New Concepts in Imaging: Optical and Statistical Models
D. Mary, C. Theys and C. Aime (eds)
EAS Publications Series, **59** *(2013) 157–187*

PRINCIPLES OF IMAGE RECONSTRUCTION IN INTERFEROMETRY

É. Thiébaut[1]

Abstract. Image reconstruction from interferometric data is an inverse problem. Owing to the sparse spatial frequency coverage of the data and to missing Fourier phase information, one has to take into account not only the data but also prior constraints. Image reconstruction then amounts to minimizing a joint criterion which is the sum of a likelihood term to enforce fidelity to the data and a regularization term to impose the priors. To implement strict constraints such as normalization and non-negativity, the minimization is performed on a feasible set. When the complex visibilities are available, image reconstruction is relatively easy as the joint criterion is convex and finding the solution is similar to a deconvolution problem. In optical interferometry, only the power-spectrum and the bispectrum can be measured and the joint criterion is highly multi-modal. The success of an image reconstruction algorithm then depends on the choice of the priors and on the ability of the optimization strategy to find a good solution among all the local minima.

The best angular resolution of a telescope is given by the diffraction limit λ/D (with D the diameter of the primary mirror and λ the wavelength). For an astronomical interferometer, this limit is λ/B (with B the separation of the telescopes projected in a plane perpendicular to the line of sight). In the optical, the largest telescopes have a diameter $D \approx 10\,\mathrm{m}$; thus, with baselines up to $B \approx 600\,\mathrm{m}$, astronomical interferometers resolve much smaller angular scales, below the milliarcsecond in the H band ($1.65\,\mu\mathrm{m}$). This unrivaled resolution has however a cost: an interferometer measures only a single spatial frequency per baseline, while a monolithic telescope harvests all spatial frequencies (up to its diffraction limits) in a single exposure. The data collected by an interferometer are thus very sparse and image reconstruction is a mandatory tool to build an image in spite of the voids in the spatial frequency coverage.

[1] Centre de Recherche Astronomique de Lyon, Université Claude Bernard Lyon I, École Normale Supérieure de Lyon, France; e-mail: `eric.thiebaut@univ-lyon1.fr`

DOI: 10.1051/eas/1359009

Inverse problem approach is a very powerful tool for extracting meaningful information from available data. In particular, it is the method of choice for image reconstruction from interferometric observables. A power of the inverse approach is to relax the constraint that the model of the observables be invertible and thus let us exploit a realistic model. To benefit from this potential, the data model has to be wisely written knowing the instrument and making relevant approximations. The direct model of the interferometric observable is developed in the first sections of this paper. From the instantaneous output of an interferometer (Sect. 1), time averaging (Sect. 2) yields the expression of the complex visibilities integrated during an exposure. In the most simple case, that is when complex visibilities are directly measurable, image reconstruction amounts to solving a deconvolution problem (Sect. 3). In optical interferometry, atmospheric turbulence introduces unknown random optical path perturbations which prevent to directly measure complex visibilities and imposes to integrate observables such as the powerspectrum and the bispectrum which are insensitive to such perturbations (Sect. 4).

Owing to the sparsity of the interferometric data and to the missing of part of the Fourier phases, prior information must be taken into account to solve the image reconstruction problem in a stable and robust way. Without loss of generality, image reconstruction can be stated as an optimization problem over a feasible set (Sect. 5). The penalty to minimize is the sum of a likelihood term (Sect. 6) which enforces fidelity to the measurements and a regularization term (Sect. 7) which favors the priors. Finally, it remains to design an optimization algorithm to effectively solve the image reconstruction problem (Sect. 8).

1 Instantaneous output of an interferometer

In its simplest form, a stellar interferometer (see Fig. 1) consists in two telescopes (or antennae for an array of radio-telescopes) pointing at the astronomical target and coherently recombined. By varying the optical path delay between the two arms of the interferometer, one observes interference fringes. The contrast of the fringes and their phase are the amplitude and phase of the so-called *complex visibility* which is related to the observed object by:

$$V_{j_1,j_2}(\lambda, t) = g^*_{j_1}(\lambda, t)\, g_{j_2}(\lambda, t)\, \widehat{I}_\lambda(\boldsymbol{b}_{j_1,j_2}(t)/\lambda) \tag{1.1}$$

with j_1 and j_2 the indexes of the interfering telescopes, λ the wavelength, t the time, $g_j(\lambda, t)$ the instantaneous complex amplitude transmission for the jth telescope, $g^*_j(\lambda, t)$ its complex conjugate, $\widehat{I}_\lambda(\boldsymbol{\nu})$ the angular Fourier transform of the specific brightness distribution $I_\lambda(\boldsymbol{\theta})$ of the observed object in angular direction $\boldsymbol{\theta}$, and $\boldsymbol{b}_{j_1,j_2}(t)$ the projected *baseline*:

$$\boldsymbol{b}_{j_1,j_2}(t) = \boldsymbol{r}_{j_2}(t) - \boldsymbol{r}_{j_1}(t)$$

where $\boldsymbol{r}_j(t)$ is the position of the jth telescope projected on a plane perpendicular to the line of sight. The amplitude of the complex transmission $g_j(\lambda, t)$ accounts for

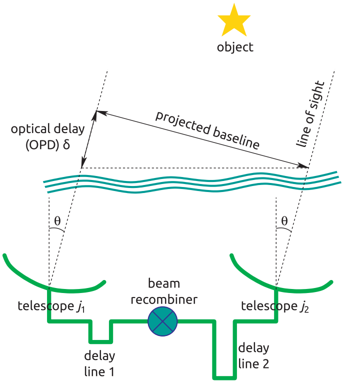


Fig. 1. Interferometer.

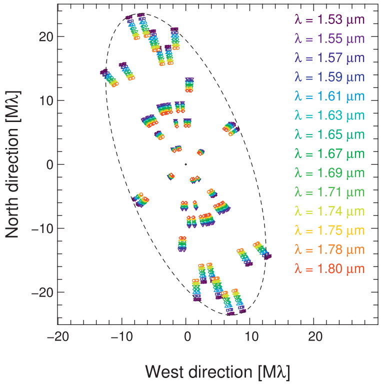


Fig. 2. (u, v) coverage with IOTA 3-telescope interferometer in the H band (from: Lacour *et al.* 2008). The spatial frequencies $\boldsymbol{\nu}$ are given as the projected baselines $\boldsymbol{b}$ in mega-wavelength units (symbol Mλ) that is $10^6 \times \boldsymbol{b}/\lambda = 10^6 \times \boldsymbol{\nu}$.

the efficiency of the transfer of the light from the jth telescope to the recombiner, the phase of $g_j(\lambda, t)$ accounts for the optical delay along this travel.

Equation (1.1) shows that a stellar interferometer samples the Fourier transform of the brightness distribution $\widehat{I}_\lambda(\boldsymbol{\nu})$ at the spatial frequency:

$$\boldsymbol{\nu}_{j_1,j_2}(\lambda, t) = \boldsymbol{b}_{j_1,j_2}(t)/\lambda = \big(\boldsymbol{r}_{j_2}(t) - \boldsymbol{r}_{j_1}(t)\big)/\lambda\,.$$

A single exposure yields one measurement of $\widehat{I}_\lambda(\boldsymbol{\nu})$ per pair of recombined telescopes per spectral channel. For N_{tel} telescopes in a non-redundant configuration, there is a maximum of $N_{\mathsf{tel}}\,(N_{\mathsf{tel}} - 1)/2$ different baselines. Thanks to Earth rotation, the sampling of the spatial frequencies – the so-called (u, v) plane – by a given configuration of telescopes varies with the time, this is called *super-synthesis*. The sampled frequencies also depend on the wavelength: the longer the wavelength the shorter the sampled frequency. Because of the limited number of telescopes for current optical interferometers ($2 \leq N_{\mathsf{tel}} \leq 6$), even by combining all these possible measurements, the sampling of the (u, v) plane remains very sparse and uneven (*cf.* Fig. 2).

2 Averaging during exposures

The previous equations consider the instantaneous and monochromatic case: they are given for continuously varying time t, wavelength λ and projected telescope positions $\boldsymbol{r}_j(t)$. In practice, a finite number of measurements are obtained for given exposure times, spectral channels and telescope combinations. In the sequel, we use the index m to label the available data: for the m-th measurement (possibly complex), the exposure time is denoted t_m, λ_m is the effective wavelength of the spectral channel and there are up to three interfering telescopes numbered $j_{m,1}$, $j_{m,2}$ and $j_{m,3}$. Of course different measurements, say m and m', may have the same observing times ($t_{m'} = t_m$) or may share the same telescopes and the same spectral channel.

Because of the finite exposure time and spectral bandwidth, the instantaneous and monochromatic complex visibility in Equation (1.1) must be averaged to give the *effective* complex visibility:

$$V_m = \langle V_{j_{m,1},j_{m,2}}(\lambda,t)\rangle_m = \big\langle g^*_{j_{m,1}}(\lambda,t)\, g_{j_{m,2}}(\lambda,t)\, \widehat{I}_\lambda\big(\boldsymbol{b}_{j_{m,1},j_{m,2}}(t)/\lambda\big)\big\rangle_m \tag{2.1}$$

where $\langle\ldots\rangle_m$ denotes averaging (or integrating) during the exposure and inside the spectral channel corresponding to the m-th measurement:

$$\langle f(\lambda,t)\rangle_m \overset{\text{def}}{=} \frac{1}{\Delta t_m}\int_{t_m-\Delta t_m/2}^{t_m+\Delta t_m/2} \frac{1}{\Delta\lambda_m}\int s_m(\lambda)\, f(\lambda,t)\,\mathrm{d}\lambda\,\mathrm{d}t \tag{2.2}$$

with Δt_m the duration of the exposure, $s_m(\lambda)$ the transmission of the spectral channel, and $\Delta\lambda_m \overset{\text{def}}{=} \int s_m(\lambda)\,\mathrm{d}\lambda$ the effective spectral bandwidth.

To measure interference patterns, the effective bandwidth $\Delta\lambda_m$ must be such that the complex amplitude transmissions are approximately constant in each spectral channel and the exposure duration Δt_m must be short enough to neglect the temporal variation of the baselines. Under these conditions, the double integral which results from combining Equations (2.1) and (2.2) becomes separable:

$$\begin{aligned} V_m &= \frac{1}{\Delta t_m}\int_{t_m-\Delta t_m/2}^{t_m+\Delta t_m/2} \frac{1}{\Delta\lambda_m}\int s_m(\lambda)\, g^*_{j_{m,1}}(\lambda,t)\, g_{j_{m,2}}(\lambda,t) \\ &\qquad\qquad \times \widehat{I}_\lambda\big(\boldsymbol{b}_{j_{m,1},j_{m,2}}(t)/\lambda\big)\;\mathrm{d}\lambda\,\mathrm{d}t \\ &\approx \frac{1}{\Delta t_m}\int_{t_m-\Delta t_m/2}^{t_m+\Delta t_m/2} g^*_{j_{m,1}}(\lambda_m,t)\, g_{j_{m,2}}(\lambda_m,t)\,\mathrm{d}t \\ &\quad\times \frac{1}{\Delta\lambda_m}\int s_m(\lambda)\,\widehat{I}_\lambda\big(\boldsymbol{b}_{j_{m,1},j_{m,2}}(t_m)/\lambda_m\big)\;\mathrm{d}\lambda \\ &= \widehat{h}_m\,\widehat{I}_m(\boldsymbol{\nu}_m) \end{aligned} \tag{2.3}$$

with:

$$\widehat{h}_m \overset{\text{def}}{=} \frac{1}{\Delta t_m}\int_{t_m-\Delta t_m/2}^{t_m+\Delta t_m/2} g^*_{j_{m,1}}(\lambda_m,t)\, g_{j_{m,2}}(\lambda_m,t)\,\mathrm{d}t\,, \tag{2.4}$$

$$\widehat{I}_m(\boldsymbol{\nu}) \overset{\text{def}}{=} \frac{1}{\Delta\lambda_m}\int s_m(\lambda)\,\widehat{I}_\lambda(\boldsymbol{\nu})\,\mathrm{d}\lambda \tag{2.5}$$

$$\approx \widehat{I}_{\lambda_m}(\boldsymbol{\nu}) \tag{2.6}$$

$$\boldsymbol{\nu}_m \overset{\text{def}}{=} \boldsymbol{b}_m/\lambda_m\,, \tag{2.7}$$

$$\boldsymbol{b}_m \overset{\text{def}}{=} \boldsymbol{r}_{j_{m,2}}(t_m) - \boldsymbol{r}_{j_{m,1}}(t_m), \tag{2.8}$$

respectively the effective interferometric transfer function, the Fourier transform of the specific brightness distribution integrated in the spectral channel, the spatial frequency and the effective baseline for the m-th observed complex visibility. The approximation in Equation (2.6) applies for spectral bandwidths narrower than the spectral features of the specific brightness distribution. To simplify the notations but without loss of generality, we will assume that this is the case in what follows.

When the complex amplitude transmissions are stable during and exposure, the effective interferometric transfer function can be further simplified:

$$\widehat{h}_m \approx g^*_{j_{m,1}}\, g_{j_{m,2}} \tag{2.9}$$

where:

$$g_{j_{m,i}} \overset{\text{def}}{=} \frac{1}{\Delta t_m} \int_{t_m - \Delta t_m/2}^{t_m + \Delta t_m/2} g_{j_{m,i}}(\lambda_m, t)\,\mathrm{d}t \approx g_{j_{m,i}}(\lambda_m, t_m). \tag{2.10}$$

Thus, for monochromatic observations with an interferometer composed of N_{tel} telescopes and under stable observing conditions, the effective transfer function only depends on $N_{\mathsf{tel}} - 1$ complex numbers (one complex amplitude transmission can be chosen arbitrarily) per exposure while there are $N_{\mathsf{tel}}\,(N_{\mathsf{tel}} - 1)/2$ measured complex visibilities. Depending on the number of interfering telescopes, the amount of information needed to estimate the transfer function may be much smaller than the amount of measurements. This open the possibility to perform *self-calibration* (Cornwell & Wilkinson 1981; Schwab 1980).

3 Easy case: image reconstruction $\sim$ deconvolution

Considering only complex visibilities for a given effective wavelength λ, we can combine them to form the distribution:

$$\widehat{d}_\lambda(\boldsymbol{\nu}) \overset{\text{def}}{=} \sum_{m \in \mathbb{S}_\lambda} V_m\, \delta(\boldsymbol{\nu} - \boldsymbol{\nu}_m) \tag{3.1}$$

with $\mathbb{S}_\lambda = \{m\colon \lambda_m = \lambda\}$ and $\delta(\cdot)$ the Dirac's distribution. Using the definition of the observed complex visibilities V_m in Equation (2.3) and the approximation in Equation (2.6), $\widehat{d}_\lambda(\boldsymbol{\nu})$ can be expanded as follows:

$$\begin{aligned}
\widehat{d}_\lambda(\boldsymbol{\nu}) &= \sum_{m \in \mathbb{S}_\lambda} \widehat{h}_m\, \widehat{I}_{\lambda_m}(\boldsymbol{\nu}_m)\, \delta(\boldsymbol{\nu} - \boldsymbol{\nu}_m) \\
&= \widehat{I}_\lambda(\boldsymbol{\nu}) \sum_{m \in \mathbb{S}_\lambda} \widehat{h}_m\, \delta(\boldsymbol{\nu} - \boldsymbol{\nu}_m) \\
&= \widehat{I}_\lambda(\boldsymbol{\nu})\, \widehat{h}_\lambda(\boldsymbol{\nu}), \qquad (3.2)
\end{aligned}$$

with:

$$\widehat{h}_\lambda(\boldsymbol{\nu}) = \sum_{m \in \mathbb{S}_\lambda} \widehat{h}_m\, \delta(\boldsymbol{\nu} - \boldsymbol{\nu}_m). \tag{3.3}$$

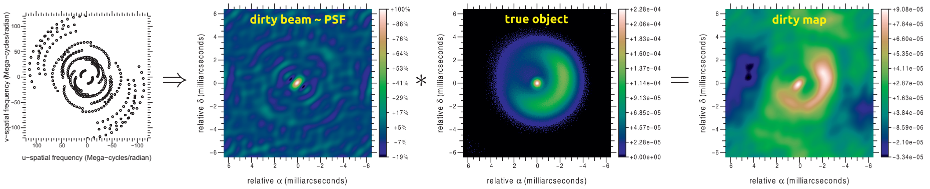


Fig. 3. From *left* to *right*: spatial frequency sampling, dirty beam, object brightness distribution and dirty image.

Taking the inverse Fourier transform of $\widehat{d}_\lambda(\boldsymbol{\nu})$, we obtain a 2D angular distribution called the *dirty image*:

$$\begin{aligned} d_\lambda(\boldsymbol{\theta}) &\overset{\text{def}}{=} \iint \widehat{d}_\lambda(\boldsymbol{\nu})\, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle\boldsymbol{\theta},\boldsymbol{\nu}\rangle}\, \mathrm{d}^2\boldsymbol{\nu} \\ &= \iint \widehat{h}_\lambda(\boldsymbol{\nu})\, \widehat{I}_\lambda(\boldsymbol{\nu})\, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle\boldsymbol{\theta},\boldsymbol{\nu}\rangle}\, \mathrm{d}^2\boldsymbol{\nu} \\ &= (h_\lambda * I_\lambda)\,(\boldsymbol{\theta}) \end{aligned} \tag{3.4}$$

where $\langle\boldsymbol{\theta},\boldsymbol{\nu}\rangle$ is the 2D scalar product of $\boldsymbol{\theta}$ by $\boldsymbol{\nu}$ and the symbol $*$ denotes the convolution product of the brightness distribution $I_\lambda(\boldsymbol{\theta})$ by the so-called *dirty beam*:

$$\begin{aligned} h_\lambda(\boldsymbol{\theta}) &\overset{\text{def}}{=} \iint \widehat{h}_\lambda(\boldsymbol{\nu})\, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle\boldsymbol{\theta},\boldsymbol{\nu}\rangle}\, \mathrm{d}^2\boldsymbol{\nu} \\ &= \sum_{m\in\mathbb{S}_\lambda} \widehat{h}_m\, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle\boldsymbol{\theta},\boldsymbol{\nu}_m\rangle}. \end{aligned} \tag{3.5}$$

In words, the dirty image $d_\lambda(\boldsymbol{\theta})$, synthesized from the observed complex visibilities, is simply the convolution of the specific brightness distribution $I_\lambda(\boldsymbol{\theta})$ by the dirty beam $h_\lambda(\boldsymbol{\theta})$. Figure 3 shows, for given (u,v)-coverage and observed object, the resulting dirty beam and dirty image. The dirty beam $h_\lambda(\boldsymbol{\theta})$ is the analogous of the point spread function (PSF) in conventional imaging; it is however not a probability density function, in particular, when super-synthesis is exploited, $h_\lambda(\boldsymbol{\theta})$ is not a normalized non-negative distribution (*cf.* the negative lobes of the dirty beam in Fig. 3).

To summarize, when the observables are the complex visibilities V_m and the transfer function $\widehat{h}_m$ properly calibrated, Equation (3.4) shows that image reconstruction amounts to a deconvolution problem (Cornwell 1995). There are however many unmeasured values – the voids in the coverage of the (u,v)-plane – thus the problem is, at least, ill-posed and other constraints than the data are required to warrant the uniqueness and the stability of the solution. The principles of image reconstruction developed in the remaining sections of this paper can be applied to solve this inverse problem.

So far, no considerations have been made regarding the quality of the measurements which may be very variable. In practice, *regridding* techniques (Sramek & Schwab 1989; Thompson & Bracewell 1974) are implemented to synthesize a dirty image with proper weighting of the data according to their confidence levels. By inverse Fourier transforming the expression of $\widehat{d}_\lambda(\boldsymbol{\nu})$ given by Equation (3.1), the dirty image can be directly synthesized from the complex visibilities data V_m:

$$d_\lambda(\boldsymbol{\theta}) = \sum_{m \in \mathbb{S}_\lambda} V_m \, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle \boldsymbol{\theta}, \boldsymbol{\nu}_m \rangle}. \tag{3.6}$$

To take into account the variable quality of the measurements, one can use statistical weights and synthesize the dirty image as:

$$d_\lambda(\boldsymbol{\theta}) = \sum_{m \in \mathbb{S}_\lambda} w_m \, V_m \, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle \boldsymbol{\theta}, \boldsymbol{\nu}_m \rangle}. \tag{3.7}$$

where the weights w_m are computed according to the variance of the noise. The corresponding dirty beam then writes:

$$h_\lambda(\boldsymbol{\theta}) = \sum_{m \in \mathbb{S}_\lambda} w_m \, \widehat{h}_m \, \mathrm{e}^{+\mathrm{i}\,2\,\pi\,\langle \boldsymbol{\theta}, \boldsymbol{\nu}_m \rangle}. \tag{3.8}$$

The somewhat idealized case considered here is relevant for radio-astronomy for which the complex amplitude transmissions $g_j(\lambda, t)$ are stable during an exposure and can be calibrated. We will see next (Sect. 4) that, due to the atmospheric turbulence, these assumptions cannot be made in the optical where the situation is much more involved. In terms of complexity, an intermediate situation arises when the transfer function $\widehat{h}_m$ cannot be calibrated. *Self calibration* methods (Cornwell & Wilkinson 1981) have been developed to cope with this case and consist in jointly recovering the complex amplitude transmissions $g_j(\lambda_m, t_m)$, see Equation (2.10), and the image of the object from uncalibrated complex visibilities. Self calibration is the analogous of *blind deconvolution* in conventional imaging (Campisi & Egiazarian 2007).

4 The effects of turbulence

The *atmospheric turbulence* induces random variations of the refractive index along the path traveled by the light (Roddier 1981). These fluctuations affect the modulus and the phase of the complex transmissions $g_j(\lambda, t)$ during an exposure. For instance, for an instrument like Amber (Petrov *et al.* 2007), the modulus $|g_j(\lambda, t)|$ fluctuates due to the boiling of the speckle pattern in the focal plane of the telescopes which changes the amount of coherent light injected in the optical fibers which feed the instrument and perform the spatial filtering. The turbulence also induces random delays in the optical path which affect the phase $\phi_j(\lambda, t)$ of $g_j(\lambda, t)$. The variations of the modulus of the complex transmissions can be estimated or calibrated, *e.g.* by the photometric channels of Amber. But it is much more difficult to estimate the phase errors. The situation is about to improve with the development of recombinators with phase reference (Delplancke *et al.* 2003) but, for now, there are no reliable means to estimate the phase $\phi_j(\lambda, t)$. This has a profound impact on the kind of measurements provided by an optical interferometer.

Because of the fluctuations of the complex transmissions $g_j(\lambda, t)$ during an exposure, the approximation in Equation (2.9) no longer applies: the effective transfer function $\widehat{h}_m$ is given by Equation (2.4). Then, if the fluctuations of the phase $\phi_j(\lambda, t)$ of $g_j(\lambda, t)$ are too important during the exposure, the integrand in Equation (2.4) becomes randomly distributed around zero and the averaging during the exposure yields:

$$\widehat{h}_m \approx 0. \tag{4.1}$$

This means that the complex visibilities cannot be measured when the unknown random phase fluctuations are too large during an exposure. This is the case at optical wavelengths. Even if the phase fluctuations are not so important, the effective transfer function cannot be described by a small number of complex transmissions. This forbids the use of self-calibration to guess the effective transfer function: in order to directly exploit the mean complex visibilities, $\widehat{h}_m$ must be calibrated simultaneously to the observations. For these reasons, astronomers have to integrate observables which are *insensitive to phase delay errors*.

Using very short exposure durations, typically ~1 ms, compared to the evolution time of the atmospheric effects, the instantaneous complex visibilities can be measured but with unknown phase terms. The interferometric observables are then computed by forming, from simultaneously observed complex visibilities, quantities which are insensitive to the phase of the complex transmissions. These observables are the *powerspectrum*:

$$\begin{aligned} P_m &\overset{\text{def}}{=} \langle |V_{j_{m,1},j_{m,2}}(\lambda, t)|^2 \rangle_m \\ &\approx \underbrace{\langle |g_{j_{m,1}}(\lambda, t)|^2 \, |g_{j_{m,2}}(\lambda, t)|^2 \rangle_m}_{>0} \, |\widehat{I}_{\lambda_m}(\boldsymbol{\nu}_m)|^2 \end{aligned} \tag{4.2}$$

and the *bispectrum*:

$$\begin{aligned} B_m &\overset{\text{def}}{=} \big\langle V_{j_{m,1},j_{m,2}}(\lambda, t) \, V_{j_{m,2},j_{m,3}}(\lambda, t) \, V_{j_{m,3},j_{m,1}}(\lambda, t) \big\rangle_m \\ &\approx \underbrace{\langle |g_{j_{m,1}}(\lambda, t)|^2 \, |g_{j_{m,2}}(\lambda, t)|^2 \, |g_{j_{m,3}}(\lambda, t)|^2 \rangle_m}_{>0} \, \widehat{I}^{(3)}_{\lambda_m}(\boldsymbol{\nu}_m, \boldsymbol{\nu}'_m) \end{aligned} \tag{4.3}$$

where:

$$\begin{aligned} \boldsymbol{\nu}_m &= (\boldsymbol{r}_{j_{m,2}}(t_m) - \boldsymbol{r}_{j_{m,1}}(t_m))/\lambda_m \,, \\ \boldsymbol{\nu}'_m &= (\boldsymbol{r}_{j_{m,3}}(t_m) - \boldsymbol{r}_{j_{m,2}}(t_m))/\lambda_m \,, \end{aligned}$$

and:

$$\widehat{I}^{(3)}_{\lambda}(\boldsymbol{\nu}, \boldsymbol{\nu}') \overset{\text{def}}{=} \widehat{I}_{\lambda}(\boldsymbol{\nu}) \, \widehat{I}_{\lambda}(\boldsymbol{\nu}') \, \widehat{I}^{*}_{\lambda}(\boldsymbol{\nu} + \boldsymbol{\nu}') \tag{4.4}$$

is the bispectrum of the brightness distribution of the object. To be able to measure the powerspectrum, given by Equation (4.2), two telescopes ($j_{m,1}$ and $j_{m,2}$) have to be coherently recombined; while, to measure the bispectrum, given by Equation (4.3), three telescopes ($j_{m,1}$, $j_{m,2}$ and $j_{m,3}$) have to be coherently recombined.

Note that, being non-linear quantities, the empirical powerspectrum and bispectrum have bias terms which are not shown here to simplify the equations. Dainty & Greenaway (1979) and Wirnitzer (1985) give the expressions of unbiased estimators for the powerspectrum and for the bispectrum respectively at low light levels (photon counting mode).

5 Inverse problem approach for image reconstruction

Given the interferometric observables, we want to recover an image, that is an approximation of the object specific brightness distribution at a given wavelength. Before going into the details of a method to tackle this problem, we can anticipate a number of issues and make some preliminary remarks. (i) Due to voids in the spatial frequency coverage, we are dealing with very *sparse data* (with typically a few tens of baselines, see Fig. 2). (ii) Avoiding the turbulence effects implies to use *non-linear data* (powerspectrum or bispectrum) which is more difficult to fit than, say, the complex visibilities. (iii) Compared to the $N_{\mathsf{tel}}\,(N_{\mathsf{tel}}-1)/2$ sampled frequencies per exposure, the powerspectrum provides no Fourier phase information while the bispectrum only provides $(N_{\mathsf{tel}}-1)\,(N_{\mathsf{tel}}-2)/2$ *phase closures*, so there are missing phase data (with only 3 telescopes, 2/3rd of the phases are missing). (iv) There may be calibration problems which means that there are additional unknown factors in the data.

For the sake of simplicity, we will consider in the following the case of monochromatic image reconstruction (at a given wavelength λ) and assume that we are working with debiased and calibrated data. That is, all the effective transfer functions are assumed to be equal to unity and the main problem is to deal with the sparsity of the data, the missing Fourier phase information and the non-linearity of the estimators. The possible types of measurements that may be available are:

- complex visibilities: $V_m \approx \widehat{I}_\lambda(\boldsymbol{\nu}_m)$;
- powerspectrum data: $P_m \approx \left|\widehat{I}_\lambda(\boldsymbol{\nu}_m)\right|^2$;
- bispectrum data: $B_m \approx \widehat{I}^{(3)}_\lambda(\boldsymbol{\nu}_m, \boldsymbol{\nu}'_m)$;

where the $\approx$ symbol is used because of omitted error terms.

As all measured quantities are related to the Fourier transform of the specific brightness distribution, we first need a model of the complex visibilities. This is the subject of Section 5.1.

On the one hand, due to the noise, exactly fitting the data is pointless and we expect some discrepancy between actual data and their model given the sought image. On the other hand, owing to the amount of missing information (sparse sampling of the spatial frequencies and, perhaps, only partial Fourier phase information), the data alone cannot uniquely define an image: additional *priors* are required. Image reconstruction is then a compromise between fidelity to the data and to the priors; the different formulations of this inverse problem are introduced in Section 5.2.

We will see that solving the image restoration problem amounts to minimizing the sum of two terms: a likelihood term to enforce data fidelity and a regularization term to promote agreement with the priors. Bayesian inference (Sect. 5.3) can be invoked to formally derive these terms. Practical derivation of the likelihood term is discussed in Section 6. The regularization is developed in Section 7. At least because of the necessary flexibility of the regularization[2], choosing the regularization and its tuning parameters is needed. This is briefly discussed in Section 7.3.

Finally it remains to effectively solve the problem, that is to find the best image parameters which minimize the given penalized likelihood. Numerical optimization is introduced in Section 8.

5.1 Image and complex visibilities models

Because of the noise and of the limited number of measurements, it is hopeless to aim at recovering the specific brightness distribution $I_\lambda(\boldsymbol{\theta})$ of the observed object exactly. Instead, a realistic objective is to seek for a good estimate of an approximation $i(\boldsymbol{\theta})$ of $I_\lambda(\boldsymbol{\theta})$ which depends on a finite number of parameters. To that end, the specific brightness distribution in angular direction $\boldsymbol{\theta}$ can be approximated by:

$$i(\boldsymbol{\theta}) \overset{\text{def}}{=} \sum\nolimits_n x_n \, b_n(\boldsymbol{\theta}) \approx I_\lambda(\boldsymbol{\theta}) \tag{5.1}$$

with $\{b_n(\boldsymbol{\theta})\colon \mathbb{R}^2 \mapsto \mathbb{R}\}_{n=1}^{N}$ a basis of functions and $\boldsymbol{x} \in \mathbb{R}^N$ the *image parameters*. This general parametrization accounts, for instance, for a pixelized image, for a wavelet decomposition, etc.. For image reconstruction, it may be the most convenient to use a shift-invariant basis of functions defined by:

$$b_n(\boldsymbol{\theta}) = b(\boldsymbol{\theta} - \boldsymbol{\theta}_n) \tag{5.2}$$

where $b(\boldsymbol{\theta})\colon \mathbb{R}^2 \mapsto \mathbb{R}$ is a single basis function and $\mathbb{G} = \{\boldsymbol{\theta}_n \in \mathbb{R}^2 \mid n = 1, \ldots, N\}$ is a grid of evenly spaced positions. If $b(\boldsymbol{\theta})$ is an interpolation function (Thévenaz *et al.* 2000), then the image parameters sample the brightness distribution:

$$x_n = i(\boldsymbol{\theta}_n) \approx I_\lambda(\boldsymbol{\theta}_n).$$

The advantage of approximating the specific brightness distribution by the linear expansion $i(\boldsymbol{\theta})$ given in Equation (5.1) is that its exact Fourier transform is also linear with respect to the image parameters $\boldsymbol{x}$:

$$\widehat{\imath}(\boldsymbol{\nu}) = \sum\nolimits_n x_n \, \widehat{b}_n(\boldsymbol{\nu}) \approx \widehat{I}_\lambda(\boldsymbol{\nu}), \tag{5.3}$$

[2] Such flexibility is required because the object of interest is unknown.

where the hat $\widehat{\ }$ denotes the Fourier transformed distribution and $\boldsymbol{\nu}$ is the spatial frequency conjugate of the angular position $\boldsymbol{\theta}$. For any sampled spatial frequency $\boldsymbol{\nu}_m$ the model complex visibility thus writes:

$$y_m \overset{\text{def}}{=} \widehat{\imath}(\boldsymbol{\nu}_m) = \sum\nolimits_n \widehat{b}_n(\boldsymbol{\nu}_m)\, x_n = \sum\nolimits_n H_{m,n}\, x_n \approx \widehat{I}_\lambda(\boldsymbol{\nu}_m)\,,$$

with $H_{m,n} = \widehat{b}_n(\boldsymbol{\nu}_m)$. In matrix notation:

$$\boldsymbol{y} = \mathbf{H}\cdot\boldsymbol{x}, \tag{5.4}$$

where $\boldsymbol{y} \in \mathbb{C}^M$ collects the model complex visibilities at all sampled frequencies and $\mathbf{H} \in \mathbb{C}^{M\times N}$ is a sub-sampled Fourier transform operator. The memory requirement to store the coefficients of the operator $\mathbf{H}$ and the computer time needed to apply $\mathbf{H}$ (or its adjoint) both scale as $O(M \times N)$. Fast approximations of $\mathbf{H}$ based on the FFT can be used (Fessler & Sutton 2003; Potts *et al.* 2001) when $M \times N$ is too large. To use these fast approximations, the image model must be defined with shift-invariant basis functions, see Equation (5.2), on an evenly spaced grid $\mathbb{G}$.

5.2 Inverse problem formulations

As stated before, image reconstruction is a compromise between various constraints resulting from the measurements and from prior knowledge. The first of these constraints is that the image must be *compatible with the available data*. This is asserted by comparing the measurements with their model given the image parameters $\boldsymbol{x}$. To keep the maximum flexibility and since the model of all the measured quantities depend on the model complex visibilities $\boldsymbol{y} = \mathbf{H}\cdot\boldsymbol{x}$, we postulate that, to be compatible with the measurements, the image parameters must satisfy the following criterion:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \le \eta \tag{5.5}$$

where $f_{\mathsf{data}}(\boldsymbol{y})\colon \mathbb{C}^M \mapsto \mathbb{R}_+$ is a measure of the distance between the model complex visibilities $\boldsymbol{y} = \mathbf{H}\cdot\boldsymbol{x}$ and the actual data. The threshold η is chosen to set how close to the data should be the model. As $f_{\mathsf{data}}(\boldsymbol{y})$ is a distance, the smaller η the closer the model to the data. However taking $\eta = 0$ would mean that the model should exactly match the data and thus *fit the noise* which is undesirable. So we always want $\eta > 0$, depending on the exact definition of $f_{\mathsf{data}}(\boldsymbol{y})$, the value of the threshold may vary with, *e.g.*, the noise level and the number of measurements.

The level of agreement with the prior knowledge can be expressed in the same spirit by specifying a distance $f_{\mathsf{prior}}(\boldsymbol{x})$ and requiring that this distance be as small as possible providing that the model remains compatible with the data. Formally, this writes:

$$\boxed{\boldsymbol{x}^\star = \mathop{\arg\min}_{\boldsymbol{x}\in\mathbb{X}} f_{\mathsf{prior}}(\boldsymbol{x}) \quad \text{s.t.} \quad f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \le \eta,} \tag{5.6}$$

where the *feasible set* $\mathbb{X} \subset \mathbb{R}^N$ is introduced to impose strict constraints such as the non-negativity of the image. For instance, using bilinear interpolation for the approximation in Equation (5.1), the specific brightness distribution $i(\boldsymbol{\theta})$ is non-negative and normalized if and only if the parameters $\boldsymbol{x}$ are non-negative and their sum is equal to $\xi \stackrel{\text{def}}{=} \iint i(\boldsymbol{\theta})\,\mathrm{d}^2\boldsymbol{\theta}$:

$$i(\boldsymbol{\theta}) \geq 0 \quad \text{and} \quad \iint i(\boldsymbol{\theta})\,\mathrm{d}^2\boldsymbol{\theta} = \xi \quad \Longleftrightarrow \quad \boldsymbol{x} \in \mathbb{X}$$

with:

$$\mathbb{X} = \{\boldsymbol{x} \in \mathbb{R}^N \mid \boldsymbol{x} \geq 0, \mathbf{1}^\top \cdot \boldsymbol{x} = \xi\}, \tag{5.7}$$

where the inequality $\boldsymbol{x} \geq 0$ is taken componentwise and where $\mathbf{1}$ is the vector of $\mathbb{R}^N$ with all components equal to 1:

$$\begin{aligned} \boldsymbol{x} \geq 0 \quad &\Longleftrightarrow \quad \forall n, x_n \geq 0 \\ \mathbf{1} = (1, \ \ldots, 1)^\top \quad &\Longrightarrow \quad \mathbf{1}^\top \cdot \boldsymbol{x} = \sum\nolimits_n x_n. \end{aligned}$$

The constrained problem (5.6) is usually solved via the Lagrangian (Nocedal & Wright 2006):

$$\mathcal{L}(\boldsymbol{x};\ell) = f_{\mathsf{prior}}(\boldsymbol{x}) + \ell\, f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x})$$

with $\ell \geq 0$ the Lagrange multiplier for the inequality constraint $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \leq \eta$. Assuming that $\mathcal{L}(\boldsymbol{x};\ell)$ has a unique reachable minimum on the feasible set, we can define:

$$\boldsymbol{x}^+_{\mathcal{L}}(\ell) \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} \mathcal{L}(\boldsymbol{x};\ell)\,,$$

and seek for the value $\ell^\star \geq 0$ of the multiplier such that the solution $\boldsymbol{x}^\star = \boldsymbol{x}^+_{\mathcal{L}}(\ell^\star)$ complies with the constraints. Obviously, we want $\ell^\star > 0$ otherwise the data play no role in the determination of the solution. Intuitively, having the solution strictly closer to the data than required, *i.e.* $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}^\star) < \eta$, yields a worst value of $f_{\mathsf{prior}}(\boldsymbol{x}^\star)$ than having $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}^\star) = \eta$. Thus, unless the a priori solution:

$$\boldsymbol{x}_{\mathsf{prior}} \stackrel{\text{def}}{=} \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} f_{\mathsf{prior}}(\boldsymbol{x})$$

is such that $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}_{\mathsf{prior}}) \leq \eta$, in which case the solution is $(\boldsymbol{x}^\star, \ell^\star) = (\boldsymbol{x}_{\mathsf{prior}}, 0)$, the solution to the problem (5.6) is given by $\boldsymbol{x}^\star = \boldsymbol{x}^+_{\mathcal{L}}(\ell^\star)$ with $\ell^\star > 0$ such that $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}^\star) = \eta$.

Since the solution is obtained for a Lagrange multiplier strictly positive, we can take $\mu = 1/\ell$ and alternatively define the solution to be given by minimizing another penalty function:

$$\boldsymbol{x}^+_f(\mu) = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} f(\boldsymbol{x};\mu) \quad \text{with:} \quad f(\boldsymbol{x};\mu) = f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})\,. \tag{5.8}$$

The solution is then $\boldsymbol{x}^\star = \boldsymbol{x}^+_f(\mu^\star)$ where the optimal weight $\mu^\star > 0$ for the priors is such that $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}^\star) = \eta$. The two different formulations are equivalent and yield the same solution of the constrained problem (5.6).

We shall now see how to derive the expression of the *distances* $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x})$ and $f_{\mathsf{prior}}(\boldsymbol{x})$.

5.3 Bayesian inference

The previous considerations may found strong theoretical justification in a Bayesian framework where probabilities represent any available information. For instance, in a *maximum a posteriori* (MAP) approach, the best image parameters $\boldsymbol{x}_{\mathsf{MAP}}$ are the most likely ones given the data $\boldsymbol{z}$:

$$\boldsymbol{x}_{\mathsf{MAP}} = \arg\max_{\boldsymbol{x}} \Pr(\boldsymbol{x}|\boldsymbol{z}),$$

where $\Pr(\boldsymbol{x}|\boldsymbol{z})$ denotes the probability (or the probability density function) of $\boldsymbol{x}$ given $\boldsymbol{z}$. Note that the data $\boldsymbol{z}$ collects all measurements; in our case, $\boldsymbol{z}$ may include complex visibilities, powerspectra and bispectra. Using Bayes theorem[3], discarding terms which do no depend on $\boldsymbol{x}$ and noting that $-\log(p)$ is a strictly decreasing function of p yields:

$$\begin{aligned}\boldsymbol{x}_{\mathsf{MAP}} &= \arg\max_{\boldsymbol{x}} \frac{\Pr(\boldsymbol{z}|\boldsymbol{x})\ \Pr(\boldsymbol{x})}{\Pr(\boldsymbol{z})}\\ &= \arg\max_{\boldsymbol{x}} \Pr(\boldsymbol{z}|\boldsymbol{x})\ \Pr(\boldsymbol{x})\\ &= \arg\min_{\boldsymbol{x}} -\log(\Pr(\boldsymbol{z}|\boldsymbol{x})) - \log(\Pr(\boldsymbol{x})).\end{aligned}$$

Hence:

$$\boldsymbol{x}_{\mathsf{MAP}} = \arg\min_{\boldsymbol{x}} f_{\boldsymbol{z}|\boldsymbol{x}}(\boldsymbol{x}) + f_{\boldsymbol{x}}(\boldsymbol{x}), \tag{5.9}$$

with:

$$f_{\boldsymbol{z}|\boldsymbol{x}}(\boldsymbol{x}) = -\log(\Pr(\boldsymbol{z}|\boldsymbol{x})) \tag{5.10}$$

$$f_{\boldsymbol{x}}(\boldsymbol{x}) = -\log(\Pr(\boldsymbol{x})). \tag{5.11}$$

In words, the MAP solution $\boldsymbol{x}_{\mathsf{MAP}}$ is a compromise between maximizing the likelihood of the data $\boldsymbol{z}$ given the model parameters $\boldsymbol{x}$ and maximizing the prior probability of the model. Said otherwise, the compromise is between fitting the data, *i.e.* minimize $f_{\boldsymbol{z}|\boldsymbol{x}}(\boldsymbol{x})$, and agreement with prior knowledge, *i.e.* minimize $f_{\boldsymbol{x}}(\boldsymbol{x})$.

Finally, the solution $\boldsymbol{x}^{+}_{f}(\mu)$ of the problem (5.8) is also the MAP solution $\boldsymbol{x}_{\mathsf{MAP}}$ if we take:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) = c'_0 + c_1\, f_{\boldsymbol{z}|\boldsymbol{x}}(\boldsymbol{x}) \tag{5.12}$$

$$\mu\, f_{\mathsf{prior}}(\boldsymbol{x}) = c''_0 + c_1\, f_{\boldsymbol{x}}(\boldsymbol{x}) \tag{5.13}$$

with c'_0, c''_0 and $c_1 > 0$ any suitable real constants. From this close relationship, we deduce a possible way to define the penalty functions $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x})$ and $f_{\mathsf{prior}}(\boldsymbol{x})$. This is the subject of the next two sections.

[3] Bayes theorem states that the joint probability of A and B writes:

$$\Pr(A,B) = \Pr(A)\ \Pr(B|A) = \Pr(B)\ \Pr(A|B).$$

6 Likelihood of the data

Ideally, the likelihood should be strictly based on the noise statistics of the data:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) \overset{\text{def}}{=} c_0' - c_1 \log(\Pr(\boldsymbol{z}|\mathbf{H}\cdot\boldsymbol{x})).$$

If the measurements have Gaussian statistics, then for $c_1 = 2$ and for an appropriate choice of c_0', the likelihood term is a so-called χ^2 given by:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) = [\boldsymbol{z} - \widetilde{\boldsymbol{z}}(\mathbf{H}\cdot\boldsymbol{x})]^\top \cdot \mathbf{W} \cdot [\boldsymbol{z} - \widetilde{\boldsymbol{z}}(\mathbf{H}\cdot\boldsymbol{x})],$$

where $\widetilde{\boldsymbol{z}}(\mathbf{H}\cdot\boldsymbol{x})$ is the model of the measurements $\boldsymbol{z}$ and $\mathbf{W}$ is a weighting matrix equal to the inverse of the covariance of the measurements: $\mathbf{W} = \operatorname{Cov}\{\boldsymbol{z}\}^{-1}$. Our notation accounts for the fact that the model of the measurements only depends on the model complex visibilities $\mathbf{H}\cdot\boldsymbol{x}$ and assumes that all measurements are real valued (any complex valued data has to be considered as a pair of reals).

A first difficulty is that the statistics of real interferometric measurements is not well known and may not be Gaussian at all. For instance, Figure 4 shows the empirical distribution of bispectrum data. At low signal to noise ratio (SNR), the distribution may be well approximated by a Gaussian distribution; while, at high SNR, the *banana shaped* distribution of the data suggests that the amplitude and phase of the complex bispectrum may be independent variables. Figure 5 shows that this banana shaped distribution can only be grossly approximated by a Gaussian with respect to the real and imaginary parts of the bispectrum data.

A second difficulty is that not all statistical information is provided with the data. Generally, only estimates of the error bar (standard deviation) of each measurement is available. In particular no information is stored about the correlation of the measurements. This is the case of data stored into the OI-FITS format, a data exchange standard for optical interferometry (Pauls *et al.* 2005). Without any measured correlations, one is obliged to assume that measurements are independent variables (for the powerspectrum data) or pairs of variables (for complex data like the complex visibilities and the bispectra). The likelihood term is then a sum of terms, one for each independent subset of data:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) = \sum_m f_m(\boldsymbol{z}_m - \widetilde{\boldsymbol{z}}_m(\mathbf{H}\cdot\boldsymbol{x}))$$

where each elementary datum $\boldsymbol{z}_m$ is either a real or a pair of reals (amplitude and phase or real and imaginary parts of a complex measurement).

In the most simple case, the data consists in independent calibrated complex visibilities with independent and identically distributed (i.i.d.) real and imaginary parts (the so-called Goodman approximation, Goodman 1985). The likelihood term then writes:

$$f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x}) = \sum_m w_m \, |\boldsymbol{z}_m - (\mathbf{H}\cdot\boldsymbol{x})_m|^2$$

with $w_m = 1/\operatorname{Var}\{\operatorname{Re}\{\boldsymbol{z}_m\}\} = 1/\operatorname{Var}\{\operatorname{Im}\{\boldsymbol{z}_m\}\}$ and $|\boldsymbol{z}_m - (\mathbf{H}\cdot\boldsymbol{x})_m|$ the modulus of the complex residuals. In matrix notation and providing the Argand representation[4] of the complex visibilities is used, the likelihood can be put in the form of a quadratic cost function with respect to the unknowns $\boldsymbol{x}$:

$$f_{\mathsf{data}}(\mathbf{H}{\cdot}\boldsymbol{x}) = (\boldsymbol{z} - \mathbf{H}{\cdot}\boldsymbol{x})^{\top} \cdot \mathbf{W} \cdot (\boldsymbol{z} - \mathbf{H}{\cdot}\boldsymbol{x})$$

where $\mathbf{W}$ is block diagonal matrix with 2×2 blocks. This is suitable for radio-astronomy data but not for current optical interferometers. See, for instance, Meimon *et al.* (2005a) and Thiébaut (2008) for various approximate expressions of the likelihood term. Note that Goodman approximation would give circular isocontours in Figure 5.

For complex data $\boldsymbol{z}_m = \rho_m \, \exp(\mathrm{i}\,\varphi_m)$ in polar form with independent modulus and phase, Meimon *et al.* (2005a) suggested to use a quadratic approximation of the likelihood:

$$f_m(\mathbf{H}\cdot\boldsymbol{x}) = \boldsymbol{e}_m(\mathbf{H}\cdot\boldsymbol{x})^{\top} \cdot \begin{pmatrix} W_m^{\mathsf{rr}} & W_m^{\mathsf{ri}} \\ W_m^{\mathsf{ri}} & W_m^{\mathsf{ii}} \end{pmatrix} \cdot \boldsymbol{e}_m(\mathbf{H}\cdot\boldsymbol{x}), \tag{6.1}$$

with the weights:

$$W_m^{\mathsf{rr}} = \frac{\cos^2\varphi_m}{\mathrm{Var}\{\rho_m\}} + \frac{\sin^2\varphi_m}{\rho_m^2\;\mathrm{Var}\{\varphi_m\}}, \tag{6.2}$$

$$W_m^{\mathsf{ri}} = \left(\frac{1}{\mathrm{Var}\{\rho_m\}} - \frac{1}{\rho_m^2\;\mathrm{Var}\{\varphi_m\}}\right) \cos\varphi_m\;\sin\varphi_m\,, \tag{6.3}$$

$$W_m^{\mathsf{ii}} = \frac{\sin^2\varphi_m}{\mathrm{Var}\{\rho_m\}} + \frac{\cos^2\varphi_m}{\rho_m^2\;\mathrm{Var}\{\varphi_m\}}, \tag{6.4}$$

and the complex residuals:

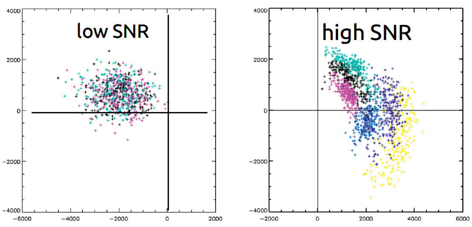
$$\boldsymbol{e}_m(\mathbf{H}\cdot\boldsymbol{x}) = \begin{pmatrix} \rho_m\;\cos\varphi_m - \widetilde{\rho}_m(\mathbf{H}{\cdot}\boldsymbol{x})\;\cos\widetilde{\varphi}_m(\mathbf{H}{\cdot}\boldsymbol{x}) \\ \rho_m\;\sin\varphi_m - \widetilde{\rho}_m(\mathbf{H}{\cdot}\boldsymbol{x})\;\sin\widetilde{\varphi}_m(\mathbf{H}{\cdot}\boldsymbol{x}) \end{pmatrix} \tag{6.5}$$

where the tilde indicates the model of a given measurement. The expression of the likelihood in Equation (6.1) can be used for complex visibilities V_m or bispectrum data B_m in polar form as provided by OI-FITS format. However note that this yields a non-quadratic penalty for the bispectrum.

Some algorithms ignore the measured amplitudes of the bispectrum and only consider the bispectrum phase $\beta_m = \arg(B_m)$ to provide Fourier phase information for the image reconstruction, the Fourier amplitude information being provided by the powerspectrum data. In this case, practical expressions of the likelihood with respect to such kind of data must be introduced. In MiRA algorithm (Thiébaut 2008), powerspectrum data are treated as independent Gaussian variables, the likelihood for the measured powerspectrum P_m then writes:

$$f_m(\mathbf{H}\cdot\boldsymbol{x}) = \frac{\left(P_m - \widetilde{P}_m(\mathbf{H}\cdot\boldsymbol{x})\right)^2}{\mathrm{Var}\{P_m\}}. \tag{6.6}$$

[4] Real and imaginary parts.



(triple product of FKV0509)

Source : C. Hummel *et al.* <http://www.mrao.cam.ac.uk/~jsy1001/exchange/complex/complex.html>

Fig. 4. Empirical distribution of complex bispectrum data at low (left and high (right) signal to noise ratio (SNR). Horizontal axis is the real part, vertical axis is the imaginary part.

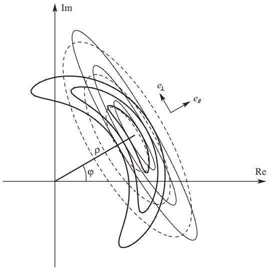


Fig. 5. Convex quadratic approximations of the true distribution of errors for a complex measurement. Thick lines: χ^2 isocontours (at 1, 2 and 3 rms levels) for a complex data with independent amplitude and phase. Dashed lines: isocontours for the global quadratic approximation. Thin lines: isocontours for the local quadratic approximation. (Meimon *et al.* 2005a).

In order to account for phase wrapping and to avoid excessive non-linearity, the term related to the phase closures data is defined by MiRA to be the weighted quadratic distance between the complex phasors rather than between the phases closures:

$$f_m(\mathbf{H}\cdot\boldsymbol{x}) = \frac{1}{\mathrm{Var}\{\beta_m\}}\left|\mathrm{e}^{\mathrm{i}\,\beta_m} - \mathrm{e}^{\mathrm{i}\,\widetilde{\beta}_m(\mathbf{H}\cdot\boldsymbol{x})}\right|^2 . \tag{6.7}$$

In the limit of small phase closure errors, the penalty becomes:

$$f_m(\mathbf{H}\cdot\boldsymbol{x}) \approx \frac{\left[\beta_m - \widetilde{\beta}_m(\mathbf{H}\cdot\boldsymbol{x})\right]^2}{\mathrm{Var}\{P_m\}} \tag{6.8}$$

which is readily the χ^2 term that would be obtained for Gaussian phase statistics. This justifies the weighting used in Equation (6.7). Other expressions of the likelihood with respect to phase data have been proposed to cope with the phase wrapping (Haniff 1991; Lannes 2001) but, in practice, they give penalties which slow down or even prevent the convergence of the optimization algorithm.

For optical interferometry which only provides powerspectrum and bispectrum data, the likelihood term $f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x})$ is highly non-quadratic, *e.g.* see Equations (6.6) and (6.7). This will give rise to optimization issues when fitting the data. Before tackling these issues, let us discuss the second penalty term, that is the regularization.

7 Regularization

In principle, the regularization penalty could be derived from Bayesian considerations (see Sect. 5.3):

$$\mu\, f_{\mathsf{prior}}(\boldsymbol{x}) = c_0'' - c_1\, \log(\mathrm{Pr}(\boldsymbol{x})). \tag{7.1}$$

with c_0'' any real constant and $c_1 > 0$ the same constant as in the previous section. However, introducing a prior probability density function of the parameters which is sufficiently general for all possible observed objects would yield highly uninformative priors which do not really help finding a satisfying image. To be effective, the regularization has to be more restrictive which implies to make more specific assumptions about the object brightness distribution. Besides, even if we knew the object quite exactly, we would like that the prior penalty be at least insensitive to the observing conditions, thus to the position of the object, its orientation and its distance (*i.e.* its angular size and its integrated brightness).

7.1 Simple quadratic regularization

Let us examine the consequences of these elementary considerations. To simplify our reasoning, we consider the pixel-oriented image model:

$$x_n = i(\boldsymbol{\theta}_n) \approx I_\lambda(\boldsymbol{\theta}_n),$$

and assume that the parameters $\boldsymbol{x}$ follow a Gaussian distribution[5], then:

$$\begin{aligned} f_{\boldsymbol{x}}(\boldsymbol{x}) &= -\log(\mathrm{Pr}(\boldsymbol{x})) \\ &= c + (1/2)\, (\boldsymbol{x} - \overline{\boldsymbol{x}})^\top \cdot \mathbf{C}_{\boldsymbol{x}}^{-1} \cdot (\boldsymbol{x} - \overline{\boldsymbol{x}}) \end{aligned} \tag{7.2}$$

where $\overline{\boldsymbol{x}} = \mathrm{E}\{\boldsymbol{x}\}$ is the expected value of $\boldsymbol{x}$, $\mathbf{C}_{\boldsymbol{x}} = \mathrm{Cov}(\boldsymbol{x})$ its covariance and

$$c = \frac{1}{2}\,\log\left[\det\left(\frac{\mathbf{C}_{\boldsymbol{x}}}{2\,\pi}\right)\right]$$

is a constant due to the normalization of $\mathrm{Pr}(\boldsymbol{x})$ and which does not depend on $\boldsymbol{x}$.

From the principle that the regularization shall be shift-invariant, the covariance $(\mathbf{C}_{\boldsymbol{x}})_{n,n'}$ between the nth and the n'th pixels must only depend on their relative position $\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n'}$; moreover, since the regularization shall be isotropic, it must only depend on the relative distance $\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n'}\|$. This is also true for the inverse of the covariance matrix, thus:

$$\left(\mathbf{C}_{\boldsymbol{x}}^{-1}\right)_{n,n'} = \alpha\, \zeta(\|\boldsymbol{\theta}_n - \boldsymbol{\theta}_{n'}\|/\Omega) \tag{7.3}$$

[5] Which cannot be really true because of the non-negativity and, perhaps, normalization constraints.

where $\alpha > 0$ is a scaling factor, ζ: $\mathbb{R}_+ \mapsto \mathbb{R}$ is a function of the relative angular separation between the pixels and Ω is a typical angular size. From the requirements that the prior shall not depend on the absolute brightness of the object, nor on its angular size, the factor α shall scale as the reciprocal of the square of the object brightness and Ω shall scale as the angular size of the object.

As the regularization shall be shift-invariant, the mean must not depend on the pixel index, hence:

$$\overline{\boldsymbol{x}} = \beta\,\mathbf{1}, \tag{7.4}$$

where β is the mean pixel brightness. Noting that:

$$\begin{aligned} (x_n - x_{n'})^2 &= [(x_n - \beta) - (x_{n'} - \beta)]^2 \\ &= (x_n - \beta)^2 + (x_{n'} - \beta)^2 - 2\,(x_n - \beta)\,(x_{n'} - \beta) \\ \Longrightarrow \quad (x_n - \beta)\,(x_{n'} - \beta) &= (1/2)\left[(x_n - \beta)^2 + (x_{n'} - \beta)^2 - (x_n - x_{n'})^2\right], \end{aligned}$$

the prior penalty $f_{\boldsymbol{x}}(\boldsymbol{x})$ in Equation (7.2) writes:

$$\begin{aligned} f_{\boldsymbol{x}}(\boldsymbol{x}) &= c + (1/2)\,(\boldsymbol{x} - \beta\,\mathbf{1})^\top \cdot \mathbf{C}_{\boldsymbol{x}}^{-1} \cdot (\boldsymbol{x} - \beta\,\mathbf{1}) \\ &= c + \frac{1}{2} \sum_{n,n'} \left(\mathbf{C}_{\boldsymbol{x}}^{-1}\right)_{n,n'} (x_n - \beta)\,(x_{n'} - \beta) \\ &= c + \frac{\mu_0}{2} \sum_{n} (x_n - \beta)^2 + \frac{1}{2} \sum_{n<n'} \mu_{n,n'}\,(x_n - x_{n'})^2 \end{aligned} \tag{7.5}$$

with:

$$\mu_0 = \sum_{n} \left(\mathbf{C}_{\boldsymbol{x}}^{-1}\right)_{n,n'} = \sum_{n'} \left(\mathbf{C}_{\boldsymbol{x}}^{-1}\right)_{n,n'} \tag{7.6}$$

$$\mu_{n,n'} = -\left(\mathbf{C}_{\boldsymbol{x}}^{-1}\right)_{n,n'} \tag{7.7}$$

where the two equivalent expressions for μ_0 come from the fact that the covariance matrix is symmetrical and so is its inverse.

Taking $c_1 = 2$ (as for the likelihood in Sect. 6) and $c_0'' = -c_1\,c$, yields the quadratic regularization term:

$$\mu\,f_{\mathsf{prior}}(\boldsymbol{x}) = \mu_0 \sum_{n} (x_n - \beta)^2 + \sum_{n<n'} \mu_{n,n'}\,(x_n - x_{n'})^2\,. \tag{7.8}$$

These simple and general considerations lead us to the quadratic regularization in Equation (7.8) which has the required properties (shift-invariance, isotropy, etc.) and which is parametrized by so-called *hyper-parameters*: α, β (both related to the object brightness), Ω (the size of the object) and ζ: $\mathbb{R}_+ \mapsto \mathbb{R}$ the relative weighting function. If we take $\mu = \mu_0 > 0$, $\beta = 0$ and $\mu_{n,n'} = 0$, $\forall(n, n')$, then we obtain the most simple form of Tikhonov's regularization (Tikhonov & Arsenin 1977):

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum_{n} x_n^2 = \|\boldsymbol{x}\|_2^2.$$

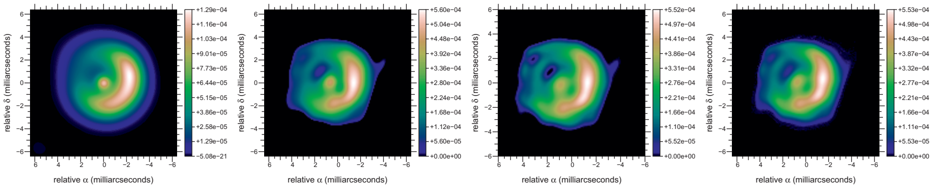


Fig. 6. Image reconstruction with various types of regularization. From *left* to *right*: (a) original object smoothed to the resolution of the interferometer (FWHM $\sim 15\,\mathrm{mas}$); (b) reconstruction with a quadratic regularization given by Equation (7.9) and which imposes a compact field of view; (c) reconstruction with edge-preserving regularization as in Equation (7.10); (d) reconstruction with maximum entropy regularization as in Equation (7.12). All reconstructions by algorithm MiRA (Thiébaut 2008) and from the powerspectrum and the phase closures data of the 2004' Imaging Beauty Contest (Lawson *et al.* 2004).

Whereas if we take $\mu_0 = 0$ and $\mu_{n,n'} \geq 0$ a decreasing function of the distance between the nth and the n'th pixels, then we obtain a regularization which favors solutions where nearby pixels have similar values hence the smoothness of the restored image.

7.2 A marketplace for regularization

The Gaussian assumption for the prior distribution of the image parameters yields quadratic regularizations, like the one in Equation (7.8), which are easy to minimize numerically. However such regularizations alone[6] are not very efficient to interpolate missing data when dealing with sparse interferometric data. They are also not the best choice to restore some features of the observed objects, in particular point-like sources or sharp edges. Non-quadratic regularizations have been proposed which may be more suitable for sparse data and images with sharp structures.

The most useful regularizations for image restoration are shift-invariant, (approximately) isotropic and parametrized by a few hyper-parameters. However, in the case of optical interferometry data where the observables (powerspectrum and bispectrum) are insensitive to the position of the object, it may be useful to introduce a shift-variant regularization to fix this degeneracy (see the *compactness* regularization below proposed by le Besnerais *et al.* 2008).

It is impossible to give an exhaustive list of regularizations, but for image restoration, in particular from interferometric data, the following prior penalties have been used with some success:

Quadratic smoothness is imposed by minimizing the differences between close pixels. This is achieved with:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \|\mathbf{D} \cdot \boldsymbol{x}\|_2^2$$

[6] Without the strict constraints imposed by the feasible set $\mathbb{X}$.

where $\mathbf{D}$ is a finite difference operator. For instance, in 1-D:

$$(\mathbf{D}\cdot\boldsymbol{x})_n = x_{n+1} - x_n$$

and in 2-D:

$$(\mathbf{D}\cdot\boldsymbol{x})_{n_1,n_2} = \left(\begin{array}{c} x_{n_1+1,n_2} - x_{n_1,n_2} \\ x_{n_1,n_2+1} - x_{n_1,n_2} \end{array}\right).$$

This regularization is specific instance of Equation (7.8) with $\mu_0 = 0$ and

$$\mu_{n,n'} = \mu\,[\delta(n_1+1-n'_1)\,\delta(n_2-n'_2) + \delta(n_1-n'_1)\,\delta(n_2+1-n'_2)].$$

A similar result can be obtained with:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \|\boldsymbol{x} - \mathbf{S}\cdot\boldsymbol{x}\|_2^2$$

where $\mathbf{S}$ is a smoothing operator.

Compactness can be achieved with

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum\nolimits_n w_n^{\mathsf{prior}} x_n^2, \tag{7.9}$$

where $w_n^{\mathsf{prior}} \geq 0$ are given weights. If the weights increase with the distance to a given position (for instance, $w_n^{\mathsf{prior}} \propto \|\boldsymbol{\theta}_n\|^\beta$ with $\beta > 0$), this regularization favors a compact brightness distribution with its flux concentrated around this position. In the Fourier domain, this yields *spectral smoothness* which may be very helpful to interpolate the voids in the (u,v)-coverage.

If the weights are all strictly positive, it can be shown (le Besnerais *et al.* 2008) that the default solution:

$$\boldsymbol{x}^{\mathsf{prior}} \overset{\text{def}}{=} \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} \sum\nolimits_n w_n^{\mathsf{prior}} \boldsymbol{x}_n^2$$

on the feasible set $\mathbb{X}$ given in Equation (5.7) is simply:

$$x_n^{\mathsf{prior}} \propto 1/w_n^{\mathsf{prior}}$$

where the constant of proportionality is such that the normalization constraint is satisfied.

Non-linear smoothness can be imposed with the following general expression:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum\nolimits_n \sqrt{\|\nabla x_n\|^2 + \epsilon^2} \tag{7.10}$$

where $\|\nabla x_n\|^2$ is the squared magnitude of the spatial gradient in the image at nth pixel and $\epsilon \geq 0$. Taking $\epsilon = 0$ yields the so-called *total variation* (TV) regularization which favors flat regions separated by sharp edges (Rudin *et al.* 1992). Otherwise, taking $\epsilon > 0$ yields *edge-preserving smoothness* (Charbonnier *et al.* 1997) which behaves as a quadratic smoothness prior in region where the spatial gradient of the image is smaller than ϵ in magnitude, while preserving sharp edges elsewhere. The actual expression in Equation (7.10) depends on the approximation of the spatial gradient which is usually implemented via a finite difference operator: $\nabla x_n = \mathbf{D}_n \cdot \boldsymbol{x}$ (Chambolle *et al.* 2011). There are also other possibilities to achieve edge-preserving regularization (see *e.g.*, Charbonnier *et al.* 1997).

Spatial sparsity can be imposed thanks to separable ℓ_p norms:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = \sum\nolimits_n |x_n|^p, \tag{7.11}$$

with $p \geq 0$. If $p < 1$, minimizing the ℓ_p norm favors sparse distribution, while $p = 2$ corresponds to regular *Tikhonov regularization* (Tikhonov & Arsenin 1977) and favors flat distributions. Note that p must be greater or equal 1 to have a convex criterion. Taking the smallest such p, that is $p = 1$, is what is advocated in *compress sensing* (Donoho 2006).

Maximum entropy methods (MEM) have been proposed for radio-astronomy and exploit a separable non-linear regularization with the general form:

$$f_{\mathsf{prior}}(\boldsymbol{x}) = -\sum\nolimits_n h(x_n|\overline{x}_n). \tag{7.12}$$

Here the prior is to assume that the image is drawn toward a prior model $\overline{\boldsymbol{x}}$ according to a non-quadratic potential h, called the *entropy.* Various entropy terms have been proposed in the literature (Narayan & Nityananda 1986):

$$\begin{array}{ll} \text{MEM-sqrt:} & h(x|\overline{x}) = \sqrt{x}; \\ \text{MEM-log:} & h(x|\overline{x}) = \log(x); \\ \text{MEM-prior:} & h(x|\overline{x}) = x - \overline{x} - x\,\log\left(x/\overline{x}\right). \end{array}$$

Being separable, the expression in Equation (7.12) assumes that the pixel values are uncorrelated. To impose some level of smoothness in the solution, Horne (1985) has proposed a non-separable MEM regularization by defining the prior model $\overline{\boldsymbol{x}}$ as a smoothed version of the model $\boldsymbol{x}$, for instance: $\overline{\boldsymbol{x}} = \mathbf{S} \cdot \boldsymbol{x}$ with $\mathbf{S}$ a smoothing operator.

7.3 Choosing and tuning the regularization

As we have just seen, there are many different possible expressions for the regularization term. Since the exact statistics of the sought object is seldom known, the regularization has to be chosen on the basis of general properties that one expect to see in the sought image. In the case of interferometric imaging, Renard *et al.* (2011) have compared the regularization methods presented in the previous section. As expected they concluded that the best prior depends on the object of interest. However, non-linear smoothness, in Equation (7.10), and compactness combined with non-negativity constraints, in Equation (7.9), are the most

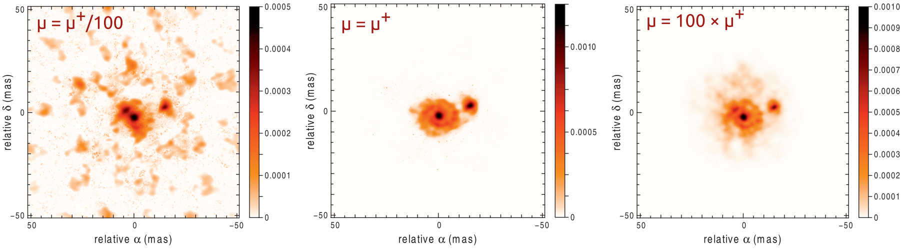


Fig. 7. Image reconstruction with ℓ_2 compactness and for various levels of regularization. The optimal regularization level is μ^+ (source: Renard *et al.* 2011).

successful regularizations in general. Figure 6 shows that images restored with different types of regularization are fairly similar. This is a general observation: providing there are sufficient data and the hyper-parameters are correctly set (see below), the restored image either succeeds to approximate the object or clearly fails (Renard *et al.* 2011). In practice, it is fruitful to exploit the variety of regularization types to determine which one is most adapted to the object of interest. Comparing images obtained under different priors is also useful to disentangle between artifacts and real features. One must however keep in mind that, among other properties, the priors must be able to lift the degeneracies of the inverse problem and to regularize it, that is to warrant a unique and stable solution with respect to small perturbations such as those due to the noise.

In addition to the choice of the form of the regularization itself, there are tuning parameters: the weight μ of the regularization, and perhaps some other hyper-parameters (*e.g.* the relaxation parameter ϵ in the edge-preserving regularization below). Ideally one would like to set these hyper-parameters automatically according to some objective criterion. Although several unsupervised methods have been proposed for setting the hyper-parameters, this is still a vivid research subject and no methods is at the same time robust and easy to apply. When there are few hyper-parameters, visual assessment of the result is often sufficient to correctly set these parameters. For instance, Figure 7 shows the effects of tuning the level of regularization μ. Compared to the optimal setting (central panel in Fig. 7), if the weight of the regularization is too small, many artifacts due to the voids in the (u, v) coverage contaminate the image (left panel in Fig. 7). On the contrary, if the weight of the regularization is too important, the image becomes too flat (right panel in Fig. 7). Although this depends on the particular regularization implemented.

8 Optimization strategy

We have seen that image reconstruction amounts to solving:

$$\min_{\boldsymbol{x}\in\mathbb{X}} \underbrace{\{\mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\mathbf{H}\cdot\boldsymbol{x})\}}_{f(\boldsymbol{x})} \cdot \tag{8.1}$$

In the case of optical interferometric data, this constrained optimization problem depends on a very large number of parameters (the image pixels), is highly non-linear[7] and multi-modal (has multiple minima). Solving such a problem requires *global optimization* or a good starting point followed by continuous optimization. It is remarkable that existing image reconstruction algorithms implement not only different priors but also different strategies to search the solution.

CLEAN (Högbom 1974) was initially developed for radio-interferometry (*i.e.* for complex visibility data) and exploits a matching pursuit algorithm to iteratively build the image by modifying a single pixel at every iteration. The *building-blocks* method (Hofmann & Weigelt 1993) is an adaptation of the CLEAN algorithm to deal with bispectrum data. The assumption made by these two methods is that the object of interest mainly consists in point-like sources. Using the regularization given by Equation (7.11) with $p = 1$ (*i.e.* taking the ℓ_1 norm of the pixels as the prior penalty) yields a similar result and produces a spatially sparse solution. Introducing such a continuous regularization, although not smooth, gives the opportunity to use optimization strategies much more efficient than matching pursuit algorithms (Thiébaut *et al.* 2012).

Wisard (Meimon *et al.* 2005b) implements a kind of self-calibration strategy alternating between (i) estimating the missing Fourier phases given the object and the phase closures to complete the data and produce pseudo-complex visibility data, and (ii) image reconstruction given these pseudo-data and the priors.

MACIM (Markov Chain Imager, Ireland *et al.* 2008) generates a stochastic sampling of the posterior probability

$$\Pr(\boldsymbol{x}|\boldsymbol{z}) \propto \Pr(\boldsymbol{z}|\boldsymbol{x}) \; \Pr(\boldsymbol{x})$$

by means of a Monte-Carlo Markov Chain (MCMC) algorithm. The image samples can then be used to find the mode of the distribution (which gives the most likely solution) or to compute the posterior mean of the sought image (which gives the image which minimizes the mean quadratic error). For large size problems, MCMC may however take prohibitive computational time to generate good samples of the posterior distribution.

Wipe (Lannes *et al.* 1997), BSMEM (Baron & Young 2008; Buscher 1994) and MiRA (Thiébaut 2008) directly minimize the penalty in Equation (8.1) by means of non-linear conjugate gradient algorithm, sub-space method (Skilling and Bryan 1984) or quasi-Newton methods (Nocedal & Wright 2006). These optimization algorithms can deal with non-linear penalties with very large number of parameters and, possibly, with constraints such as non-negativity. A change of variables can be introduced to implement the normalization constraint (le Besnerais *et al.* 2008). To my knowledge, Wipe can only cope with complex visibility data and has not been adapted to deal with optical interferometry data.

In an attempt to unify direct optimization and self-calibration approaches to solve the image reconstruction problem (8.1), we describe next another

[7] Which means that the joint criterion $f(\boldsymbol{x})$ is *non-quadratic.*

optimization strategy that can be adapted to any type of data and priors. The method follows the Alternating Direction Method of Multipliers (ADMM, Gabay & Mercier 1976) and consists in alternatively minimizing the two terms $f_{\mathsf{prior}}(\boldsymbol{x})$ and $f_{\mathsf{data}}(\boldsymbol{y})$ subject to the constraint $\boldsymbol{y} = \mathbf{H} \cdot \boldsymbol{x}$.

8.1 Augmented Lagrangian

Solving the image reconstruction problem (8.1) by *direct minimization* is exactly the same as solving the *constrained problem*:

$$\min_{\boldsymbol{x} \in \mathbb{X}, \boldsymbol{y}} \{\mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y})\} \quad \text{s.t.} \quad \mathbf{H} \cdot \boldsymbol{x} = \boldsymbol{y} \tag{8.2}$$

where the *model complex visibilities* $\boldsymbol{y} = \mathbf{H} \cdot \boldsymbol{x}$ have been explicitly introduced as *auxiliary variables.* This will allow us to treat separately the specificity of $f_{\mathsf{prior}}(\boldsymbol{x})$ and $f_{\mathsf{data}}(\boldsymbol{y})$, in particular their non linearity or lack of smoothness.

A standard approach to solve the constrained problem (8.2) is to use the Lagrangian of the problem:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}) = \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) + \boldsymbol{u}^{\top} \cdot (\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y}),$$

with $\boldsymbol{u}$ the Lagrange multipliers associated to the constraints $\mathbf{H} \cdot \boldsymbol{x} = \boldsymbol{y}$. For a solution $\{\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star}, \boldsymbol{u}^{\star}\}$ of the problem, the necessary conditions of optimality, the so-called Karush-Kuhn-Tucker (KKT) conditions, write:

$$\mathbf{H} \cdot \boldsymbol{x}^{\star} = \boldsymbol{y}^{\star} \tag{8.3}$$

$$\mathbf{0} \in \partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star}, \boldsymbol{u}^{\star}) \tag{8.4}$$

$$\mathbf{0} \in \partial_{\boldsymbol{y}} \mathcal{L}(\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star}, \boldsymbol{u}^{\star}) \tag{8.5}$$

where ∂ denotes the subdifferential operator Boyd *et al.* (2010) which only contains the gradient of its argument if it is differentiable. For instance, if the Lagrangian is differentiable with respect to variables $\boldsymbol{x}$, the second KKT condition in Equation (8.4) becomes:

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^{\star}, \boldsymbol{y}^{\star}, \boldsymbol{u}^{\star}) = \mathbf{0}.$$

Using the Lagrangian involves searching the optimal multipliers $\boldsymbol{u}^{\star}$ such that minimizing the Lagrangian with respect to the variables $(\boldsymbol{x}, \boldsymbol{y})$ given the multipliers yields a solution matching the constraints. However, finding the optimal multipliers requires to solve a system of M (the number of observed baselines) non-linear equations which is much more involved than finding a single root as required by the constrained problem in Section 5.2.

Unless a closed form solution exists, it is easier to solve the constrained problem (8.2) by using the *augmented Lagrangian* (Hestenes 1969; Powell 1969):

$$\mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \rho) = \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}) + (\rho/2)\, \|\mathbf{H} \cdot \boldsymbol{x} - \boldsymbol{y}\|_2^2 \tag{8.6}$$

with $\rho > 0$ the weight of the augmented penalty to reinforce the constraints. Obviously for any variables matching the constraints, *i.e.* such that $\mathbf{H} \cdot \boldsymbol{x} = \boldsymbol{y}$, the Lagrangian and the augmented Lagrangian are equal; thus they both yield the same solution. Solving the constrained problem (8.2) via the augmented Lagrangian however has a number of practical advantages compared to using the Lagrangian: (i) it provides an explicit update formula for the multipliers (see Eq. (8.7) in Algorithm 1), (ii) it owns strong convergence properties for ρ large enough even for non-smooth penalties, (iii) it can be exploited to derive a simple yet efficient algorithm based on alternate minimization (see Algorithm 2).

Solving the image reconstruction problem (8.2) via the augmented Lagrangian and simply considering the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ as a single group of variables yields the following algorithm:

Algorithm 1: *Augmented Lagrangian algorithm for solving (8.2).* Choose initial multipliers $\boldsymbol{u}_0$. Then, for $k = 0, 1, \ldots$, repeat the following steps until convergence:

1. Choose augmented penalty parameter $\rho_k > 0$ and improve the variables:

$$\left\{\boldsymbol{x}_{k+1}, \boldsymbol{y}_{k+1}\right\} \approx \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X},\boldsymbol{y}} \mathcal{L}_{\mathsf{A}}\left(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}_k; \rho_k\right).$$

2. Update the multipliers:

$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \rho_k \left(\mathbf{H}\cdot\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\right). \blacksquare \tag{8.7}$$

8.2 Alternating direction method of multipliers

Algorithm 1 involves minimizing the likelihood and the regularization at the same time which has not much practical interest compared to directly minimizing Equation (8.1) with respect to $\boldsymbol{x}$. The minimization becomes easier if one considers the penalties $f_{\mathsf{prior}}(\boldsymbol{x})$ and $f_{\mathsf{data}}(\boldsymbol{y})$ separately. To that end, Step 1 of Algorithm 1 can be implemented thanks to alternating minimization, for instance:

$$\boldsymbol{x}_{k+1} = \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{u}_k; \rho_k),$$

followed by

$$\boldsymbol{y}_{k+1} = \operatorname*{arg\,min}_{\boldsymbol{y}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{u}_k; \rho_k).$$

This imposes to choose an initial value $\boldsymbol{y}_0$ for the auxiliary variables $\boldsymbol{y}$. If an initial image $\boldsymbol{x}_0$ is available, the order of updating $\boldsymbol{x}$ and $\boldsymbol{y}$ can be exchanged. Alternating minimization yields the following algorithm:

Algorithm 2: *Alternate Direction Method of Multipliers (ADMM) algorithm for solving (8.2).* Choose initial multipliers $\boldsymbol{u}_0$ and initial complex visibilities $\boldsymbol{y}_0$. Then, for $k = 0, 1, \ldots$, repeat the following steps until convergence:

1. **Image Reconstruction Step.** Choose the augmented penalty parameter $\rho_k > 0$ and approximately find the best image given the complex visibilities and the Lagrange multipliers:
$$\boldsymbol{x}_{k+1} \approx \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} \mathcal{L}_{\mathsf{A}}\left(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{u}_k; \rho_k\right).$$
2. **Self Calibration Step.** Approximately find the best complex visibilities given the image and the Lagrange multipliers:
$$\boldsymbol{y}_{k+1} \approx \operatorname*{arg\,min}_{\boldsymbol{y}} \mathcal{L}_{\mathsf{A}}\left(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{u}_k; \rho_k\right).$$
3. **Updating of the Lagrange Multipliers.** Apply the following formula to update the multipliers:
$$\boldsymbol{u}_{k+1} = \boldsymbol{u}_k + \rho_k \left(\mathbf{H}\cdot\boldsymbol{x}_{k+1} - \boldsymbol{y}_{k+1}\right). \blacksquare \tag{8.8}$$

Before going into the details of the algorithm, let us remark that by elementary manipulations, the augmented Lagrangian can be rewritten as:

$$\begin{aligned}\mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{u}; \rho) &= \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) + \boldsymbol{u}^{\top}\cdot(\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y}) + \frac{\rho}{2}\,\|\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y}\|_2^2 \\ &= \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + f_{\mathsf{data}}(\boldsymbol{y}) + \frac{\rho}{2}\,\|\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y} + \boldsymbol{u}/\rho\|_2^2 - \frac{1}{2\,\rho}\,\|\boldsymbol{u}\|_2^2. \end{aligned} \tag{8.9}$$

8.2.1 Image reconstruction step

Discarding in Equation (8.9) terms which do not depend on the variables $\boldsymbol{x}$, Step 1 of Algorithm 2 consists in improving $\boldsymbol{x}$ given the other variables and writes:

$$\begin{aligned}\boldsymbol{x}_{k+1} &= \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}, \boldsymbol{y}_k, \boldsymbol{u}_k; \rho_k) \\ &= \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} \mu\, f_{\mathsf{prior}}(\boldsymbol{x}) + (\rho_k/2)\,\|\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{y}_k + \boldsymbol{u}_k/\rho_k\|_2^2 \\ &= \operatorname*{arg\,min}_{\boldsymbol{x}\in\mathbb{X}} (\mu/\rho_k)\, f_{\mathsf{prior}}(\boldsymbol{x}) + (1/2)\,\|\mathbf{H}\cdot\boldsymbol{x} - \boldsymbol{v}_k\|_2^2 \end{aligned} \tag{8.10}$$

$$\text{with:}\quad \boldsymbol{v}_k = \boldsymbol{y}_k - \boldsymbol{u}_k/\rho_k\,, \tag{8.11}$$

which is the analogous of *image reconstruction* given *pseudo-complex visibilities* $\boldsymbol{v}_k = \boldsymbol{y}_k - \boldsymbol{u}_k/\rho_k$ with i.i.d. Gaussian noise of variance $\propto \mu/\rho_k$. Note that, if the feasible set is just $\mathbb{R}^N$, the right hand side of Equation (8.10) is the value returned by the proximity operator[8] of $(\mu/\rho_k)\, f_{\mathsf{prior}}$ at $\boldsymbol{v}_k$ (Combettes & Pesquet 2011).

[8] The proximity operator of $f\colon \mathbb{R}^N \mapsto \mathbb{R}$ is defined by:
$$\operatorname{prox}_f(\boldsymbol{v}) = \operatorname*{arg\,min}_{\boldsymbol{x}} \left\{ f(\boldsymbol{x}) + (1/2)\,\|\boldsymbol{x} - \boldsymbol{v}\|_2^2 \right\}.$$

Depending on the particular regularization $f_{\mathsf{prior}}(\boldsymbol{x})$, a specific algorithm may be designed to efficiently solve this problem. If the regularization is quadratic, Equation (8.10) is a large scale quadratic problem which can be solved by existing methods like the *gradient projection conjugate gradient* algorithm (GPCG by Moré & Toraldo 1991). Otherwise, for a number of non smooth $f_{\mathsf{prior}}(\boldsymbol{x})$, there exist closed form solutions of Equation (8.10) with $\mathbb{X} = \mathbb{R}^N$ (Combettes & Pesquet 2011) which can be adapted to account for non negativity constraint (Thiébaut *et al.* 2012).

8.2.2 Updating the complex visibilities

Discarding in Equation (8.9) terms which do not depend on the auxiliary variables $\boldsymbol{y}$, Step 2 of Algorithm 2 consists in improving $\boldsymbol{y}$ given the other variables and writes:

$$\begin{aligned}
\boldsymbol{y}_{k+1} &= \arg\min_{\boldsymbol{y}} \mathcal{L}_{\mathsf{A}}(\boldsymbol{x}_{k+1}, \boldsymbol{y}, \boldsymbol{u}_k; \rho_k) \\
&= \arg\min_{\boldsymbol{y}} f_{\mathsf{data}}(\boldsymbol{y}) + (\rho_k/2)\,\|\mathbf{H}\cdot\boldsymbol{x}_{k+1} - \boldsymbol{y} + \boldsymbol{u}_k/\rho_k\|_2^2 \\
&= \arg\min_{\boldsymbol{y}} f_{\mathsf{data}}(\boldsymbol{y}) + (\rho_k/2)\,\|\boldsymbol{y} - \boldsymbol{w}_k\|_2^2 \qquad (8.12)\\
\text{with:}\quad \boldsymbol{w}_k &= \mathbf{H}\cdot\boldsymbol{x}_{k+1} + \boldsymbol{u}_k/\rho_k \qquad (8.13)
\end{aligned}$$

which enforces the complex visibilities $\boldsymbol{y}$ to be a compromise between the actual data and the *shifted* model complex visibilities $\boldsymbol{w}_k = \mathbf{H}\cdot\boldsymbol{x}_{k+1} + \boldsymbol{u}_k/\rho_k$. If there are missing data (for instance, incomplete Fourier phases when working with the bispectrum or the phase closures and the powerspectrum), this step is nevertheless a well posed problem thanks to the augmented term $(\rho_k/2)\,\|\boldsymbol{y} - \boldsymbol{w}_k\|_2^2$.

8.3 Conclusions about optimization strategy

Steps 1 and 2 of Algorithm 2 are the analogous of the image reconstruction and self-calibration steps in self-calibration methods (Cornwell & Wilkinson 1981; Meimon *et al.* 2005b; Schwab 1980). However, to really mimic these latter methods, these steps should be carried out in Algorithm 2 with the Lagrange multipliers always equal to zero. Formally, this means that standard self-calibration methods do not consistently solve a well defined optimization problem. This is not the case of the proposed approach where the self-calibration step accounts for the Lagrange multipliers which are associated to the constraints that $\mathbf{H}\cdot\boldsymbol{x} = \boldsymbol{y}$.

Although global optimization is in principle required to solve Equation (8.1), the most successful algorithms proposed for optical interferometry BSMEM (Baron & Young 2008) and MiRA (Thiébaut 2008) use direct optimization. They however implement numerical optimization algorithms designed for smooth penalties[9].

[9] *Smooth* means here twice continuously differentiable.

Thanks to the variable splitting trick, Algorithm 2 handles separately the specificities of $f_{\mathsf{prior}}(\boldsymbol{x})$ and $f_{\mathsf{data}}(\boldsymbol{y})$. As a consequence, it can efficiently cope with non-smooth penalties such as the ones used to impose sparsity. Moreover, the augmented penalty term introduces a simple quadratic term which regularizes the minimization of $f_{\mathsf{prior}}(\boldsymbol{x})$ and that of $f_{\mathsf{data}}(\boldsymbol{y})$. This makes theses sub-optimization problems well posed and may speed up their numerical solving.

9 Summary and perspectives

After describing the type of measurements which can be acquired with an interferometer and the specific issues due to the turbulence. We addressed the inverse problem of synthesizing an image from these data. The inverse approach provided us a useful framework to derive a kind of recipe for image reconstruction. This recipe involves:

1. A **direct model** of the observables $\boldsymbol{z}$ given the image parameters $\boldsymbol{x}$. This model implements an approximation of the brightness distribution $I_\lambda(\boldsymbol{\theta})$ and its Fourier transform $\widehat{I}_\lambda(\boldsymbol{\nu})$ from which is derived the linear relationship $\boldsymbol{y} = \mathbf{H} \cdot \boldsymbol{x}$ between the sampled complex visibilities $y_m = \widehat{I}_\lambda(\boldsymbol{\nu}_m)$ and the image parameters.

2. A **criterion** to be minimized to determine a unique and stable solution. This criterion takes the form $f(\boldsymbol{x}) = f_{\mathsf{data}}(\mathbf{H} \cdot \boldsymbol{x}) + \mu\, f_{\mathsf{prior}}(\boldsymbol{x})$ and reflects the compromise between fidelity to the data, *i.e.* minimizing $f_{\mathsf{data}}(\mathbf{H} \cdot \boldsymbol{x})$, and to the priors, *i.e.* minimizing $f_{\mathsf{prior}}(\boldsymbol{x})$. The hyper-parameter $\mu > 0$ is used to tune this trade-off. Eventually, a feasible set $\mathbb{X}$ can be introduced to account for strict constraints such as non negativity or normalization of the solution.

3. An **optimization strategy** to solve the constrained optimization problem.

The same general framework can been used to describe most (if not all) interferometric image reconstruction algorithms (le Besnerais *et al.* 2008; Thiébaut & Giovannelli 2010; Thiébaut 2009) so the issues encountered while cooking the recipe are also general and have their counterparts in all proposed methods.

In this short presentation, we mainly focused on the so-called *analysis approach* to reconstruct a non-parametric model of the brightness distribution. An alternative, the *synthesis approach*, is to describe the image as the combination of a number of elementary atoms (Elad *et al.* 2007). In the synthesis approach, the regularization is achieved by imposing to use the smallest number of atoms to explain the data. As described in our presentation, this sparsity constraint may be introduced via an ℓ_1 norm penalty and the problem solved by specific algorithms to cope with continuous but non-smooth criteria. It is also possible to try to mimic the effects of using an ℓ_0 norm penalty with *greedy algorithms*. The CLEAN algorithm (Högbom 1974) mentioned in Section 8 can be seen as a precursor of the synthesis approach where the atoms have all the same shape (they are point-like sources) which are only allowed to have different brightnesses and positions.

The ADMM strategy implemented by Algorithm 2 was introduced for pedagogical proposes to make a link between constrained optimization and self-calibration methods and to exhibit some of the issues of solving the optimization part of the image restoration problem. We have argued that the proposed strategy is more consistent than existing self-calibration methods and more flexible than using algorithms restricted to smooth penalties. Introducing variables splitting and ADMM strategy was also motivated by the effectiveness of a similar approach for multi-spectral interferometric data. In this case, the reconstruction algorithm was designed to deal with complex visibilities and exploits structured sparsity regularization to favor point-like sources in the image (Thiébaut *et al.* 2012). To deal with current optical interferometry data, it remains to demonstrate whether such an approach has the ability to find a path to a good solution at a lower cost than a stochastic global optimization method like MACIM (Ireland *et al.* 2008).

As mentioned along this presentation, optimization is not the only direction of research to improve interferometric imaging. Perhaps first of all, multi-spectral image reconstruction is now required to fully exploit the spectral resolution of the existing interferometers. Indeed, it has been clearly demonstrated that spatio-spectral regularization drastically improves the quality of the restored images (Soulez *et al.* 2008). Hence existing algorithms must be extensively modified to globally account for multi-variate data and not just reused to perform independent reconstructions at given wavelengths (le Bouquin *et al.* 2009). In spite of its unrivaled angular resolution, stellar interferometry is not as popular as, say adaptive optics, in the astronomical community. This is partially due to the difficulty to interpret the interferometric data. Making state of the art image reconstruction algorithms available to non-specialists may be a good way to promote interferometric observations. To that end, the methods must be not only robust but also relatively easy to use. Developing unsupervised methods to automatically tune the hyper-parameters of image reconstruction algorithms is therefore of particular interest.

References

Baron, F., & Young, J.S., 2008, Image reconstruction at cambridge university. In Society of Photo-Optical Instrumentation Engineers (SPIE) Conf. Ser., Vol. 7013, 70133X, `DOI: 10.1117/12.789115`

Boyd, S., Parikh, N., Chu, E., Peleato, Bo., & Eckstein, J., 2010, Distributed optimization and statistical learning via the alternating direction method of multipliers, Foundations and Trends in Machine Learning, 3, 1, `DOI: 10.1561/2200000016`, `http://www.stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf`

Buscher, D.F., 1994, Direct maximum-entropy image reconstruction from the bispectrum, ed. J.G. Robertson & W.J. Tango, IAU Symp. 158: Very High Angular Resolution Imaging, p. 91

Campisi, P., & Egiazarian, K., 2007, Blind image deconvolution: theory and applications (CRC Press), ISBN 9780849373671

Chambolle, A., Levine, S.E., & Lucier, B.J., 2011, SIAM J. Imaging Sciences, 4, 277

Charbonnier, P., Blanc-Féraud, L., Aubert, G., & Barlaud, M., 1997, IEEE Trans. Image Process., 6, 298

Combettes, P.L., & Pesquet, J.-C., 2011, Proximal splitting methods in signal processing, chapter Fixed-Point Algorithms for Inverse Problems in Science and Engineering (Springer, New York), 185

Cornwell, T., 1995, Imaging concepts, ed. J.A. Zensus, P.J. Diamond & P.J. Napier, ASP Conf. Ser. 82, 39

Cornwell, T.J., & Wilkinson, P.N., 1981, MNRAS, 196, 1067

Dainty, J.C., & Greenaway, A.H., 1979, J. Opt. Soc. Am., 69, 786

Delplancke, F., Derie, F., Paresce, F., *et al.*, 2003, Ap&SS, 286, 99,

Donoho, D., 2006, Comm. Pure Appl. Math., 59, 907

Elad, M., Milanfar, P., & Rubinstein, R., 2007, Inverse Probl., 23, 947

Fessler, J.A., & Sutton, B.P., 2003, IEEE Trans. Signal Process., 51, 560

Gabay, D., & Mercier, B., 1976, Comput. Math. Applications, 2, 17

Goodman, J.W., 1985, Statistical Optics (John Wiley & Sons), ISBN 0-471-01502-4

Haniff, C., 1991, J. Opt. Soc. Am. A, 8, 134

Hestenes, M.R., 1969, J. Optimiz. Theory Applications, 4, 303

Hofmann, K.-H., & Weigelt, G., 1993, A&A, 278, 328

Horne, K., 1985, MNRAS, 213, 129

Högbom, J.A., 1974, A&AS, 15, 417

Ireland, M.J., Monnier, J., & Thureau, N., 2008, Monte-Carlo imaging for optical interferometry, ed. J.D. Monnier, M. Schöller & W.C. Danchi, Advances in Stellar Interferometry, Vol. 6268, p. 62681T1, SPIE, `DOI: 10.1117/12.670940`

Lacour, S., Meimon, S., Thiébaut, É., *et al.*, 2008, A&A, 485, 561

Lannes, A., Anterrieu, E., & Maréchal, P., 1997, A&AS, 123, 183

Lannes, A., 2001, J. Opt. Soc. Am. A, 18, 1046

Lawson, P.R., Cotton, W.D., Hummel, C.A., *et al.*, 2004, BAAS, 36, 1605

le Besnerais, G., Lacour, S., Mugnier, L.M., *et al.*, 2008, IEEE J. Selected Topics Signal Process., 2, 767

le Bouquin, J.-B., Lacour, S., Renard, S., *et al.*, 2009, A&A, 496, L1

Meimon, S., Mugnier, L.M., & le Besnerais, G., 2005a, J. Opt. Soc. Am. A, 22, 2348

Meimon, S., Mugnier, L.M., & le Besnerais, G., 2005b, Opt. Lett., 30, 1809

Moré, J., & Toraldo, G., 1991, SIAM J. Optim., 1, 93, `http://locus.siam.org/SIOPT/volume-01/art_0801008.html`

Narayan, R., & Nityananda, R., 1986, ARA&A, 24, 127

Nocedal, J., & Wright, S.J., 2006, Numerical Optimization, 2nd edition (Springer Verlag), `http://www.zla-ryba.cz/NumOpt.pdf`

Pauls, T.A., Young, J.S., Cotton, W.D., & Monnier, J.D., 2005, PASP, 117, 1255

Petrov, R.G., Malbet, F., *et al.*, 2007, A&A, 464, 1

Potts, D., Steidl, G., & Tasche, M., 2001, Modern Sampling Theory: Mathematics and Applications, chapter Fast Fourier transforms for nonequispaced data: A tutorial (Birkhauser, Boston), 249

Powell, M.J.D., 1969, Optimization, chapter A method for nonlinear constraints in minimization problems (Academic Press), 283

Renard, S., Thiébaut, É., & Malbet, F., 2011, A&A, 533, A64

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

Rudin, L.I., Osher, S., & Fatemi, E., 1992, Physica D, 60, 259

Schwab, F., 1980, Proc. SPIE, 231, 18

Skilling, J., & Bryan, R.K., 1984, MNRAS, 211, 111

Soulez, F., Thiébaut, É., Gressard, A., Dauphin, R., & Bongard, S., 2008, Heterogeneous multidimensional data deblurring, In 16th European Signal Processing Conference (EUSIPCO), Lausanne, Suisse, http://hal-ujm.ccsd.cnrs.fr/ujm-00293660/en/

Sramek, R., & Schwab, F., 1989, Imaging, ed. Richard A. Perley, Frederic R., Schwab & Alan H. Bridle, Synthesis Imaging in Radio Astronomy, Vol. 6, 117

Thiébaut, É., & Giovannelli, J.-F., 2010, IEEE Signal Process. Mag., 27, 97

Thiébaut, É., 2008, MiRA: an effective imaging algorithm for optical interferometry, ed. Françoise Delplancke Markus Schöller, William C. Danchi. Astronomical Telescopes and Instrumentation, Vol. 7013, 70131I–1, SPIE

Thiébaut, É., 2009, New Astron. Rev., 53, 312

Thiébaut, É., Soulez, F., & Denis, L., 2012, accepted for publication in J. Opt. Soc. Am. A, http://arxiv.org/abs/1209.2362

Thompson, A.R., & Bracewell, R.N., 1974, AJ, 79, 11

Thévenaz, P., Blu, T., & Unser, M., 2000, IEEE Trans. Medical Imag., 19, 739, http://bigwww.epfl.ch/publications/thevenaz0002.html

Tikhonov, A.N., & Arsenin, V.Y., 1977, Solution of Ill-posed Problems, Scripta Series in Mathematics (Winston & Sons, Washington), ISBN 0-470-99124-0

Wirnitzer, B., 1985, J. Opt. Soc. Am. A, 2, 14