

Unified Image Reconstruction Description

Opticon WP4 Deliverable 4.1

John Young and Eric Thiébaud

rev 0.2

January 3, 2017



Abstract

This document provides background information to astronomers wishing to use image reconstruction algorithms specifically developed for optical (visible/IR) interferometry, such as MiRA, WISARD, and BSMEM. It should be read in conjunction with the algorithm-specific documentation and cookbooks that will also be provided as outputs of this Joint Research Activity. We describe image reconstruction from interferometric data with a focus on the specific problems encountered at optical wavelengths. The challenging issues in image reconstruction from interferometric data are introduced in the context of a general inverse problem approach. This framework is then used to describe the algorithms specifically developed for optical interferometry. Other topics of interest to users of these algorithms are also covered, including general descriptions of the inputs, outputs and user interactions needed for successful image reconstruction.

1 Introduction

The technique of aperture synthesis imaging, when applied at optical (visible and near-infrared) wavelengths, offers an invaluable capability to image astrophysical processes on milliarcsecond angular scales. The state-of-the-art optical interferometers capable of performing such observations, for example VLTI ([Glindemann et al. 2000](#)) and CHARA ([ten Brummelaar et al. 2005](#)), are now mature instruments, and have yielded a number of published images. However, the available image reconstruction software tools are relatively difficult for inexperienced astronomers to use, and this is a significant (but by no means the only) factor limiting the number of model-independent images that are published.

This document has been produced as part of a Joint Research Activity (JRA), funded by the EU Seventh Framework Programme, which aims to provide user-friendly image reconstruction software, accessible to the entire community, and compatible with the instrumental capabilities that will be offered by VLTI and CHARA a few years hence. The main outputs of the JRA will be modified versions of existing algorithms with a common user interface, test input datasets representative of a range of interferometric instruments and observation types, and cookbooks that guide new users through various example image reconstruction tasks.

The purpose of this document is to provide relevant background information for end-users of the image reconstruction software and cookbooks, and also for researchers preparing the other JRA outputs. Specifically, we will introduce a lexical and mathematical vocabulary for describing the algorithms, and provide high-level descriptions of the algorithms with a focus on their inputs, outputs, similarities and differences. To inform the design of the common user interface, we also enumerate common failure modes and explain the nature of any user interventions that may be required to ensure successful image reconstruction. A more detailed presentation of the mathematical background is given by [Thiébaud \(2013\)](#).

Multi-telescope interferometers provide sparse measurements of the Fourier transform of the brightness distribution of the observed objects (Section 2). An inverse problem methodology (Section 3) can be used to reconstruct images from such sparsely-sampled Fourier data. At optical wavelengths, additional problems arise due to incomplete Fourier phase information, and to the non-linearity of the direct model of the data in terms of the image parameters. These issues have led to the development of specific algorithms based on similar inverse problem approaches (Section 4). The input parameters required by these algorithms are summarised in Section 4.3.

In the following sections, key terms are highlighted (here is an example of the typography: **complex visibility**). Where several synonyms are in common usage, the highlighted one is preferred for the JRA outputs.

2 Interferometric Data

The instantaneous and monochromatic output of an optical interferometer is the so-called **complex visibility** $V_{j_1, j_2}(t)$ of the interference fringes between the j_1 -th and the j_2 -th telescopes at instant t (Monnier 2003):

$$V_{j_1, j_2}(t) = g_{j_1}(t)^* g_{j_2}(t) \hat{I}(\boldsymbol{\nu}_{j_1, j_2}(t)) . \quad (1)$$

Here $\hat{I}(\boldsymbol{\nu})$ is the Fourier transform of $I(\boldsymbol{\theta})$, the brightness distribution of the observed object in angular direction $\boldsymbol{\theta}$, and $g_j(t)$ is the complex amplitude throughput for the light from the j -th telescope. The spatial frequency $\boldsymbol{\nu}_{j_1, j_2}(t) = (\mathbf{r}_{j_2}(t) - \mathbf{r}_{j_1}(t))/\lambda$ is determined by the projection of the baseline between the pair of telescopes (j_1, j_2) onto the plane perpendicular to the line of sight, and is inversely proportional to the observing wavelength λ (see Fig. 1).

These equations assume that the diameters of the telescopes are much smaller than their projected separation and that the object is an incoherent light source. An interferometer therefore provides sparse measurements of the Fourier transform of the brightness distribution of the observed object. The top-left panel of Fig. 2 shows an example of the sampling of spatial frequencies by an interferometer.

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

$$V_{j_1, j_2, m}^{\text{data}} = \langle V_{j_1, j_2}(t) \rangle_m + V_{j_1, j_2, m}^{\text{err}} \quad (2)$$

where $\langle \rangle_m$ denotes averaging during the m -th exposure and $V_{j_1, j_2, m}^{\text{err}}$ stands for the errors due to noise and modeling approximations. The exposure duration is short enough to consider the projected baseline $\mathbf{r}_{j_2}(t) - \mathbf{r}_{j_1}(t)$ as constant, thus:

$$\langle V_{j_1, j_2}(t) \rangle_m \simeq G_{j_1, j_2, m} \hat{I}(\boldsymbol{\nu}_{j_1, j_2, m}) \quad (3)$$

with $\boldsymbol{\nu}_{j_1, j_2, m} = \langle \boldsymbol{\nu}_{j_1, j_2}(t) \rangle_m \simeq \boldsymbol{\nu}_{j_1, j_2}(t_m)$, $t_m = \langle t \rangle_m$ the mean time of the observation and $G_{j_1, j_2, m} = \langle g_{j_1}(t)^* g_{j_2}(t) \rangle_m$ the effective **optical transfer function** (OTF). The fast variations of the instantaneous OTF are mainly due to the random optical path differences (OPD) caused by the atmospheric turbulence.

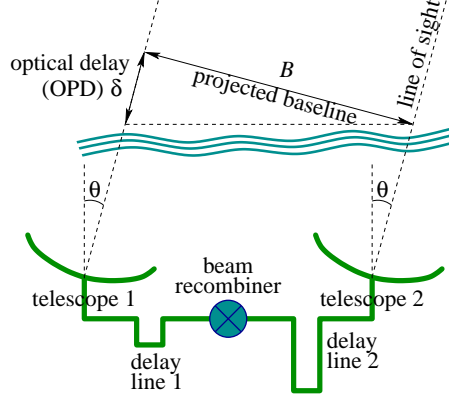


Figure 1: Geometrical layout of an interferometer. B is the projected baseline, θ is the view angle and δ is the geometrical optical path difference which is compensated by the delay lines.

At short wavelengths (visible and infrared), the phase variance exceeds a few squared radians and $G_{j_1, j_2, m} \simeq 0$, hence the object's complex visibility cannot be directly measured. A possible solution would be to compensate for the OPD errors in real time using fast delay lines. However, this requires a bright reference source close to the observed object and a *dual-star* interferometer such as VLTI/PRIMA (Delplancke et al. 2003) that can relate the phase measured on the reference source to a simultaneous measurement of the target object. An alternative approach is to integrate non-linear estimators that are insensitive to telescope-wise phase errors. This requires high acquisition rates (~ 1000 Hz in the near infrared) and involves special data processing but otherwise no special instrumentation.

Following the second approach, most current optical interferometers integrate the **power spectrum** (for $j_1 \neq j_2$):

$$S_{j_1, j_2, m} = \langle |V_{j_1, j_2}(t)|^2 \rangle_m \simeq \rho_{j_1, m} \rho_{j_2, m} |\hat{I}(\boldsymbol{\nu}_{j_1, j_2, m})|^2, \quad (4)$$

with $\rho_{j, m} = \langle |g_j(t)|^2 \rangle_m$ the mean squared modulus of the complex throughput of the j -th telescope during the m -th exposure. The transfer function $\rho_{j_1, m} \rho_{j_2, m}$ of the power spectrum is insensitive to the phase errors and can be estimated by simultaneous photometric calibration and, to compensate for residual effects, from the power spectrum of a reference source (a so-called *calibrator*) observed shortly before and/or afterwards. Hence measuring $S_{j_1, j_2, m}$ gives the object power spectrum $|\hat{I}(\boldsymbol{\nu}_{j_1, j_2, m})|^2$ but carries no Fourier phase information.

To obtain Fourier phase information, the **bispectrum** of the complex visibilities is measured:

$$\begin{aligned} B_{j_1, j_2, j_3, m} &= \langle V_{j_1, j_2}(t) V_{j_2, j_3}(t) V_{j_3, j_1}(t) \rangle_m \\ &\simeq \rho_{j_1, m} \rho_{j_2, m} \rho_{j_3, m} \hat{I}(\boldsymbol{\nu}_{j_1, j_2, m}) \hat{I}(\boldsymbol{\nu}_{j_2, j_3, m}) \hat{I}(\boldsymbol{\nu}_{j_3, j_1, m}), \end{aligned} \quad (5)$$

where j_1 , j_2 and j_3 denote three different telescopes. As for the power spectrum, the bispectrum transfer function $\rho_{j_1,m} \rho_{j_2,m} \rho_{j_3,m}$ can be calibrated. Since this transfer function is real, it has no effect on the phase of the bispectrum (the so-called **closure phase**) which is equal to that of the object. However a fraction of the phase information is missing. The set of all baselines between T telescopes (in a non-redundant configuration), samples $T(T-1)/2$ different spatial frequencies but the closure phase only yields $(T-1)(T-2)/2$ linearly independent phase estimates (Monnier 2003). The smaller the number of telescopes, the larger the fraction of missing phase information. Furthermore, all information on the absolute position of the observed object is lost regardless of the number of telescopes.

In practice, obtaining the power spectrum and the bispectrum involves measuring the instantaneous complex visibilities (that is, for a very short integration time compared to the evolution of the turbulence) and averaging their power spectrum and bispectrum over the effective exposure time. Being non-linear functions of noisy variables, these quantities are biased but the biases are straightforward to remove (Gordon and Buscher 2012). To simplify the description of the algorithms, we will consider that the *de-biased* and *calibrated* power spectrum and bispectrum are available as input data for image reconstruction, thus:

$$S_{j_1,j_2,m}^{\text{data}} = |\hat{I}(\boldsymbol{\nu}_{j_1,j_2,m})|^2 + S_{j_1,j_2,m}^{\text{err}}, \quad (6)$$

$$B_{j_1,j_2,j_3,m}^{\text{data}} = \hat{I}(\boldsymbol{\nu}_{j_1,j_2,m}) \hat{I}(\boldsymbol{\nu}_{j_2,j_3,m}) \hat{I}(\boldsymbol{\nu}_{j_3,j_1,m}) + B_{j_1,j_2,j_3,m}^{\text{err}} \quad (7)$$

where $S_{j_1,j_2,m}^{\text{err}}$ and $B_{j_1,j_2,j_3,m}^{\text{err}}$ are zero-mean terms that account for noise and model errors.

3 Imaging from Sparse Fourier Data

We consider here the simplest problem of image reconstruction given sparse Fourier coefficients (the complex visibilities), assuming that the OTF has been calibrated. This discussion will inform the subsequent treatment of image reconstruction from non-linear observables in Sec. 4.

3.1 Data and Image Models

To simplify the notation, we introduce the *data vector* $\mathbf{y} \in \mathbb{C}^L$ which collates all the complex visibility measurements: $y_\ell = V_{j_1,j_2,m}^{\text{data}}$ with $\ell \sim (j_1, j_2, m)$ to denote a one-to-one mapping between index ℓ and triplet (j_1, j_2, m) for measured data. Long baseline interferometers provide data for a limited set $\mathcal{L} = \{\boldsymbol{\nu}_k\}_{k=1,\dots,K}$ of observed spatial frequencies, corresponding to the projected baselines at the times of the observations.

The image is a parametrized representation of the object brightness distri-

bution. A very general description is given by a linear expansion:

$$I(\boldsymbol{\theta}) = \sum_{n=1}^N x_n b_n(\boldsymbol{\theta}) \xrightarrow{\text{F.T.}} \hat{I}(\boldsymbol{\nu}) = \sum_{n=1}^N x_n \hat{b}_n(\boldsymbol{\nu}), \quad (8)$$

where $\{b_n(\boldsymbol{\theta})\}_{n=1,\dots,N}$ are basis functions and $\boldsymbol{x} \in \mathbb{R}^N$ are the image parameters, for instance, the values of the image *pixels*, or wavelet coefficients. In what follows, we will assume that \boldsymbol{x} are the image values on an evenly spaced grid of pixels.

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

$$\Delta\theta \ll \frac{\lambda}{2 B_{\max}} \quad (9)$$

where $B_{\max} = \max_{j_1, j_2, t} |\mathbf{r}_{j_1}(t) - \mathbf{r}_{j_2}(t)|$ is the maximum projected separation between interfering telescopes. Oversampling by a factor of at least 2 is usually used, hence the pixel size is given by: $\Delta\theta \lesssim \lambda/(4 B_{\max})$. To avoid aliasing and image truncation, the field of view must be chosen large enough and without forgetting that the reciprocal of the width of the field of view also sets the sampling step of the spatial frequencies.

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

$$V_k(\boldsymbol{x}) = \hat{I}(\boldsymbol{\nu}_k) = \sum_{n=1}^N T_{k,n} x_n, \quad (10)$$

where the coefficients of the matrix $\mathbf{T} \in \mathbb{C}^{K \times N}$ are $T_{k,n} = \hat{b}_n(\boldsymbol{\nu}_k)$. The matrix \mathbf{T} performs the Fourier transform of non-equispaced data, which is a very costly operation. This problem is not specific to interferometry, similar needs in crystallography, tomography and bio-medical imaging have led to the development of fast algorithms to approximate this operation (Potts et al. 2001). For instance:

$$\mathbf{T} \simeq \mathbf{R} \cdot \mathbf{F} \cdot \mathbf{S}, \quad (11)$$

where $\mathbf{F} \in \mathbb{C}^{N \times N}$ is the fast Fourier transform (FFT) operator, $\mathbf{R} \in \mathbb{C}^{K \times N}$ is a linear operator to interpolate the discrete Fourier transform of the image $\hat{\mathbf{x}} = \mathbf{F} \cdot \mathbf{x}$ at the observed spatial frequencies and \mathbf{S} is diagonal and compensates the field of view apodization (or spectral smoothing) caused by \mathbf{R} . Note that \mathbf{S} may also perform zero-padding of the image to reduce the frequency size and thus improve the precision.

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

$$\mathbf{y} = \mathbf{A} \cdot \mathbf{x} + \mathbf{e} \quad (12)$$

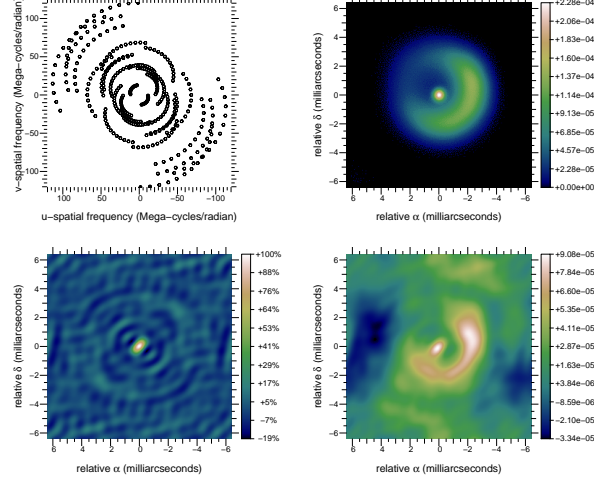


Figure 2: Top left: (u, v) coverage. Top right: observed object. Bottom left: dirty beam. Bottom right: dirty image. Object model and (u, v) coverage are from the *2004 Beauty Contest* (Lawson et al. 2004).

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

$$G_{\ell, k} = \begin{cases} g_{j_1, m}^* g_{j_2, m} & \text{if } \ell \sim (j_1, j_2, m) \in \mathcal{B}_k \\ 0 & \text{otherwise.} \end{cases} \quad (13)$$

Here \mathcal{B}_k is the set of telescope pairs and exposures yielding spatial frequency ν_k .

Applying the pseudo-inverse $\mathbf{T}^\dagger = \mathbf{S}^{-1} \cdot \mathbf{F}^{-1} \cdot \mathbf{R}^\dagger$ of \mathbf{T} to the data yields the so-called *dirty image* (see Fig. 2):

$$\hat{\mathbf{y}} = \mathbf{T}^\dagger \cdot \mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \hat{\mathbf{e}} \quad (14)$$

where $\hat{\mathbf{e}} = \mathbf{T}^\dagger \cdot \mathbf{e}$ and $\mathbf{H} = \mathbf{T}^\dagger \cdot \mathbf{G} \cdot \mathbf{T}$. Neglecting the apodization, \mathbf{H} essentially performs the convolution of the image by the **dirty beam** (see Fig. 2). From Eq. (12) and Eq. (14), image reconstruction from interferometric data can be seen either as a problem of interpolating *missing* Fourier coefficients or as deconvolution of the dirty map by the dirty beam (Giovannelli and Coulais 2005).

3.2 Inverse Problem Approach

Since many Fourier frequencies are not measured, simply fitting the data does not uniquely define the desired image. Such an ill-posed problem can be solved by an inverse problem approach (Tarantola 2005) by imposing *a priori* constraints to select a unique image among all those which are consistent with the

data. The role of the priors is to smoothly interpolate voids in the (u, v) coverage while avoiding high frequencies beyond the diffraction limit. Without loss of generality, we assume that these constraints are monitored by a penalty function $f_{\text{prior}}(\mathbf{x})$ which measures the agreement of the image with the priors: the lower $f_{\text{prior}}(\mathbf{x})$, the better the agreement. In the inverse problem framework, $f_{\text{prior}}(\mathbf{x})$ is known as the **regularization**. The parameters \mathbf{x}^+ of the image which best matches the priors while fitting the data are obtained by solving a constrained optimization problem:

$$\mathbf{x}^+ = \arg \min f_{\text{prior}}(\mathbf{x}), \text{ subject to: } \mathbf{A} \cdot \mathbf{x} = \mathbf{y}. \quad (15)$$

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

$$\mathcal{X} = \left\{ \mathbf{x} \in \mathbb{R}^N \mid \mathbf{x} \geq 0, \quad \sum_n x_n = 1 \right\} \quad (16)$$

where $\mathbf{x} \geq 0$ means: $\forall n, x_n \geq 0$. Rather than enforce precise agreement with the data, we can instead measure the distance of the model to the data by a penalty function $f_{\text{data}}(\mathbf{x})$ and replace the equality constraint in Eq. (15) with $f_{\text{data}}(\mathbf{x}) \leq \eta_{\text{data}}$. Here η_{data} is set according to the expected level of errors, in order to avoid over-fitting the data:

$$\mathbf{x}^+ = \arg \min_{\mathbf{x} \in \mathcal{X}} f_{\text{prior}}(\mathbf{x}), \text{ subject to: } f_{\text{data}}(\mathbf{x}) \leq \eta_{\text{data}}. \quad (17)$$

The Lagrangian of this constrained optimization problem can be written as:

$$\mathcal{L}(\mathbf{x}; \ell) = f_{\text{prior}}(\mathbf{x}) + \ell \left(f_{\text{data}}(\mathbf{x}) - \eta_{\text{data}} \right) \quad (18)$$

where ℓ is the Lagrange multiplier associated with the inequality constraint $f_{\text{data}}(\mathbf{x}) \leq \eta_{\text{data}}$. If the constraint is *active*¹, then $\ell > 0$ and $f_{\text{data}}(\mathbf{x}) = \eta_{\text{data}}$ at the solution \mathbf{x}^+ (Nocedal and Wright 2006). Dropping the constant η_{data} which does not depend on \mathbf{x} , the solution is obtained by solving either of the following problems:

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

where

$$f(\mathbf{x}; \mu) = f_{\text{data}}(\mathbf{x}) + \mu f_{\text{prior}}(\mathbf{x}) \quad (19)$$

is the penalty function and the **hyperparameter** $\mu = 1/\ell > 0$ has to be tuned to match the constraint $f_{\text{data}}(\mathbf{x}) = \eta_{\text{data}}$. Hence we can equivalently consider

¹Conversely, the constraint being *inactive* would imply that $\ell = 0$, which would mean that the data are useless, which is hopefully not the case...

that we are solving the problem of maximizing the agreement of the model with the data subject to the constraint that the priors be below a preset level:

$$\mathbf{x}^+ = \arg \min_{\mathbf{x} \in \mathcal{X}} f_{\text{data}}(\mathbf{x}), \text{ subject to: } f_{\text{prior}}(\mathbf{x}) \leq \eta_{\text{prior}}. \quad (20)$$

For convex penalties and providing that the Lagrange multipliers (μ and ℓ) and the thresholds (η_{data} and η_{prior}) are set consistently, image restoration is achieved by solving either of the problems in Eq. (17), Eq. (20) or by minimizing the penalty function in Eq. (19). However, choosing which of these particular problems to solve can be a deciding issue for the efficiency of the method in terms of computational time. For instance, if $f_{\text{data}}(\mathbf{x})$ and $f_{\text{prior}}(\mathbf{x})$ are both *smooth functions*, direct minimization of $f(\mathbf{x}; \mu)$ in Eq. (19) can be done by using general purpose optimization algorithms (Nocedal and Wright 2006).

3.3 Distance to the Data

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

$$f_{\text{data}}(\mathbf{x}) = \|\mathbf{y} - \mathbf{A} \cdot \mathbf{x}\|_2^2. \quad (21)$$

The distance to the data has to be defined according to the statistics of the errors $\mathbf{e} = \mathbf{y} - \mathbf{A} \cdot \mathbf{x}$ given the image model. Assuming Gaussian statistics, this leads to:

$$f_{\text{data}}(\mathbf{x}) = (\mathbf{y} - \mathbf{A} \cdot \mathbf{x})^T \cdot \mathbf{W}_{\text{err}} \cdot (\mathbf{y} - \mathbf{A} \cdot \mathbf{x}), \quad (22)$$

where the weighting matrix $\mathbf{W}_{\text{err}} = \mathbf{C}_{\text{err}}^{-1}$ is the inverse of the covariance matrix of the errors. There is a slight issue because we are dealing with complex values. Since complex numbers are just pairs of reals, complex valued vectors (such as \mathbf{y} , \mathbf{e} and $\mathbf{A} \cdot \mathbf{x}$) can be *flattened* into ordinary real vectors (with doubled size) to use standard linear algebra notation and define the covariance matrix as $\mathbf{C}_{\text{err}} = \langle \mathbf{e} \cdot \mathbf{e}^T \rangle$. This is what is assumed in Eq. (22).

Real data may however have different statistics. For instance, the **OIFITS** exchange file format for optical interferometric data assumes that the amplitude and the phase of complex data (complex visibility or bispectrum) are independent (Pauls et al. 2005). As a result, the isocontours of the corresponding log-likelihood form a *non-convex* valley in the complex plane. Meimon et al. (2005a) have proposed quadratic convex approximations of the true log-likelihood 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 $y_\ell = \rho_\ell \exp(i\varphi_\ell)$, their local quadratic approximation is written:

$$f_{\text{data}}(\mathbf{x}) = \sum_{\ell} \left\{ \frac{\text{Re}(e_\ell e^{-i\varphi_\ell})^2}{\sigma_{//,\ell}^2} + \frac{\text{Im}(e_\ell e^{-i\varphi_\ell})^2}{\sigma_{\perp,\ell}^2} \right\} \quad (23)$$

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

$$\sigma_{//,\ell}^2 = \text{Var}(\rho_\ell), \quad \sigma_{\perp,\ell}^2 = \rho_\ell^2 \text{Var}(\varphi_\ell). \quad (24)$$

3.4 Regularization

Of all the methods for constraining the ill-posed imaging problem, maximum entropy (MEM) has perhaps the longest history. MEM methods are based on the 1950s work of Jaynes on information theory; the underlying idea is to obtain the least informative image which is consistent with the data (Ables 1974). This amounts to minimizing a criterion like the one in Eq. (19) with $\mu f_{\text{prior}}(\mathbf{x}) = -S(\mathbf{x})$ where the **entropy** $S(\mathbf{x})$ measures the informational content of the image \mathbf{x} . In this framework, $f_{\text{prior}}(\mathbf{x})$ is sometimes called **negentropy**. Among all the expressions considered for the negentropy of an image, one of the most popular is (Gull and Skilling 1984):

$$f_{\text{prior}}(\mathbf{x}) = \sum_n [x_n \log(x_n/\bar{x}_n) - x_n + \bar{x}_n] \quad (25)$$

with $\bar{\mathbf{x}}$ the **default image**; that is, the one which would be recovered in the absence of any data. The default image $\bar{\mathbf{x}}$ can be taken as being a flat image, an image previously restored, an image of the same object at a lower resolution, etc.

Minimizing the joint criterion in Eq. (19) with entropy regularization has a number of issues as the problem is highly non-linear and the number of unknowns is very large (as many as there are pixels). Various methods have been proposed, but the most effective algorithm (Skilling and Bryan 1984) employs non-linear optimization in a local sub-space of search directions with the Lagrange multiplier μ tuned on the fly to match the constraint that $f_{\text{data}}(\mathbf{x}) = \eta_{\text{data}}$.

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

$$f_{\text{prior}}(\mathbf{x}) = (\mathbf{x} - \bar{\mathbf{x}})^T \cdot \mathbf{C}_{\text{prior}}^{-1} \cdot (\mathbf{x} - \bar{\mathbf{x}}) \quad (26)$$

with $\mathbf{C}_{\text{prior}}$ the prior covariance and $\bar{\mathbf{x}}$ the prior solution. One simple but effective example is the *compactness* penalty:

$$f_{\text{prior}}(\mathbf{x}) = \sum_n w_n^{\text{prior}} x_n^2, \quad (27)$$

where the weights w_n^{prior} are increasing with the distance to the center of the image thus favouring structures concentrated within this part of the image. Under strict normalization and positivity constraints and in the absence of any data, the default image given by this prior is $\bar{x}_n \propto 1/w_n^{\text{prior}}$ where the factor comes from the normalization requirement (Le Besnerais et al. 2008). This regularizer is effective because imposing a compact, positive brightness distribution gives rise to smooth interpolation in the Fourier plane.

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

$$f_{\text{prior}}(\mathbf{x}) = \sum_{n_1, n_2} \sqrt{\epsilon^2 + |\nabla \mathbf{x}|_{n_1, n_2}^2} \quad (28)$$

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

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

where $(n_1, n_2) \sim n$ are the column and row indices corresponding to the n -th pixel. The penalization in Eq. (28) behaves as a quadratic function where the magnitude of the spatial gradient is small compared to ϵ , and a linear function when the gradient is large. Hence reduction of small local variations is achieved without over-penalizing strong sharp features. In the limit $\epsilon \rightarrow 0$, edge-preserving smoothness behaves like *total variation* (Rudin et al. 1992) which has proved successful in imposing sparsity.

4 Image Reconstruction from Non-Linear Data

At optical wavelengths, the complex visibilities are not directly measurable, the available data (*c.f.* Section 2) being the power spectrum, the bispectrum and/or the closure phase. Image reconstruction algorithms that use these data can be designed to follow the same inverse problem approach as above. However, the direct model of the data is now non-linear and specific expressions to implement f_{data} have to be derived. The non-linearity has also some impact on the optimization strategy.

4.1 Data Penalty

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

$$f_{\text{data}}^{\text{ps}}(\mathbf{x}) = \sum_{m, j_1 < j_2} \frac{(S_{j_1, j_2, m}^{\text{data}} - S_{j_1, j_2, m}^{\text{model}}(\mathbf{x}))^2}{\text{Var}(S_{j_1, j_2, m}^{\text{data}})}, \quad (29)$$

with $S_{j_1, j_2, m}^{\text{model}}(\mathbf{x}) = |\hat{I}(\boldsymbol{\nu}_{j_1, j_2, m})|^2$ the model of the power spectrum. For the penalty with respect to the bispectrum data, there is the additional difficulty of dealing with complex data. The Goodman approximation (Goodman 1985) yields:

$$f_{\text{data}}^{\text{bisp}}(\mathbf{x}) = \sum_{m, j_1 < j_2 < j_3} w_{j_1, j_2, j_3, m}^{\text{bisp}} |B_{j_1, j_2, j_3, m}^{\text{data}} - B_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x})|^2 \quad (30)$$

with $B_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x}) = \hat{I}(\boldsymbol{\nu}_{j_1, j_2, m}) \hat{I}(\boldsymbol{\nu}_{j_2, j_3, m}) \hat{I}(\boldsymbol{\nu}_{j_3, j_1, m})$ the model of the bispectrum and the weights are derived from the variance of the bispectrum data. An expression similar to that in Eq. (23) can be derived for bispectrum data with

independent modulus and phase errors (Meimon et al. 2009). To account for phase wrapping, Haniff (1991) proposed to define the penalty with respect to the closure phase data as:

$$f_{\text{data}}^{\text{cl}}(\mathbf{x}) = \sum_{m, j_1 < j_2 < j_3} \frac{\text{arc}^2(\beta_{j_1, j_2, j_3, m}^{\text{data}} - \beta_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x}))}{\text{Var}(\beta_{j_1, j_2, j_3, m}^{\text{data}})} \quad (31)$$

with $\beta_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x}) = \varphi(\boldsymbol{\nu}_{j_1, j_2, m}) + \varphi(\boldsymbol{\nu}_{j_2, j_3, m}) + \varphi(\boldsymbol{\nu}_{j_3, j_1, m})$ the model of the closure phase. However, this penalty is not continuously differentiable with respect to \mathbf{x} , which can prevent the convergence of optimization algorithms. This problem can be avoided by using the complex phasors (Thiébaud 2008):

$$f_{\text{data}}^{\text{cl}}(\mathbf{x}) = \sum_{m, j_1 < j_2 < j_3} \frac{|e^{i\beta_{j_1, j_2, j_3, m}^{\text{data}}} - e^{i\beta_{j_1, j_2, j_3, m}^{\text{model}}(\mathbf{x})}|^2}{\text{Var}(\beta_{j_1, j_2, j_3, m}^{\text{data}})}, \quad (32)$$

which is approximately equal to the penalty in Eq. (31) in the limit of small closure phase errors.

Depending on which types of data are available, and assuming that the different types of data have statistically independent errors, the *total* penalty with respect to the data is simply a sum of some of the penalties given by equations (29)–(32). For instance, to fit the power spectrum and the closure phase data:

$$f_{\text{data}}(\mathbf{x}) = f_{\text{data}}^{\text{ps}}(\mathbf{x}) + f_{\text{data}}^{\text{cl}}(\mathbf{x}). \quad (33)$$

4.2 Image Reconstruction Algorithms

We now describe the image reconstruction methods that have been successful on realistic optical interferometric data and which are sufficiently mature to be used with real data. In addition to coping with sparse Fourier data, these methods were specifically designed to tackle the non-linear direct model, to account for the particular statistics of the data (Meimon et al. 2005a) and to handle the OI-FITS data format (Pauls et al. 2005). A summary of the methods is given in Tab. 1.

1) *BSMEM* algorithm (Buscher 1994; Baron and Young 2008) makes use of a Maximum Entropy Method (Section 3.4) to regularize the problem of image restoration from the measured bispectrum (hence its name). The improved BSMEM version (Baron and Young 2008) uses the Gull and Skilling entropy, see Eq. (25), and a likelihood term with respect to the complex bispectrum which assumes independent Gaussian noise statistics for the amplitude and phase of the measured bispectrum. The optimization engine is MEMSYS which implements the strategy proposed by Skilling and Bryan (1984) and automatically finds the most likely value for the hyperparameter μ . Because it makes no attempt to directly convert the data into complex visibilities, a strength of BSMEM is that it can handle any type of data sparsity (such as missing closure phases).

2) *The Building Block Method* (Hofmann and Weigelt 1993) is similar to the CLEAN method but designed for reconstructing images from bispectrum

Name	Authors	Optimization	Regularization
BSMEM	Baron, Buscher	Trust region gradient	MEM-prior
MiRA	Thiébaud	VMLM-B	Many
WISARD	Meimon, Mugnier, Le Besnerais	VMLM-B plus self-calibration	Many
MACIM	Ireland, Monnier	Simulated annealing	MEM
SQUEEZE	Baron, Monnier, Kloppenborg	Parallel tempering	Sparsity
Building Block Method	Hofmann, Weigelt	Matching pursuit	Sparsity

Table 1: Summary of the algorithms for image reconstruction from optical interferometric data described in Section 4.2.

data obtained by means of speckle or long baseline interferometry. The method proceeds iteratively to reduce a cost function $f_{\text{data}}^{\text{bisp}}$ equal to that in Eq. (30) with weights set to a constant or to an expression motivated by Wiener filtering. The minimization of the penalty is achieved by a matching pursuit algorithm which imposes sparsity of the solution. The image is given by the building block model in Eq. (8) and, at the n -th iteration, the new image $I^{[n]}(\boldsymbol{\theta})$ is obtained by adding a new building block at location $\boldsymbol{\theta}^{[n]}$ with a weight $\alpha^{[n]}$ to the previous image, so as to maintain the normalization:

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

The weight and location of the new building block is derived by minimizing the criterion $f_{\text{data}}^{\text{bisp}}$ with respect to these parameters. Strict positivity and support constraints can be trivially enforced by limiting the possible values for $\alpha^{[n]}$ and $\boldsymbol{\theta}^{[n]}$. 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.

3) *MACIM algorithm* (Ireland et al. 2008), for MArkov Chain IMager, aims at maximizing the posterior probability:

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

MACIM implements MEM regularization and a specific regularizer which favors large regions of dark space in between bright regions. For this latter regularization, $f_{\text{prior}}(\boldsymbol{x})$ is the sum of all pixels with zero flux on either side of their boundaries. MACIM attempts to maximize $\Pr(\boldsymbol{x}|\boldsymbol{y})$ by a simulated annealing algorithm with the Metropolis sampler. Although maximizing $\Pr(\boldsymbol{x}|\boldsymbol{y})$ is the same as minimizing $f_{\text{data}}(\boldsymbol{x}) + \mu f_{\text{prior}}(\boldsymbol{x})$, the use of *normalized* probabilities is required by the Metropolis sampler to accept or reject the image samples. Simulated annealing can in principle solve the global optimization problem of maximizing $\Pr(\boldsymbol{x}|\boldsymbol{y})$, but convergence can be very slow, especially for objects comprising multiple components.

4) *SQUEEZE algorithm* (Baron et al. 2010) was developed recently by Fabien Baron and John Monnier at the University of Michigan, with the collaboration of Brian Kloppenborg from the University of Denver. SQUEEZE has a very comprehensive set of features, including both Markov Chain Monte Carlo (as in MACIM) and gradient-based optimization engines, and the ability to combine geometric model-fitting with model-independent imaging. Novel capabilities such as wavelet regularization within the compressed sensing framework and imaging on spheroids are also being implemented.

5) *MiRA algorithm* (Thiébaud 2008) defines the sought image as the minimum of the penalty function in Eq. (19). Minimization is done by VMLM-B, a limited variable memory method (based on BFGS updates) with bound constraints for the positivity (Thiébaud 2002). MiRA is written in a modular way: any type of data can be taken into account by providing a function that computes the corresponding penalty and its gradient. For the moment, MiRA handles complex visibility, power spectrum and closure-phase data via penalty terms given by Eq. (23), Eq. (29) and Eq. (32). Like BSMEM, MiRA can cope with any missing data, in particular, it can be used to restore an image given only the power spectrum (*i.e.* without any Fourier phase information) with at least a 180° orientation ambiguity.

6) *WISARD algorithm* (Meimon et al. 2005b) recovers an image from power spectrum and closure phase data. It exploits a self-calibration approach (Readhead and Wilkinson 1978; Cornwell and Wilkinson 1981) to recover missing Fourier phases. Given a current estimate of the image and the closure phase data, WISARD first derives missing Fourier phase information in such a way as to minimize the number of unknowns. Then, the synthesized Fourier phases are combined with the square root of the measured power spectrum to generate pseudo complex visibility data which are fitted 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. This approach gives a unique solution for the image restoration step, although overall the global problem remains multi-modal.

MiRA and WISARD have been developed in parallel and share some common features. They use the same optimization engine (Thiébaud 2002) and means to impose positivity and normalization (Le Besnerais et al. 2008). However they differ in the way missing data is taken into account: WISARD *explicitly* solves for missing Fourier phase information; while MiRA *implicitly* accounts for any lack of information through the direct model of the data (Le Besnerais et al. 2008). Both implement many different regularizers (negentropy, quadratic or edge-preserving smoothness, compactness, total variation, etc.)

4.3 Image Reconstruction Inputs

4.3.1 OIFITS data

The algorithms described in Sec. 4.2 all accept data in the form of one or more OIFITS files. OIFITS (Pauls et al. 2005) is a registered convention based on

the FITS standard (Hanisch et al. 2001), designed for storage of calibrated optical interferometric data. Supplementing the formal definition of OIFITS, Pauls et al. (2004) give an overview of the convention and explains the design decisions taken.

An OIFITS file can store data for multiple observing targets, taken with multiple interferometers/instruments. The supported observables are the complex visibility, power spectrum (squared visibility in OIFITS parlance) and bispectrum (triple product). The bispectrum phase (closure phase) may be stored without a corresponding bispectrum amplitude. Each observable is stored in a separate type of table, and any number of (including zero) instances of each table type may be present in the file. Telescope (station) identifiers, observation times and wavebands (wavelength and spectral bandwidth) are also stored, the latter in separate tables to reduce duplication.

The typical steps involved in reading a set of OIFITS files for image reconstruction are as follows:

- Merge multiple OIFITS files specified by the user
- Extract data for the observing target, timespan and wavelengths (specified as a numerical range or set of `INSNAME` values) specified by the user.
- Determine the observation times and wavelengths of all usable data (usable meaning not flagged and comprising an appropriate combination of observables, for example powerspectra and bispectra measured simultaneously using the same telescopes and spectral channels).
- Extract the usable data for these observation times and wavelengths.

Proposals for extending OIFITS have been discussed at length by the community, and we expect that preliminary versions of these extensions will be agreed in the near future, so that they can be used in preparing the test datasets for the JRA. The planned extensions include conventions for storing wavelength-differential visibility data and the spectral energy distribution of the observed object. Wherever possible, the test datasets should also validate as standard OIFITS.

4.3.2 Image model parameters

The most common pixellated image model comprises a shift-invariant set of basis functions $b_n(\boldsymbol{\theta}) = b(\boldsymbol{\theta} - \boldsymbol{\theta}_n)$ on an equispaced two-dimensional grid $\{\boldsymbol{\theta}_n\}$, with $b(\boldsymbol{\theta})$ the pixel shape.

The pixellated model is usually parametrised by the grid spacing $\Delta\theta$ and the field of view given by the overall extent of the grid. As mentioned in Sec. 3.1, the grid spacing must oversample the diffraction limit by a sufficient factor to avoid modeling errors, bearing in mind that a degree of super-resolution in the reconstructed image can often be achieved. The field of view must be large enough to avoid aliasing. Often the precise size of the observed object is unknown, so the image size would be chosen based on the instrumental field of

Name	$f_{\text{prior}}(\mathbf{x})$	Parameters	Implemented by
Quadratic smoothness	$\ \mathbf{x} - \mathbf{S} \cdot \mathbf{x}\ ^2$		MiRA, WISARD?
MEM-sqrt	$-\sum_n \sqrt{x_n}$		MiRA
MEM-log	$-\sum_n \log(x_n)$		MiRA
MEM-prior	$\sum_n [x_n \log(x_n/\bar{x}_n) - x_n + \bar{x}_n]$	Default image $\bar{\mathbf{x}}$	MiRA, BSMEM
Compactness	$\sum_n w_n^{\text{prior}} x_n^2; w_n^{\text{prior}} = \ \boldsymbol{\theta}_n\ ^\beta$	$\beta > 0$	MiRA, WISARD
Edge-preserving smoothness (\sim TV when $\epsilon \rightarrow 0$)	$\sum_{n_1, n_2} \sqrt{\epsilon^2 + \nabla \mathbf{x} _{n_1, n_2}^2}$	$\epsilon > 0$	MiRA, WISARD

Table 2: Summary of the most useful regularization types implemented by the algorithms being developed under the JRA. The names used here are taken from [Renard et al. \(2011\)](#).

view. It is quite feasible to choose the image model parameters automatically based on the range of spatial frequencies in the data, perhaps rounding $\Delta\theta$ to a few significant figures.

4.3.3 Regularization parameters

A summary of the regularization types implemented by MiRA, WISARD and BSMEM is given in Tab. 2. [Renard et al. \(2011\)](#) compared these regularization methods plus several others. As expected they concluded that the best prior depends on the object of interest. However, edge-preserving smoothness (Eq. (28)), and compactness (Eq. (27)) were shown to be successful in most cases. Astronomers should not attach too much importance to the choice of regularizer: Figure 3 shows image reconstructions using different types of regularization, all of which are fairly similar and acceptable for publication.

The MEM-prior regularization involves a prior model, the **default image**. An informative prior model can be useful for constraining the reconstruction when the data are sparse or noisy, otherwise it just speeds up convergence. If the data consist of only powerspectra and/or bispectra, the position of the object is completely unconstrained — a non-flat default image can be used to fix the position. For other priors, a non-flat initial model used to start the optimization will serve the same purpose.

Irrespective of the particular form of $f_{\text{prior}}(\mathbf{x})$, the degree of regularization imposed is set by the value of the hyperparameter μ in Eq. (19). This should strictly be derived from Bayesian considerations but can be done quite accurately by visual inspection of the restored image. From Fig. 4, one can see the effects of under-regularization (which yields more artifacts) and over-regularization (which yields over-simplification of the image).

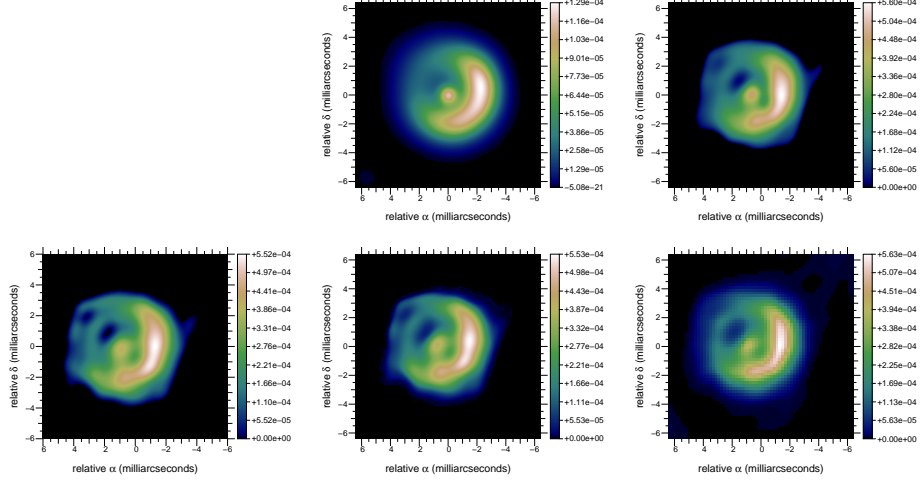


Figure 3: MiRA and BSMEM image reconstructions with various types of regularization. From top-left to bottom-right: (a) original object smoothed to the resolution of the interferometer ($\text{FWHM} \sim 15 \text{ mas}$); (b) MiRA reconstruction with a compactness quadratic regularization given by Eq. (27); (c) MiRA reconstruction with edge-preserving regularization as in Eq. (28); (d) MiRA reconstruction with MEM regularization as in Eq. (25); (e) BSMEM reconstruction with MEM regularization.

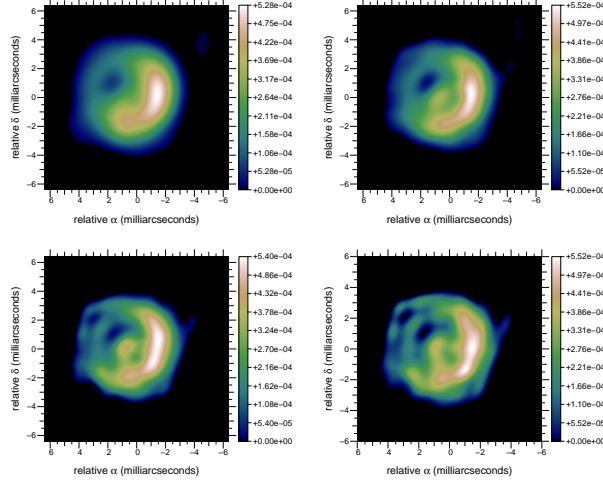


Figure 4: Image reconstruction under various regularization levels. Algorithm is MiRA with edge-preserving regularization given in Eq. (28) with $\epsilon = 10^{-4}$ and $\mu = 10^6$ (top-left), $\mu = 10^5$ (top-right), $\mu = 10^4$ (bottom-left) and $\mu = 3 \times 10^3$ (bottom-right).

5 Discussion

The main issues in image reconstruction from interferometric data are the sparsity of the Fourier measurements and the lack of part of the Fourier phase information. The inverse problem approach appears to be suitable to describe the most important existing algorithms in this context. Indeed, the image reconstruction methods can be stated as the minimization of a mixed criterion under some strict constraints such as positivity and normalization. Two different types of terms appear into this criterion: likelihood terms which enforce consistency of the model image with the data, and regularization terms which lift the inherent degeneracies of the image reconstruction problem. Hence the differences between the various algorithms lie in the kind of measurements considered, in the approximations for the direct model and for the statistics of the errors and in the prior imposed by the regularization. For non-convex criteria which occur when the OTF is 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. From a technical point of view, future developments of these algorithms will certainly focus on global optimization and unsupervised reconstruction. However, to fully exploit the existing instruments and those currently under development, the most worthwhile topics to investigate are multi-spectral imaging and making use of additional information such as a low resolution image or partial model of the observed object.

References

- J. Ables. Maximum Entropy Spectral Analysis. *Astron. & Astrophys. Suppl.*, 15:383, 1974.
- F. Baron and J. Young. Image reconstruction at Cambridge University. In *Proc. SPIE: Optical and Infrared Interferometry*, volume 7013, page 144, 2008.
- F. Baron, J. Monnier, and B. Kloppenborg. A novel image reconstruction software for optical/infrared interferometry. In *Proc. SPIE: Optical and Infrared Interferometry II*, volume 7734, 2010.
- D. Buscher. Direct maximum-entropy image reconstruction from the bispectrum. In *IAU Symp. 158: Very High Angular Resolution Imaging*, page 91, 1994.
- T. Cornwell and P. Wilkinson. A new method for making maps with unstable radio interferometers. *Monthly Notices of the RAS*, 196:1067, 1981.
- F. Delplancke, F. Derie, F. Paresce, A. Glindemann, F. Lévy, S. Lévêque, and S. Ménardi. PRIMA for the VLTI - science. *Astrophys. & Space Science*, 286: 99, 2003.

- J.-F. Giovannelli and A. Coulais. Positive deconvolution for superimposed extended source and point sources. *Astron. & Astrophys.*, 439:401, 2005.
- A. Glindemann, R. Abuter, F. Carbognani, F. Delplancke, F. Derie, A. Gennai, P. B. Gitton, P. Kervella, B. Koehler, S. A. Leveque, S. Menardi, A. Michel, F. Paresce, T. P. Duc, A. Richichi, M. Schoeller, M. Tarengi, A. Wallander, and R. Wilhelm. The VLT Interferometer: a unique instrument for high-resolution astronomy. In *Proc. SPIE: Interferometry in Optical Astronomy*, volume 4006, pages 2–12, 2000.
- J. Goodman. *Statistical Optics*. John Wiley & Sons, 1985. ISBN 0-471-01502-4.
- J. Gordon and D. Buscher. Detection noise bias and variance in the power spectrum and bispectrum in optical interferometry. *Astron. & Astrophys.*, 541:A46, 2012.
- S. Gull and J. Skilling. The maximum entropy method. In *Measurement and Processing for Indirect Imaging*, page 267, 1984.
- C. Haniff. Least-squares Fourier phase estimation from the modulo 2π bispectrum phase. *J. Opt. Soc. Am. A*, 8:134, 1991.
- R. Hanisch, A. Farris, E. Greisen, W. Pence, B. Schlesinger, P. Teuben, R. Thompson, and A. Warnock, III. Definition of the Flexible Image Transport System (FITS). *Astron. & Astrophys.*, 376:359–380, 2001.
- K.-H. Hofmann and G. Weigelt. Iterative image reconstruction from the bispectrum. *Astron. & Astrophys.*, 278:328, 1993.
- M. Ireland, J. Monnier, and N. Thureau. Monte-Carlo imaging for optical interferometry. In *Proc. SPIE: Advances in Stellar Interferometry*, volume 6268, page 62681T, 2008.
- P. Lawson, W. Cotton, C. Hummel, J. Monnier, M. Zhao, J. Young, H. Thorsteinsson, S. Meimon, L. Mugnier, G. Le Besnerais, E. Thiébaud, and P. Tuthill. An interferometry imaging beauty contest. In *Proc. SPIE: New Frontiers in Stellar Interferometry*, volume 5491, page 886, 2004.
- G. Le Besnerais, S. Lacour, L. Mugnier, E. Thiébaud, G. Perrin, and S. Meimon. Advanced imaging methods for long-baseline optical interferometry. *IEEE Journal of Selected Topics in Signal Processing*, 2:767, 2008.
- S. Meimon, L. Mugnier, and G. Le Besnerais. Convex approximation to the likelihood criterion for aperture synthesis imaging. *J. Opt. Soc. Am. A*, 22: 2348, 2005a.
- S. Meimon, L. Mugnier, and G. Le Besnerais. Reconstruction method for weak-phase optical interferometry. *Optics Lett.*, 30:1809, 2005b.

- S. Meimon, L. Mugnier, and G. Le Besnerais. Self-calibration approach for optical long-baseline interferometry imaging. *J. Opt. Soc. Am. A*, 26(1):108–120, 2009.
- J. Monnier. Optical interferometry in astronomy. *Reports on Progress in Physics*, 66:789, 2003.
- J. Nocedal and S. Wright. *Numerical Optimization*. Springer, 2006.
- T. Pauls, J. Young, W. Cotton, and J. Monnier. A data exchange standard for optical (visible/IR) interferometry. In *Proc. SPIE: New Frontiers in Stellar Interferometry*, volume 5491, page 1231, 2004.
- T. Pauls, J. Young, W. Cotton, and J. Monnier. A data exchange standard for optical (visible/IR) interferometry. *Publications of the ASP*, 117:1255, 2005.
- D. Potts, G. Steidl, and M. Tasche. Fast Fourier transforms for nonequispaced data: A tutorial. In J. Benedetto and P. Ferreira, editors, *Modern Sampling Theory: Mathematics and Applications*, page 249. Birkhäuser, 2001.
- A. Readhead and P. Wilkinson. The mapping of compact radio sources from VLBI data. *Astrophys. J.*, 223:25, 1978.
- S. Renard, E. Thiébaud, and F. Malbet. Image reconstruction in optical interferometry: benchmarking the regularization. *Astronomy & Astrophysics*, 533: A64, 2011. ISSN 0004-6361, 1432-0746.
- L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
- J. Skilling and R. Bryan. Maximum entropy image reconstruction: general algorithm. *Monthly Notices of the RAS*, 211:111, 1984.
- A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005. ISBN 0-89871-572-5.
- T. A. ten Brummelaar, H. A. McAlister, S. T. Ridgway, W. G. Bagnuolo, Jr., N. H. Turner, L. Sturmann, J. Sturmann, D. H. Berger, C. E. Ogden, R. Cadman, W. I. Hartkopf, C. H. Hopper, and M. A. Shure. First Results from the CHARA Array. II. A Description of the Instrument. *Astrophys. J.*, 628:453–465, 2005.
- É. Thiébaud. Optimization issues in blind deconvolution algorithms. In *Proc. SPIE: Astronomical Data Analysis II*, volume 4847, page 174, 2002.
- É. Thiébaud. MiRA: an effective imaging algorithm for optical interferometry. In *Proc. SPIE: Optical and Infrared Interferometry*, volume 7013, page 70131I, 2008.
- É. Thiébaud. Principles of image reconstruction in interferometry. *European Astronomical Society Publications Series*, 59:157–187, 2013.