IMAGE DECONVOLUTION

M. Bertero and P. Boccacci
DISI, Universita' di Genova,
Via Dodecaneso 35, I-16146 Genova, Italy
bertero@disi.unige.it, boccacci@disi.unige.it

Abstract Image deconvolution is a basic problem in the processing of microscopic images. It is ill-posed, as the vast majority of inverse problems, and therefore it requires an accurate modeling taking into account all the known properties of the process of image formation and acquisition. In this chapter, after a brief discussion of the ill-posedness of image deconvolution in a continuous setting, we develop a detailed statistical model which applies to the case of fluorescence microscopy. Two approximate models, denoted as the Gaussian case and the Poisson case, are also introduced. The statistical model is the starting point for a maximum likelihood approach. In the Gaussian case one re-obtains the standard least squares problem which is also ill-posed. The constraint of non-negativity is introduced and two iterative algorithms converging to nonnegative least-squares solutions are presented. The need of early stopping of the iterations to produce sensible results is discussed. Moreover, an iterative method converging to the maximum-likelihood estimates of the Poisson case is presented: it is the classical RL (or EM) method which also requires early stopping of the iterations. Finally Bayesian methods, based on the use of *a priori* statistical information on the object to be restored, are introduced and their relationship with the standard regularization theory of inverse problems is discussed.

Keywords: Image deconvolution, maximum likelihood, regularization, iterative methods

Introduction

In several domains of applied science images are degraded by blurring and noise. Blurring is a perturbation due to the imaging system (caused, for instance, by diffraction, aberrations etc.) while noise is intrinsic to the detection process. Therefore image deconvolution is basically a post-processing of the detected images aimed to reduce the disturbing effects of blurring and noise.

When formulated in a naive way, image deconvolution implies the solution of a linear equation (a linear system of algebraic equations in the discrete case), but this problem turns out to be ill-posed: the solution may not exist or may not be unique. Moreover, even if a unique solution can be found (as in the case

of discrete problems), this solution is strongly perturbed by noise propagation. The standard approach to the treatment of ill-posed problems is provided by the so-called *regularization theory* (for an introduction see, for instance, [1]).

In general, the ill-posedness implies that there are too many approximate solutions of the problem which are able to reproduce the data within a given noise level. To get a sensible solution one must reformulate the problem of image deconvolution by taking into account, as far as possible, all the known properties of image formation and acquisition as well as *a priori* information about properties of the object to be restored.

In the case of space invariance the process of image formation can be described by a *Point Spread Function* (PSF), which must be computed or measured. On the other hand, in applications such as Astronomy and Fluorescence Microscopy, images are detected by means of CCD cameras and therefore the process of image acquisition is affected by a combination of Poisson (photon counting) and Gaussian noise (read-out noise due to the amplifier). Finally *a priori* information about the unknown image, such as, for instance, non-negativity, is available.

The need of taking into account all the properties mentioned above, and, in particular, the statistical properties of the noise, implies that statistical approaches can be used for reformulating the problem of image deconvolution. *Maximum likelihood* looks as the most natural one and, in such a way, the image deconvolution problem can be reduced to the minimization of a suitable functional. Since the new problem is still ill-posed, two approaches are possible: the first is to use iterative minimization algorithms with *regularization properties*, in the sense that it is possible to control noise propagation by a suitable stopping of the iterations (such a property is called *semiconvergence*); the second is to use *a priori* information about statistical properties of the unknown object in the framework of Bayesian estimation.

In Section 2 we discuss the ill-posedness of the problem in a continuous setting and we prove the need of additional information as a way to reduce the uncertainty of the approximate solutions. Section 3 is devoted to a detailed statistical model which applies to the case of fluorescence microscopy; moreover two approximate models, denoted respectively as the Gaussian and the Poisson case, are introduced. In Section 4 the maximum likelihood approach is described and applied to the two approximate models of the previous section. In particular, in the Gaussian case we show that one re-obtains the classical least squares problem which is also ill-posed; next the additional constraint of non-negativity is introduced and two iterative methods for approximating the non-negative least squares solutions are discussed. Moreover the classical iterative method for approximating the maximum likelihood solutions in the Poisson case, the so-called Richardson-Lucy (RL) or Expectation-Maximization (EM) method, is presented and its main properties are discussed. Finally Section 5

is devoted to the Bayesian methods and their relationship with the methods provided by regularization theory.

1. Basic Properties of Image Deconvolution

The main feature of image deconvolution is ill-posedness. The practical implication of this property is that the solution of a (discrete) convolution equation is completely corrupted by an excess of noise propagation. However, the most simple way for understanding the difficulties generated by ill-posedness is to discuss the problem in a continuous setting.

Continuous Model

We assume that images are described by functions of 2 or 3 variables, $\mathbf{x} \in \mathbf{R}^n$ (n=2 in the 2D case and n=3 in the 3D case). We denote by $f_0(\mathbf{x})$ the intensity, at the point $\mathbf{x}$, of the object to be imaged, by $g_0(\mathbf{x})$ its image produced by the optical instrument before detection (also called the noiseless image) and by $g(\mathbf{x})$ the detected image.

In most imaging systems the noiseless image is approximately a linear function of the object. Therefore the imaging system defines a linear operator A such that:

$$g_0 \;=\; A\, f_0 \;\; . \tag{1}$$

If the system is space invariant, then the operator A is a *convolution operator,* i. e. there exists a function $h(\mathbf{x})$ such that:

$$(A\, f_0)(\mathbf{x}) \;=\; \int \; h(\mathbf{x}-\mathbf{x}')f_0(\mathbf{x}')\; d(\mathbf{x}' \;\; . \tag{2}$$

The function $h(\mathbf{x})$ is the PSF of the imaging system. Models of 3D PSFs for different kinds of microscopes are discussed in the accompanying chapter *Image Formation in Fluorescence Microscopy* by G. Vicidomini.

If we denote the Fourier Transform (FT) of a function by the corresponding capital letter, then from the *convolution theorem* and Eqs. (1)- (2), we get:

$$G_0(\omega) \;=\; H(\omega)\; F_0(\omega) \;, \tag{3}$$

$\omega \in \mathbf{R}^n$ being the coordinates in Fourier space, also called *space frequencies*.

The function $H(\omega)$ is the *transfer function* (TF), which describes the behaviour of the imaging system in the frequency domain. If the TF is zero outside a bounded domain $\mathbf{\Omega}$, the PSF is said to be *band-limited* and the set $\mathbf{\Omega}$ is called the *band* of the imaging system. It is obvious that the noiseless image g_0 is also band-limited.

In most cases the PSF has the following properties:

$$i) \;\; h(\mathbf{x}) \;\geq\; 0 \;\;;\;\; ii) \;\; \int h(\mathbf{x})\; d\mathbf{x} \;<\; +\infty \;\; . \tag{4}$$

The first property implies that the noiseless image g_0 is also non-negative, while the second property implies that the imaging system is a low-pass filter. In fact, the Riemann-Lebesgue theorem implies that the TF $H(\omega)$ is bounded and continuous and that it tends to zero when $|\omega| \rightarrow +\infty$.

Ill-Posedness of Image Deconvolution

The detected images are corrupted by several kinds of noise (this point will be discussed in the next section). In the continuous setting this situation is modeled by writing the relation between the detected image g and the noiseless image g_0 in the following form:

$$g(\mathbf{x}) = g_0(\mathbf{x}) + w(\mathbf{x}) = (A\, f_0)(\mathbf{x}) + w(\mathbf{x}) \ . \tag{5}$$

This expression is not related to specific assumptions about the noise; in particular it does not mean that we are assuming additive noise. The term $w(\mathbf{x})$ is just the difference between the detected and the noiseless image.

In terms of the FTs we have:

$$G(\omega) = H(\omega)\, F_0(\omega) + W(\omega) \ , \tag{6}$$

and therefore, even if the noiseless image is band-limited, the detected image, in general, is not, because $W(\omega)$ may not be zero where $H(\omega) = 0$ (*out-of-band noise*).

Given the detected image g and the PSF h, the problem of *image deconvolution* is to determine a sound estimate f of f_0. If the noise term w is small, then the most natural approach is to look for a solution of the linear equation:

$$Af = g \ , \tag{7}$$

which, in Fourier space, becomes:

$$H(\omega)\, F(\omega) = G(\omega) \ . \tag{8}$$

This elementary equation clarifies the difficulties of image deconvolution. In fact, if the PSF is band-limited while the noise is not, the equation is inconsistent outside the band of the instrument. In other words, *no solution exists*! Moreover, again in the case of a band-limited system, even if a solution exists, *the solution is not unique*: one can add to it an object whose FT is zero over the band and takes arbitrary values outside the band. Such a pathological object is sometimes called an *invisible object*, because its image is zero even if it is not zero. Finally, even if a solution exists and is unique, a small variation of the noise in points of the band where the TF is small can modify completely the solution, as we can understand by substituting Eq. (6) into Eq. (8) and solving for F; the result is given by:

$$F(\omega) = F_0(\omega) + \frac{W(\omega)}{H(\omega)} \ . \tag{9}$$

All these remarks imply that the problem is *ill-posed*, because the solution may not exist, may not be unique and may not depend continuously on the data.

However, it is obvious that the true object f_0 is an approximate solution of Eq. (7), in the sense that Af_0 is not exactly g but is close to g. This remark suggests to investigate the set of the approximate solutions of Eq. (7), i. e. the set of the objects whose noiseless images approximate g within a given error level. The difficulty is that this set is too broad: it contains both sensible and crazy objects. To find methods for extracting the sensible ones one must reformulate the problem by taking into account all the available information both on the process of image acquisition (noise) and on the object itself (*a priori* information, such as non-negativity).

2. A Statistical Model

The first step in the reformulation of image deconvolution is to model the noise corrupting the data. We restrict the analysis to the case of fluorescence microscopy where the noise is a random process intimately related to photon emission and acquisition, so that the best framework is provided by a discrete setting. For the sake of generality we consider the 3D case; the reduction to the 2D case is obvious.

If we assume magnification one, then the object and the image are defined in the same volume which can be partitioned into the same number of voxels with the same size. The latter is defined by the acquisition process: it is given by the size of the elements of the CCD camera in the lateral directions, and by the scanning distance in the axial direction. The voxels can be characterized by a multi-index $\mathbf{n}$, which is the triple of indexes $\{n_1, n_2, n_3\}$, defining the position of the voxel within the image volume and taking respectively N_1, N_2, and N_3 values. This triple will be denoted by $\mathbf{N}$. Moreover, given a physical quantity $h(\mathbf{x})$, its value at the voxel $\mathbf{n}$, also called *voxel value*, is denoted by $h(\mathbf{n})$. It is given, for instance, by the integral of $h(\mathbf{x})$ over the voxel volume.

Photons, emitted in the object volume, are collected by the microscope and detected in the image volume. This is a statistical process. We denote by $\xi(\mathbf{n}')$ the Random Variable (RV) describing the statistical distribution of the number of photons emitted at voxel $\mathbf{n}'$ and collected by the microscope during the integration time T. Then the *first basic assumption* is the following:

- $\xi(\mathbf{n}')$ is a Poisson RV, with expected value $f(\mathbf{n}')$, i. e.

$$P_{\xi(\mathbf{n}')}(m) \;=\; \frac{e^{-f(\mathbf{n}')} f(\mathbf{n}')^m}{m!} \tag{10}$$

  is the probability of the emission of m photons at voxel $\mathbf{n}'$;

- the RVs $\xi(\mathbf{n}')$ and $\xi(\mathbf{n}'')$, corresponding to different voxels, are statistically independent.

Next we must provide a statistical interpretation of the sampled PSF, whose voxel values are denoted by $h(\mathbf{n})$. If it is normalized in such a way that:

$$\sum_{\mathbf{n}} h(\mathbf{n}) \;=\; 1 \;, \tag{11}$$

and if we assume space invariance, then $h(\mathbf{n}-\mathbf{n}')$ *is the probability that a photon emitted at voxel* $\mathbf{n}'$ *of the object volume is collected at voxel* $\mathbf{n}$ *of the image volume*. Moreover, if we assume, as usual, a periodic extension of the voxel values of f and h, so that their cyclic convolution can be computed by means of FFT, we denote by A the block-circulant matrix defined by:

$$(Af)(\mathbf{n}) \;=\; \sum_{\mathbf{n}'} h(\mathbf{n}-\mathbf{n}')f(\mathbf{n}') \;. \tag{12}$$

Let $\xi(\mathbf{n},\mathbf{n}')$ be the RV describing the statistical distribution of the number of photons emitted at voxel $\mathbf{n}'$ and collected at voxel $\mathbf{n}$, then the *second basic assumption* is the following:

- $\xi(\mathbf{n},\mathbf{n}')$ is a Poisson RV with expected value given by:

$$E\{\xi(\mathbf{n},\mathbf{n}')\} \;=\; h(\mathbf{n}-\mathbf{n}')f(\mathbf{n}') \;; \tag{13}$$

- for any fixed $\mathbf{n}''$, and $\mathbf{n} \neq \mathbf{n}'$, $\xi(\mathbf{n},\mathbf{n}'')$ and $\xi(\mathbf{n}',\mathbf{n}'')$ are statistically independent.

We remark that the previous assumptions are consistent. In fact, it is obvious that the following relationship holds true between the RVs we have introduced:

$$\xi(\mathbf{n}') \;=\; \sum_{\mathbf{n}} \xi(\mathbf{n},\mathbf{n}') \;. \tag{14}$$

Since, for fixed $\mathbf{n}'$, the RVs $\xi(\mathbf{n},\mathbf{n}')$ are independent and Poisson distributed, it follows that $\xi(\mathbf{n}')$ is also Poisson distributed and that its expected value is the sum of the expected values:

$$E\{\xi(\mathbf{n}')\} \;=\; \sum_{\mathbf{n}} E\{\xi(\mathbf{n},\mathbf{n}')\} \;=\; \sum_{\mathbf{n}} h(\mathbf{n}-\mathbf{n}')f(\mathbf{n}') \;=\; f(\mathbf{n}') \;, \tag{15}$$

where the normalization condition of Eq. (11) has been used. This result is in agreement with the first assumption.

The previous analysis concerns the emission process. Let us consider now the detection process in the image volume, where photons are detected by means of a CCD camera. *If we assume efficiency one*, then all the photons collected by the microscope are detected by the CCD camera and therefore the statistical properties of the detected photons can be deduced from the statistical

properties of the emitted ones. If we denote by $\eta_{obj}(\mathbf{n})$ the RV describing the photons emitted by the object and detected at voxel $\mathbf{n}$, we have;

$$\eta_{obj}(\mathbf{n}) \;=\; \sum_{\mathbf{n}'} \xi(\mathbf{n},\mathbf{n}') \;\; . \tag{16}$$

Thanks to the previous assumptions, this RV is Poisson distributed with an expected value given by:

$$E\{\eta_{obj}(\mathbf{n})\} \;=\; \sum_{\mathbf{n}'} h(\mathbf{n}-\mathbf{n}')f(\mathbf{n}') \;=\; (Af)(\mathbf{n}) \;\; . \tag{17}$$

Moreover, the RVs $\eta_{obj}(\mathbf{n})$ associated to different voxels are statistically independent. We will denote by $g_{obj}(\mathbf{n})$ the number of photons which are actually detected at voxel $\mathbf{n}$, i. e. the realization of the RV $\eta_{obj}(\mathbf{n})$.

Now we can consider the following simplified version of the model developed in [2] for data acquired by a CCD camera. If we denote by $g(\mathbf{n})$ the value at voxel $\mathbf{n}$) of the detected image, then this is the realization of a RV $\eta(\mathbf{n})$ which is given by:

$$\eta(\mathbf{n}) \;=\; \eta_{obj}(\mathbf{n}) + \eta_{bkg}(\mathbf{n}) + \zeta_{ron}(\mathbf{n}) \;\; , \tag{18}$$

where:

- $\eta_{obj}(\mathbf{n})$ is the number of object photons given in Eq. (16), with expected value given in Eq. (17);
- $\eta_{bkg}(\mathbf{n})$ is the number of photons due to background emission; it is an independent Poisson RV with expected value $b(\mathbf{n})$, which, in most cases, can be assumed to be constant; in general it can be estimated and, with the detected image $g(\mathbf{n})$ and the PSF $h(\mathbf{n})$, is one of the data of the problem;
- $\zeta_{ron}(\mathbf{n})$ is the additive Read-Out Noise (RON); it is an independent Gaussian process, whose expected value and variance can also be estimated.

We denote by η the set of the independent RVs $\eta(\mathbf{n})$, by g the set of their realizations $g(\mathbf{n})$ and by f the set of the object expected values $f(\mathbf{n})$. Since the RVs $\eta(\mathbf{n})$ are statistically independent and each of them is the sum of a Poisson RV with expected value $(Af)(\mathbf{n}) + b(\mathbf{n})$ and of an independent Gaussian process, the joint probability density of the RVs η for a given f is given by:

$$P_\eta(g|f) \;= \tag{19}$$

$$= \prod_{\mathbf{n}} \sum_{m=0}^{+\infty} \; e^{-[(Af)(\mathbf{n})+b(\mathbf{n})]} \;\; \frac{[(Af)(\mathbf{n})+b(\mathbf{n})]^m}{m!} \, P_{ron}(g(\mathbf{n})-m) \;\; ,$$

where $P_{ron}(u)$ is the probability density of the RON. In the frequent case of white Gaussian noise, with expected value r and standard deviation σ, we have:

$$P_{ron}(u) \;=\; \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-r)^2}{2\sigma^2}} \quad . \tag{20}$$

A useful, even if very rough, approximation is obtained by neglecting the photon noise, so that the randomness of the image is only due to additive white noise. In such a case, if we assume, for simplicity, a RON with zero expected value, then we have:

$$P_\eta(g|f) \;=\; \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{|N|} exp\left\{-\frac{1}{2\sigma^2}||g-(Af+b)||_2^2\right\} \quad , \tag{21}$$

where $|N| = N_1 N_2 N_3$ and $||.||_2$ denotes the Euclidean norm:

$$||h||_2 \;=\; \left(\sum_{\mathbf{n}} |h(\mathbf{n})|^2\right)^{1/2} \quad . \tag{22}$$

Another approximation, in general more accurate than the previous one and frequently used both in Microscopy and Astronomy, is obtained by neglecting the RON with respect to the photon noise. Since in such an approximation each detected value $g(\mathbf{n})$ should be an integer number, we can write:

$$P_\eta(g|f) \;=\; \prod_{\mathbf{n}} e^{-[(Af)(\mathbf{n})+b(\mathbf{n})]} \frac{[(Af)(\mathbf{n})+b(\mathbf{n})]^{g(\mathbf{n})}}{g(\mathbf{n})!} \quad . \tag{23}$$

In the following, the approximation, where only the RON is considered, will be called the **Gaussian case**, while that, where only the photon noise is considered, will be called the **Poisson case**.

3. Maximum Likelihood Methods

In Statistics the Maximum Likelihood (ML) approach provides the most frequently used methods for parameter estimation. Its application to image deconvolution is based on the knowledge of the random properties of the detected image, i. e. on the knowledge of the probability density $P_\eta(g|f)$. If the detected image g, the PSF h, and the background b are given, then $P_\eta(g|f)$ is a function only of f and the problem of image deconvolution becomes the problem of estimating these unknown parameters. In such a situation the ML estimator answers to the following question:

Which object f is most likely to produce the detected image g ?

Definition 1 - For a given detected image g the *likelihood function* is the function of the object f defined by:

$$L_g(f) \;=\; P_\eta(g|f) \;\; . \tag{24}$$

Definition 2 - A ML-estimate of the object f is any object f_{LM} which maximizes the likelihood function:

$$f_{ML} \;=\; arg\, max_f \; L_g(f) \;\; . \tag{25}$$

Remark - In the case of image deconvolution this problem is, in general, ill-posed. We will justify this statement in a particular case.

In all practical applications to imaging the probability density of η for a given f is the product of a very large number of factors, so that it is useful to introduce the following log-likelihood function:

$$l_g(f) \;=\; log\; L_g(f) \;\; . \tag{26}$$

Since the logarithm is an increasing and concave function, the set of the maximum points of $l_g(f)$ coincides with the set of the ML-estimates. However, to clarify the relationship with the standard approach to Inverse Problems provided by regularization theory, it is convenient to consider functionals of the following type:

$$J_g(f) \;=\; \; -\; log\; L_g(f) \;+\; (terms\; depending\; only\; on\; g) \;\; . \tag{27}$$

The terms depending on g are added or subtracted, in some cases, to obtain well-known discrepancy functionals, as we will show in the following.

In conclusion, the ML-estimates are solutions of a variational problem of the following type:

$$f_{ML} \;=\; arg\, min_f \; J_g(f) \;\; . \tag{28}$$

In such an approach it is obvious that it is possible to introduce additional constraints such as non-negativity. This case is quite important and it deserves additional comments.

The non-negative ML-estimates are solutions of the problem:

$$f^+_{ML} \;=\; arg\, min_{f \geq 0} \; J_g(f) \;\; . \tag{29}$$

If the functional $J_g(f)$ is convex, then necessary and sufficient conditions for a minimum point are provided by the *Kuhn-Tucker (KT) conditions* (see, for instance, [3]):

$$f_{ML}(\mathbf{n}) \nabla_f \; J_g(f) \mid_{f=f_{ML}} (\mathbf{n}) \;=\; 0 \;\; , \tag{30}$$

$$f_{ML}(\mathbf{n}) \;\geq\; 0 \quad , \quad \nabla_f \; J_g(f) \mid_{f=f_{ML}} (\mathbf{n}) \;\geq\; 0 \;\; . \tag{31}$$

These conditions imply that, in the voxels where f^+_{ML} is zero, the gradient of the functional is non-negative and that f^+_{ML} is non-negative in the voxels where the gradient is zero. In general these solutions are zero in a large part of the voxels and therefore they are not reliable from the physical point of view.

The Gaussian Case

In the Gaussian case the likelihood function is given by Eq. (21) and therefore the functional $J_g(f)$ is given by:

$$J_g(f) \;=\; ||g \;-\; (Af + b)||_2^2 \;= ||Af \;-\; g_s||_2^2 \;, \tag{32}$$

where g_s is the subtracted image defined by:

$$g_s \;=\; g \;-\; b \;. \tag{33}$$

It follows that the ML-method is equivalent to the classical Least-Squares (LS) method, which is just the starting point of regularization theory [1]. Any minimum point of this functional is called a *LS-solution* and is denoted by f_{LS}; it is also a solution of the *Euler equation*:

$$A^T A f_{LS} \;=\; A^T g \quad , \tag{34}$$

which can be easily obtained by zeroing the gradient of the functional. Here A^T denotes the transposed of the matrix A.

Since A is a block-circulant matrix, Eq. (34) is diagonalized by the Discrete Fourier Transform (DFT) and we obtain:

$$|H(\mathbf{k})|^2 \; F_{LS}(\mathbf{k}) \;=\; H^*(\mathbf{k}) \; G_s(\mathbf{k}) \;, \tag{35}$$

where $H(\mathbf{k})$ is the DFT of $h(\mathbf{n})$, etc., and the * denotes complex conjugation.

This equation is always consistent, and therefore a solution always exists. But in the Fourier voxels $\mathbf{k}$ where $H(\mathbf{k}) = 0$, the value of the DFT of the LS-solution is not determined, and therefore, in such a case, the solution is not unique; on the other hand, in the voxels $\mathbf{k}$ where $H(\mathbf{k}) \neq 0$, the value of the DFT of the LS-solution is given by:

$$F_{LS}(\mathbf{k}) \;=\; \frac{G_s(\mathbf{k})}{H(\mathbf{k})} \;. \tag{36}$$

These values coincide with those provided by the linear equation $Af = g_s$.

We conclude that the LS-problem is ill-posed, first because the solution may not be unique, as indicated above, and second because the LS-solutions are strongly perturbed by noise. Uniqueness can be restored by looking for the LS-solution with minimal Euclidean norm, also called *generalized solution*; it

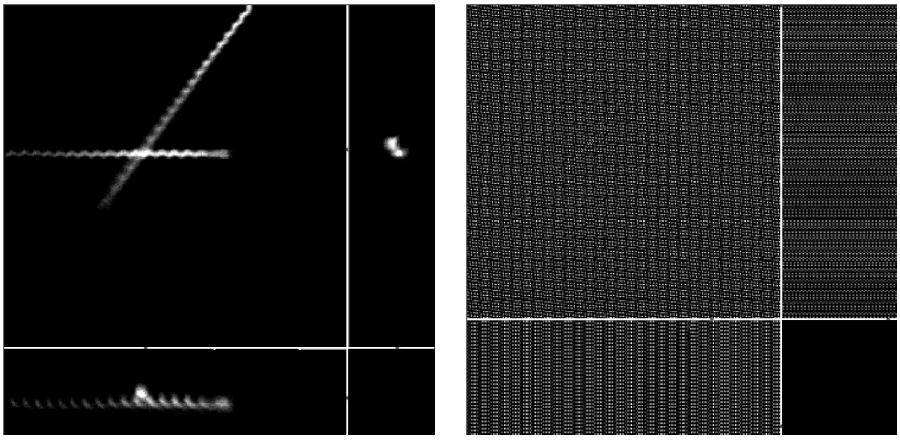

Figure 1. Illustration of the numerical instability of the generalized solution. Left panel: three orthogonal sections of a two photon 3D image of the mature sperm head of the Octopus Eledone Cyrrhosa (details are given in the text) - Right panel: the "reconstruction" provided by the generalized solution.

is obtained by setting $F_{LS}(\mathbf{k}) = 0$ in the Fourier voxels where $H(\mathbf{k}) = 0$ and therefore is given by:

$$f_{LS}(\mathbf{n}) \;=\; \frac{1}{|N|} \sum_{H(\mathbf{k})\neq 0} \frac{G_s(\mathbf{k})}{H(\mathbf{k})}\, e^{i2\pi \mathbf{n}\cdot\frac{\mathbf{k}}{\mathbf{N}}} \quad . \tag{37}$$

In Figure 1 we give the result obtained by means of this solution in the case of an image recorded with a two-photon microscope. The object is the mature sperm head of the Octopus Eledone Cirrhoses. The image is obtained with a $100X/1.4NA$ oil immersion lens, "open pinhole" condition, excitation at $720nm$, filter emission at $450nm$; it is a cube of $256 \times 256 \times 64$ 8bits voxel values. The pixel dimension is $137nm$ while the plane distance is $200nm$. In the left panel of Figure 1 a transversal section and two orthogonal axial sections of the image are shown. The PSF used for deconvolution is that modeled in the chapter on *Image Formation*. In the right panel the result provided by the generalized solution is given. The image has been completely destroyed by noise propagation.

As it was already observed in Section 2, the numerical instability of the solution given in Eq. (37) is due to the small value of $H(\mathbf{k})$ which amplify the noise corrupting the corresponding components of the subtracted image. This effect is quantified by the *condition number* of the problem which is given by:

$$\alpha \;=\; \frac{H_{max}}{H_{min}} \;, \tag{38}$$

where H_{max}, and H_{min} are respectively the largest and smallest non-zero value of $|H(\mathbf{k})|$. In fact, if δg_s is a small variation of the subtracted image and δf_{LS} is the corresponding variation of the generalized solution, the following inequality holds true:

$$\frac{\|\delta f_{LS}\|_2}{\|f_{LS}\|_2} \ \leq \ \alpha \frac{\|\delta g_s\|_2}{\|g_s\|_2} \ , \tag{39}$$

which easily follows from the inequalities (obtained by means of the Parseval equality for the DFT):

$$\|\delta f_{LS}\|_2 \ \leq \ \frac{1}{H_{min}} \|\delta g_s\|_2 \ , \ \ \|g_s\|_2 \ \leq \ H_{max} \|f_{LS}\|_2 \ . \tag{40}$$

The problem is ill-conditioned when $\alpha \ >> \ 1$ and a discrete ill-conditioned problem is typically the result of the discretization of an ill-posed problem [1].

The condition number gives also a measure of the *uncertainty* characterizing the approximate LS-solutions. In fact, the set of all the objects f compatible with the data within a given noise level ϵ is given by:

$$S_{LS}^{(\epsilon)} \ = \ \{ \ f \mid \|Af - g_s\|_2 \ \leq \ \epsilon \} \ , \tag{41}$$

and, by means of Parseval equality, we find the following equation for its boundary in Fourier space:

$$\frac{1}{|N|} \sum_{\mathbf{k}} |H(\mathbf{k}) F(\mathbf{k}) - G_s(\mathbf{k})|^2 \ = \ \epsilon^2 \ , \tag{42}$$

or also:

$$\sum_{\mathbf{k}} \frac{\epsilon^2 |H(\mathbf{k})|^2}{|N|} |F(\mathbf{k}) - \frac{G_s(\mathbf{k})}{H(\mathbf{k})}|^2 \ = \ 1 \ . \tag{43}$$

Therefore the boundary is an ellipsoid with center the generalized solution and half-axis inversely proportional to $|H(\mathbf{k})|$; since the condition number is just the ratio between the largest and smallest half-axis, it provides a measure of the extent of the set of the approximate solutions.

Since this set is too broad, one can try to identify a subset of interest by means of additional constraints. If we use non-negativity, then we can define non-negative LS-solutions as follows:

$$f_{LS}^{+} \ = \ arg \, min_{f \geq 0} \, \|Af - g_s\|_2 \ . \tag{44}$$

The KT conditions of Eqs. (30)-(31) become now:

$$f_{LS}(\mathbf{n}) \left\{ (A^T A f_{LS})(\mathbf{n}) \ - \ (A^T g_s)(\mathbf{n}) \right\} \ = \ 0 \ , \tag{45}$$

$$f_{LS}(\mathbf{n}) \ \geq \ 0 \ , \ \ (A^T A f_{LS})(\mathbf{n}) \ - \ (A^T g_s)(\mathbf{n}) \ \geq \ 0 \ . \tag{46}$$

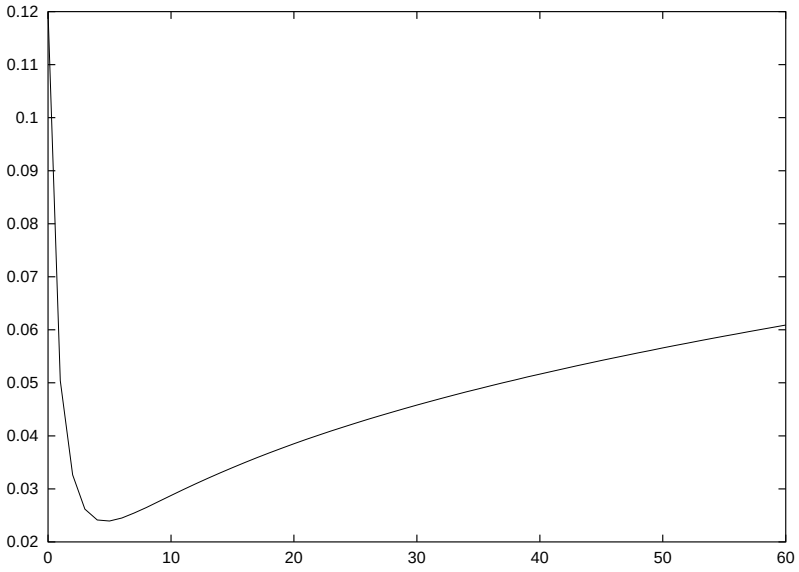

Figure 2. Example of the behaviour of the reconstruction error as a function of the number of iterations in the case of an iterative method with the semiconvergence property. The optimal number of iterations is defined by the minimum of the curve.

The solution of this problem is also affected by numerical instability and may not be unique. However it is possible to introduce iterative algorithms converging to the non-negative LS-solutions and having the following *semiconvergence* property (see [1] for a discussion): *first the iterates approach a sensible solution and then go away*.

In Figure 2 we plot a typical example of the behaviour of the reconstruction error, as a function of the number of iterations, for one of these iterative algorithms. If $f^{(k)}$ is the result of the $k-th$ iteration and f_0 is the unknown object, then the relative reconstruction error is defined by:

$$\rho^{(k)} \;=\; \frac{||f^{(k)} - f_0||_2}{||f_0||_2} \;\; . \tag{47}$$

As follows from Figure 2, the reconstruction error has a minimum after a suitable number of iterations. Therefore the algorithm must not be pushed to convergence and an early stopping is required. It is obvious that the optimal number of iterations, corresponding to the minimum of the reconstruction error, can be computed only in the case of numerical simulations. Its estimation in the case of real data is a difficult problem and must be investigated for any particular application of image deconvolution. As a general rule one can say

that, for a given imaging system, the optimal number of iterations depends on the Signal-to-Noise Ratio (SNR) (which depends on the integration time) and decreases for decreasing values of the SNR. It is obvious that the quality of the restoration is also decreasing; however, lower quality reconstructions can be obtained at a lower computational cost.

A first example of these iterative methods is provided by the **projected Landweber method**, which is basically a gradient method with a projection, at each iteration, on the closed cone of the non-negative objects. If we denote by P_+ this projection, which is defined by:

$$(P_+ f)(\mathbf{n}) \;=\; \begin{cases} f(\mathbf{n}) & \text{if } f(\mathbf{n}) > 0 \\ 0 & \text{if } f(\mathbf{n}) \leq 0, \end{cases} \tag{48}$$

then the iterative algorithm is as follows:

$$\begin{aligned} &i) \quad give \;\; f^{(0)} \;\geq\; 0; \\ &ii) \quad given \;\; f^{(k)}, \; compute: \\ &\qquad f^{(k+1)} \;=\; P_+ \left\{ f^{(k)} \;+\; \tau \left(A^T g_s \;-\; A^T A f^{(k)} \right) \right\} \;\; , \end{aligned} \tag{49}$$

where τ is a relaxation parameter which must be selected by the user within the following range of values:

$$0 \;<\; \tau \;<\; \frac{2}{H_{max}^2} \;\; . \tag{50}$$

The convergence of this algorithm is proved in [4]. Moreover, by taking the limit in Eq. (49), it is easy to verify that the limit satisfies the KT conditions and therefore is a solution of the constrained LS-problem. The semiconvergence of the algorithm is discussed in [1]: it can be proved in the non-projected case (namely when the projection P_+ is replaced by the identity) and is an experimental result in the projected case, namely it can be verified by means of numerical experiments.

In the practical applications, a frequent choice for the initial guess is $f^{(0)} = 0$ because, in the case of non-uniqueness, this choice presumably provides the minimal norm solution (this result holds true in the non-projected case). Moreover, the computational burden of the algorithm can be estimated by remarking that each iteration requires the computation of two FFTs.

A second example is provided by the **Iterative Space Reconstruction Algorithm (ISRA)**:

$$\begin{aligned} &i) \quad give \;\; f^{(0)} \;\geq\; 0; \\ &ii) \quad given \;\; f^{(k)}, \; compute: \\ &\qquad f^{(k+1)} \;=\; f^{(k)} \frac{A^T g_s}{A^T A f^{(k)}} \;\; , \end{aligned} \tag{51}$$

where the quotient of two images is defined voxel by voxel.

The convergence of this algorithm is proved in [5]. By taking the limit in Eq. (51), it is easy to verify that it satisfies the first of the KT conditions. The semiconvergence of the algorithm is an experimental result derived from numerical simulations.

In the practical applications the iterations are, in general, initialized with a constant image. Moreover the computational burden of each iteration coincides with that of the projected Landweber method.

The Poisson Case

In the Poisson case the likelihood function is given by Eq. (23), and one can obtain the following expression of the function $J_g(f)$:

$$J_g(f) \;=\; \sum_{\mathbf{n}} \left\{ g(\mathbf{n})\; ln\; \frac{g(\mathbf{n})}{(Af+b)(\mathbf{n})} \;+\; [(Af+b)(\mathbf{n}) - g(\mathbf{n})] \right\} \quad . \tag{52}$$

This function is a generalization of the Kullback-Leibler divergence and is also called the Csiszar I-divergence [6]; it provides a measure of the discrepancy between the detected image g and the computed image $Af+b$, associated with f. It is defined for nonnegative values of g and positive values of $Af+b$. If the PSF is nonnegative, then the natural domain of the functional is the closed cone of the nonnegative objects.

The functional $J_g(f)$ is convex, as it can be shown by verifying that the Hessian matrix is positive semi-definite (it is positive definite if and only if the equation $Af = 0$ implies $f = 0$, i. e. in the case of uniqueness of the solution of the deconvolution equation). Moreover the functional takes nonnegative values (as follows from the elementary inequality $a\, lna - a\, lnx + x - a \;\geq\; 0$, which holds true for fixed $a > 0$ and any $x > 0$) and therefore is bounded from below. These two properties imply that all the minima, namely the ML-solutions f_{ML}, are global and that they satisfy the KT conditions which, as follows from the computation of the gradient, take the following form:

$$f_{ML}(\mathbf{n}) \left\{ 1 - \left(A^T \frac{g}{Af_{ML}+b} \right)(\mathbf{n}) \right\} \;=\; 0 \;, \tag{53}$$

$$f_{ML}(\mathbf{n}) \;\geq\; 0 \;,\; 1 - \left(A^T \frac{g}{Af_{ML}+b} \right)(\mathbf{n}) \;\geq\; 0 \;, \tag{54}$$

where the normalization condition of Eq. (11) has been used and, again, the quotient of two images is defined voxel by voxel.

An iterative method for the computation of the ML-solutions was proposed by several authors, in particular by Richardson [7] and Lucy [8] for the deconvolution of astronomical images, and by Shepp and Vardi [9] for emission tomography. As shown in [9], this method is related to a general approach for

the solution of ML problems, known as *Expectation-Maximization* (EM). For these reasons, the algorithm is known as *Richardson-Lucy* (RL) method in Astronomy and as EM method in tomography. We will use the first name; it is as follows:

$$\begin{aligned} &i) \quad give\ f^{(0)} \geq 0; \\ &ii) \quad given\ f^{(k)},\ compute: \\ &\qquad f^{(k+1)} = f^{(k)} A^T \left(\frac{g}{A f^{(k)} + b} \right) \ . \end{aligned} \tag{55}$$

It is evident that, *if the image g and the PSF h are nonnegative, then each iterate is also nonnegative.*

The convergence of this algorithm is proved in [10] and [11] (an incomplete proof is also given in [9]). By taking the limit in Eq. (55), it is easy to verify that it satisfies the first of the KT conditions. The semiconvergence of the algorithm is an experimental result derived from numerical simulations [1].

The computation of one iterate of this method is more expensive than the computation of one iterate of ISRA or of the projected Landweber method; in fact it requires the computation of four FFTs: two for the denominator and two for applying the transposed matrix to the quotient. Moreover the convergence is slow, so that, in general, a large number of iterations is required for reaching the optimal solution.

For these reasons it is important to find methods for accelerating the convergence. One approach is proposed in [12] where an *acceleration exponent* is introduced and Eq. (55) is modified as follows:

$$f^{(k+1)} = f^{(k)} \left(A^T \frac{g}{A f^{(k)} + b} \right)^{\omega} \ , \tag{56}$$

with $\omega > 1$; the authors show that a reduction in the number of iterations by a factor ω should be expected. Numerical simulations indicate that in the case $\omega = 2$ one has (semi)convergence with a reduction in the number of iterations by a factor of 2 (remark that the computational cost of one iteration is not significantly increased with respect to the standard case $\omega = 1$).

However there is an important point which must be taken into account. A nice property of the RL method is that, if $b = 0$ and the PSF satisfies the normalization condition of Eq. (11), then all the iterates have the following property:

$$\sum_{\mathbf{n}} f^{(k)}(\mathbf{n}) = \sum_{\mathbf{n}} g(\mathbf{n}) \ . \tag{57}$$

The physical meaning of this relation is obvious: the photon content of each iterate $f^{(k)}$ coincides with the total number of detected photons.

If the background is not zero, then this condition is not automatically satisfied and therefore it can be introduced as a constraint in the minimization

problem. In the case of the accelerated algorithm of Eq. (56) this constraint is necessary in order to guarantee convergence. As shown in [13] this is equivalent to normalize each iterate in such a way that:

$$\sum_{\mathbf{n}} \left\{ f^{(k)} + b \right\} (\mathbf{n}) \;=\; \sum_{\mathbf{n}} g(\mathbf{n}) \;\; . \tag{58}$$

Therefore the iterative algorithm must include also this normalization step.

4. Bayesian Methods and Regularization

The ill-posedness (ill-conditioning) of ML-problems is generated by a lack of information on the object f, in particular by a lack of information at the frequencies corresponding to small values of the TF $H(\mathbf{k})$. A remedy can be the use of additional deterministic information (the classical example of non-negativity has been already discussed) or the use of additional statistical properties of f.

In a complete probabilistic approach it is assumed that both the object f and the detected image g are realizations of RVs denoted respectively by ξ and η, and that the problem is solved if we are able to estimate their joint probability density $P_{\xi\eta}(f, g)$.

As we discussed in Section 2, the conditional probability density of η given f can be deduced from known statistical properties of the noise. However, the marginal probability density of ξ, $P_{\xi}(f)$, in general is not known. One can guess this probability density using his knowledge, or ignorance, about f. The model used is usually called a *prior*.

Once the marginal distribution of ξ is given, the joint probability density can be obtained from *Bayes formula*:

$$P_{\xi\eta}(f, g) \;=\; P_{\eta}(g|f)\; P_{\xi}(f) \;\; . \tag{59}$$

Then, from the other Bayes formula, one obtains the conditional probability density of ξ given g:

$$P_{\xi}(f|g) \;=\; \frac{P_{\eta}(g|f)\; P_{\xi}(f)}{P_{\eta}(g)} \;\; , \tag{60}$$

which is just the solution of the image deconvolution problem in the Bayesian approaches. In fact from this probability density one can compute all the desired quantities concerning the restored image. More precisely, instead of computing a unique restored image one can compute the probability of any possible restored image.

However, in all applications, it is necessary to show at least one restored image and this can be provided by the *Maximum A Posteriori* (MAP) estimate, which is the object maximizing the *a posteriori* conditional probability :

$$f_{MAP} \;=\; arg\; max_f\; P_{\xi}(f|g) \;\; . \tag{61}$$

Introducing also in this case the log-function and neglecting the term independent of f, one finds:

$$f_{MAP} \;=\; arg\,max_f \;\{l_g(f) \;+\; ln\,P_\xi(f)\} \quad , \tag{62}$$

or, in terms of the functional $J_g(f)$ introduced in Section 2:

$$f_{MAP} \;=\; arg\,min_f \;\{J_g(f) \;-\; ln\,P_\xi(f)\} \quad . \tag{63}$$

Therefore the term $-\,ln\,P_\xi(f)$ plays the role of a regularization functional.

The most frequently used priors are of the Gibbs type, i. e.:

$$P_\xi(f) \;=\; C\;exp\,\{-\mu\Omega(f)\} \quad , \tag{64}$$

where μ is a parameter, controlling the amount of regularization (in regularization theory it is called the *regularization parameter*), and $\Omega(f)$ is a functional describing prior information about the object to be estimated [14]. In such a case the MAP problem takes the following form:

$$f_{MAP} \;=\; arg\,min_f \;\{J_g(f) \;+\; \mu\;\Omega(f)\} \quad . \tag{65}$$

Examples of Gibbs priors are given by the following functionals:

$$i) \qquad \Omega(f) \;=\; \|f\|_2^2 \;\; (white\;noise\;prior)\;; \tag{66}$$

$$ii) \qquad \Omega(f) \;=\; \|\Delta f\|_2^2 \;\; (smoothing\;prior)\;; \tag{67}$$

$$iii) \qquad \Omega(f) \;=\; \|f\|_1 \;\; (impulse\;noise\;prior)\;; \tag{68}$$

$$iv) \qquad \Omega(f) \;=\; \||\nabla f|\| \;\; (total\;variation\;prior)\;. \tag{69}$$

Moreover, in the Gaussian case, the functional of Eq. (65) becomes:

$$J_{g,\mu}(f) \;= \|Af \;-\; g_s\|_2^2 \;+\; \mu\Omega(f) \quad , \tag{70}$$

and we obtain the basic functional of the classical regularization theory of inverse problems. Therefore all functional analytic methods developed for this theory (see, for instance, [15]) apply also to the investigation of the MAP solutions in the Gaussian case.

In the particular case of the prior $i)$ we get the classical *Tikhonov regularization method*, which, in our case, consists in minimizing, for each value of the regularization parameter μ, the following functional:

$$J_{g,\mu}(f) \;= \|Af \;-\; g_s\|_2^2 \;+\; \mu\|f\|_2^2 \;\; = \tag{71}$$

$$= \;\; \frac{1}{|N|}\sum_{\mathbf{k}} \Big\{|H(\mathbf{k})F(\mathbf{k}) - G_s(\mathbf{k})|^2 \;+\; \mu\;|F(\mathbf{k})|^2\Big\} \quad .$$

The unique object minimizing this functional is called *regularized solution*, it is denoted by $f^{(\mu)}$ and is also the unique solution of the Euler equation:

$$(A^TA \;+\; \mu I)f^{(\mu)} \;=\; A^Tg_s \quad , \tag{72}$$

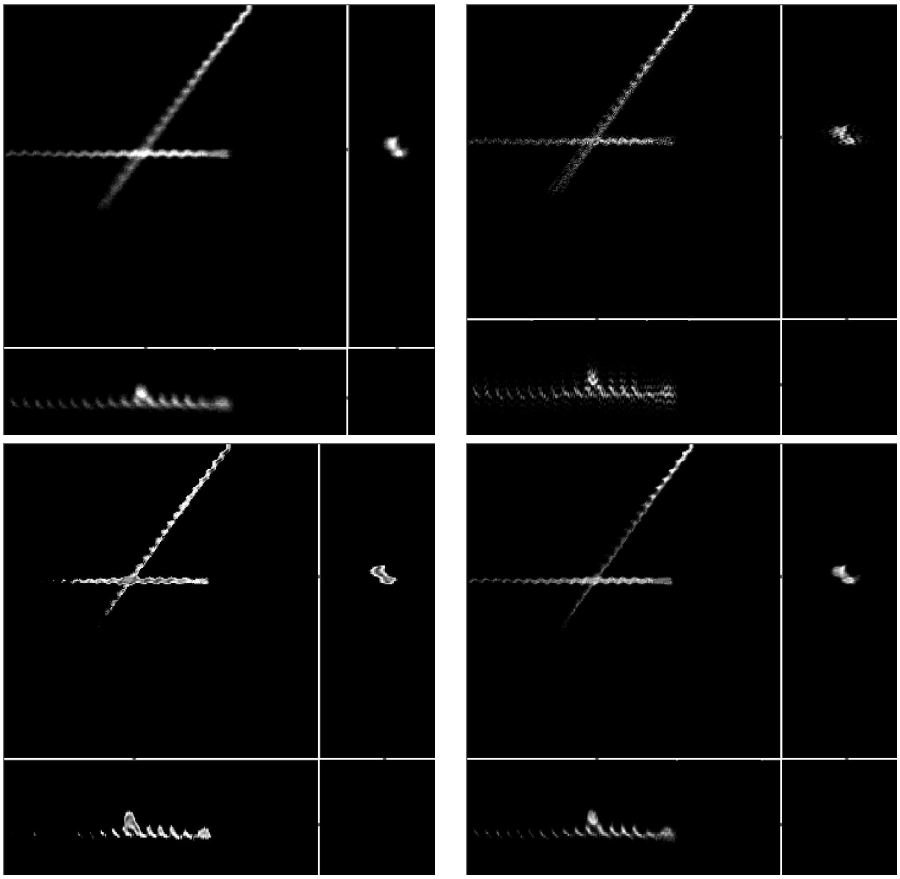

Figure 3. Comparison of the results provided by three different deconvolution methods in the case of the image of Figure 1 - Top-left panel: the original image - Top right panel: the reconstruction provided by the Tikhonov method with $\mu = 10^{-3}$ - Down-left panel: the reconstruction provided by the projected Landweber method with 50 iterations and relaxation parameter $\tau = 1.8$ - Down-right panel: the reconstruction provided by the RL (EM) method with 22 iterations.

which must be compared with Eq. (34). The DFT of the regularized solution is given by:

$$F^{(\mu)}(\mathbf{k}) \;=\; \frac{H^*(\mathbf{k})}{|H(\mathbf{k})|^2 + \mu} G_s(\mathbf{k}) \;\; . \tag{73}$$

For $\mu = 0$ we re-obtain the generalized solution while we obtain 0 in the limit $\mu \to \infty$. The problem is to estimate the optimal value of μ, a problem similar to that of the optimal number of iterations in the iterative methods described above. If we define a relative restoration error as in Eq. (47), then it can be

proved that, as a function of μ, $\rho^{(\mu)}$ first decreases, goes through a minimum and then increases, with a behaviour similar to that shown in Figure 2. Several criteria for the choice of the regularization parameter have been proposed (see again [1] or [15]).

In Figure 3 we compare the result provided by the Tikhonov method with those provided by the projected Landweber and RL methods in the case of the image of Figure 1. The values of the parameters used in the reconstructions are given in the caption. We point out that Tikhonov reconstruction is affected by artifacts generated by the well-known Gibbs effect, related to the truncation of Fourier series in the case of functions with sharp edges. In fact, as follows from Eq. (73), Tikhonov regularization is equivalent to a linear low-pass filter. The Gibbs artifacts appear as "oscillations" around the edges and introduce negative values on the background. Therefore a reduction of these effects should be provided by the constraint of non-negativity. The reconstructions obtained with the iterative methods implementing this constraint confirms this statement. The best reconstruction is obtained with the RL method which, in this case, is a method consistent with the type of noise corrupting the image.

In the Poisson case the MAP estimates are obtained by minimizing the functional:

$$\begin{aligned} J_{g,\mu}(f) &= \sum_{\mathbf{n}} \left\{ g(\mathbf{n}) ln \frac{g(\mathbf{n})}{(Af+b)(\mathbf{n})} + [(Af+b)(\mathbf{n}) - g(\mathbf{n})] \right\} + \quad (74) \\ &+ \mu\, \Omega(f) \ . \end{aligned}$$

A complete theory has not yet been developed in this case, even if several partial results are contained in the scientific literature. Moreover, several iterative methods have been proposed for the minimization of these functionals. A unified approach is proposed in [13].

5. Concluding Remarks

As outlined in this Chapter, image deconvolution is a difficult problem. First, it requires an accurate modeling of image formation, including both a model of the PSF and a model of the noise. Second, accurate reconstruction algorithms must be designed, coherent with the model of image formation. In this Chapter we have only discussed the two approximate models we called the Gaussian case and the Poisson case. But in some circumstances (faint objects), the more accurate model of Eq. 19 should be used. Methods applying to this case have been proposed [16]; their utility in Microscopy must still be demonstrated.

Software packages containing tools for image deconvolution are available but, in general, their use is not easy; moreover, in the 3D case the computational burden and the storage requirements can be excessive. In such a sit-

uation, it may be difficult for a user to improve the images of his/her microscope by means of deconvolution methods. From this point of view, the project "Power Up Your Microscope" looks interesting. It provides a free service based on web tools (see www.powermicroscope.com): the microscopist uploads the images and the characteristics of the microscope; the latter are used by the system for the computation of the PSF which is required for the deconvolution of the images; finally, after notification, he/she can downloads the deconvolved images. In this way he/she has easy access to refined methods for image deconvolution.

Acknowledgments

We thank G. Vicidomini for kindly providing the images and the reconstructions of the figures of this Chapter.

References

[1] Bertero M., and Boccacci P. *Introduction to Inverse Problems in Imaging*. Bristol: IOP Publishing, 1998

[2] Snyder D. L.,Hammoud A. M., and White R. L. Image recovery from data acquired with a charge-coupled-device camera. J. Opt. Soc. Am. 1993; A 10: 1014-1023

[3] Avriel M. *Nonlinear Programming: Analysis and Methods*. New York: Prentice Hall, 1976

[4] Eicke B. Iteration methods for convexly constrained ill-posed problems in Hilbert space. Num. Funct. Anal. Optim. 1992; 13: 413-429

[5] De Pierro A. R. On the convergence of the iterative space reconstruction algorithm for volume ECT. IEEE Trans. Med. Imaging 1987; 6: 124-125

[6] Csiszar I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. Annals of Stat. 1991; 19: 2032-2066

[7] Richardson W. H. Bayesian-based iterative method of image restoration. J. Opt. Soc. Am. 1972; 62: 55-59

[8] Lucy L. An iterative technique for the rectification of observed distribution. Astron. J. 1974; 79: 745-754

[9] Shepp L. A. and Vardi Y. Maximum likelihood reconstruction for emission and transmission tomography. IEEE Trans. Med. Imaging 1982; 1: 113-122

[10] Lange K., and Carson R. EM reconstruction algorithm for emission and transmission tomography. J. Comp. Assisted Tomography 1984; 8: 306-316

[11] Vardi Y., Shepp L. A., and Kaufman L. A statistical model for Positron Emission Tomography. J. Am. Stat. Association 1985; 80: 8-20

[12] Lanteri H., Roche M., Cuevas O., and Aime C. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints. Signal Processing 2001; 81: 945-974

[13] Lanteri H., Roche M., and Aime C. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. Inverse Problems 2002; 18: 1397-1419

[14] Kaipio J. P., and Somersalo E. *Computational and Statistical Methods for Inverse Problems*. Berlin: Springer, 2004

[15] Engl W. H., Hanke M., and Neubauer A. *Regularization of Inverse Problems*. Dordrecht: Kluwer, 1996

[16] Lanteri H., and Theys C. Restoration of Poisson images with Gaussian noise - Application th astrophysical data. Signal Processing 2005; in press