INTRODUCTION TO IMAGE RECONSTRUCTION AND INVERSE PROBLEMS

Eric Thiébaut
Observatoire de Lyon (France)
thiebaut@obs.univ-lyon1.fr

Abstract   Either because observed images are blurred by the instrument and transfer medium or because the collected data (e.g. in radio astronomy) are not in the form of an image, image reconstruction is a key problem in observational astronomy. Understanding the fundamental problems underlying the deconvolution (noise amplification) and the way to solve for them (regularization) is the prototype to cope with other kind of inverse problems.

Keywords: image reconstruction, deconvolution, inverse problems, regularization.

1.1   Image formation

1.1.1   Standard image formation equation

For an incoherent source, the observed image $y(\mathbf{s})$ in a direction $\mathbf{s}$ is given by the standard image formation equation:

$$y(\mathbf{s}) = \int h(\mathbf{s}|\mathbf{s}')\, x(\mathbf{s}')\, \mathrm{d}\mathbf{s}' + n(\mathbf{s}) \tag{1}$$

where $x(\mathbf{s}')$ is the object brightness distribution, $h(\mathbf{s}|\mathbf{s}')$ is the instrumental point spread function (PSF), and $n(\mathbf{s})$ accounts for the noise (source and detector). The PSF $h(\mathbf{s}|\mathbf{s}')$ is the observed brightness distribution in the direction $\mathbf{s}$ for a point source located in the direction $\mathbf{s}'$.

Figure 1 shows the simulation of a galaxy observed with a PSF which is typical of an adaptive optics system. The noisy blurred image in Fig. 1 will be used to compare various image reconstruction methods described in this course.
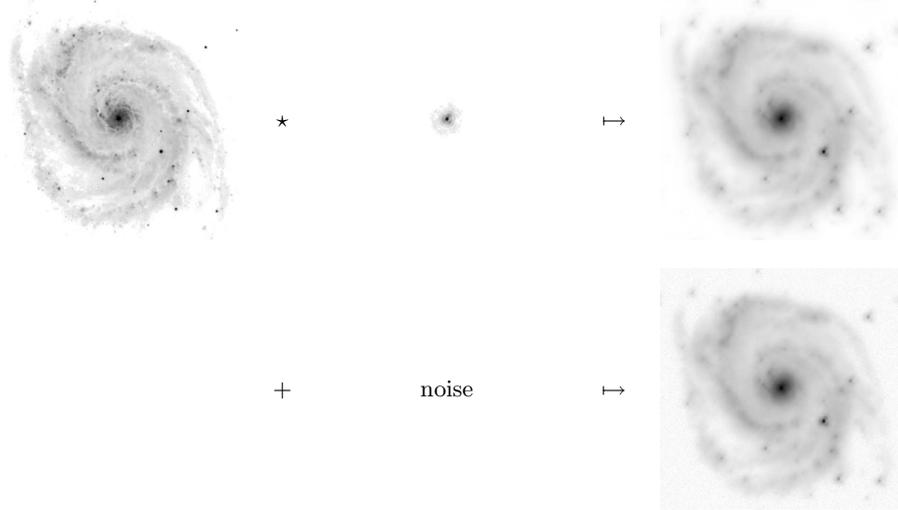
Figure 1. Simulation of the image of a galaxy. From top-left to bottom-right: true object brightness distribution, PSF (instrument + transfer medium), noiseless blurred image, noisy blurred image. The PSF image was kindly provided by F. and C. Roddier.

1.1.2 Discrete form of image formation equation

For discretized data (e.g. pixels), Eq. (1) can be written in matrix form:

$$\mathbf{y} = \mathbf{H} \cdot \mathbf{x} + \mathbf{n} \qquad (2)$$

where $\mathbf{H}$ is the response matrix, $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{n}$ are the data, object brightness distribution and noise vectors. These vectors have the same layout; for instance, the data vector corresponding to a $N{\times}N$ image writes:

$$\mathbf{y} = [y(1,1), y(2,1), \dots, y(N,1), y(1,2), \dots, y(N,2), \dots, y(N,N)]^{\mathrm{T}}.$$

1.1.3 Shift invariant PSF

Within the isoplanatic patch of the instrument and transfer medium, the PSF can be assumed to be shift invariant:

$$h(\mathbf{s}|\mathbf{s}') = h(\mathbf{s} - \mathbf{s}'), \qquad (3)$$

for $\|\mathbf{s} - \mathbf{s}'\|$ small enough. In this case, the image formation equation involves a convolution product:

$$y(\mathbf{s}) = \int h(\mathbf{s} - \mathbf{s}')\, x(\mathbf{s}')\, \mathrm{d}\mathbf{s}' + n(\mathbf{s}) \qquad (4\text{a})$$

$$\xrightarrow{\text{FT}} \quad \hat{y}(\mathbf{u}) = \hat{h}(\mathbf{u})\, \hat{x}(\mathbf{u}) + \hat{n}(\mathbf{u}) \qquad (4\text{b})$$

where the hats denote Fourier transformed distributions and $\mathbf{u}$ is the spatial frequency. The Fourier transform $\hat{h}(\mathbf{u})$ of the PSF is called the modulation transfer function (MTF).

In the discrete case, the convolution by the PSF is diagonalized by using the discrete Fourier transform (DFT):

$$\hat{y}_u = \hat{h}_u\,\hat{x}_u + \hat{n}_u \tag{5}$$

where index u means u-th spatial frequency $\mathbf{u}$ of the discrete Fourier transformed array.

1.1.4 Discrete Fourier transform

The 1-D discrete Fourier transform (DFT) of a vector $\mathbf{x}$ is defined by:

$$\hat{x}_u = \sum_{k=0}^{N-1} x_k\,\mathrm{e}^{-2\,\mathrm{i}\,\pi\,u\,k/N} \quad \overset{\mathrm{FT}}{\longleftarrow} \quad x_k = \frac{1}{N}\sum_{u=0}^{N-1} \hat{x}_u\,\mathrm{e}^{+2\,\mathrm{i}\,\pi\,u\,k/N}$$

where $\mathrm{i} \equiv \sqrt{-1}$ and N is the number of elements in $\mathbf{x}$. The $N \times N$ discrete Fourier transform (2-D DFT) of $\mathbf{x}$ writes:

$$\hat{x}_{u,v} = \sum_{k,l} x_{k,l}\,\mathrm{e}^{-2\,\mathrm{i}\,\pi\,(u\,k+v\,l)/N} \quad \overset{\mathrm{FT}}{\longleftarrow} \quad x_{k,l} = \frac{1}{N^2}\sum_{u,v} \hat{x}_u\,\mathrm{e}^{+2\,\mathrm{i}\,\pi\,(u\,k+v\,l)/N}$$

where N is the number of elements along each dimension. Using matrix notation:

$$\hat{\mathbf{x}} = \mathbf{F}\cdot\mathbf{x} \quad \overset{\mathrm{FT}}{\longleftarrow} \quad \mathbf{x} = \mathbf{F}^{-1}\cdot\hat{\mathbf{x}} = \frac{1}{N_{\mathrm{pixel}}}\mathbf{F}^{\mathrm{H}}\cdot\hat{\mathbf{x}}$$

where N_{pixel} is the number of elements in $\mathbf{x}$, and $\mathbf{F}^{\mathrm{H}}$ is the conjugate transpose of $\mathbf{F}$ given by (1-D case):

$$F_{u,k} \equiv \exp(-2\,\mathrm{i}\,\pi\,u\,k/N)\,.$$

In practice, DFT's can be efficiently computed by means of fast Fourier transforms (FFT's).

1.2 Direct inversion

Considering the diagonalized form (5) of the image formation equation, a very tempting solution is to perform straightforward direct inversion in the Fourier space and then Fourier transform back to get the

deconvolved image. In other words:

$$\mathbf{x}^{(\text{direct})} = \text{FFT}^{-1}\left(\frac{\text{FFT}\left(\begin{array}{c}\end{array}\right)}{\text{FFT}\left(\begin{array}{c}\end{array}\right)}\right) =$$
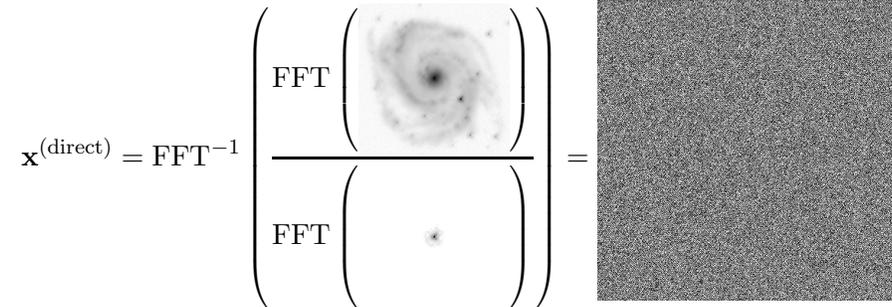
The result is rather disappointing: the deconvolved image is even *worse* than the observed one! It is easy to understand what happens if we consider the Fourier transform of the direct solution:

$$\hat{x}_u^{(\text{direct})} = \hat{y}_u/\hat{h}_u = \hat{x}_u + \hat{n}_u/\hat{h}_u \tag{6}$$

which shows that, in the direct inversion, the perfect (but unknown) solution $\hat{x}_u$ get corrupted by a term $\hat{n}_u/\hat{h}_u$ due to noise. In this latter term, division by small values of the MTF $\hat{h}_u$, for instance at high frequencies where the noise dominates the signal (see Fig. 2a), yields very large distorsions. Such noise amplification produces the high frequency artifacts displayed by the direct solution.

Instrumental transmission (convolution by the PSF) is always a smoothing process whereas noise is usually non-negligible at high frequencies, the noise amplification problem therefore *always* arises in deconvolution. This is termed as *ill-conditioning* in inverse problem theory.

1.3   Truncated frequencies

A crude way to eliminate noise amplification is to use a suitable cutoff frequency u_{cutoff} and to write the solution as:

$$\hat{x}_u^{(\text{cutoff})} = \begin{cases} \dfrac{\hat{y}_u}{\hat{h}_u} & \text{for } |u| < u_{\text{cutoff}} \\[2ex] 0 & \text{for } |u| \geq u_{\text{cutoff}} \end{cases} \tag{7}$$

In our example, taking $u_{\text{cutoff}} = 80$ frequels guarantees that the noise remains well below the signal level (see Fig. 2a); the resulting image is shown in Fig. 3a. The improvement in resolution and quality is clear but the stiff truncature of frequencies is responsible of ripples which can be seen around bright sources in Fig. 3a. A somewhat smoother effect can be obtained by using other kind of inverse filters like the Wiener filter described in the next section.
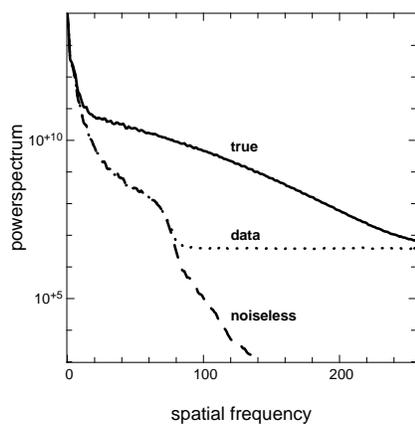
Figure 2a. Profiles of the power-spectra of the true brightness distribution, the noiseless blurred image, and the actual data (noisy and blurred image). Clearly the noise dominates after frequency $\simeq 80$ frequels.
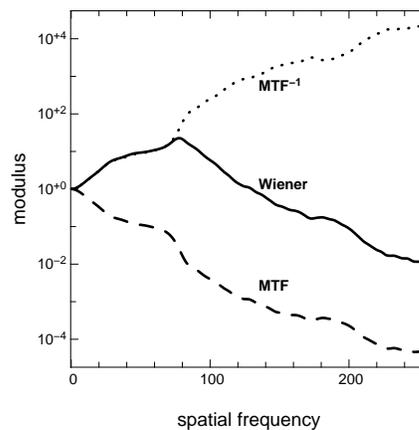
Figure 2b. Profiles of the modulation transfer function (MTF), its inverse and Wiener inverse-filter.
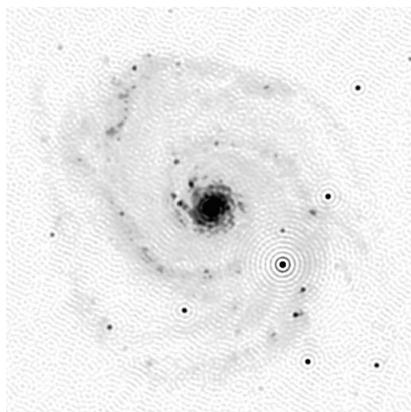
Figure 3a. Restored image using a simple cutoff frequency ($u_{\text{cutoff}} = 80$) in the deconvolution.
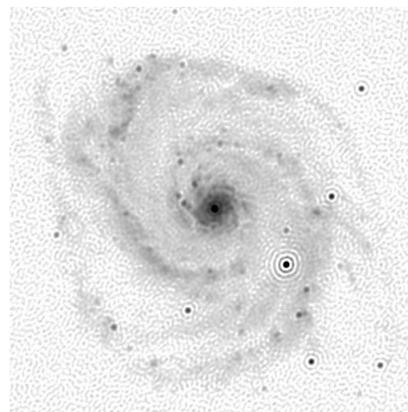
Figure 3b. Image obtained by using Wiener inverse-filter.

1.4   Wiener inverse-filter

The Wiener inverse-filter is derived from the following two criteria:

- The solution is given by applying a linear filter $\mathbf{f}$ to the data and the Fourier transform of the solution writes:

$$\hat{x}_u^{(\text{Wiener})} \equiv \hat{f}_u^{(\text{Wiener})} \, \hat{y}_u \,. \tag{8}$$

- The expected value of the quadratic error with respect to the true object brightness distribution must be as small as possible:

$$\mathbf{f}^{(\text{Wiener})} = \arg\min_{\mathbf{f}} \mathrm{E}\left\{ \left\| \mathbf{x}^{(\text{Wiener})} - \mathbf{x}^{(\text{true})} \right\|^2 \right\} \tag{9}$$

where $\mathrm{E}\{x\}$ is the expected value of x. For those not familiar with this notation,

$$\arg\min_{\mathbf{f}} \phi(\mathbf{f})$$

is the element $\mathbf{f}$ which minimizes the value of the expression $\phi(\mathbf{f})$.

To summarize, Wiener inverse-filter is the linear filter which insures that the result is as close as possible, on average and in the least squares sense, to the true object brightness distribution.

In order to derive the expression of the filter, we write the expected value ϵ of the quadratic error:

$$\epsilon \equiv \mathrm{E}\left\{ \left\| \mathbf{x}^{(\text{Wiener})} - \mathbf{x}^{(\text{true})} \right\|^2 \right\} = \mathrm{E}\left\{ \sum_k \left(x_k^{(\text{Wiener})} - x_k^{(\text{true})} \right)^2 \right\}$$

by Parseval's theorem (in its discrete form):

$$\epsilon = \frac{1}{N_{\text{pixel}}} \mathrm{E}\left\{ \sum_u \left| \hat{x}_u^{(\text{Wiener})} - \hat{x}_u^{(\text{true})} \right|^2 \right\} = \frac{1}{N_{\text{pixel}}} \mathrm{E}\left\{ \sum_u \left| \hat{f}_u \, \hat{y}_u - \hat{x}_u^{(\text{true})} \right|^2 \right\} \,.$$

The extremum of ϵ (actually a minimum) is reached for $\mathbf{f}$ such that:

$$\frac{\partial \epsilon}{\partial \mathrm{Re}\{\hat{f}_u\}} = 0 \quad \text{and} \quad \frac{\partial \epsilon}{\partial \mathrm{Im}\{\hat{f}_u\}} = 0, \quad \forall u \,.$$

Since the two partial derivatives of ϵ have real values, the minimum of ϵ obeys:

$$\frac{\partial \epsilon}{\partial \mathrm{Re}\{\hat{f}_u\}} + \mathrm{i}\, \frac{\partial \epsilon}{\partial \mathrm{Im}\{\hat{f}_u\}} = 0, \quad \forall u$$

$$\Longleftrightarrow \quad \mathrm{E}\left\{ \hat{y}_u^\star \left(\hat{f}_u \, \hat{y}_u - \hat{x}_u \right) \right\} = 0, \quad \forall u$$

$$\Longleftrightarrow \quad \mathrm{E}\left\{ (\hat{h}_u \, \hat{x}_u + \hat{n}_u)^\star \left(\hat{f}_u \, (\hat{h}_u \, \hat{x}_u + \hat{n}_u) - \hat{x}_u \right) \right\} = 0, \quad \forall u$$

$$\Longleftrightarrow \quad (|\hat{h}_u|^2 \, \hat{f}_u - \hat{h}_u^\star) \, \mathrm{E}\{|\hat{x}_u|^2\} + \hat{f}_u \, \mathrm{E}\{|\hat{n}_u|^2\} = 0, \quad \forall u$$

where $z^\star$ denotes the complex conjugate to z. The last expression has been obtained after some simplifications which apply when the signal and the noise are *uncorrelated* and when the noise is centered: $\mathrm{E}\{\hat{n}_u\} = 0$. The final relation can be solved to obtain the expression of the Wiener inverse-filter:

$$\hat{f}_u^{(\mathrm{Wiener})} = \frac{\hat{h}_u^\star}{\left|\hat{h}_u\right|^2 + \dfrac{\mathrm{E}\{|\hat{n}_u|^2\}}{\mathrm{E}\{|\hat{x}_u|^2\}}} \ . \tag{10}$$

Figure 2b and Eq. (10) show that the Wiener inverse-filter is close to the direct inverse-filter for frequencies of high signal-to-noise ratio (SNR), but is strongly attenuated where the SNR is poor:

$$\hat{f}_u^{(\mathrm{Wiener})} \simeq \begin{cases} 1/\hat{h}_u & \text{for } \mathrm{SNR}_u \gg 1 \\[2ex] 0 & \text{for } \mathrm{SNR}_u \ll 1 \end{cases} \quad \text{with } \mathrm{SNR}_u \equiv \sqrt{\frac{\left|\hat{h}_u\right|^2 \mathrm{E}\{|\hat{x}_u|^2\}}{\mathrm{E}\{|\hat{n}_u|^2\}}} \ .$$

The Wiener filter therefore avoids noise amplification and provides the best solution according to some *quality criterion*. We will see that these features are common to all other methods which correctly solve the deconvolution inverse problem. The result of applying Wiener inverse-filter to the simulated image is shown in Fig. 3b.

Wiener inverse-filter however yields, possibly, unphysical solution with *negative values* and *ripples* around sharp features (e.g. bright stars) as can be seen in Fig. 3b. Another drawback of Wiener inverse-filter is that spectral densities of noise and signal are usually *unknown* and must be guessed from the data. For instance, for white noise and assuming that the spectral density of object brightness distribution follows a simple parametric law, e.g. a power law, then:

$$\begin{cases} \mathrm{E}\{|\hat{n}_u|^2\} = \text{constant} \\[2ex] \mathrm{E}\{|\hat{x}_u|^2\} \propto \|u\|^{-\beta} \end{cases} \implies \hat{f}_u = \frac{\hat{h}_u^\star}{\left|\hat{h}_u\right|^2 + \alpha\,\|u\|^\beta} \ , \tag{11}$$

where $\|u\|$ is the length of the u-th spatial frequency. We are left with the problem of determining the best values for α and β; this can be done by means of Generalized Cross-Validation (GCV, see related section).

1.5 Maximum likelihood

Maximum likelihood methods are commonly used to estimate parameters from noisy data. Such methods can be applied to image restoration, possibly with additional constraints (e.g., positivity). Maximum likelihood methods are however not appropriate for solving ill-conditioned inverse problems as will be shown in this section.

1.5.1   Unconstrained maximum likelihood

The maximum likelihood (ML) solution is the one which maximizes the probability of the data $\mathbf{y}$ given the model among all possible $\mathbf{x}$:

$$\mathbf{x}^{(\mathrm{ML})} = \arg\max_{\mathbf{x}} \Pr\{\mathbf{y}|\mathbf{x}\} = \arg\min_{\mathbf{x}} \phi_{\mathrm{ML}}(\mathbf{x}) \tag{12a}$$

where

$$\phi_{\mathrm{ML}}(\mathbf{x}) \propto -\log(\Pr\{\mathbf{y}|\mathbf{x}\}) + \text{constant} \tag{12b}$$

is the likelihood penalty related to the log-likelihood up to an additive constant (and also up to a multiplicative strictly positive factor).

Unconstrained ML for Gaussian noise. In the case of Gaussian noise, the log-likelihood reads:

$$-\log(\Pr\{\mathbf{y}|\mathbf{x}\}) = \frac{1}{2}(\mathbf{H}\cdot\mathbf{x} - \mathbf{y})^{\mathrm{T}}\cdot\mathbf{W}\cdot(\mathbf{H}\cdot\mathbf{x} - \mathbf{y}) + \text{constant}$$

where $\mathbf{H}\cdot\mathbf{x}$ is the model of the data (see Eq. (2)) and the weighting matrix $\mathbf{W}$ is the inverse of the covariance matrix of the data:

$$\mathbf{W} = \mathrm{Cov}(\mathbf{y})^{-1}. \tag{13}$$

Taking $\eta = 2$ and dropping any additive constant, the likelihood penalty $\phi_{\mathrm{Gauss}}(\mathbf{x})$ writes:

$$\phi_{\mathrm{Gauss}}(\mathbf{x}) \equiv (\mathbf{H}\cdot\mathbf{x} - \mathbf{y})^{\mathrm{T}}\cdot\mathbf{W}\cdot(\mathbf{H}\cdot\mathbf{x} - \mathbf{y}). \tag{14}$$

The minimum of $\phi_{\mathrm{Gauss}}(\mathbf{x})$ obeys:

$$\left.\frac{\partial\phi_{\mathrm{Gauss}}(\mathbf{x})}{\partial x_k}\right|_{\mathbf{x}^{(\mathrm{ML})}} = 0, \quad \forall k.$$

The gradient of the penalty is:

$$\nabla\phi_{\mathrm{Gauss}}(\mathbf{x}) = 2\,\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot(\mathbf{H}\cdot\mathbf{x} - \mathbf{y}), \tag{15}$$

then $\mathbf{x}^{(\mathrm{ML})}$ solves of the so-called *normal equations*:

$$\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot(\mathbf{H}\cdot\mathbf{x}^{(\mathrm{ML})} - \mathbf{y}) = 0 \tag{16}$$

which yields a unique minimum provided that $\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\mathbf{H}$ be positive definite (by construction this matrix is positive but may not be definite). The solution of the normal equations is:

$$\mathbf{x}^{(\mathrm{ML})} = (\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\mathbf{H})^{-1}\cdot\mathbf{H}^{\mathrm{T}}\cdot\mathbf{W}\cdot\mathbf{y} \tag{17}$$

which is the well known solution of a weighted linear least squares problem[1]. However we have not yet proven that the ML solution is able to smooth out the noise. We will see that this is not the case in what follows.

Unconstrained ML for Gaussian white noise. For Gaussian stationary noise, the covariance matrix is diagonal and proportional to the identity matrix:

$$\mathrm{Cov}(\mathbf{y}) = \mathrm{diag}(\sigma^2) = \sigma^2\,\mathbf{I} \quad \Longleftrightarrow \quad \mathbf{W} \equiv \mathrm{Cov}(\mathbf{y})^{-1} = \sigma^{-2}\,\mathbf{I}\,.$$

Then the ML solution reduces to:

$$\mathbf{x}^{(\mathrm{ML})} = (\mathbf{H}^{\mathrm{T}} \cdot \mathbf{H})^{-1} \cdot \mathbf{H}^{\mathrm{T}} \cdot \mathbf{y}\,.$$

Taking the discrete Fourier transform of $\mathbf{x}^{(\mathrm{ML})}$ yields:

$$\hat{x}_u^{(\mathrm{ML})} = \frac{\hat{h}_u^{\star}\,\hat{y}_u}{\left|\hat{h}_u\right|^2} = \frac{\hat{y}_u}{\hat{h}_u} \tag{18}$$

which is exactly the solution obtained by the direct inversion of the diagonalized image formation equation. As we have seen before, this is a *bad solution* into which the noise is amplified largely beyond any acceptable level. The reader must not be fooled by the particular case considered here for sake of simplicity (Gaussian stationary noise and DFT to approximate Fourier transforms), this disappointing result is very general: *maximum likelihood only is unable to properly solve ill-conditioned inverse problems.*

1.5.2   Constrained maximum likelihood

In the hope that additional constraints such as *positivity* (which must hold for the restored brightness distribution) may avoid noise amplification, we can seek for the *constrained maximum likelihood* (CML) solution:

$$\mathbf{x}^{(\mathrm{CML})} = \arg\min_{\mathbf{x}} \phi_{\mathrm{ML}}(\mathbf{x}) \quad \text{subject to} \quad x_j \geq 0, \quad \forall j\,. \tag{19}$$

Owing to the constraints, no direct solution exists and we must use *iterative methods* to obtain the solution. It is possible to use bound constrained version of optimization algorithms such as conjugate gradients

---

[1]For general linear least squares problems, rather than directly use Eq. (17) to find $\mathbf{x}^{(\mathrm{ML})}$, it is generally faster to solve the normal equations by Cholesky factorization of $\mathbf{H}^{\mathrm{T}} \cdot \mathbf{W} \cdot \mathbf{H}$ or to compute the least squares solution from QR or LQ factorizations (see e.g., Press et al., 1992).

or limited memory variable metric methods (Schwartz and Polak, 1997; Thiébaut, 2002) but multiplicative methods have also been derived to enforce non-negativity and deserve particular mention because they are widely used: *RLA* (Richardson, 1972; Lucy, 1974) for Poissonian noise; and *ISRA* (Daube-Witherspoon and Muehllehner, 1986) for Gaussian noise.

General non-linear optimization methods. Since the penalty may be non-quadratic, a non-linear multi-variable optimization algorithm must be used to minimize the penalty function $\phi_{\mathrm{ML}}(\mathbf{x})$. Owing to the large number of parameters involved in image restoration (but this is also true for most inverse problems), algorithms requiring a limited amount of memory must be chosen. Finally, in case non-negativity is required, the optimization method must be able to deal with bound constraints for the sought parameters. Non-linear conjugate gradients (CG) and limited-memory variable metric methods (VMLM or L-BFGS) meet these requirements and can be modified (using gradient projection methods) to account for bounds (Schwartz and Polak, 1997; Thiébaut, 2002). Conjugate gradients and variable metric methods only require a user supplied code to compute the penalty function and its gradient with respect to the sought parameters. All the images in this course where restored using VMLM-B algorithm a limited-memory variable metric method with bound constraints (Thiébaut, 2002).

Richardson-Lucy algorithm. RLA has been obtained independently by Richardson (1972) and Lucy (1974) on the basis of probabilistics considerations. Exactly the same algorithm has also been derived by others authors using the *expectation-maximization* (EM) procedure (Dempster et al., 1977). RLA is mostly known by astronomers whereas EM is mostly used in medical imaging; but, again, they are just different names for exactly the same algorithm. RLA yields the constrained maximum likelihood solution for Poissonian noise:

$$\mathbf{x}^{(\mathrm{RLA})} = \arg \min_{\mathbf{x}} \phi_{\mathrm{Poisson}}(\mathbf{x}) \quad \text{subject to} \quad x_j \geq 0, \quad \forall j \qquad (20\mathrm{a})$$

where:

$$\phi_{\mathrm{Poisson}}(\mathbf{x}) = \sum_j \left[\tilde{y}_j + \left(\log \frac{y_j}{\tilde{y}_j} - 1 \right) \right] y_j \qquad (20\mathrm{b})$$

and $\tilde{\mathbf{y}} \equiv \mathbf{H} \cdot \mathbf{x}$ is the model of the observed image. Starting with an initial guess $\mathbf{x}^{(0)}$ (for instance a uniform image), RLA improves the solution by using the recursion:

$$x_j^{(k+1)} = \left(\mathbf{H}^{\mathrm{T}} \cdot \mathbf{q}^{(k)} \right)_j x_j^{(k)} \quad \text{with} \quad q_j^{(k)} = y_j / \tilde{y}_j^{(k)} \qquad (21)$$

where $\mathbf{q}^{(k)}$ is obtained by element-wise division of the data $\mathbf{y}$ by the model $\tilde{\mathbf{y}}^{(k)} \equiv \mathbf{H} \cdot \mathbf{x}^{(k)}$ for the restored image $\mathbf{x}^{(k)}$ at k-th iteration. The following pseudo-code implements RLA given the data $\mathbf{y}$, the PSF $\mathbf{h}$, a starting solution $\mathbf{x}^{(0)}$ and a number of iterations n:

> **RLA($\mathbf{y}$, $\mathbf{h}$, $\mathbf{x}^{(0)}$, n)**
> $\hat{\mathbf{h}} := \text{FFT}(\mathbf{h})$ (store MTF)
> for $k = 0, ..., n-1$
> $\tilde{\mathbf{y}}^{(k)} := \text{FFT}^{-1}\Big(\hat{\mathbf{h}} \times \text{FFT}(\mathbf{x}^{(k)})\Big)$ (k-th model)
> $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} \times \text{FFT}^{-1}\Big(\hat{\mathbf{h}}^{\star} \times \text{FFT}(\mathbf{y}/\tilde{\mathbf{y}}^{(k)})\Big)$ (new estimate)
> return $\mathbf{x}^{(n)}$

where multiplications $\times$ and divisions / are performed element-wise and $\hat{\mathbf{h}}^{\star}$ denotes the complex conjugate of the MTF $\hat{\mathbf{h}}$. This pseudo-code shows that each RLA iteration involves 4 FFT's (plus a single FFT in the initialization to compute the MTF).

Image Space Reconstruction Algorithm. ISRA (Daube-Witherspoon and Muehllehner, 1986) is a multiplicative and iterative method which yields the constrained maximum likelihood in the case of Gaussian noise. The ISRA solution is obtained using the recursion:

$$x_j^{(k+1)} = x_j^{(k)} \frac{(\mathbf{H}^{\text{T}} \cdot \mathbf{W} \cdot \mathbf{y})_j}{(\mathbf{H}^{\text{T}} \cdot \mathbf{W} \cdot \mathbf{H} \cdot \mathbf{x}^{(k)})_j} \tag{22}$$

A straightforward implementation of ISRA is:

> **ISRA($\mathbf{y}$, $\mathbf{h}$, $\mathbf{W}$, $\mathbf{x}^{(0)}$, n)**
> $\hat{\mathbf{h}} := \text{FFT}(\mathbf{h})$ (store MTF)
> $\mathbf{r} := \text{FFT}^{-1}\Big(\hat{\mathbf{h}}^{\star} \times \text{FFT}(\mathbf{W} \cdot \mathbf{y})\Big)$ (store numerator)
> for $k = 0, ..., n-1$
> $\tilde{\mathbf{y}}^{(k)} := \text{FFT}^{-1}\Big(\hat{\mathbf{h}} \times \text{FFT}(\mathbf{x}^{(k)})\Big)$ (k-th model)
> $\mathbf{s}^{(k)} := \text{FFT}^{-1}\Big(\hat{\mathbf{h}}^{\star} \times \text{FFT}(\mathbf{W} \cdot \tilde{\mathbf{y}}^{(k)})\Big)$ (k-th denominator)
> $\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} \times \mathbf{r}/\mathbf{s}^{(k)}$ (new estimate)
> return $\mathbf{x}^{(n)}$

which, like RLA, involves 4 FFT's per iteration. In the case of stationary noise ($\mathbf{W} = \sigma^{-2}\mathbf{I}$), ISRA can be improved to use only 2 FFT's per iteration:

> **ISRA($\mathbf{y}$, $\mathbf{h}$, $\mathbf{x}^{(0)}$, n)**
> $\hat{\mathbf{h}} := \text{FFT}(\mathbf{h})$ (compute MTF)
> $\mathbf{r} := \text{FFT}^{-1}\Big(\hat{\mathbf{h}}^{\star} \times \text{FFT}(\mathbf{y})\Big)$ (store numerator)
> for $k = 0, ..., n-1$

$$\mathbf{s}^{(k)} := \mathrm{FFT}^{-1}\left(\left|\hat{\mathbf{h}}\right|^2 \times \mathrm{FFT}(\mathbf{x}^{(k)})\right) \qquad (k\text{-th denominator})$$

$$\mathbf{x}^{(k+1)} := \mathbf{x}^{(k)} \times \mathbf{r}/\mathbf{s}^{(k)} \qquad\qquad (\text{new estimate})$$

return $\mathbf{x}^{(n)}$

Multiplicative algorithms (ISRA, RLA and EM) are very popular (mostly RLA in astronomy) because they are very simple to implement and their very first iterations are very efficient. Otherwise their convergence is much slower than other optimization algorithms such as constrained conjugate gradients or variable metric. For instance, the result shown in Fig. 4c was obtained in 30 minutes of CPU time on a 1 GHz Pentium III laptop by VMLM-B method, against more than 10 hours by Richardson-Lucy algorithm.

Multiplicative methods are only useful to find a non-negative solution and, owing to the multiplication in the recursion, they leave unchanged any pixel that happens to take zero value[2].

At the cost of deriving specialized algorithms, multiplicative methods can be generalized to other expressions of the penalty to account for different noise statistics[3] (Lantéri et al., 2001) and can even be used to explicitly account for regularization (Lantéri et al., 2002).

1.5.3 Maximum likelihood summary

Constraints such as positivity may help to improve the sought solution (at least any un-physical solution is avoided) because it plays the role of a floating support (thus limiting the effective number of significant pixels). This kind of constraints are however only effective if there are enough *background pixels* and the improvement is mostly located near the edges of the support. In fact, there may be good reasons to *not* use non-negativity: e.g. because there is no significant background, or because a background bias exists in the raw image and has not been properly removed.

Figures 4b and 4c show that neither unconstrained nor non-negative maximum likelihood approaches are able to recover a usable image. Deconvolution by unconstrained/constrained maximum likelihood yields noise amplification — in other words, the maximum likelihood solution remains ill-conditioned (i.e. a *small* change in the data due to noise can produce arbitrarily *large* changes in the solution): *regularization* is needed.

---

[2]This *feature* is however useful to implement support constraints when the support is known in advance.

[3]If the exact statistics of the noise is unknown, assuming stationary Gaussian noise is however more robust than Poissonian (Lane, 1996).

When started with a *smooth* image, iteratives maximum likelihood algorithms can achieve some level of regularization by early stopping of the iterations *before* convergence (see e.g. Lantéri et al., 1999). In this case, the regularized solution is not the maximum likelihood one and it also depends on the initial solution and the number of performed iterations. A better solution is to explicitly account for additional regularization constraints in the penalty criterion. This is explained in the next section.

1.6 Maximum *a posteriori* (MAP)

1.6.1 Bayesian approach

We have seen that the maximum likelihood solution:

$$\mathbf{x}^{(\mathrm{ML})} = \arg \max_{\mathbf{x}} \Pr\{\mathbf{y}|\mathbf{x}\} \,,$$

which maximizes the probability of the data given the model, is unable to cope with noisy data. Intuitively, what we rather want is to find the solution which has maximum probability given the data $\mathbf{y}$. Such a solution is called the maximum *a posteriori* (MAP) solution and reads:

$$\mathbf{x}^{(\mathrm{MAP})} = \arg \max_{\mathbf{x}} \Pr\{\mathbf{x}|\mathbf{y}\} \,. \tag{23}$$

From Baye's theorem:

$$\Pr\{\mathbf{x}|\mathbf{y}\} = \frac{\Pr\{\mathbf{y}|\mathbf{x}\} \, \Pr\{\mathbf{x}\}}{\Pr\{\mathbf{y}\}} \,,$$

and since the probability of the data $\mathbf{y}$ alone does not depend on the unknowns $\mathbf{x}$, we can write:

$$\mathbf{x}^{(\mathrm{MAP})} = \arg \max_{\mathbf{x}} \Pr\{\mathbf{y}|\mathbf{x}\} \, \Pr\{\mathbf{x}\} \,.$$

Taking the log-probabilities:

$$\mathbf{x}^{(\mathrm{MAP})} = \arg \min_{\mathbf{x}} \Big[-\log(\Pr\{\mathbf{y}|\mathbf{x}\}) - \log(\Pr\{\mathbf{x}\}) \Big] \,.$$

Finally, the maximum *a posteriori* solves:

$$\mathbf{x}^{(\mathrm{MAP})} = \arg \min_{\mathbf{x}} \phi_{\mathrm{MAP}}(\mathbf{x}) \tag{24a}$$

where the *a posteriori* penalty reads:

$$\phi_{\mathrm{MAP}}(\mathbf{x}) = \phi_{\mathrm{ML}}(\mathbf{x}) + \phi_{\mathrm{prior}}(\mathbf{x}) \tag{24b}$$

with:

$$\phi_{\mathrm{ML}}(\mathbf{x}) = -\eta \log \Pr\{\mathbf{y}|\mathbf{x}\} + \text{constant} \qquad (24c)$$
$$\phi_{\mathrm{prior}}(\mathbf{x}) = -\eta \log \Pr\{\mathbf{x}\} + \text{constant} \qquad (24d)$$

where $\eta > 0$. $\phi_{\mathrm{ML}}(\mathbf{x})$ is the *likelihood* penalty and $\phi_{\mathrm{prior}}(\mathbf{x})$ is the so-called *a priori* penalty. These terms are detailed in what follows. Nevertheless, we can already see that the only difference with the maximum likelihood approach, is that the MAP solution must minimize an additional term $\phi_{\mathrm{prior}}(\mathbf{x})$ which will be used to account for regularization.

1.6.2 MAP: the likelihood penalty term

We have seen that minimizing the likelihood penalty $\phi_{\mathrm{ML}}(\mathbf{x})$ enforces *agreement with the data*. Exact expression of $\phi_{\mathrm{ML}}(\mathbf{x})$ should depends on the known statistics of the noise. However, if the statistics of the noise is not known, using a least-squares penalty is more robust (Lane, 1996). In the following, and for sake of simplicity, we will assume Gaussian stationnary noise:

$$\phi_{\mathrm{ML}}(\mathbf{x}) = (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})^{\mathrm{T}} \cdot \mathbf{W} \cdot (\mathbf{H} \cdot \mathbf{x} - \mathbf{y})$$
$$= \frac{1}{\sigma^2} \sum_k ((\mathbf{H} \cdot \mathbf{x})_k - y_k)^2 \qquad (25a)$$
$$= \frac{1}{N_{\mathrm{pixel}}\, \sigma^2} \sum_u |\hat{h}_u\, \hat{x}_u - \hat{y}_u|^2. \qquad (25b)$$

where the latter expression has been obtained by (discrete) Parseval's theorem.

1.6.3 MAP: the *a priori* penalty term

The a priori penalty $\phi_{\mathrm{prior}}(\mathbf{x}) \propto -\log \Pr\{\mathbf{x}\}$ allows us to account for additional constraints not carried out by the data alone (i.e. by the likelihood term). For instance, the prior can enforce agreement with some preferred (e.g. smoothness) and/or exact (e.g. non-negativity) properties of the solution. At least, the prior penalty is responsible of regularizing the inverse problem. This implies that the prior must provide information where the data alone fail to do so (in particular in regions where the noise dominates the signal or where data are missing). Not all prior constraints have such properties and the enforced *a priori* must be chosen with care. Taking into account additional a priori constraints has also some drawbacks: it must be realized that the solution will be biased toward the prior.

1.6.4   Needs of a tunable *a priori*

Usually the functional form of the *a priori* penalty $\phi_{\text{prior}}(\mathbf{x})$ is chosen from *qualitative* (non-deterministic) arguments, for instance:

> "*Since we know that noise mostly contaminates the higher frequencies, we should favor the smoothest solution among all the solutions in agreement with the data.*"

Being qualitative, the relative strength of the *a priori* with respect to the likelihood must therefore be adjustable. This can be achieved thanks to an *hyperparameter* μ:

$$\phi_{\text{MAP}}(\mathbf{x}) = \phi_\mu(\mathbf{x}) \equiv \phi_{\text{ML}}(\mathbf{x}) + \mu\,\phi_{\text{prior}}(\mathbf{x})\,. \tag{26}$$

Alternatively μ can be seen as a Lagrange multiplier introduced to solve the constrained problem: minimize $\phi_{\text{prior}}(\mathbf{x})$ subject to $\phi_{\text{ML}}(\mathbf{x})$ be equal to some value (see Gull's method below). Also note that there can be more than one hyperparameter. For instance, α and β in our parameterized Wiener filter in Eq. (11) can be seen as such hyperparameters.

1.6.5   Smoothness prior

Because the noise usually contaminates the high frequencies, smoothness is a very common regularization constraint. Smoothness can be enforced if $\phi_{\text{prior}}(\mathbf{x})$ is some measure of the roughness of the sought distribution $\mathbf{x}$, for instance (in 1-D):

$$\phi_{\text{roughness}}(\mathbf{x}) = \sum_j [x_{j+1} - x_j]^2. \tag{27}$$

The roughness can also be measured from the Fourier transform $\hat{\mathbf{x}}$ of the distribution $\mathbf{x}$:

$$\phi_{\text{prior}}(\mathbf{x}) = \sum_u w_u\,|\hat{x}_u|^2$$

with spectral weights $w_u \geq 0$ and being a non-decreasing function of the spatial frequency, e.g.:

$$w_u = |\mathbf{u}|^\beta \quad \text{with} \quad \beta \geq 0\,.$$

The regularized solution is easy to obtain in the case of *Gaussian white noise* if we choose a *smoothness* prior measured in the Fourier space. In this case, the MAP penalty writes:

$$\begin{aligned} \phi_\mu(\mathbf{x}) &= \phi_{\text{Gauss}}(\mathbf{x}) + \mu\,\phi_{\text{prior}}(\mathbf{x}) \\ &= (\mathbf{H}\cdot\mathbf{x} - \mathbf{y})^{\text{T}} \cdot \mathbf{W} \cdot (\mathbf{H}\cdot\mathbf{x} - \mathbf{y}) + \mu \sum_u w_u\,|\hat{x}_u|^2 \end{aligned}$$

(a) true brightness distribution   **(b) observed image (blurred and noisy)**   **(c) constrained maximum likelihood (Lucy–Richardson**

(d) Wiener   **(e) unconstrained MAP**   **(f) constrained MAP with L1–L2 norm**

(g) constrained MAP (under–regularized)   **(h) constrained MAP**   **(i) constrained MAP (over–regularized)**
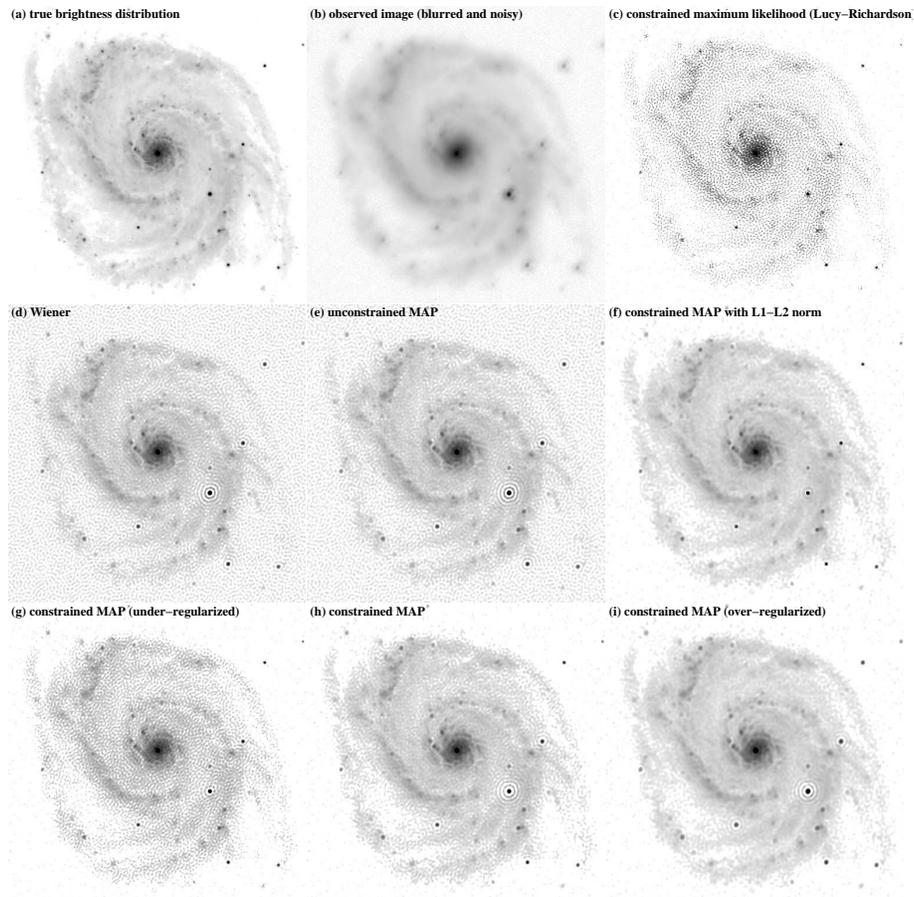
Figure 4.   Comparison of deconvolution results by various methods and regularization levels. From top-left to bottom-right: (a) true object, (b) observed image, (c) non-negative maximum likelihood (ML) solution, (d) solution by Wiener inverse-filter, (e) unconstrained maximum *a posteriori* (MAP) solution with quadratic smoothness, (f) non-negative MAP solution with $\ell_1 - \ell_2$ norm, (g) under-regularized non-negative MAP solution with quadratic smoothness, (h) non-negative MAP solution with quadratic smoothness, (i) over-regularized non-negative MAP solution with quadratic smoothness. The regularization level μ_{GCV} has been obtained from generalized cross-validation (GCV).

$$= \frac{1}{\sigma^2} \sum_j ((\mathbf{H} \cdot \mathbf{x})_j - y_j)^2 + \mu \sum_u w_u \, |\hat{x}_u|^2$$

$$= \frac{1}{N_{\text{pixel}} \, \sigma^2} \sum_u |\hat{h}_u \, \hat{x}_u - \hat{y}_u|^2 + \mu \sum_u w_u \, |\hat{x}_u|^2$$

of which the *complex* gradient is:

$$\frac{\partial \phi_\mu(\mathbf{x})}{\partial \text{Re}(\hat{x}_u)} + \mathrm{i} \, \frac{\partial \phi_\mu(\mathbf{x})}{\partial \text{Im}(\hat{x}_u)} = \frac{2}{N_{\text{pixel}} \, \sigma^2} \, \hat{h}_u^\star \, (\hat{h}_u \, \hat{x}_u - \hat{y}_u) + 2 \, \mu \, w_u \, \hat{x}_u \, .$$

The root of this expression is the MAP solution:

$$\hat{x}_u^{[\mu]} \equiv \frac{\hat{h}_u^\star \, \hat{y}_u}{|\hat{h}_u|^2 + \mu \, N_{\text{pixel}} \, \sigma^2 \, w_u} \tag{28}$$

Taking $\mu \, N_{\text{pixel}} \, \sigma^2 \, w_u = \text{E}\{|\hat{n}_u|^2\}/\text{E}\{|\hat{x}_u|^2\}$ or $w_u = \|u\|^\beta$ and $\alpha = \mu \, N_{\text{pixel}} \, \sigma^2$, this solution is identical to the one given by Wiener inverse-filter in Eq. (11). This shows that Wiener approach is a particular case in MAP framework.

1.6.6 Other kind of regularization terms

There exist many different kind of regularization which enforce different constraints or similar constraints but in a different way.

For instance, the smoothness regularization in Eq. (27) is quadratic with respect to the sought parameters and is a particular case of Tikhonov regularization:

$$\phi_{\text{Tikhonov}}(\mathbf{x}) = (\mathbf{D} \cdot \mathbf{x})^{\text{T}} \cdot (\mathbf{D} \cdot \mathbf{x})$$

where $\mathbf{D}$ is either the identity matrix or some differential operator. Quadratic regularization may also be used in the case of a Gaussian prior or to find a solution satisfying some known correlation matrix (Tarantola and Valette, 1982):

$$\phi_{\text{GP}}(\mathbf{x}) = (\mathbf{x} - \mathbf{p})^{\text{T}} \cdot \mathbf{C}_{\mathbf{x}}^{-1} \cdot (\mathbf{x} - \mathbf{p})$$

where $\mathbf{C}_{\mathbf{x}}$ is the assumed correlation matrix (when $\mathbf{p} = 0$) or covariance matrix with respect to the prior $\mathbf{p}$ which is also the default solution when there are no data.

The well-known maximum entropy method (MEM) can be implemented thanks to a non-quadratic regularization term which is the so-called negentropy:

$$\phi_{\text{MEM}}(\mathbf{x}) = \sum_j \left[p_j - x_j + x_j \, \log \left(\frac{x_j}{p_j} \right) \right]$$

where **p** is some prior distribution.

Non-quadratic $\ell_1 - \ell_2$ norms applied to the spatial gradient may be used to enforce a smoothness constraints but avoid ripples around point-like sources or sharp edges. In effect, the $\ell_1 - \ell_2$ norm will prevent small differences of intensity between neighbor pixels but put less constraints for large differences. The effectiveness of using such a norm to measure the roughness of the sought image is shown in Fig. 4f which no longer exhibits ripples around the bright stars.

1.7   Choosing the hyperparameter(s)

In order to finally solve our inverse problem, we have to choose an adequate level of regularization. This section presents a few methods to select the value of μ. Titterington et al. (1985) have made a comparison of the results obtained from different methods for choosing the value of the hyperparameters.

1.7.1   Gull's approach

For Gaussian noise, the MAP solution is given by minimizing:

$$\phi_\mu(\mathbf{x}) = \chi^2(\mathbf{x}) + \mu\,\phi_{\mathrm{prior}}(\mathbf{x})$$
$$\text{where:}\quad \chi^2 \equiv [\mathbf{m}(\mathbf{x}) - \mathbf{y}]^{\mathrm{T}} \cdot \mathbf{W} \cdot [\mathbf{m}(\mathbf{x}) - \mathbf{y}]\,,$$

where $\mathbf{m}(\mathbf{x})$ is the model. For a perfect model, the expected value of χ^2 is equal to the number of measurements: $\mathrm{E}\{\chi^2\} = N_{\mathrm{data}}$. If the regularization level is too small, the MAP model will tend to overfit the data resulting in a value of χ^2 smaller than its expected value. On the contrary, if the regularization level is too high, the MAP model will be too biased by the a priori and χ^2 will be larger than its expected value. These considerations suggest to choose the weight μ of the regularization such that:

$$\chi^2(\mathbf{x}^{[\mu]}) = \mathrm{E}\{\chi^2\} = N_{\mathrm{data}}$$

where $\mathbf{x}^{[\mu]}$ is the MAP solution obtained for a given μ. In practice, this choice tends to *oversmooth* the solution. In fact, the model being derived from the data, it is always biased toward the data and, even for a correct level of regularization, the expected value of χ^2 must be less than N_{data}. For a parametric model with M free parameters adjusted so as to minimize χ^2, the correct formula is:

$$\mathrm{E}\{\chi^2\} = N_{\mathrm{data}} - M\,,$$

the difference $N_{\mathrm{data}} - M$ is the so-called number of degrees of freedom. This formula cannot be directly applied to our case because the parameters of our model are correlated by the regularization and $M \ll N_{\mathrm{param}}$.

Gull (1989), considering that the prior penalty is used to control the effective number of free parameters, stated that $M \simeq \mu \, \phi_{\mathrm{prior}}(\mathbf{x})$. Following Gull's approach, the hyperparameter μ and the MAP solution are obtained by:

$$\mathbf{x}^{(\mathrm{Gull})} = \arg \min_{\mathbf{x}} \phi_\mu(\mathbf{x}) \quad \text{subject to:} \ \phi_\mu(\mathbf{x}) = N_{\mathrm{data}} \tag{29}$$

with $\phi_\mu(\mathbf{x}) = \chi^2(\mathbf{x}) + \mu \, \phi_{\mathrm{prior}}(\mathbf{x})$ and also possibly subject to $x_j \geq 0, \forall j$. This method is rarely used because it requires to have a very good estimation of the absolute noise level.

1.7.2 Cross-validation

Cross-Validation methods make use of the fact that it is possible to estimate missing measurements when the solution of an inverse problem is obtained.

Ordinary cross-validation. Let:

$$\mathbf{x}^{[\mu,k]} = \arg \min \phi_\mu(\mathbf{x}|\mathbf{y}^{[k]}) \quad \text{where} \quad \mathbf{y}^{[k]} \equiv \{y_j : j \neq k\} \tag{30}$$

be the regularized solution obtained from the *incomplete data set* where the k-th measurement is missing; then:

$$\tilde{y}^{[\mu,k]} \equiv (\mathbf{H} \cdot \mathbf{x}^{[\mu,k]})_k \tag{31}$$

is the predicted value of the missing data y_k. The *cross-validation* penalty is the weighted sum of the quadratic difference between the predicted value and the real measurement:

$$\mathrm{CV}(\mu) \equiv \sum_k \frac{(\tilde{y}^{[\mu,k]} - y_k)^2}{\sigma_k^2} \tag{32}$$

where σ_k^2 is the variance of the k-th measurement (noise is assumed to be uncorrelated). For a given value of the hyperparameter, $\mathrm{CV}(\mu)$ measures the statistical ability of the inversion to predict the value of missing data. A good choice for the hyperparameter is the one that minimizes $\mathrm{CV}(\mu)$ since it would achieve the best prediction capability. $\mathrm{CV}(\mu)$ can be re-written into a more workable expression:

$$\mathrm{CV}(\mu) = \sum_k \frac{(\tilde{y}_k^{[\mu]} - y_k)^2}{\sigma_k^2 \, (1 - a_{k,k}^{[\mu]})^2} \tag{33}$$

where:

$$\tilde{\mathbf{y}}^{[\mu]} = \mathbf{H} \cdot \mathbf{x}^{[\mu]} \tag{34}$$

is the model for a given μ and $\mathbf{a}^{[\mu]}$ is the so-called *influence matrix*:

$$a_{k,l}^{[\mu]} = \frac{\partial \tilde{y}_k^{[\mu]}}{\partial y_l} \; . \tag{35}$$

Generalized cross-validation.   To overcome some problems with ordinary cross-validation, Golub et al. (1979) have proposed the generalized cross-validation (GCV) which is a weighted version of CV:

$$\text{GCV}(\mu) = \sum_k w_k^{[\mu]} \frac{(\tilde{y}^{[\mu,k]} - y_k)^2}{\sigma_k^2} = \frac{\sum_k (\tilde{y}_k^{[\mu]} - y_k)^2 / \sigma_k^2}{\left[1 - \frac{1}{N_{\text{pixel}}} \sum_k a_{k,k}^{[\mu]} \right]^2} \; . \tag{36}$$

(G)CV in the case of deconvolution.   In our case, i.e. *gaussian white noise* and *smoothness* prior, the MAP solution is:

$$\hat{x}_u^{[\mu]} \equiv \frac{\hat{h}_u^\star \, \hat{y}_u}{|\hat{h}_u|^2 + \mu \, r_u} \quad \text{with} \quad r_u = N_{\text{pixel}} \, \sigma^2 \, w_u.$$

the Fourier transform of the corresponding model is:

$$\hat{\tilde{y}}_u^{[\mu]} \equiv \hat{h}_u \, \hat{x}_u^{[\mu]} = \frac{|\hat{h}_u|^2 \, \hat{y}_u}{|\hat{h}_u|^2 + \mu \, r_u} = q_u^{[\mu]} \, \hat{y}_u \quad \text{with} \quad q_u^{[\mu]} = \frac{|\hat{h}_u|^2}{|\hat{h}_u|^2 + \mu \, r_u}$$

and all the diagonal terms of the influence matrix are *identical*:

$$a_{k,k}^{[\mu]} = \frac{\partial \tilde{y}_k^{[\mu]}}{\partial y_k} = \frac{1}{N_{\text{pixel}}} \sum_u q_u^{[\mu]} \; .$$

In our case, i.e. gaussian white noise and smoothness prior, CV and GCV have the same expression:

$$\text{CV}(\mu) = \text{GCV}(\mu) = \frac{\sum_u \left(1 - q_u^{[\mu]}\right)^2 |\hat{y}_u|^2}{\left[1 - \frac{1}{N_{\text{pixel}}} \sum_u q_u^{[\mu]}\right]^2}$$

which can be evaluated for different values of μ in order to find its minimum.

1.8   Myopic and blind deconvolution

So far, we considered image deconvolution assuming that the PSF was *perfectly known*. In practice, this is rarely the case. For instance, when the PSF is measured by a calibration procedure, it is corrupted

by some level of *noise.* Moreover, if the observing conditions change, the calibrated PSF can mismatch the actual PSF. It may even be the case that the PSF cannot be properly calibrated at all, because it is varying too rapidly, or because there is no time or no means to do such a calibration. What can we do to cope with that?

In this case, since the unknown are the object brightness distribution $\mathbf{x}$ and the actual PSF $\mathbf{h}$, the MAP problem has to be restated as:

$$\{\mathbf{x}, \mathbf{h}\}^{(\mathrm{MAP})} = \arg \max_{\{\mathbf{x}, \mathbf{h}\}} \Pr\{\mathbf{x}, \mathbf{h} | \mathbf{y}\} \tag{37}$$

where the data are $\mathbf{y} = \{\mathbf{y}_{\mathrm{obj}}, \mathbf{y}_{\mathrm{PSF}}\}$, $\mathbf{y}_{\mathrm{obj}}$ being the observed image of the object and $\mathbf{y}_{\mathrm{PSF}}$ being the calibration data. Expanding the previous equation:

$$
\begin{aligned}
\{\mathbf{x}, \mathbf{h}\}^{(\mathrm{MAP})} &= \arg \max_{\{\mathbf{x}, \mathbf{h}\}} \Pr\{\mathbf{x}, \mathbf{h} | \mathbf{y}\} \\
&= \arg \max_{\{\mathbf{x}, \mathbf{h}\}} \frac{\Pr\{\mathbf{y} | \mathbf{x}, \mathbf{h}\} \, \Pr\{\mathbf{x}, \mathbf{h}\}}{\Pr\{\mathbf{y}\}} \\
&= \arg \max_{\{\mathbf{x}, \mathbf{h}\}} \frac{\Pr\{\mathbf{y} | \mathbf{x}, \mathbf{h}\} \, \Pr\{\mathbf{x}\} \, \Pr\{\mathbf{h}\}}{\Pr\{\mathbf{y}\}} \\
&= \arg \max_{\{\mathbf{x}, \mathbf{h}\}} \Pr\{\mathbf{y} | \mathbf{x}, \mathbf{h}\} \, \Pr\{\mathbf{x}\} \, \Pr\{\mathbf{h}\} \\
&= \arg \min_{\{\mathbf{x}, \mathbf{h}\}} \left(-\log \Pr\{\mathbf{y} | \mathbf{x}, \mathbf{h}\} - \log \Pr\{\mathbf{x}\} - \log \Pr\{\mathbf{h}\} \right)
\end{aligned}
$$

we find that the sought PSF and object brightness distribution are a minimum of:

$$\phi_{\mathrm{myopic}}(\mathbf{x}, \mathbf{h}) = \phi_{\mathrm{ML}}(\mathbf{y} | \mathbf{x}, \mathbf{h}) + \mu_{\mathrm{obj}} \, \phi_{\mathrm{obj}}(\mathbf{x}) + \mu_{\mathrm{PSF}} \, \phi_{\mathrm{PSF}}(\mathbf{h}) \tag{38}$$

where $\phi_{\mathrm{ML}}(\mathbf{y} | \mathbf{x}, \mathbf{h}) \propto -\log \Pr\{\mathbf{y} | \mathbf{x}, \mathbf{h}\}$ is the likelihood penalty and where $\phi_{\mathrm{obj}}(\mathbf{x}) \propto -\log \Pr\{\mathbf{x}\}$ and $\phi_{\mathrm{psf}}(\mathbf{h}) \propto -\log \Pr\{\mathbf{h}\}$ are regularization terms enforcing the a priori constraints for the sought distributions.

Assuming Gaussian noise and if the calibration data is given by an image of a point-like source, the MAP criterion writes:

$$
\begin{aligned}
\phi_{\mathrm{myopic}}(\mathbf{x}, \mathbf{h}) = \;& (\mathbf{h} \odot \mathbf{x} - \mathbf{y}_{\mathrm{obj}})^{\mathrm{T}} \cdot \mathbf{W}_{\mathrm{obj}} \cdot (\mathbf{h} \odot \mathbf{x} - \mathbf{y}_{\mathrm{obj}}) \\
& + (\mathbf{h} - \mathbf{y}_{\mathrm{PSF}})^{\mathrm{T}} \cdot \mathbf{W}_{\mathrm{PSF}} \cdot (\mathbf{h} - \mathbf{y}_{\mathrm{PSF}}) \\
& + \mu_{\mathrm{obj}} \, \phi_{\mathrm{obj}}(\mathbf{x}) + \mu_{\mathrm{PSF}} \, \phi_{\mathrm{PSF}}(\mathbf{h})
\end{aligned}
\tag{39}
$$

where $\odot$ denotes convolution and where $\mathbf{W}_{\mathrm{obj}}$ and $\mathbf{W}_{\mathrm{PSF}}$ are the weighting matrices for the object and PSF images respectively.

Solving such a *myopic* deconvolution problem is much more difficult because its solution is highly non-linear with respect to the data. In

effect, whatever are the expressions of the regularization terms, the criterion to minimize is no longer quadratic with respect to the parameters (due to the first likelihood term). Nevertheless, a much more important point to care of is that unless enough constraints are set by the regularization terms, the problem may not have a unique solution.

A possible algorithm for finding the solution of the myopic problem is to proceed by successive regularized deconvolutions. At every stage, a new estimate of the object is obtained by a first regularized deconvolution given the data, the constraints and an estimate of the PSF, then another regularized deconvolution is used to obtain a new estimate of the PSF given the constraints, the data and the previous estimate of the object brightness distribution:

$$
\begin{aligned}
\mathbf{x}^{(k+1)} \;=\; & \arg\min_{\mathbf{x}} \Big[(\mathbf{h}^{(k)} \odot \mathbf{x} - \mathbf{y}_{\mathrm{obj}})^{\mathrm{T}} \cdot \mathbf{W}_{\mathrm{obj}} \cdot (\mathbf{h}^{(k)} \odot \mathbf{x} - \mathbf{y}_{\mathrm{obj}}) \\
& + \mu_{\mathrm{obj}}\, \phi_{\mathrm{obj}}(\mathbf{x}) \Big] \\
\mathbf{h}^{(k+1)} \;=\; & \arg\min_{\mathbf{h}} \Big[(\mathbf{h} \odot \mathbf{x}^{(k+1)} - \mathbf{y}_{\mathrm{obj}})^{\mathrm{T}} \cdot \mathbf{W}_{\mathrm{obj}} \cdot (\mathbf{h} \odot \mathbf{x}^{(k+1)} - \mathbf{y}_{\mathrm{obj}}) \\
& + (\mathbf{h} - \mathbf{y}_{\mathrm{PSF}})^{\mathrm{T}} \cdot \mathbf{W}_{\mathrm{PSF}} \cdot (\mathbf{h} - \mathbf{y}_{\mathrm{PSF}}) \\
& + \mu_{\mathrm{PSF}}\, \phi_{\mathrm{PSF}}(\mathbf{h}) \Big]
\end{aligned}
$$

where $\mathbf{x}^{(k)}$ and $\mathbf{h}^{(k)}$ are the sought distributions at k-th iteration. Such an iterative algorithm does reduce the global criterion $\phi_{\mathrm{MAP}}(\mathbf{x}, \mathbf{h})$ but the final solution depends on the initial guess $\mathbf{x}^{(0)}$ or $\mathbf{h}^{(0)}$ unless the regularization terms warrant unicity.

Myopic deconvolution has enough flexibility to account for different cases depending on the signal-to-noise ratio of the measurements:

- In the limit $\mathbf{W}_{\mathrm{PSF}} \to +\infty$, the PSF is perfectly characterized by the calibration data (i.e. $\mathbf{h} \leftarrow \mathbf{y}_{\mathrm{PSF}}$) and myopic deconvolution becomes identical to **conventional deconvolution**.

- In the limit $\mathbf{W}_{\mathrm{PSF}} \to 0$ or if no calibration data are available, myopic deconvolution becomes identical to **blind deconvolution** which involves to find the PSF and the brightness distribution of the object from only an image of the object.

Stated like this, conventional and blind deconvolution appear to be just two extreme cases of the more general myopic deconvolution problem. We however have seen that conventional deconvolution is easier to perform than myopic deconvolution and we can anticipate that blind deconvolution must be far more difficult.
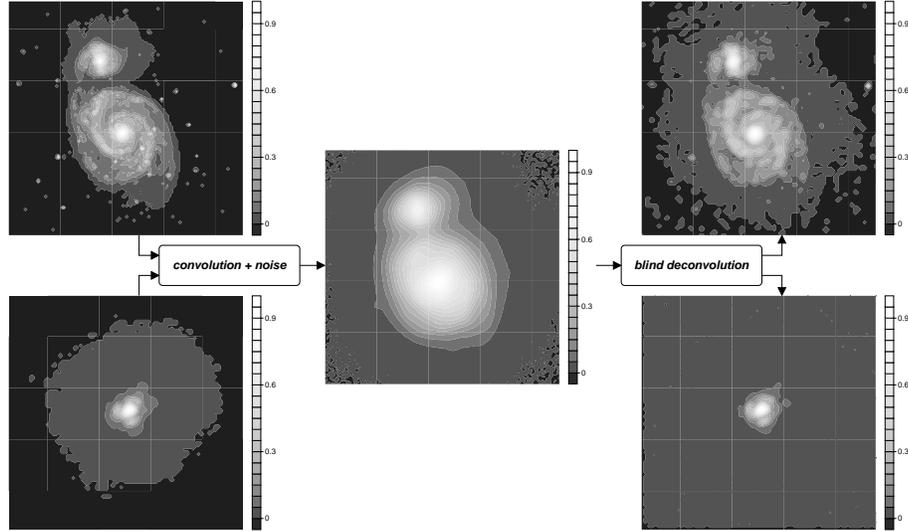
Figure 5. Example of blind deconvolution in action. Left: true object brightness and PSF. Middle: simulation of corresponding observed image. Right: the two components found by blind deconvolution. Source: Thiébaut, 2002.

Nevertheless a number of blind deconvolution algorithms have been devised which are able to notably improve the quality of real (i.e. noisy) astronomical images (e.g. Ayers and Dainty, 1988; Lane, 1992; Thiébaut and Conan, 1995). For instance and following the MAP approach, blind deconvolution involves the minimization of the join criterion (Thiébaut and Conan, 1995; Thiébaut, 2002):

$$
\begin{aligned}
\phi_{\mathrm{blind}}(\mathbf{x}, \mathbf{h}) \quad = \quad & (\mathbf{h} \odot \mathbf{x} - \mathbf{y}_{\mathrm{obj}})^{\mathrm{T}} \cdot \mathbf{W}_{\mathrm{obj}} \cdot (\mathbf{h} \odot \mathbf{x} - \mathbf{y}_{\mathrm{obj}}) \\
& + \mu_{\mathrm{obj}}\, \phi_{\mathrm{obj}}(\mathbf{x}) + \mu_{\mathrm{PSF}}\, \phi_{\mathrm{PSF}}(\mathbf{h})\,.
\end{aligned} \tag{40}
$$

Figure 5 shows an example of blind deconvolution by the resulting algorithm applied to simulated data. Of course the interest of blind deconvolution is not restricted to astronomy and it can be applied to other cases for which the instrumental response cannot be properly calibrated for instance in medical imaging (see Fig. 6a and Fig. 6b).

1.9   Concluding remarks

We have seen how to properly solve for the inverse problem of image deconvolution. But all the problems and solutions discussed in this course are not specific to image restoration and apply for other problems.
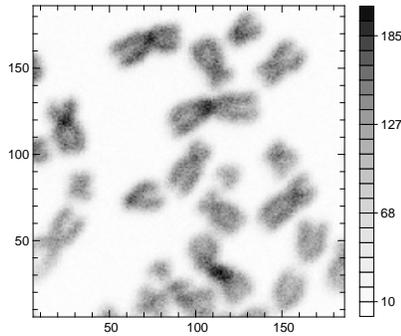
Figure 6a. Microscopic image of chromosomes (courtesy Jean-Claude Bernengo from the Centre Commun de Quantimétrie, Université Claude Bernard, Lyon, France).
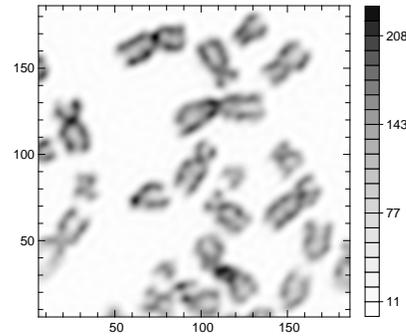
Figure 6b. Microscopic image of chromosomes improved by blind deconvolution.

Inverse problems are very common in experimental and observational sciences. Typically, they are encountered when a large number of parameters (as many as or more than measurements) are to be retrieved from measured data assuming a model of the data – also called the direct model. Such problems are ill-conditioned in the sense that a simple inversion of the direct model applied directly to the data yields a solution which exhibits significant, or even dominant, features which are completely different for a small change of the input data (for instance due to a different realization of the noise). Since the objective constraints set by the data alone are not sufficient to provide a unique and satisfactorily solution, additional subjective constraints must be taken into account. Enforcing such a priori constraints in order to make the inverse problem well-conditioned is termed regularization.

In addition to the mathematical requirement that regularization is effectively able to supplement the lack of information carried by the data alone, it is important that the regularization constraints be physically relevant because the regularized solution will be biased toward the a priori. For instance, the smoothness constraints is efficient for avoiding noise amplification in deconvolution but it will give a solution which is systematically smoother than the observed object. For that reason, in the case of stellar field images, maximum entropy regularization may be preferred to smoothness constraints. However, the important point is more what type of constraints is set by the regularization rather than how exactly this is implemented. There are many different ways to

measure the roughness (first or second derivatives, ...) but they shall give solutions which differ only for details.

The level of regularization must be tuned with care: too high and the result will be excessively biased toward the a priori which means that not all the informational contents of data is extracted; too low and the solution will be corrupted by artifacts due to the amplification of the noise. A number of methods have been devised to objectively find the good level of regularization. Among others, generalized cross validation (GCV) chooses the level of regularization for which the solution of the inverse problem has the best capability to predict missing measurements.

Generally, solving inverse problems can be stated as constrained optimization of some criterion, the so-called penalty function. There is therefore a strong link between inverse problems and non-linear constrained optimization methods.

Acknowledgments

All the results shown in this chapter were processed with *Yorick* which is a free data processing software available for Unix, MacOS/X and MS-Windows. The yorick home site is `ftp://ftp-icf.llnl.gov/pub/Yorick`.

The book "*Inverse Problem Theory*" (Tarantola, 1987) is a very good introduction to the subject (the first part of the book, "*Discrete Inverse Problem*", is freely downloadable at: `http://www.ipgp.jussieu.fr/~tarant`).

References

Ayers, G. R. and Dainty, J. C. (1988). Iterative blind deconvolution and its applications. *Opt. Lett.*, 13(7):547–549.

Daube-Witherspoon, M. E. and Muehllehner, G. (1986). An iterative image space reconstruction algorithm suitable for volume ect. *IEEE Trans. Med. Imaging*, 5(2):61–66.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *J. R. Stat. Soc. B*, 39:1–37.

Golub, Gene H., Heath, Michael, and Wahba, Grace (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21:215–223.

Gull, S. F. (1989). *Maximum Entropy and Bayesian Methods*, chapter Developments in maximum entropy data analysis, pages 53–72. Kluwer Academic.

Lane, R. G. (1992). Blind deconvolution of speckle images. *J. Opt. Soc. Am. A*, 9(9):1508–1514.

Lane, R. G. (1996). Methods for maximum-likelihood deconvolution. *J. Opt. Soc. Am. A*, 13(10):1992–1998.

Lantéri, H., Soummer, R., and Aime, C. (1999). Comparison between ISRA and RLA algorithms. Use of a Wiener Filter based stopping criterion. *A&AS*, 140:235–246.

Lantéri, H., Roche, M., and Aime, C. (2002). Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inverse Problems*, 18:1397–1419.

Lantéri, H., Roche, M., Cuevas, O., and Aime, C. (2001). A general method to devise maximum-likelihood signal restoration multiplicative algorithms with nonnegativity constraints. *Signal Processing*, 81:945–974.

Lucy, L. B. (1974). An iterative technique for the rectification of observed distributions. *ApJ*, 79(6):745–754.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992). *Numerical Recipes in C*. Cambridge University Press, 2nd edition.

Richardson, W. H. (1972). Bayesian-based iterative method of image restauration. *J. Opt. Soc. Am.*, 62(1):55–59.

Schwartz, A. and Polak, E. (1997). Family of projected descent methods for optimization problems with simple bounds. *Journal of Optimization Theory and Applications*, 92(1):1–31.

Tarantola, A. (1987). *Inverse Problem Theory*. Elsevier.

Tarantola, A. and Valette, B. (1982). Generalized nolinear inverse problems solved using the least squares criterion. *Reviews of Geophysics and Space Physics*, 20(2):219–232.

Thiébaut, E. and Conan, J.-M. (1995). Strict a priori constraints for maximum likelihood blind deconvolution. *J. Opt. Soc. Am. A*, 12(3):485–492.

Thiébaut, Eric (2002). Optimization issues in blind deconvolution algorithms. In Starck, Jean-Luc and Murtagh, Fionn D., editors, *Astronomical Data Analysis II*, volume 4847, pages 174–183. SPIE.

Titterington, D. M. (1985). General structure of regularization procedures in image reconstruction. *Astron. Astrophys.*, 144:381–387.