!/,k}} + \frac{\mathrm{Im}\left(e_k\, \mathrm{e}^{-\mathrm{i}\,\varphi_k}\right)^2}{\sigma^2_{\perp,k}} \right\} \tag{51}$$

where $e_k = d_k - m_k(\boldsymbol{x};\boldsymbol{g})$ is the complex residual and the variances along and perpendicular to the complex datum vector are:

$$\sigma^2_{/\!/,k} = \mathrm{Var}(\rho_k) \tag{52}$$

$$\sigma^2_{\perp,k} = \rho_k^2\ \mathrm{Var}(\varphi_k)\,. \tag{53}$$

Note that the Goodman model is retrieved when $\rho_k^2\ \mathrm{Var}(\varphi_k) = \mathrm{Var}(\rho_k)$.

2.9. Maximum entropy methods

In statistical physics, the *entropy* is the logarithm of the probability $\log \Pr(\boldsymbol{x})$. Hence, minimizing $f_{\mathsf{prior}}(\boldsymbol{x}) \propto -\log \Pr(\boldsymbol{x})$ under the constraint of fitting the data can be interpreted as a *maximum entropy method* (MEM for short). In

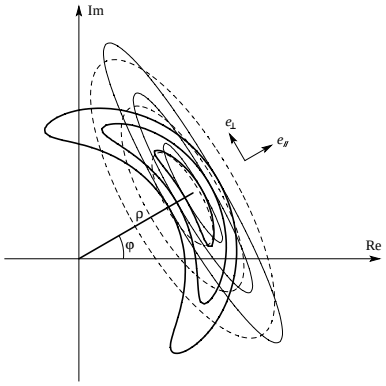

Figure 5: Convex quadratic approximations of complex data. Thick lines: χ^2 isocontours (at 1, 2 and 3 rms levels) for a complex data with independent amplitude and phase. Dashed lines: isocontours for the global quadratic approximation. Thin lines: isocontours for the local quadratic approximation. Note that Goodman approximation would give circular isocontours in this figure.

words, MEM is Bayesian approach taken literally. It directly follows from the Sections 2.6 and 2.7 that the MEM solution is then given by minimizing a criterion like the one in Eq. (46). There is nothing specific here: in particular the penalty $f_{\mathsf{prior}}(\boldsymbol{x})$ (sometimes called *negentropy*) can take any mathematical form. However, the term *maximum entropy methods* is generally used for non-quadratic prior penalties. Narayan and Nityananda (1986) reviewed MEM for radio-interferometry imaging and argued that only non-quadratic priors can interpolate missing Fourier data. Le Besnerais et al. (2008) have nonetheless shown that good image reconstruction can be achieved from interferometric data using a simple quadratic prior whereas strictly accounting for positivity. Indeed, the key point for good Fourier interpolation is the compactness of the brightness distribution which is imposed by the positivity as it plays the role of a floating support (*e.g.* see Fig. 6b).

One of the most popular prior penalty usually considered for MEM is:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum_j \left[\bar{x}_j - x_j + x_j \, \log(x_j/\bar{x}_j)\right] \tag{54}$$

where $\bar{\boldsymbol{x}}$ is the *default image*, that is, the one which would be recovered in absence of any data. The default image can be taken as being a flat image, an image previously restored under a stronger prior (perhaps at a lower resolution), *etc.* When strict normalization apply for the sought image (and providing the prior image is normalized), the negentropy simplifies to:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum_j x_j \, \log(x_j/\bar{x}_j) \, . \tag{55}$$

Omitting the prior image and to give an idea of the zoological diversity in the world of MEM's, the following other non-quadratic negentropy were also

considered:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = -\sum_j \log(x_j)\,. \tag{56}$$

or:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = -\sum_j \sqrt{x_j} \tag{57}$$

Except for Eq. (57), these MEM penalties forbid negative pixel values (providing the prior map is non-negative): there is no needs to explicitly impose positivity. This is sometimes put forward by the proponents of these methods.

Also note that, these MEM penalties are *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 (hence, some *smoothness*), the prior can be chosen to depend on the sought image. For instance: $\bar{\boldsymbol{x}} = \mathbf{P} \cdot \boldsymbol{x}$ where $\mathbf{P}$ is some averaging/smoothing linear operator. This type of *floating prior* has been used to loosely enforce constraints such as radial symmetry (Horn, 1985).

Once chosen an expression for the negentropy and a value for μ (or for η_{data}), image reconstruction is carried by minimizing the joint criterion in Eq. (46). This task has a number of issues as the problem is highly non-linear and as there may be a very large number of unknowns (as many as there are pixels). Various methods have been proposed, but the most effective algorithm is due to Skilling and Bryan (1984) who sought for the solution by a non-linear optimization in a local sub-space of search directions with the Lagrange multiplier μ tuned on the fly to match the constraint set by η_{data}.

2.10. Clean *method*

The proponents of maximum entropy methods claim that only non-linear[6] regularization (*i.e.* such as the entropy) can helps to solve for the degeneracies in image synthesis problem; that is, interpolate the complex visibility for missing spatial frequencies (Narayan and Nityananda, 1986). In fact, the necessary condition for an effective regularization in this context is not so much to avoid noise amplification (though some kind of smoothness constraint may be necessary to avoid excessive *superesolution*, *i.e.* extrapolating the u-v coverage) but mostly to set constraints so that Fourier interpolation of missing frequencies is done smoothly. Since smoothness along spatial frequencies is qualitatively equivalent to compactness in the image plane, a means to achieve smooth Fourier interpolation is to favor compactness of the restored image. This is explicitly the underlying idea in Clean algorithm (Högbom, 1974) which attempts to obtain

[6] Meaning *non-quadratic* here. The minimum of a quadratic penalty being obtained by solving a linear system of equations (the so-called *normal equations*, see Section 2.5), the corresponding optimization problem is sometimes categorized as being a *linear problem*, which is rather confusing for non-specialists.

an image which is compatible with the data and is made of a minimal number of point-like sources.

CLEAN method is of the family of the *matching pursuit* algorithms, it is an iterative method which works as follows. Given the data in the form of a dirty image, the location of the brightest source needed to best explain the data is searched. The sought image is 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 which is searched for evidence of another point-like source. After convergence, (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 PSF) to have a resolution in agreement with the extension of the u-v coverage. Obviously CLEAN is aimed at restoring images mainly made of point-like sources, star clusters for instance. However, once most compact sources have been removed, the residual dirty image is essentially due to the remaining extended sources which may be smooth enough to not be too distorted by the convolution by the dirty beam. The residual dirty image is therefore added to the clean image to produce a final image consisting in compact sources (convolved by the clean beam) plus smooth extended components. Schwarz (1978) has formalized CLEAN and studied its convergence to show that it is equivalent to an iterative deconvolution with early stopping.

In the general framework proposed in this tutorial, it is possible to emulate the CLEAN method by using a specific regularization penalty which imposes to have the minimum number of significant pixels in the restored image. First, we observe that the ℓ_0-norm of $\boldsymbol{x}$:

$$\|\boldsymbol{x}\|_0 = \sum_j c_0(x_j) \quad \text{with: } c_0(x) = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases} \tag{58}$$

is simply equals to the number of non-zero values in $\boldsymbol{x}$. Hence, using the ℓ_0-norm as a regularization:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \|\boldsymbol{x}\|_0 \tag{59}$$

leads to seek for the image compatible with the data and which has the minimum number of significant pixels. Using the ℓ_0-norm as a regularization criterion is formally correct but has a profound impact on the strategy to solve the image restoration problem. Indeed, the ℓ_0-norm is not a smooth function and direct minimization of $f(\boldsymbol{x};\mu)$ in Eq. (46) by a general purpose optimization algorithm is not possible; special methods must be devised (the matching pursuit is a possibility). To have an idea of how difficult is the problem, first note that the number $P = \|\boldsymbol{x}\|_0$ of bright pixels plays the role of the tuning parameter, then realize that seeking for the minimum ℓ_0-norm solution involves solving the following combinatorial problem: *Which are the significant (non-zero) pixels?* Owing to the number of parameters, this is a very difficult task as there are

$$\binom{N}{P} = \frac{N!}{(N-P)!\,P!}\,,$$

different ways to choose P non-zero pixels out of N pixels and, for typical values of N an P, this is a really huge number.[7]

Compressive sensing is a signal processing approach aimed at dealing with sparse data. One of the nice results demonstrated by compressive sensing is that, in most practical cases, the minimum ℓ_0-norm solution can be obtained by using an ℓ_1-norm regularization (Candes et al., 2006):

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \|\boldsymbol{x}\|_1 \stackrel{\text{def}}{=} \sum_j |x_j| \,. \tag{60}$$

The great advantage is that, apart for a singularity at $x_j = 0$, the ℓ_1-norm is a continuous function which is much easier to minimize than the ℓ_0-norm. However note that using an ℓ_1 norm regularization has no effect when combined with positivity and normalization constraints. The algorithm of Giovannelli and Coulais (2005) (see Sect. 2.11) can be seen as an improved CLEAN method which is explicitly based on the minimization of a criterion.

2.11. Other methods

This section briefly summarizes promising (though less popular than MEM and CLEAN) image reconstructions methods that have been proposed to solve the image synthesis problem in astronomy.

Wakker and Schwarz (1988) improved over CLEAN by using a multi resolution approach where two maps (at low and high resolution) are simultaneously build by a CLEAN-like algorithm. Starck et al. (1994) have generalized this approach by using a wavelet expansion to describe the image — which could be formally expressed in terms of Eq. (10) — and achieved *multiresolution deconvolution* of the interferometric data by a matching pursuit algorithm similar to the CLEAN method applied to the wavelet coefficients. Their solution satisfies positivity and support constraints.

The WIPE method by Lannes et al. (1997) is a regularized fit of the interferometric data under positivity and support constraints. The synthesized image is given by Eq. (11) (using an equally-spaced grid) and WIPE explicitly set the effective resolution by means of the building block function $b(\Delta\boldsymbol{\theta})$ (the so-called *neat beam*) with an additional penalty term to avoid spurious high frequencies. The image parameters are estimated by WIPE as the ones that minimize:

$$f_{\mathsf{Wipe}}(\boldsymbol{x}) = \sum_k w_k \, |\hat{b}_k \, d_k - (\mathbf{A} \cdot \boldsymbol{x})_k|^2 + \sum_{k, |\nu_k| > \nu_{\mathsf{eff}}} |(\mathbf{F} \cdot \boldsymbol{x})_k|^2 \tag{61}$$

where $\boldsymbol{d}$ are the calibrated complex visibility data, $\hat{\boldsymbol{b}}$ is the Fourier transform of the neat beam, $\mathcal{L}$ is the list of observed frequencies (*cf.* Section 2.2), ν_{eff} is an effective cutoff frequency (which depends on the extension of the u-v coverage:

[7] For instance, for a 128×128 pixel image, there are $\sim 5.7 \times 10^{2310}$ possibilities to choose the location of 10 % significant pixels and only(!) $\sim 2 \times 10^{397}$ possibilities to choose the location of 1 % significant pixels.

$\nu_{\mathsf{eff}} > \sup_{\nu\in\mathcal{L}}|\nu|$), $\mathbf{F}$ is the Fourier transform operator, $\mathbf{A}$ is the model matrix given by Eq. (14), *i.e.* accounting for the sub-sampled Fourier transform and the neat beam, and $\boldsymbol{w}$ are the weights:

$$w_k = \frac{1/\sigma_k^2}{\sum_{k'} 1/\sigma_{k'}^2} \tag{62}$$

where $\sigma_k^2 = |\hat{b}_k|^2 \operatorname{Var}(\operatorname{Re}(d_k)) = |\hat{b}_k|^2 \operatorname{Var}(\operatorname{Im}(d_k))$, *i.e.* Goodman approximation is assumed. In the criterion minimized by WIPE, Eq. (61), we can identify a likelihood term, *cf.* Eq. (50), and a regularization term to cope with the ill-posedness of the problem, there is no hyper parameter to tune the level of this latter term. Finally, WIPE performs the optimization by a conjugate gradient search with a stopping criterion derived from the analysis of the conditioning of the regularized problem. This analysis is built up during the iterations.

Recently, Giovannelli and Coulais (2005) have proposed a kind of multi-resolution method which restore an image $\boldsymbol{x} = \boldsymbol{x}_{\mathsf{e}} + \boldsymbol{x}_{\mathsf{p}}$ made of two maps: $\boldsymbol{x}_{\mathsf{e}}$ for *extended* structures and $\boldsymbol{x}_{\mathsf{p}}$ for *point-like* components. The map $\boldsymbol{x}_{\mathsf{e}}$ is regularized by imposing smoothness whereas the map $\boldsymbol{x}_{\mathsf{p}}$ is regularized by imposing sparseness. With additional strict positivity and (optionally) support constraints, it turns out that the two kinds of regularization can be coded by quadratic penalties. Their method amounts to minimize the following penalty:

$$\begin{aligned} f_{\mathsf{mix}}(\boldsymbol{x}_{\mathsf{e}}, \boldsymbol{x}_{\mathsf{p}}) = {} & \|\boldsymbol{d} - \mathbf{S}\cdot\mathbf{F}\cdot(\boldsymbol{x}_{\mathsf{e}} + \boldsymbol{x}_{\mathsf{p}})\|^2 + \lambda_{\mathsf{s}}\, \boldsymbol{c}^{\mathrm{T}}\cdot\boldsymbol{x}_{\mathsf{p}} + \epsilon_{\mathsf{s}}\, \|\boldsymbol{x}_{\mathsf{p}}\|^2 \\ & + \lambda_{\mathsf{c}}\, \|\mathbf{D}\cdot\boldsymbol{x}_{\mathsf{e}}\|^2 + \epsilon_{\mathsf{m}}\, (\boldsymbol{c}^{\mathrm{T}}\cdot\boldsymbol{x}_{\mathsf{e}})^2 \end{aligned} \tag{63}$$

where $\boldsymbol{d}$ are the calibrated complex visibilities, $\mathbf{S}$ and $\mathbf{F}$ are the sub-sampling and Fourier transform operators (*cf.* Section 2.5), λ_{s}, ϵ_{s}, λ_{c} and ϵ_{m} are the tuning parameters of the regularization terms, $\mathbf{D}$ is a local finite difference operator, for instance (in 1-D to simplify the notation):

$$\|\mathbf{D}\cdot\boldsymbol{x}\|^2 = \sum_n (x_{n+1} - x_n)^2 \tag{64}$$

and $\boldsymbol{c} = (1, 1, \ldots, 1)^{\mathrm{T}}$ is a vector with all components set to one, hence:

$$\boldsymbol{c}^{\mathrm{T}}\cdot\boldsymbol{x} = \sum_n x_n\,. \tag{65}$$

There are 4 tuning parameters of the regularization terms in Eq. (63): $\lambda_{\mathsf{s}} \geq 0$ and $\epsilon_{\mathsf{s}} > 0$ control the sparseness of $\boldsymbol{x}_{\mathsf{p}}$, $\lambda_{\mathsf{c}} > 0$ controls the level of smoothness in the extended map $\boldsymbol{x}_{\mathsf{e}}$, and $\epsilon_{\mathsf{m}} > 0$ (or $\epsilon_{\mathsf{m}} \geq 0$ if there is a support constraint) insures strict convexity of the regularization with respect to $\boldsymbol{x}_{\mathsf{e}}$. Finally, Giovannelli and Coulais (2005) extensively use circulant approximations to implement a very fast constrained optimization method targeted at solving their problem.

3. Image synthesis in spite of turbulence

3.1. Complex OTF due to turbulence

Without a phase reference and because of the atmospheric turbulence, it is not possible to completely calibrate the instantaneous transmission gains of

an interferometer. The OPD errors due to atmospheric turbulence follow independent Gaussian distributions. Hence, during m-th exposure (which is much longer than the typical evolution time of the turbulence), the mean complex throughput for j-th telescope is:

$$\langle g_j(t)\rangle_m = g_{j,m}\,\mathrm{e}^{-\frac{1}{2}\,\sigma_\phi^2}\,, \tag{66}$$

where $\sigma_\phi^2 = (2\,\pi/\lambda)^2\ \mathrm{Var}(\delta)$ is the variance of the phase and $g_{j,m} = \exp(\mathrm{i}\,\phi_{j,m})$ with $\phi_{j,m} = (2\,\pi/\lambda)\,\langle\delta_j(t)\rangle_m$ the mean phase during the exposure. To simplify the discussion to come, it has been assumed that the variance of the OPD is the same for all telescopes and exposures and that the interferometer is perfect in other respects (static aberrations and geometrical OPD are exactly compensated and amplitude attenuations are properly calibrated). Then, Eq. (66) comes directly from the well known expectation:

$$\left\langle \mathrm{e}^{\mathrm{i}\,\phi}\right\rangle = \mathrm{e}^{\mathrm{i}\,\langle\phi\rangle}\,\mathrm{e}^{-\frac{1}{2}\,\sigma_\phi^2}\,, \tag{67}$$

for a phase which follows a Gaussian distribution: $\phi \sim \mathcal{N}(\langle\phi\rangle, \sigma_\phi^2)$. When the variance σ_ϕ^2 is small (as assumed in Section 2), Eq. (66) yields Eq. (6). At least for telescopes very distant from each other, the OPD errors due to atmospheric turbulence are independent and the mean short exposure OTF for the complex visibility measured by j_1-th and j_2-th telescopes becomes:

$$\begin{aligned}\left\langle g_{j_1}^\star(t)\,g_{j_2}(t)\right\rangle_m &= \langle g_{j_1}(t)\rangle_m\ \langle g_{j_2}(t)\rangle_m \\ &= g_{j_1,m}^\star\,g_{j_2,m}\,\mathrm{e}^{-\sigma_\phi^2}\,.\end{aligned} \tag{68}$$

In Eq. (66) and (68), the mean complex throughputs have an unknown phase (which depends on telescope and exposure time) and an amplitude set by the phase variance which scales as the OPD variance times λ^{-2}. At short wavelengths (optical), the phase variance exceeds a few squared radians and the mean OTF during exposures is negligible. The object's complex visibility cannot be directly measured and other estimators that are insensitive to this effect must be measured (*cf.* Section 3.3). On the other hand, at long wavelengths (radio), the phase variance is small, $\exp(-\sigma_\phi^2) \simeq 1$, it however remains the problem of deriving the unknown phasors $g_{j,m}$. This issue can be solved by means of *self calibration* (*cf.* Section 3.2).

3.2. Self calibration

For slow or weak turbulence, the variance of phase errors during the exposure time is small and the complex throughputs are almost equal to unknown phasors. Then, the likelihood term writes:

$$f_{\mathsf{data}}(\boldsymbol{x},\boldsymbol{g}) = \sum_k \sum_{(j_1,j_2,m)\in\mathcal{B}_k} w_{j_1,j_2,m}\,\left|V_{j_1,j_2,m}^{\mathsf{data}} - g_{i_1,m}^\star\,g_{j_2,m}\,V_k(\boldsymbol{x})\right|^2 \tag{69}$$

where:

$$g_{j,m} \simeq \mathrm{e}^{\mathrm{i}\,\phi_{j,m}}\,. \tag{70}$$

The problem is to not only derive the image, that is, the parameters $\boldsymbol{x}$, but also the unknown phases $\phi_{j,m}$. Since the complex throughputs and the sought image are statistically independent, standard Bayesian analysis (*cf.* Section 2.6) yields the following problem to solve:

$$(\boldsymbol{x}_{\mathsf{MAP}}, \boldsymbol{g}_{\mathsf{MAP}}) = \underset{\boldsymbol{x},\boldsymbol{g}}{\arg\min} \left\{ f_{\mathsf{data}}(\boldsymbol{x}, \boldsymbol{g}) + \mu_{\mathsf{img}}\, f^{\mathsf{img}}_{\mathsf{prior}}(\boldsymbol{x}) + \mu_{\mathsf{gain}}\, f^{\mathsf{gain}}_{\mathsf{prior}}(\boldsymbol{g}) \right\} \tag{71}$$

where $\mu_{\mathsf{img}}\, f^{\mathsf{img}}_{\mathsf{prior}}(\boldsymbol{x})$ and $\mu_{\mathsf{gain}}\, f^{\mathsf{gain}}_{\mathsf{prior}}(\boldsymbol{g})$ are the regularization terms for the image parameters and for the complex throughputs respectively. The latter can be derived from prior statistics about the turbulence (Roddier, 1981). Strict constraints such as the positivity of the image may have to be taken into account in Eq. (71). In addition to the issue of proper setting of the priors, direct minimization of the criterion in Eq. (71) is a difficult task because: (i) the problem is no longer convex with respect to the complete set of parameters $(\boldsymbol{x}, \boldsymbol{g})$ (hence *global* optimization is required); (ii) the parameters are heterogeneous and scale very differently which adds to the bad conditioning of the problem (hence slowing down the convergence). Fortunately, to solve for the first issue, a simpler strategy than true global optimization has proven effectiveness (and incidentally also discards the second issue). This method is based on the following insights. First, given the data and the complex throughputs, robust image reconstruction can be performed by one of the image synthesis methods described in Section 2. Second, given the data and an estimate of the object brightness distribution, there are more constraints than unknown when seeking for the complex gains. This problem, although non-convex, may be easier to solve. These considerations suggest an alternate scheme to find both the image and the complex gains. This approach is called *self-calibration* because it uses the current estimate of the image as a reference source to calibrate the complex gains (the sought object serving itself as a reference source). The self-calibration algorithm writes:

Step 1. *Initialization.* Choose regularization (the f_{prior}'s and the μ's) and initial image $\boldsymbol{x}^{(0)}$; set $n = 1$.

Step 2. *Self-calibration step.* Given the image $\boldsymbol{x}^{(n-1)}$, find the *best* complex throughputs $\boldsymbol{g}^{(n)}$ by solving:

$$\boldsymbol{g}^{(n)} = \underset{\boldsymbol{g}}{\arg\min} \left\{ f_{\mathsf{data}}\big(\boldsymbol{x}^{(n-1)}, \boldsymbol{g}\big) + \mu_{\mathsf{gain}}\, f^{\mathsf{gain}}_{\mathsf{prior}}(\boldsymbol{g}) \right\} .$$

Step 3. *Image reconstruction step.* Apply image synthesis algorithm to recover a new image estimate given the data and the complex throughputs:

$$\boldsymbol{x}^{(n)} = \underset{\boldsymbol{x}}{\arg\min} \left\{ f_{\mathsf{data}}\big(\boldsymbol{x}, \boldsymbol{g}^{(n)}\big) + \mu_{\mathsf{img}}\, f^{\mathsf{img}}_{\mathsf{prior}}(\boldsymbol{x}) \right\} .$$

Step 4. *Convergence test.* If algorithm has converged, terminate with solution $\boldsymbol{x}^{(n)}$; otherwise; increment n and loop to Step 2.

It is worth noting that any image synthesis methods (*e.g.* CLEAN, MEM, *etc.* described in Section 2) can be used to perform the image reconstruction step. Also, by exchanging the order of the self-calibration and image reconstruction steps, the algorithm can be started with given initial gains $\boldsymbol{g}^{(0)}$ instead of an initial image $\boldsymbol{x}^{(0)}$. In general, the self-calibration step is non-convex, it is however usually solved by means of local minimization starting with the previous estimate of the complex throughputs (more on this below). Even if the 2 steps, self-calibration and image reconstruction, are both convex, the global problem is however non-convex. Hence, the final solution depends on algorithm initialization.

Self-calibration was initially proposed by Readhead and Wilkinson (1978) to derive missing Fourier phase information from phase closure data, the technique was later improved by Cotton (1979). Schwab (1980) used self-calibration to estimate the complex gains $\boldsymbol{g}$ by minimizing a criterion similar to $f_{\mathsf{data}}(\boldsymbol{x}, \boldsymbol{g})$ in Eq. (48) by means of non-linear optimization. Schwab's self-calibration was further improved by Cornwell and Wilkinson (1981) who introduced some priors for the complex gains, that is, the term $\mu_{\mathsf{gain}}\, f^{\mathsf{gain}}_{\mathsf{prior}}(\boldsymbol{g})$ in the penalty minimized in Step 2 above. However, for most authors, no prior about the throughputs is assumed, hence $\mu_{\mathsf{gain}} = 0$. For more conventional imaging problems such as deconvolution of blurred images, the so-called *blind deconvolution* method (Campisi and Egiazarian, 2007) is very similar to the objective of self-calibration and is used to improve the quality of degraded images when the point spread function is unknown.

Although, following Schwab (1980), $f_{\mathsf{data}}(\boldsymbol{x}, \boldsymbol{g})$ in Eq. (48) can be minimized with respect to $\boldsymbol{g}$ by a general non-linear optimization algorithm (Nocedal and Wright, 2006; Bonnans et al., 2006), the self calibration step can be solved by a very simple yet effective algorithm proposed by Lacour et al. (2007) and which can be derived as follows. The optimal complex gains are a stationary point of f_{data} in Eq. (48), hence:

$$
\begin{aligned}
& \frac{\partial f_{\mathsf{data}}(\boldsymbol{x}, \boldsymbol{g})}{\partial g_{j,m}} = 0 \\
\iff \quad & \sum_{(j',k):\,(j',j,m)\in\mathcal{B}_k} w_{j',j,m} \left(V^{\mathsf{data}}_{j',j,m} - g^{\star}_{j',m}\, g_{j,m}\, V_k \right) g_{j',m}\, V_k{}^{\star} \\
& + \sum_{(j',k):\,(j,j',m)\in\mathcal{B}_k} w_{j,j',m} \left(V^{\mathsf{data}}_{j,j',m}{}^{\star} - g_{j,m}\, g^{\star}_{j',m}\, V_k{}^{\star} \right) g^{\star}_{j',m}\, V_k = 0\,,
\end{aligned}
$$

which, factorizing out $g_{j,m}$, can be re-expressed in the form of a set of fixed point equations with respect to the complex gains:

$$
g_{j,m} = \frac{\displaystyle\sum_{(j',k):\,(j',j,m)\in\mathcal{B}_k} w_{j',j,m}\, g_{j',m}\, V_k{}^{\star}\, V^{\mathsf{data}}_{j',j,m} + \sum_{(j',k):\,(j,j',m)\in\mathcal{B}_k} w_{j,j',m}\, g^{\star}_{j',m}\, V_k\, V^{\mathsf{data}}_{j,j',m}{}^{\star}}{\displaystyle\sum_{(j',k):\,(j',j,m)\in\mathcal{B}_k} w_{j',j,m}\, |g_{j',m}|^2\, |V_k|^2 + \sum_{(j',k):\,(j,j',m)\in\mathcal{B}_k} w_{j,j',m}\, |g_{j',m}|^2\, |V_k|^2}\,. \tag{72}
$$

Using some initial guess for the complex gains in the right hand side of Eq. (72) provides a new estimate (the left hand side) of the gains. This operation can

be iterated until convergence which occurs after a few iterations (Lacour et al., 2007). Intuitively, one can recognize in the right hand side of Eq. (72) a weighted sum of estimates of the gain $g_{j,m}$ providing that all the other gains are known. This kind of algorithm is very similar to the Gauss-Seidel iterative method for solving a set of linear equations.

3.3. Measurements in optical interferometry

For fast or strong turbulence, the mean transmission for the complex visibilities during an exposure is negligible. There are two possibilities to overcome this problem. The first solution consists in integrating non-linear estimators that are insensitive to telescope-wise phase errors. This requires high acquisition rates ($\sim 1000\,\mathrm{Hz}$ in K-band) and involves special data processing but otherwise no special instrumentation. We discuss this solution in the following. A second solution is to compensate the OPD errors in real time thanks to fast delay lines and to a simultaneous phase reference. This requires a dedicated instrumentation and that a bright reference source be available in the vicinity of the observed object (*e.g.* Prima facility at VLTI, Delplancke et al., 2003). The use of a phase reference is also needed when observing faint objects.

To overcome loss in visibility transmission due to fast varying OPD errors, current optical interferometers integrate the powerspectrum:

$$\begin{aligned} S_{j_1,j_2,m} &= \langle |V_{j_1,j_2}(t)|^2 \rangle_m \\ &= \left\langle \left| g^\star_{j_1}(t)\, g_{j_2}(t)\, \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2}(t)\right) \right|^2 \right\rangle_m \\ &= \left\langle \left| g_{j_1}(t) \right|^2 \right\rangle_m \left\langle \left| g_{j_2}(t) \right|^2 \right\rangle_m \left\langle \left| \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2}(t)\right) \right| \right\rangle^2 \\ &= \rho_{j_1,m}\, \rho_{j_2,m}\, |\hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)|^2 , \end{aligned} \tag{73}$$

where it has been assumed that the complex throughputs are independent (and that $j_1 \neq j_2$), the mean frequency $\boldsymbol{\nu}_{j_1,j_2,m}$ is defined in Eq. (8), and

$$\rho_{j,m} = \left\langle \left| g_j(t) \right|^2 \right\rangle_m \tag{74}$$

is the mean squared modulus of the complex throughput of j-th telescope during m-th exposure. By construction, the mean squared moduli of the complex throughput are insensitive to the phase errors and hence is the powerspectrum. Unlike that of the complex visibility, the transfer function $\rho_{j_1,m}\, \rho_{j_2,m}$ of the powerspectrum is not negligible. This transfer function can be estimated by simultaneous photometric calibration and, to compensate for remaining static effects, from the powerspectrum of a reference source (a so-called *calibrator*). This means that the object powerspectrum $|\hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)|^2$ 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 powerspectrum), the bispectrum of the complex visibilities is measured:

$$\begin{aligned} B_{j_1,j_2,j_3,m} &= \langle V_{j_1,j_2}(t)\, V_{j_2,j_3}(t)\, V_{j_3,j_1}(t) \rangle_m \\ &= \rho_{j_1,m}\, \rho_{j_2,m}\, \rho_{j_3,m}\, \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)\, \hat{I}\left(\boldsymbol{\nu}_{j_2,j_3,m}\right)\, \hat{I}\left(\boldsymbol{\nu}_{j_3,j_1,m}\right) , \end{aligned} \tag{75}$$

where it has been assumed that the complex throughputs are independent (and that j_1, j_2 and j_3 denote three different telescopes). As for the powerspectrum, the transfer function $\rho_{j_1,m}\,\rho_{j_2,m}\,\rho_{j_3,m}$ of the bispectrum can be calibrated. However, since this transfer function is real, it has no effect on the phase of the bispectrum, the so-called *phase closure*, which is equal to that of the object alone:

$$\begin{aligned}\beta_{j_1,j_2,j_3,m} &= \arg\left(B_{j_1,j_2,j_3,m}\right) \\ &= \arg\left(\hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)\,\hat{I}\left(\boldsymbol{\nu}_{j_2,j_3,m}\right)\,\hat{I}\left(\boldsymbol{\nu}_{j_3,j_1,m}\right)\right)\,. \end{aligned} \tag{76}$$

The (calibrated) powerspectrum provides measurements of the modulus of the object's complex visibility whereas the bispectrum (or the phase closure) provides information about the phase of the complex visibility. However, some phase information is missing. In effect, from all the interferences between T telescopes (in a non redundant 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. For a small number of telescopes, the measurable phase information is much more sparse than the modulus information. In particular, there is no direct mean to derive estimates of the object complex visibility at the sampled frequencies given only the powerspectrum and the bispectrum (or the phase closure). In the limit of a very large number of telescopes the ratio of missing information is reduced. However, at least the information of absolute position of the observed object is not available.

In practice, measuring the powerspectrum and the bispectrum involves measuring the instantaneous complex visibilities (that is, for a very short integration time compared to the evolution of turbulence) and averaging their powerspectrum and bispectrum over the short exposure time (which is short enough to consider that the observed spatial frequencies remain unchanged by Earth rotation). Being non-linear functions of noisy variables, the powerspectrum and the bispectrum are biased. The biases are however easy to estimate and discard. To simplify the description of the algorithms, we consider in what follows that the unbiased and calibrated powerspectrum and bispectrum are available as input data for image reconstruction, hence:

$$S^{\mathsf{data}}_{j_1,j_2,m} = |\hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)|^2 + \text{error}\,, \tag{77}$$

$$B^{\mathsf{data}}_{j_1,j_2,j_3,m} = \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}\right)\,\hat{I}\left(\boldsymbol{\nu}_{j_2,j_3,m}\right)\,\hat{I}\left(\boldsymbol{\nu}_{j_3,j_1,m}\right) + \text{error}\,. \tag{78}$$

where the error terms account for noise and model error but are assumed to be centered (remaining biases, if any, are negligible with respect to the standard deviation of the errors). Note that the spatial frequencies involved in the bispectrum make a *closed triangle*:

$$\boldsymbol{\nu}_{j_1,j_2,m} + \boldsymbol{\nu}_{j_2,j_3,m} + \boldsymbol{\nu}_{j_3,j_1,m} = 0\,.$$

Instead of the bispectrum data, we can consider the phase closure data:

$$\beta^{\mathsf{data}}_{j_1,j_2,j_3,m} = \text{arc}\left(\varphi\left(\boldsymbol{\nu}_{j_1,j_2,m}\right) + \varphi\left(\boldsymbol{\nu}_{j_2,j_3,m}\right) + \varphi\left(\boldsymbol{\nu}_{j_3,j_1,m}\right) + \text{error}\right)\,, \tag{79}$$

where $\varphi(\boldsymbol{\nu}) = \arg[\hat{I}(\boldsymbol{\nu})]$ is the phase of the Fourier transform of the object brightness distribution and arc( ) returns it argument wrapped in the range $(-\pi, +\pi]$.

Given the (non-linear) relationships between the image model and the data, an image reconstruction algorithm can be designed following the inverse problem approach (*cf.* Sections 2.6 and 2.7) providing we derive analytical expressions for the terms f_{data} and f_{prior} which penalize the discrepancy of the sought image with the data and with the priors respectively. The same regularization penalties as in Section 2 can be used for the prior term and we discuss below the different possibilities for the f_{data} penalty.

3.4. Likelihood terms

In principle, statistics of the powerspectrum, the bispectrum, and phase closure yield analytical expressions for the likelihood term f_{data}. A priori their statistics cannot be Gaussian: the powerspectrum is a positive quantity, the phase closure is wrapped in the range $(-\pi, +\pi]$, *etc.* However, most algorithms make use of quadratic penalties with respect to the measurements which, in a strict Bayesian viewpoint, implies Gaussian statistics. Another assumption generally made is the independence of the measurements which leads to *separable* penalties as those given in Eqs. (80), (82), (84), and (85) below.

With this approximation, the penalty with respect to the powerspectrum data writes:

$$f_{\mathsf{data}}^{\mathsf{ps}}(\boldsymbol{x}) = \sum_{m, j_1 < j_2} \frac{\left(S_{j_1,j_2,m}^{\mathsf{data}} - S_{j_1,j_2,m}^{\mathsf{model}}(\boldsymbol{x})\right)^2}{\mathrm{Var}\left(S_{j_1,j_2,m}^{\mathsf{data}}\right)}\,, \tag{80}$$

where, according to Eq. (77), the model of the powerspectrum reads:

$$S_{j_1,j_2,m}^{\mathsf{model}}(\boldsymbol{x}) = \left|\hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m}; \boldsymbol{x}\right)\right|^2. \tag{81}$$

To account for phase wrapping and for the variable quality in phase measurements, Haniff (1991) has proposed to define the penalty with respect to the phase closure data as:

$$f_{\mathsf{data}}^{\mathsf{cl}}(\boldsymbol{x}) = \sum_{m, j_1 < j_2 < j_3} \frac{\mathrm{arc}^2\left(\beta_{j_1,j_2,j_3,m}^{\mathsf{data}} - \beta_{j_1,j_2,j_3,m}^{\mathsf{model}}(\boldsymbol{x})\right)}{\mathrm{Var}\left(\beta_{j_1,j_2,j_3,m}^{\mathsf{data}}\right)}\,. \tag{82}$$

where, according to Eq. (79), the model of the phase closure reads:

$$\beta_{j_1,j_2,j_3,m}^{\mathsf{model}}(\boldsymbol{x}) = \varphi\left(\boldsymbol{\nu}_{j_1,j_2,m}; \boldsymbol{x}\right) + \varphi\left(\boldsymbol{\nu}_{j_2,j_3,m}; \boldsymbol{x}\right) + \varphi\left(\boldsymbol{\nu}_{j_3,j_1,m}; \boldsymbol{x}\right)\,. \tag{83}$$

This penalty is however not continuously differentiable with respect to the image parameters, which can prevent the convergence of optimization algorithms. To overcome this problem, the complex phasors can be used instead (Thiébaut, 2008):

$$f_{\mathsf{data}}^{\mathsf{cl}}(\boldsymbol{x}) = \sum_{m, j_1 < j_2 < j_3} \frac{\left|\mathrm{e}^{\mathrm{i}\,\beta_{j_1,j_2,j_3,m}^{\mathsf{data}}} - \mathrm{e}^{\mathrm{i}\,\beta_{j_1,j_2,j_3,m}^{\mathsf{model}}(\boldsymbol{x})}\right|^2}{\mathrm{Var}\left(\beta_{j_1,j_2,j_3,m}^{\mathsf{data}}\right)}\,, \tag{84}$$

which, in the limit of small phase closure errors, is approximately equal to the penalty in Eq. (82).

If the bispectrum data are to be used, there is the additional difficulty to deal with complex data. Assuming isotropic distribution of errors for the real and imaginary parts of the bispectrum, the following penalty with respect to the bispectrum data can be derived:

$$f^{\mathsf{bisp}}_{\mathsf{data}}(\boldsymbol{x}) = \sum_{m,j_1<j_2<j_3} w^{\mathsf{bisp}}_{j_1,j_2,j_3,m} \left| B^{\mathsf{data}}_{j_1,j_2,j_3,m} - B^{\mathsf{model}}_{j_1,j_2,j_3,m}(\boldsymbol{x}) \right|^2 . \tag{85}$$

where, according to Eq. (78), the model of the bispectrum writes:

$$B^{\mathsf{model}}_{j_1,j_2,j_3,m}(\boldsymbol{x}) = \hat{I}\left(\boldsymbol{\nu}_{j_1,j_2,m};\boldsymbol{x}\right)\, \hat{I}\left(\boldsymbol{\nu}_{j_2,j_3,m};\boldsymbol{x}\right)\, \hat{I}\left(\boldsymbol{\nu}_{j_3,j_1,m};\boldsymbol{x}\right) , \tag{86}$$

and, it follows from the assumption of isotropy that, the weights in Eq. (85) are:

$$w^{\mathsf{bisp}}_{j_1,j_2,j_3,m} = \mathrm{Var}\big(B^{\mathsf{data}}_{j_1,j_2,j_3,m}\big)^{-1} , \tag{87}$$

with $\mathrm{Var}\big(B^{\mathsf{data}}_{j_1,j_2,j_3,m}\big) \overset{\text{def}}{=} \mathrm{Var}\left(\mathrm{Re}\left(B^{\mathsf{data}}_{j_1,j_2,j_3,m}\right)\right) = \mathrm{Var}\left(\mathrm{Im}\left(B^{\mathsf{data}}_{j_1,j_2,j_3,m}\right)\right)$. This is similar to the Goodman (1985) prescription for the complex visibility.

Depending on which set of data is available, and assuming that the different type 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 Eq. (80), (82), (84) and/or Eq. (85). For instance, if one wants to fit the powerspectrum and the phase closure data, then:

$$f_{\mathsf{data}}(\boldsymbol{x}) = f^{\mathsf{ps}}_{\mathsf{data}}(\boldsymbol{x}) + f^{\mathsf{cl}}_{\mathsf{data}}(\boldsymbol{x}) . \tag{88}$$

Note that existing algorithms can impose specific expressions for $f_{\mathsf{data}}(\boldsymbol{x})$ and $f_{\mathsf{prior}}(\boldsymbol{x})$.

3.5. Image reconstruction from the powerspectrum and bispectrum data

Most algorithms specifically designed to cope with optical interferometry data can be formally described in the inverse problem framework derived so far. That is, they can be specified in terms of a criterion to optimize, perhaps under some strict constraints, and an optimization strategy. The algorithms (given in alphabetical order) which are summarized in what follows were also the competitors of the three *Image Beauty Contests* (Lawson et al., 2004, 2006; Cotton et al., 2008) which provided a comparison of their performances on the basis of realistic simulated data provided in the form of OI-FITS files (Pauls et al., 2005).

3.5.1. BSMEM algorithm

BSMEM was developed by Buscher (1994), the algorithm makes use of a Maximum Entropy Method (*cf.* Section 2.9) to regularize the problem of restoring an image from the measured bispectrum (hence its name). From the quite informal description of the algorithm (Buscher, 1994; Lawson et al., 2006), it can

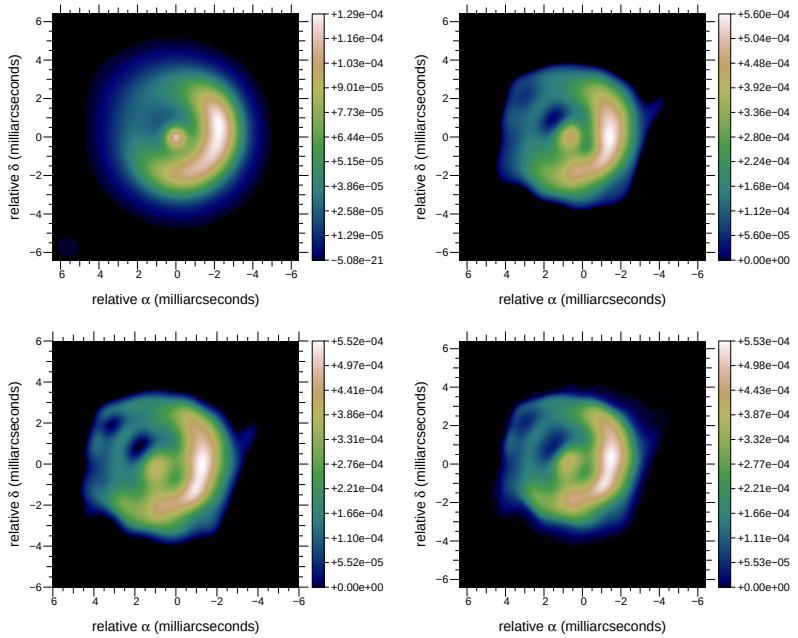

Figure 6: Image reconstruction with various types of regularization. From top-left to bottom-right: (a) original object smoothed to the resolution of the interferometer (FWHM ~ 15 mas); (b) reconstruction with a quadratic regularization given by Eq. (93) and which imposes a compact field of view; (c) reconstruction with edge-preserving regularization as in Eq. (95); (d) reconstruction with maximum entropy regularization as in Eq. (54). All reconstructions by algorithm MIRA and from the powerspectrum and the phase closures.

be guessed that BSMEM regularization penalty is given by Eq. (54) and that the measurements (powerspectra, bispectrum amplitudes, and phase closures) are independently exploited by means of a sum of likelihood penalties similar to those in Eqs. (80), (82), and (85). The optimization engine in BSMEM is MEMSYS which implements the strategy proposed by Skilling and Bryan (1984). The default image in BSMEM is derived from a first reconstruction with, *e.g.*, a uniform prior. Because it makes no attempt to directly convert the data into complex visibilities, a strength of BSMEM is that it can handle any types of data sparseness (such as missing closures). Thus, in principle, BSMEM could be used to restore images when Fourier phase data are completely missing (see Fig. 8).

3.5.2. The Building Block Method

The *Building Block Method* (Hofmann and Weigelt, 1993) was developed for reconstructing images from data such as the bispectrum obtained by means of speckle or long baseline interferometry. The Building Block Method is iterative, the objective of each step being to reduce a cost function $f_{\mathsf{data}}^{\mathsf{bisp}}$ equals to that in Eq. (85) with weights set to a constant or to:

$$w_{j_1,j_2,j_3,m}^{\mathsf{bisp}}(\boldsymbol{x}) = \frac{\left|B_{j_1,j_2,j_3,m}^{\mathsf{data}} - B_{j_1,j_2,j_3,m}^{\mathsf{model}}(\boldsymbol{x})\right|^2}{\left|B_{j_1,j_2,j_3,m}^{\mathsf{data}} - B_{j_1,j_2,j_3,m}^{\mathsf{model}}(\boldsymbol{x})\right|^2 + \mathrm{Var}\left(B_{j_1,j_2,j_3,m}^{\mathsf{data}}\right)}\,, \tag{89}$$

with $\mathrm{Var}\left(B_{j_1,j_2,j_3,m}^{\mathsf{data}}\right)$. Those weights therefore depends on the sought parameters $\boldsymbol{x}$ and, according to Hofmann and Weigelt (1993), the particular expression in Eq. (89) was motivated by Wiener filtering... The regularization is insured by a sparseness constraint and minimization of the penalty is achieved by a matching pursuit algorithm. The image is given by the building block model in Eq. (11) and, at k-th iteration, the new image $I^{[k]}(\boldsymbol{\theta})$ is obtained by adding a new building block at location $\boldsymbol{\theta}^{[k]}$ with a weight $\alpha^{[k]}$ to the previous image:

$$I^{[k]}(\boldsymbol{\theta}) = I^{[k-1]}(\boldsymbol{\theta}) + \alpha^{[k]}\, b(\boldsymbol{\theta} - \boldsymbol{\theta}^{[k]}) \tag{90}$$

or, if strict normalization is applied:

$$I^{[k]}(\boldsymbol{\theta}) = (1 - \alpha^{[k]})\, I^{[k-1]}(\boldsymbol{\theta}) + \alpha^{[k]}\, b(\boldsymbol{\theta} - \boldsymbol{\theta}^{[k]})\,. \tag{91}$$

The weight and location of the new building block is derived by minimizing the criterion $f_{\mathsf{data}}^{\mathsf{bisp}}$ with respect to these parameters. Strict positivity and support constraint can be (trivially) enforced by limiting the possible values for $\alpha^{[k]}$ and $\boldsymbol{\theta}^{[k]}$ respectively. To improve convergence, the method allows to *remove* building blocks (that is, add blocks with negative weights). A number of improvements have been implemented to speed up the computation, *e.g.* by adding/removing more than one block at a time. To avoid super resolution artifacts, the final image is convolved with a smoothing function with size set according to the spatial resolution of the instrument. The *Building Block Method* is very similar to the CLEAN method but designed for a different kind of data.

3.5.3. MACIM algorithm

MACIM (for MArkov Chain IMager) was developed by Ireland et al. (2008) and implements a global optimization strategy in a strict Bayesian framework. As explained in Section 2.6, the probability of the model given the data writes:

$$\Pr(\boldsymbol{x}|\boldsymbol{d}) \propto \exp\left(-\frac{1}{2}\, f_{\mathsf{data}}(\boldsymbol{x}) - \frac{\mu}{2}\, f_{\mathsf{prior}}(\boldsymbol{x})\right) ,$$

where the likelihood depends on the kind of data considered (powerspectrum and bispectrum). MACIM implements MEM regularization and a specific regularizer which favor large regions of dark space in-between bright regions. The penalty for this latter regularization is the sum of all pixel boundaries with zero flux on either side of the pixel boundary. MACIM attempts to maximize $\Pr(\boldsymbol{x}|\boldsymbol{d})$ by a simulated annealing algorithm with the Metropolis sampler. Although maximizing $\Pr(\boldsymbol{x}|\boldsymbol{d})$ is the same as minimizing $f_{\mathsf{data}}(\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})$, the use of probabilities is required by the Metropolis sampler to accept or reject the image samples. Theoretically, strict Bayesian point of view 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). In principle, simulated annealing is able to solve the global optimization problem of maximizing $\Pr(\boldsymbol{x}|\boldsymbol{d})$. However, the convergence of this kind of Monte-Carlo method can be very slow and depends on the parameters which set the temperature reduction law.

3.5.4. Mira algorithm

MIRA (Thiébaut, 2008) explicitly defines the sought image as the solution of a constrained optimization problem as the one in Eq. (47). To compute the model of the complex visibilities from the current image, MIRA uses the *exact* linear transform, Eq. (14), or Fourier interpolation (*cf.* Sect. 2.3) to speed up computations or to deal with large data sets. Minimization is carried on by a limited variable memory method (based on BFGS updates) with bound constraints (for the positivity) (Thiébaut, 2002). Since this method does not implement any global optimization strategy, the image restored by MIRA depends on the initial image. The strict normalization and positivity constraints are implemented by introducing the hidden variables $\boldsymbol{z}$, the image parameters being given by the non-linear change of variables:

$$x_n = \frac{z_n}{\sum_{n'} z_{n'}} \tag{92}$$

which insure that $\sum_n x_n = 1$. The joint penalty in Eq. (46) is then minimized with respect to $\boldsymbol{z}$ and subject to $\boldsymbol{z} \geq 0$ which also insures positivity of the image. MIRA is written in a modular way: basically any type of data can be taken into account providing a function that computes the corresponding penalty and its gradient with respect to the complex visibility is coded, currently complex visibility, powerspectrum and closure-phase data are possible via likelihood terms given by Eq. (80) and Eq. (84). Also many different regularizers are built into

MIRA (negentropy, quadratic of edge-preserving smoothness, total variation, *etc.*) and provisions are made to implement custom priors. For instance, a *compactness* prior is achieved by a very simple quadratic penalty:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum_n w_n^{\mathsf{prior}}\, x_n^2\,, \tag{93}$$

where the weights are increasing with the distance to the optical axis thus favoring structures concentrated within the center of the field of view. It has been shown (Le Besnerais et al., 2008) that, in the absence of any data, the default image yielded by this prior is: $\bar{x}_n \propto 1/w_n^{\mathsf{prior}}$, where the factor comes from the normalization requirement. For instance:

$$w_{\mathsf{prior}}(\boldsymbol{\theta}) = 1 + \frac{\theta_1^2 + \theta_2^2}{4\,\Gamma^2}\,, \tag{94}$$

yields a default image with radial symmetry and Lorentzian shape with full width at half minimum (FWHM) equals to Γ. Although simple, this regularizer can be very effective as shown by Fig. 6.

Since MIRA accounts for any available data, it can cope with missing data, in particular, it can be used to restore an image given only the powerspectrum (*i.e.* without any Fourier phase information) with at least a 180° orientation ambiguity. An example of such a reconstruction with no phase data is shown in Fig. 8.

3.5.5. Wisard algorithm

WISARD (Meimon et al., 2005b) algorithm recovers an image from powerspectrum and phase closure data. It shares some similarities with the self-calibration method (*cf.* Section 3.2) 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 to minimize the number of unknowns. Then, the synthesized Fourier phases are combined to the squared root of the measured powerspectrum to generate pseudo complex visibility data which are fit by the image restoration step. This step is performed by using the chosen regularization and a penalty with respect to the pseudo complex visibility data. However to account for a more realistic approximation of the distribution of complex visibility errors, WISARD make uses of a quadratic penalty which is different from the usual Goodman approximation (Meimon et al., 2005a). For the image restoration step, WISARD uses the same constrained optimization method as MIRA and also the same change of variable to impose the normalization. Taken separately, the image restoration step is a convex problem with a unique solution, the self-calibration step is certainly not strictly convex but does not seem to pose insurmountable problems (like in original self-calibration method). Nevertheless the global problem is multi-modal and, at least in difficult cases, the final solution depends on the initial guess.

There are many possible regularizers built into WISARD, for instance, the

edge-preserving smoothness prior is implemented by:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum_{n_1,n_2} c(\epsilon^{-1}\,|\nabla \boldsymbol{x}|_{n_1,n_2}) \tag{95}$$

where $\epsilon > 0$ is a chosen threshold, $|\nabla \boldsymbol{x}|$ is the magnitude of the spatial gradient in the image:

$$|\nabla \boldsymbol{x}|_{n_1,n_2} = \sqrt{(x_{n_1+1,n_2} - x_{n_1,n_2})^2 + (x_{n_1,n_2+1} - x_{n_1,n_2})^2}\,, \tag{96}$$

and $c(\,)$ is an $\ell_2 - \ell_1$ cost function such as:

$$c(z) = |z| - \log(1 + |z|)\,. \tag{97}$$

In words, the penalization in Eq. (95) behaves as a quadratic (resp. linear) function where the magnitude of the spatial gradient is small (resp. large) compared to ϵ. Thus reduction of small local variations in the sought image whereas not penalizing too much strong sharp features is achieved by this regularization penalty.

Note that MIRA and WISARD have been developed in parallel and therefore share some common features. They use the same optimization engine, and implement similar regularizations (edge-preserving smoothness, compactness). However they differ in the way missing data is taken into account. WISARD takes a self-calibration approach to *explicitly* solve for missing Fourier phase information; whereas MIRA which only use forward modeling (that is, from the image parameters to the data) *implicitly* take into account any lack of information. A more detailed comparison of the two algorithms has been done by Le Besnerais et al. (2008).

3.6. Comparison of algorithms on simulated data

The above described algorithms have been compared on simulated data during *Interferometric Beauty Contests* (Lawson et al., 2004, 2006; Cotton et al., 2008). The results of the contest were very encouraging. Although being quite different algorithms, BSMEM, the Building Block Method, MIRA and WISARD gives good image reconstructions where the main features of the object of interest can be identified in spite of the rather sparse u-v coverage (compared to what is usually available by radio-interferometry), the lack of some Fourier phase information and the non-linearities of the measurements. BSMEM and MIRA appear to be the most successful algorithms (they respectively won the first two and last one *Interferometric Beauty Contests*).

With their tuning parameters and, for some of them, the requirement to start the algorithm with an initial image, these algorithms still need some (modest) expertise to be used successfully. But this is quite manageable if one does not assume that image reconstruction is done by a *black box* algorithm and realizes that it is really a data processing task which requires user feedback and interaction. For instance, the tuning of the regularization level can be

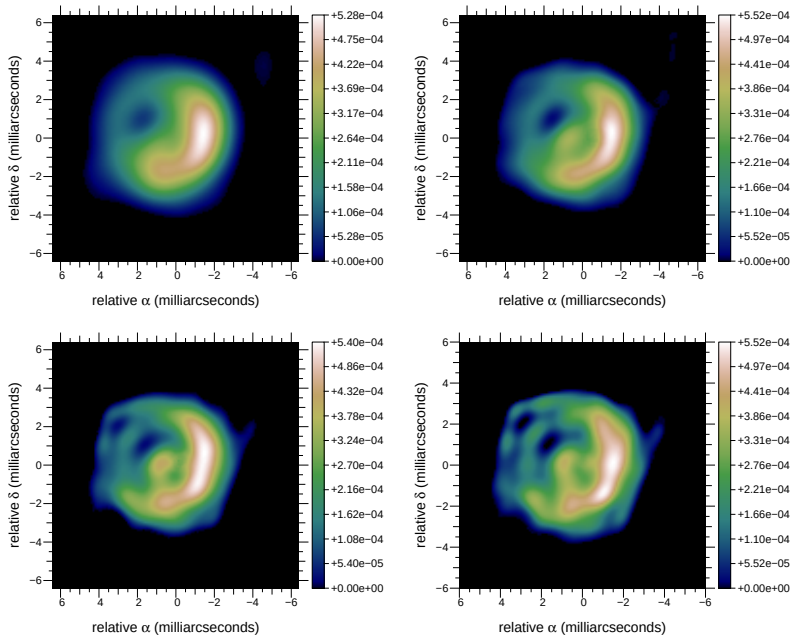

Figure 7: Image reconstruction under various regularization levels. Algorithm is MIRA with edge-preserving regularization given in Eq. (95) with $\epsilon = 10^{-4}$ and $\mu = 10^{-6}$ (top-left), $\mu = 10^{-5}$ (top-right), $\mu = 10^{-4}$ (bottom-left) and $\mu = 3 \times 10^{-3}$ (bottom-right).

derived from Bayesian considerations but can also almost be done by visual inspection of the restored image. From Fig. 7, 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. Indeed, the ~ 50 students of the *2008' VLTI Summer School* held in Keszthely were all able to obtain acceptable results with MIRA after a few trials and errors.

Figure 6 shows image reconstructions from one of the data sets of the *2004' Beauty Contest* (Lawson et al., 2004) and with different types of regularization. The most obvious result is that the synthesized images do not greatly differ and are all quite acceptable approximations of the reality (compare for instance with the dirty image in Fig. 4). Hence, providing 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 that really matters not the specific expression of the penalty imposing the prior.

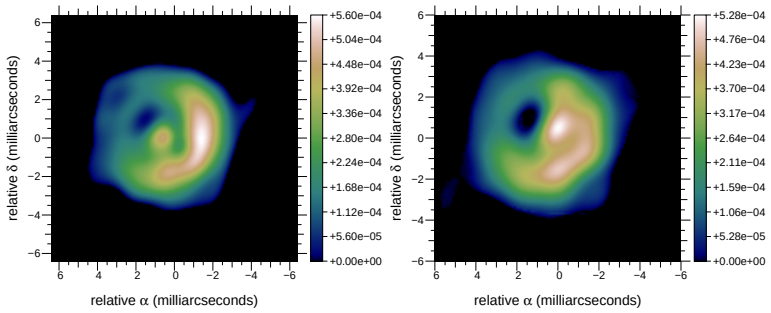

Figure 8: Image reconstruction without any Fourier phase information. The initial image must not be symmetrical and was generated with uniformly distributed random pixel values.

4. Discussion

The general framework of inverse problem approach has been used to formally analyze the requirements of image restoration algorithms for optical interferometry. In particular, the ability to account for prior information is critical to supplement missing data. Indeed, effective algorithms must be able to perform some kind of Fourier interpolation to overcome the sparseness of the u-v coverage and to recover the Fourier phases while this information is only partially available. Unknown complex throughputs and non-linearities in the measurements are other issues that must be overcame by the image restoration algorithms.

This framework has been used to describe the most important existing algorithms as being more or less directly related to the optimization of a mixed criterion under some strict constraints such as positivity and normalization. Two different type of terms appear into this criterion: likelihood terms which enforce agreement of the model image with the actual data and regularization terms which maintain the image close to the priors. Hence, the differences between the various algorithms lie in the kind of measurements considered, the assumed data model and the type of prior. For non-convex criteria which occur when the complex throughputs are unknown or when non-linear 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 multi-modal problems, the reading of the excellent books by Nocedal and Wright (2006) and Bonnans et al. (2006) is recommended for a good understanding of numerical optimization methods. For instance, the method of Skilling and Bryan (1984) integrates most components of a successful optimization strategy (preconditioning, automatic tuning of the regularization level, *etc.*).

Compared to radio-interferometry, image synthesis in optical interferometry turns out to be a task of higher difficulty due to the much sparser u-v coverage and to the non-linearities. For low signal to noise data (*i.e.* faint objects), the best can be extracted from the measurements only if the likelihood term is

a good approximation of the true one. This may rules out regridding (which introduces correlations usually ignored) and the Goodman model of noise for complex data which are both used in radio-interferometry to simplify image restoration. Available algorithms are however now ready for image reconstruction from *real* data. These algorithms are not fully automated black boxes: at least some tuning parameters and the type of regularization are left to the user choice. A general understanding of the mechanisms involved into image restoration algorithms is mandatory to fully exploit any of these methods. This is also needed to analyze possible artifacts in the synthesized images.

There is plenty of possible future developments for image restoration methods. To mention only a few: reconstruction of 3-D multi-wavelength images, reconstruction from multiple data sets (to account for a lower resolution image or for calibration data). Also another possible source of inspiration could be provided by the literature in medical imaging since tomographic data share many common properties with interferometric data (basically it consists in sparse samples in the Fourier u-v plane).

A. A practical example with MiRA

MiRA (Multi-aperture Image Reconstruction Algorithm) was specifically developed for image reconstruction from optical interferometric data. The software is written in Yorick[8] with some extensions and is freely distributed [9]. To use MiRA, you must have installed recent versions of Yorick (version $\geq$ 2.1.05 or CVS version), Yeti[10] (version $\geq$ 6.2.3) and OptimPack[11] (version $\geq$ 1.3).

In what follows, examples of image reconstruction with MIRA are shown with comments to explain the different steps. The examples are performed into an interactive Yorick session, hence commands are typed at Yorick prompt. Yorick's syntax is very similar to that of C. Documentation for a particular function or procedure, says *cmd*, is obtained by typing `help, cmd` and then the Return key.

Typically, the different stages are:

1. Load Mira software:

   ```
   include, "mira.i";
   ```

2. Load input data into opaque object:

   ```
   ws = mira_new(MIRA_HOME + "data/data1.oifits");
   ```

   where `MIRA_HOME` is a global variable defined by MIRA with the name of the directory where MIRA software is installed (with a final `"/"`). Note that `+` is used for string concatenation in Yorick. Instead of `ws`, another

[8] http://yorick.sourceforge.net/
[9] http://www-obs.univ-lyon1.fr/labo/perso/eric.thiebaut/mira.html
[10] http://www-obs.univ-lyon1.fr/labo/perso/eric.thiebaut/yeti.html
[11] http://www-obs.univ-lyon1.fr/labo/perso/eric.thiebaut/optimpack.html

variable name can be used to store the opaque object (not forgetting to change the code below accordingly).

3. Configure for image reconstruction:

```
mira_config, ws, xform="exact", dim=100,
             pixelsize=0.4*MIRA_MILLIARCSECOND;
```

`xform` is the linear transform to use to approximate the Fourier transform on an irregularly spaced grid, here `"exact"` means to use the *exact* but slow transform, another possibility is `xform="fft"` to use Fast Fourier Transform and bi-linear interpolation of the spatial frequencies (*cf.* Sect. 2.3); `dim` is the number of pixels along one side of the image, the image is square; `pixelsize` is the angular size of the pixels in radians, `MIRA_MILLIARCSECOND` is another global variable set by MIRA.

4. Choose a regularization method for the term $f_{\mathsf{prior}}(\boldsymbol{x})$:

```
rgl = rgl_new("smoothness");
```

which returns another opaque object that stores the type and the parameters of the regularization method. There are many other regularization methods built into MIRA and which can be listed by the command `rgl_info`.

5. Create an initial image which is a point-like object:

```
dim = mira_get_dim(ws);          // get image dimension
img0 = array(double, dim, dim);  // creates DIMxDIM array
img0(dim/2, dim/2) = 1.0;        // set central pixel
```

where `//` marks a comment until the end of line.

6. Attempt an image reconstruction (starting from the point-like object):

```
img1 = mira_solve(ws, img0, maxeval=500, verb=1, xmin=0.0,
                  normalization=1, regul=rgl, mu=1e6);
```

where `maxeval` set the maximum number of evaluations of the penalty function $f_{\mathsf{data}}(\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})$, `verb` set the frequency of the displayed information, `xmin` set the minimum value of the image pixels and is used here to impose positivity, `normalization` set the normalization of the image, `regul` is the regularizer, and `mu` is the regularization weight (the parameter μ).

7. The image obtained by the previous command may not be a satisfying reconstruction because: (i) the algorithm has not yet converged; (ii) the regularization method or the regularization parameters (*e.g.* μ) are not correct; (iii) the image is not centered in the field of view (the data is insensitive to a shift of the image); *etc.* To continue the iterations, it is possible to resume the reconstruction starting with the image obtained so far:

```
img1 = mira_solve(ws, img1, maxeval=500, verb=1, xmin=0.0,
                  normalization=1, regul=rgl, mu=1e6);
```

(note that `img1` has been used in place of `img0` as the initial image). It is also possible to use different parameters or, for instance, to continue the reconstruction with the recentered image:

```
img1 = mira_solve(ws, mira_recenter(img1), maxeval=500,
                  verb=1, xmin=0.0, normalization=1,
                  regul=rgl, mu=1e6);
```

Generally a few calls to `mira_solve` with different starting points and different regularization settings will be needed to find a *good* solution.

8. As a final example, let us change the regularization to use the $\ell_2 - \ell_1$ smoothness constraint defined in Eq. (95):

```
// Define regularization:
rgl = rgl_new("xsmooth");
rgl_config, rgl, "cost",      "cost_l2l1",
                 "threshold", 1e-4,
                 "dimlist",   dimsof(img0);

// Initial reconstruction (starting from a the previous image):
img2 = mira_solve(ws, img1, maxeval=2000, verb=1, xmin=0.0,
                  normalization=1, regul=rgl, mu=3e5);
```

References

Bonnans, J., Gilbert, J.-C., Lemaréchal, C., Sagastizábal, C., 2006. Numerical Optimization, 2nd Edition. Universitext. Springer.

Buscher, D. F., 1994. Direct maximum-entropy image reconstruction from the bispectrum. In: Robertson, J. G., Tango, W. J. (Eds.), IAU Symp. 158: Very High Angular Resolution Imaging. pp. 91–+.

Campisi, P., Egiazarian, K. (Eds.), 2007. Blind image deconvolution: theory and applications. CRC Press.

Candes, E. J., Romberg, J., Tao, T., 2006. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Transactions on Information Theory 52 (2), 489– 509.
URL http://arxiv.org/pdf/math.NA/0409186.pdf

Cornwell, T. J., Wilkinson, P. N., 1981. A new method for making maps with unstable radio interferometers. Mon. Not. Royal Astron. Soc. 196, 1067–1086.

Cotton, W., Monnier, J., Baron, F., Hofmann, K.-H., Kraus, S., Weigelt, G., Rengaswamy, S., Thibaut, E., Lawson, P., Jaffe, W., Hummel, C., Pauls, T., Henrique, Tuthill, P., Young, J., 2008. 2008 imaging beauty contest. In: Markus Schöller, William C. Danchi, F. D. (Ed.), Astronomical Telescopes and Instrumentation. Vol. 7013. SPIE, p. 70131N.
URL ftp://ftp.cv.nrao.edu/NRAO-staff/bcotton/Beauty08/Beauty08.pdf

Cotton, W. D., 1979. A method of mapping compact structure in radio sources using VLBI observations. Astron. J. 84, 1122–1128.

Delplancke, F., Derie, F., Paresce, F., Glindemann, A., Lévy, F., Lévêque, S., Ménardi, S., 2003. Prima for the vlti - science. Astrophys. & Space Science 286, 99–104.

Frigo, M., Johnson, S. G., 2005. The design and implementation of FFTW3. Proceedings of the IEEE 93 (2), 216–231, special issue on "Program Generation, Optimization, and Platform Adaptation".

Giovannelli, J.-F., Coulais, A., 2005. Positive deconvolution for superimposed extended source and point sources. Astron. & Astrophys. 439, 401–412.

Goodman, J. W., 1985. Statistical Optics. John Wiley & Sons.

Haniff, C., 1991. Least-squares fourier phase estimation from the modulo 2π bispectrum phase. J. Opt. Soc. Am. A 8 (1), 134–140.

Hofmann, K.-H., Weigelt, G., 1993. Iterative image reconstruction from the bispectrum. Astron. & Astrophys. 278, 328–339.

Högbom, J. A., 1974. Aperture synthesis with a non-regular distribution of interferometer baselines. Astron. & Astrophys., Suppl. 15, 417–426.

Horn, K., 1985. Images of accretion discs – i. the eclipse mapping method. Mon. Not. Royal Astron. Soc. 213, 129–141.

Ireland, M., Monnier, J., Thureau, N., 2008. Monte-carlo imaging for optical interferometry. SPIE.

Labeyrie, A., 1975. Interference fringes obtained on VEGA with two optical telescopes. Astrophys. J., Let. 196, L71–L75.

Lacour, S., Meimon, S., Thiébaut, E., Perrin, G., Verhoelst, T., Pedretti, E., Schuller, P. A., Mugnier, L., Monnier, J., Berger, J., Haubois, X., Poncelet, A., Besnerais, G. L., Eriksson, K., Millan-Gabet, R., Lacasse, M., Traub, W., 2008. The limb-darkened arcturus - imaging with the iota/ionic interferometer. Astron. & Astrophys. accepted for publication, 561–570.

Lacour, S., Thiébaut, E., Perrin, G., 2007. High dynamic range imaging with a single-mode pupil remapping system: a self-calibration algorithm for highly redundant interferometric arrays. Mon. Not. Royal Astron. Soc. 374, 832–846.

Lacour, S., Thiébaut, E., Perrin, G., Meimon, S., Haubois, X., Pedretti, E., Ridgway, S., Monnier, J., Berger, J.-P., Schuller, P., Woodruff, H., Poncelet, A., Coroller, H. L., Millan-Gabet, R., Lacasse, M., Traub, W., 2009. The pulsation of χ cygni imaged by optical interferometry: a novel technique to derive distance and mass of mira stars. Astrophys. J. 707, 632–643.

Lannes, A., Anterrieu, E., Maréchal, P., 1997. Clean and wipe. Astron. & Astrophys., Suppl. 123, 183–198.

Lawson, P. R., Cotton, W. D., Hummel, C. A., Baron, F., Young, J. S., Kraus, S., Hofmann, K.-H., Weigelt, G. P., Ireland, M., Monnier, J. D., Thiébaut, E., Rengaswamy, S., Chesneau, O., 2006. The 2006 interferometry image beauty contest. In: Monnier, J. D., Schöller, M., Danchi, W. C. (Eds.), Advances in Stellar Interferometry. Vol. 6268 of Proceedings of the SPIE.

Lawson, P. R., Cotton, W. D., Hummel, C. A., Monnier, J. D., Zhao, M., Young, J. S., Thorsteinsson, H., Meimon, S. C., Mugnier, L., Le Besnerais, G., Thibaut, E., Tuthill, P. G., 2004. The 2004 optical/ir interferometry imaging beauty contest. American Astronomical Society Meeting Abstracts 205, –+.

Le Besnerais, G., Lacour, S., Mugnier, L. M., Thiébaut, E., Perrin, G., Meimon, S., 2008. Advanced imaging methods for long-baseline optical interferometry. IEEE Journal of Selected Topics in Signal Processing 2 (4).

Le Bouquin, J.-B., Lacour, S., Renard, S., Thibaut, E., Merand, A., 2009. Pre-maximum spectro-imaging of the mira star t lep with amber/vlti. Astron. & Astrophys. 496, L1–L4.

Meimon, S., Mugnier, L. M., Besnerais, G. L., 2005a. Convex approximation to the likelihood criterion for aperture synthesis imaging. J. Opt. Soc. Am. A 22, 2348–2356.

Meimon, S., Mugnier, L. M., Besnerais, G. L., 2005b. Reconstruction method for weak-phase optical interferometry. Optics Letters 30, 1809–1811.

Monnier, J., 2003. Optical interferometry in astronomy. Reports on Progress in Physics 66 (5), 789–857.

Narayan, R., Nityananda, R., 1986. Maximum entropy image restoration in astronomy. Annual Rev. of Astron. & Astrophys. 24, 127–170.

Nocedal, J., Wright, S. J., 2006. Numerical Optimization, 2nd Edition. Springer Verlag.

Pauls, T. A., Young, J. S., Cotton, W. D., Monnier, J. D., 2005. A data exchange standard for optical (visible/ir) interferometry. Publications of the ASP 117, 1255–1262.

Perrin, G., 2009. VLTI Science Highlights. In: Proc. Astrophysics and Space Science: Science with the VLT in the ELT Era. pp. 81–87.

Potts, D., Steidl, G., Tasche, M., 2001. Modern Sampling Theory: Mathematics and Applications. Ch. Fast Fourier transforms for nonequispaced data: A tutorial, pp. 249–274.

Quirrenbach, A., 2001. Optical Interferometry. Annual Rev. of Astron. & Astrophys. 39 (1), 353–401.

Quirrenbach, A., 2009. The development of astronomical interferometry. Experimental Astronomy 26, 49–63.

Readhead, A. C. S., Wilkinson, P. N., 1978. The mapping of compact radio sources from VLBI data. Astrophys. J. 223, 25–36.

Roddier, F., 1981. The effects of atmospheric turbulence in optical astronomy. Vol. 19. North-Holland Publishing Company, Amsterdam, pp. 281–376.

Schwab, F. R., 1980. Adaptive calibration of radio interferometer data. In: Rhodes, W. T. (Ed.), Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series. Vol. 231 of Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference. pp. 18–25.

Schwarz, U. J., 1978. Mathematical-statistical description of the iterative beam removing technique (method clean). Astron. & Astrophys. 65, 345–+.

Skilling, J., Bryan, R. K., 1984. Maximum entropy image reconstruction: general algorithm. Mon. Not. Royal Astron. Soc. 211, 111–124.

Sramek, R. A., Schwab, F. R., 1989. Imaging. In: Perley, R. A., Schwab, F. R., Bridle, A. H. (Eds.), Synthesis Imaging in Radio Astronomy. Vol. 6 of Astronomical Society of the Pacific Conference Series. pp. 117–+.

Starck, J.-L., Bijaoui, A., Lopez, B., Perrier, C., 1994. Image reconstruction by the wavelet transform applied to aperture synthesis. Astron. & Astrophys. 283, 349–360.

Thiébaut, E., 2002. Optimization issues in blind deconvolution algorithms. In: Starck, J.-L., Murtagh, F. D. (Eds.), Astronomical Data Analysis II. Vol. 4847. SPIE, pp. 174–183.

Thiébaut, E., 2008. Mira: an effective imaging algorithm for optical interferometry. In: Markus Schöller, William C. Danchi, F. D. (Ed.), Astronomical Telescopes and Instrumentation. Vol. 7013. SPIE, p. 70131I.

Thompson, A. R., Bracewell, R. N., 1974. Interpolation and Fourier transformation of fringe visibilities. Astron. J. 79, 11–+.

Tikhonov, A. N., Arsenin, V. Y., 1977. Solution of Ill-posed Problems. Scripta Series in Mathematics. Winston & Sons, Washington.

Wakker, B. P., Schwarz, U. J., 1988. The Multi-Resolution CLEAN and its application to the short-spacing problem in interferometry. Astron. & Astrophys. 200, 312–322.