Dipartimento di Informatica e
Scienze dell'Informazione

Three-Dimensional Image Restoration in Fluorescence Microscopy

by

Giuseppe Vicidomini

Theses Series **DISI-TH-2008-06**

DISI, Università di Genova
v. Dodecaneso 35, 16146 Genova, Italy http://www.disi.unige.it/

Università degli Studi di Genova

Dipartimento di Informatica e
Scienze dell'Informazione

Dottorato di Ricerca in Informatica

Ph.D. Thesis in Computer Science

Three-Dimensional Image Restoration in Fluorescence Microscopy

by

Giuseppe Vicidomini

May, 2008

Dottorato in Scienze e Tecnologie dell'Informazione e della Comunicazione
Indirizzo Informatica
Dipartimento di Informatica e Scienze dell'Informazione
Università degli Studi di Genova

DISI, Univ. di Genova
via Dodecaneso 35
I-16146 Genova, Italy
http://www.disi.unige.it/

Ph.D. Thesis in Computer Science (S.S.D. INF/01)

Submitted by Vicidomini Giuseppe
DISI Univ. di Genova
vicidomini@disi.unige.it

Date of submission: May 2008

Title: Three-Dimensional Image Restoration in Fluorescence Microscopy

Supervisor: Patrizia Boccacci
DISI, Università di Genova
boccacci@disi.unige.it

Advisor: Alberto Diaspro
DIFI, Università di Genova
diaspro@fisica.unige.it

Ext. Reviewer: Nunzio Alberto Borghese
DSI, Università di Milano
borghese@dsi.unimi.it

Ext. Reviewer: Henri Lantéri
UMR 6525 Astrophysique, Université de Nice
hlanteri@unice.fr

To my family.

Abstract

Fluorescence microscopy uniquely provides noninvasive imaging of biological specimen at tissue, cellular and even molecular levels. For this reason, it represents one of the most used techniques in the life sciences. Moreover, this technique is capable of collecting a series of bi-dimensional images at different depths inside the specimen (optical sectioning) in order to obtain a three-dimensional image of the object. Principally, fluorescence microscopy images suffer from two physical limitation: noise and blur. The problem of image restoration in three-dimensional fluorescence microscopy, is to find the best estimate of the object function (that describes the concentration of fluorescent dye in the specimen), starting from the degraded 3D image obtained by the microscope system. Under suitable physical approximations, image restoration problem can be formulated as a linear inverse and ill-posed problem. It is well-known that the main difficulty in the treatment of ill-posed problems is the non-continuous dependence of the solution on the data. To get a sensible solution one must reformulate the problem of image restoration by taking into account, as far as possible, all the known properties of the image formation process as well as image detection. This means that an accurate physical-mathematical modeling is a necessary prerequisite for obtaining a correct formulation of the problem. For this reason a statistical formulation of the image restoration problem can be followed. Since the new problem is still ill-posed, one needs to use also information (prior-information) about the properties of the object to be restored. Bayesian approach enables the inclusion of available prior-information in the image restoration process.

The aim of the thesis is to investigate the image restoration process to different three-dimensional fluorescence microscopy techniques, like wide-field, confocal and two-photon excitation. In the first part of the thesis a more accurate physical-mathematical modeling of the image formation process is integrated in the image restoration problem. In the second part a new methods to describe prior-information of the object to be included in the Bayesian approach is presented. The original contribution of the Thesis includes the following points:

- a more accurate point spread function model is applied in the image restoration process. Such a point spread function model is based on the vectorial theory of light and it is able to take into account spherical aberration phenomena involved in the image formation process;

- a new method to model statistical properties of the object is presented. Modeling the field object as a Markov random field, and thus forcing the object distribution to be a Gibbs distribution, aids Bayesian approach to include statistical properties of the object in the image restoration process. Different potential functions in the Gibbs distribution are used to bring out different desired effects in the reconstructed image, like smoothing or edge-preserving. Moreover, fuzzy logic framework is used to create further suitable potential functions;
- the general split-gradient method is extended to the three-dimensional fluorescence microscopy case to produce suitable iterative methods for the solution of the image restoration problem.

Comparisons between the different derived methods and the most popular methods are presented, using both synthetic and real data sets.

Table of Contents

List of Figures

List of Tables

Notation

Main symbols

f	object function
f_0	true unknown object function
$\hat{f}$	estimate object function
g	detected image function
b	background function
g_0	ideal image
w	difference between detected and ideal image functions
k	point spread function
$\mathbf{f}, \mathbf{g}, \mathbf{b}, \mathbf{k}, \hat{\mathbf{f}}, \mathbf{f}_0$	vectors obtained from discretization of $f, g, b, k, \hat{f}$ and f_0, respectively
$\mathbf{f}^*$	ML- or MAP-solution
$\mathbf{f}^{(i)}$	ML- or MAP-estimate at iteration i
F, G, G_0, K, F_0	Fourier transform of f, g, g_0, k and f_0, respectively
A	convolution operator
$\mathbf{A}$	convolution matrix
$J_0(\mathbf{f}; \mathbf{g})$	discrepancy functional
$J_\mu(\mathbf{f}; \mathbf{g})$	regularized functional
$J_R(\mathbf{f})$	regularization functional or penalization term
$\mu = 1/\beta$	regularization parameter
$\nabla_{\mathbf{f}} J$	gradient of the functional J with respect to $\mathbf{f}$
$\mathscr{N}_{\mathbf{n}}$	set of sites neighboring $\mathbf{n}$
∇_F	fuzzy-logic derivative
φ	potential function
$KL_D(\mathbf{c}, \mathbf{d})$	Kullback-Leibler disvergence of $\mathbf{d}$ from $\mathbf{c}$
τ	reciprocal of the photon conversion factor
$\mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f})$	likelihood function
$P_{\mathbf{F}}(\mathbf{f})$	prior
$P_{\mathbf{F}}(\mathbf{f}\|\mathbf{g})$	*a-posteriori* conditional probability

Deterministic representation

Object (specimen) can be considered as a three-dimensional matrix with dimension $N_1 \times N_2 \times N_3$. Each element represents a voxel and is characterized by a triple of indexes $\mathbf{n} = (n_1, n_2, n_3)$ (matrix notation), thus:

$$\mathbf{f} = \big[\mathbf{f}(\mathbf{n})\big] = \big[\mathbf{f}(n_1, n_2, n_3)\big] \text{ with } n_1 = 0...N_1 - 1, \quad n_2 = 0...N_2 - 1 \text{ and } n_3 = 0...N_3 - 1. \quad (1)$$

Moreover, images can be considered as one-dimensional vectors with dimension $N \times 1$, where $N = N_1 \times N_2 \times N_3$. In this case each voxel is characterized by a single index n (vector notation), thus:

$$\mathbf{f} = \big[\mathbf{f}(n)\big] \text{ with } n_1 = 0...N - 1, \quad (2)$$

where $N = N_1 \times N_2 \times N_3$.
To switch from matrix notation to vectorial notation we use the lexicographic ordering. This means that transformation of indexes from matrix notation to vectorial notation and viceversa is made as follows:

$$\begin{cases} n &= n_3 N_1 N_2 + n_2 N1 + n1 \\ n_3 &= n/(N_1 N_2) \\ n_2 &= \big(n\%(N_1 N_2)\big)/N1 \\ n_1 &= \big(n\%(N_1 N_2)\big)\%N1 \end{cases} \quad (3)$$

where / denotes the integer division, and % denotes the reminder of the integer division.
In this thesis we will use principally matrix notation and we advise the reader when vectorial notation is used. However, we believe that vector representation of the images is more appropriated for a discrete formulation of the image process formation, in particular when we describe blurring of the system using a convolution matrix. For this reason we refer to the images with the term vector also when we are using matrix notation.
Similar discussion can be made for the image vectors $\mathbf{g}$, the point spread function vectors $\mathbf{k}$ and background vectors $\mathbf{b}$.
Moreover, it is important to remark that in this thesis, products and quotients of vectors will be intended in the Hadamard sense, i. e. component by component:

$$(\mathbf{cd})(\mathbf{n}) = \mathbf{c}(\mathbf{n})\mathbf{d}(\mathbf{n}), \quad \Big(\frac{\mathbf{c}}{\mathbf{d}}\Big)(\mathbf{n}) = \frac{\mathbf{c}(\mathbf{n})}{\mathbf{d}(\mathbf{n})}. \quad (4)$$

Stochastic representation

Object (specimen) $\mathbf{f}$ can be considered as a realization of a multi-valued random variable $\mathbf{F}$ consisting of the set of $\mathbf{F}(\mathbf{n})$ random variables associated to each site (voxel) $\mathbf{n}$:

$$\mathbf{F} = \big\{\mathbf{F}(\mathbf{n}), \mathbf{n} \in S\big\}, \quad (5)$$

where S denote the set of sites (voxels). To safe in complexity in this thesis we will denote:

$$P_{\mathbf{F}}(\mathbf{f}) = P_{\mathbf{F}}(\mathbf{F} = \mathbf{f}) \quad (6)$$

$$P_{\mathbf{F(n)}}(\mathbf{f(n)}) = P_{\mathbf{F(n)}}(\mathbf{F(n)} = \mathbf{f(n)}). \quad (7)$$

Similar discussion can be made for the image $\mathbf{g}$ and its associated multi-valued random variables $\mathbf{G}$.

Acronyms

$2D$	bi-dimensional
$3D$	three-dimensional
AFP	actual focus position
CCD	coupled charge device
$CLSM$	confocal laser scanning microscope
EM	expectation maximization
FFT	fast Fourier transform
FT	Fourier transform
$FWHM$	full-width-half maximum
$GEFM$	general epi-fluorescence microscope
GM	Geman & McClure
GRF	Gibbs random field
HB	Huber
HL	Hebert & Leahy
HS	hyper surface
I^5M	incoherent illumination interference image interference microscope
$ICTM$	iterative constraint Tikhonov-Miller
IR	infrared
$ISRA$	iterative space reconstruction algorithm
KL	Kullback-Leibler
LS	least square
LSM	laser scanning microscope
MAP	maximum *a-posteriori*
ML	maximum likelihood
MRF	Markov random field
NA	numerical aperture
NFP	nominal focus position
OPD	optical path difference
OSL	one-step-late
OTF	optical transfer function

PET	positron emission tomography
PMT	photomultiplier tube
PSF	point spread function
QP	quadratic potential
RI	refractive index
RLM	Richardson-Lucy method
RV	random variable
SGM	split-gradient method
SNR	signal-to-noise ratio
$SPECT$	single photon emission computerized tomography
SWM	standing wave microscope
TK	Tikhonov
$TPEM$	two-photon excitation microscope
UV	ultra-violet
WFM	wide-field microscope

Introduction

> Unfortunately what is little recognized is that the most worthwhile scientific books are those in which the author clearly indicates what he does not know; for an author most hurts his readers by concealing difficulties.
>
> Evariste Galois

Fluorescence multidimensional microscopy is an essential tool for research and industry in the areas of cellular biology and molecular medicine. Modern microscopic techniques have brought the possibility to follow live cells in action, responding to various induced perturbations. These capabilities include not only detailed dynamic information about cell morphology, but highly sensitive space-temporal data about the behavior of specific proteins in cells. Experimental systems are being developed to model mechanisms in healthy and sick cell lines, and probe the various components mediating these mechanisms, thus resolving the molecular networks underline complex cellular processes.
Biological structures and proteins can be selectively marked by fluorescent molecules without altering their functionality, and the microscope can be used to produce a map of this fluorescence distribution. However, fluorescence microscopy imaging properties and measurements imperfection distort the original three-dimensional distribution, in jargon of image restoration the so called object, thereby restricting a quantitative analysis. Image restoration is the operation that mitigates the distortions created by the microscope, allowing the user to do not fall in error. For this reason it has attracted great interest in the past decay.
This thesis is completely devoted to the image restoration problem associated to three-dimensional fluorescence microscopy.
To obtain a restored object from a given image, first it is necessary to have a mathematical and physical model that links the distribution of fluorescence molecules inside the specimen, to the observed data, in other word a model for the image formation process of the system. A general image formation model depends on different factors: on the nature of

the object analyzed, on the physical phenomena measured, on the system of detection utilized, on the geometry of the system, etc. However, in fluorescence microscopy principally two types of process, contribute to the image formation: a deterministic process and a stochastic process.
The deterministic process is related to the characteristic of the specimen and to the microscope's optics. Under some assumptions it can be modeled by a linear operator or better by a convolution operator. Therefore it can be described entirely by the response of the microscope system to a point like source, the so called point spread function.
The stochastic process is related to the detection system, hence to the noise introduced during this step. This process can be modeled using the statistical theory. Principally, a Gaussian or a Poisson statistic is used, depending on the kind of noise that it is assumed to be crucial during the detection step. For this reason image restoration process can be treated from a statistical point of view.
When the image formation model is ready, or in jargon of inverse problem theory, the forward model is ready, the image restoration problem lies simply in the inversion of such model. However, this problems turn out to be ill-posed, loosely speaking a suitable solution is very hard to find. In general, the ill-posedness implies that there are too many approximate solutions of the problem which are able to reproduce the data image within a given noise level. Therefore it is necessary to regularize the solution by imposing *a-priori* constraints, derived from information about properties of the object to be restored. The statistical Bayesian approach enables to integrate the available prior knowledge in the restoration problem. This approach is called maximum a-posteriori method and is that adopted in this thesis. In general, the prior information can be introduced by modeling the object as a realization of a Markov random field. Markov random field represents an elegant mathematical tool to bring out various desired statistical properties of the object, by means of the choice of properly so called potential functions. The final result is that the image restoration problem becomes a constraint minimization problem of a particular functional.
The last step is to find a good algorithm to solve the minimization problem. A good algorithm must provide a suitable solution in the fast way possible and must be easily adaptable to the different constraints impose to the solution, e.g. noise assumption and a-priori information. A class of algorithms corresponding to these requirements can be obtained by means of a general approach known as split-gradient method.
In conclusion to obtain a suitable restored object starting from a microscope image it is necessary to define:

- a properly image formation model;
- a properly object model;
- a properly algorithm.

The aim of this thesis is to develop a general toolbox for image restoration in three-dimensional fluorescence microscopy, able to satisfy all these requisites. The thesis is organized as follows.

Chapter 1 is devoted to a general introduction of the image restoration problem associated to fluorescence microscopy. A description of the fluorescence physical phenomena is given and the advantages and drawbacks of its application in microscopy are exploited. The various three-dimensional fluorescence techniques used in life-science laboratories nowadays are described.

Chapter 2 describes the image formation process for a fluorescence microscope. We discuss the assumptions made to model the microscope as a linear and shift-invariant system. Under this assumption the image formation process can be modeled by a convolution of the object (specimen) with the point spread function of the system. Moreover a discretization of this model is derived in order to put it in a statistical framework. Statistical framework is successively used to include in the model information about the noise introduced during the recording process.

Chapter 3 is devoted to the modeling of the point spread function. A rigorous vectorial diffraction theory is used to derive an integral representation of the point spread function for the various microscopy techniques introduced in Chapter 1. The most common abberation sources are taken into account to derive realistic point spread functions. This model is utilized to compare the resolution power of the different microscopy techniques.

In Chapter 4 the image restoration problem is approached from a statistical point of view and by likelihood method is transformed in a constraint minimization problem of a suitable functional. Classical Richardon-Lucy and iterative space reconstruction algorithms are derived respectively assuming Poisson and Gaussian distribution of the noise introduced during the recording process. However, it is well-known that these algorithms are not able to find a suitable solutions since, loosely speaking, they must to look for it in a too broad set of possible solutions. Regularization by means of Bayesian approach is introduced to include in the restoration process prior information of the object to be restored. Therefore, the set of possible solutions is reduced. The image restoration problem is still transformed in a constraint minimization problem of a regularized functional.

In Chapter 5 Markov random field theory is introduced to model object prior information required by the Bayesian approach. In general, the prior information can be introduced by regarding the object as a realization of a Markov random field. Therefore, the probability distribution of the object is obtained using the equivalence of Markov random field and Gibbs random fields. In such a way the object follows a Gibbs distribution, which can be described in terms of a potential function. Choosing an appropriate potential function different suitable properties of the object can be imposed on the solution. A simple and

well-known regularization is based on the assumption that objects are made of smooth regions, separated by sharp edges. This is called edge-preserving regularization and requires particulary potential functions. In this Chapter we consider several edge-preserving potential functions; i.e convex, non-convex and fuzzy-logic based.

In Chapter 6 we introduce the general split-gradient method to derive iterative algorithms for the constraints minimization of the regularized functionals obtained combining maximum *a-posteriori* approach to Markov random field modeling. An iterative algorithm very easy to implement is derived for each potential function previously proposed. Advantages and drawbacks of the spit-gradient methods are discussed.

Chapter 7 is devoted to validate the derived algorithms for a particular application, such as confocal image restoration. Numerical simulations are performed to give a quantitative and qualitative comparison of such algorithms. Moreover a qualitative comparison is obtained using also real biological images.

Chapter 1

Three-Dimensional Fluorescence Microscopy

> Most of fundamental ideas of science are essentially simple, and may, as a rule, be expressed in a language comprehensible to everyone.
>
> Albert Einstein

The discipline of cell biology was launched by the invention of the light microscope in 1590, when Zacharias Janseen placed two lenses in a tube. Despite its importance, however this invention has been overshadowed by newer microscopy approaches, such as the electron microscopy.
Recently, several new techniques have brought the light microscope back to cutting edge of biological research. One of these innovations is the development of multicolor fluorescent dye molecules (also known as fluorophore, fluorochrome, fluorescent probe). These fluorescent dyes used as selective stains or markers have contributed to make light microscopy an indispensable technology for the observation of biological specimen, especially for live cell observation. Fluorescent dyes can be linked to a variety of biological structures, allowing specific cell structure to stand out from their background. A fluorescent dye molecule absorbs a photon and quickly re-emits a photon of lower energy. For example the green-fluorescent BODIPY FL absorbs light at a wavelength of 507 nm and emits light at 517 nm. Fluorescence microscopy has now maturated to the point that is an essential tools in any life science laboratory.
The goal of this Chapter is to give to the reader the basic knowledge about fluorescence microscopy technique, and the motivations that push the researchers to solve the image restoration process associated to it. Most of the concepts that we introduce will be treated

deep inside in the next Chapters. Moreover, we introduce some aspects that are not strictly correlated to the aim of this thesis, but that can help the reader to understand some basic assumptions and some practical limits for the solution of the image restoration problem.

In Section 1.1 we bring out the physical phenomena of fluorescence and we explore its properties and its drawbacks. In Section 1.2 we describe how fluorescence principle can be coupled to optical light microscopy to obtain a reliable three-dimensional imaging system. In particular we describe the widely used three-dimensional microscopy techniques nowadays, i.e. wide-field, confocal and two-photon excitation. Moreover, their principal limitations are pointed out. In Section 1.3 image restoration is proposed as suitable solution to these limitations.

1.1 Fluorescence Phenomenon

The phenomenon called fluorescence is a specific type of a much broader class of phenomena called luminescence. There are a number of common types of luminescence including radioluminescence, electroluminescence, chemiluminescence, bioluminescence, thermoluminescence, and photoluminescence. Fluorescence is one of the two types of photoluminescence, phosphorescence is the other. By definition, luminescence is the spontaneous emission of optical radiation (infrared, visible or ultraviolet) by matter. One of the most straightforward ways fluorescence can occur is by the application of radiation, such as visible or ultraviolet light, to matter. Under these conditions a stimulus photon is absorbed by the molecule, which accepts the energy by entering into an excited state. Some of the energy is then converted into rotational and/or translational energy, and the remainder is emitted as light of a lower energy than that of the stimulating photon. This process is investigated in some detail in the rest of this Section.

1.1.1 Absorption and emission spectra

Under normal conditions and at room temperature, the electronic state of a molecule will be its lowest possible energy state, known as the ground state. Outside stimuli such as visible or ultraviolet light can put the molecule in an excited state, where one or more electrons occupy higher energy orbitals than in the ground state. The multiplicity of an electronic state is defined in terms of the spin quantum number S of the molecule. The spin quantum number is the absolute value of the sum of the electronic spins in the molecule. The multiplicity of the molecule is then defined as the quantity $2S + 1$, and may be either singlet or triplet. In a singlet state there are an equal number of electrons with negative and positive spins in the molecule, or said in another way, all the electrons spins exist in

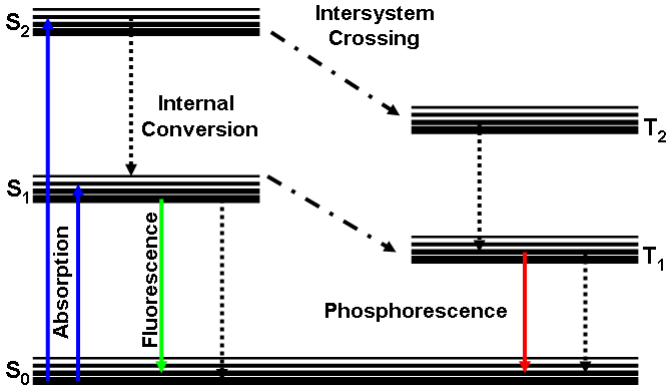

Figure 1.1: Jablonski energy diagram.

pairs. For the singlet state, S=0 and the multiplicity is therefore 1. Conversely, a triplet state is one in which there is one unpaired set of electron spins, S=1 and multiplicity is 3. Figure 1.1 represents the Jablonski energy diagram. The various energy levels of singlet states are referred to as S_0 (ground state), S_1, S_2 etc., and the triplet states are referred to as T_1, T_2 etc. Additional levels in the diagram are depicted by narrower lines which represent the vibrational energy levels of the molecule. The arrows which go between various energy levels in Figure 1.1 represent electron movement from one state to the next. Transition to an excited state is obtained by the absorption of an incident photon of sufficient energy. The incident photon must possess energy equal to the energy difference between the excited state and the ground state in order for the transition to take place. The initially excited state can be any one of a number of high energy or vibrational states. For this reason, a range of incident photon energies can produce an absorption transition. This range of possible photon energies results in a molecular absorption or excitation spectrum consisting of broad bands rather than lines (see Fig. 1.2). Because the emitted photon has less energy than the excitation photon, the wavelength of the emission is almost always longer than that of the excitation. This is known as Stokes' law.
Once a molecule has absorbed energy there are a number of routes by which it can return to ground state. A return to the ground state can be non-radiative (depicted by dotted lines in Figure 1.1). An example of a radiation-less transition is the generation of heat. Transition between energy states of the same spin state is called internal conversion. An excited molecule may make a transition to an excited triplet state. This is called an intersystem crossing since it occurs between energy states with different spin states. From an excited triplet state, an internal conversion can also occur, from which no light is emitted. However, an excited singlet or triplet state may return to the ground state via radiative decay (depicted by solid lines in Figure 1.1), and a photon is emitted. A radiative transition between a singlet state and the ground state which results in the emission of a photon

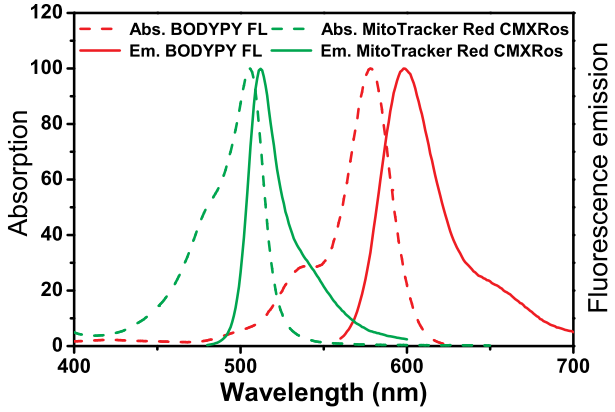

Figure 1.2: Absorption and emission spectra of MitoTracker®Red CMXros and green-fluorescence BODYPY®FL. Source: www.molecularprobes.com.

is called fluorescence. Fluorescence is nearly always a result of a transition between the lowest vibrational level of the first excited singlet state and some vibrational level of the ground state. For a radiative transition from a triplet state to the ground state, which results in the spontaneous emission of a photon, the process is called phosphorescence.
Just as with the absorption spectrum, the emission spectrum for a typical fluorescent dye involves a band of wavelengths. This is due to the fact that the transition can happen from the lowest vibrational level of S_1 to different vibrational level of the ground state.
It is the finite difference between the peaks in the excitation [1] and emission spectra, known as the Stokes shift, that makes fluorescent dyes useful in microscopy. Wavelength selective filters are inserted in the illumination and detection light paths of the microscope to separate and to collect the desired fluorescent emission from the undesired excitation light. A dye which exhibits a large Stokes shift allows for a greater ability to block unwanted radiation and to collect fluorescent emission. This results in improved performance. As can be expected, fluorescent dyes are identified and quantified by their absorption and emission spectra. As an example the absorption and emission spectra of two fluorescent dyes, re-

[1] In this thesis we will use absorption and excitation spectrum in a interchangeable manner, however a difference exists between the two terms. Excitation spectrum is the spectrum of energies emitted (fluorescence) by matter after exposure to radiation while the absorption spectrum of energies absorbed by the matter. Usually you expose the material to a large energy/wavelength range of radiation, measure everything that goes through unabsorbed and by comparing with the incident radiation, you can determine the spectrum of absorbed energies which gives you the absorption spectrum. To obtain the excitation spectrum, you expose the material with radiation, usually just a narrow range of energies, and measure the fluorescence at 90° from the direction of incidence with a detector that can simultaneously measure and sort photons of different energies. This result is an excitation spectrum.

spectively MitoTracker®Red CMXros and green-fluorescence BODYPY®FL are reported in Figure 1.2. With the proper selection of narrow band optical filters, more than one dye can be used simultaneously in a single cell to distinguish different cell characteristics.
Before switch to the description of a fluorescence microscope, three more properties of fluorophore must be introduced, which are extinction coefficient, quantum efficiency (yield) and fading phenomena [Lak99].
The efficiency with which a particular fluorophore absorbs a photon of the excitation light is a function of the molecular cross-section, and the likelihood of absorption is known as the extinction coefficient. Larger extinction coefficients indicate that the absorption of a photon (or quantum) in a given wavelength region is more likely. The quantum yield denotes the ratio of the number of quanta emitted compared to those absorbed (and is usually a value between 0.1 and 1.0). Quantum yield values below 1 are the result of the loss of energy through non-radiative pathways rather than the radiative pathway of fluorescence. The product between the quantum yield and the extinction coefficient is defined as the brightness. A high brightness is always desirable in fluorescence experiments.
A wide spectrum of conditions often come into play that ultimately affect the radiation of fluorescence emission and thus reduce the intensity. The general term for a reduction of fluorescence emission intensity is fading, a catch-all category that is usually further subdivided into quenching and photobleaching.
Photobleaching is defined as the permanent destruction of fluorescence by light-induced conversion of the fluorophore to a chemically non-fluorescent compound. The effects of photobleaching are usually reduced by regulating the light dose, or fluence, of the excitation. The fluence is defined as the intensity, or irradiance, of the excitation multiplied by the time of illumination. Traditionally, photobleaching has been controlled by reducing the fluence on the fluorophore by one of these alternatives: 1) short duration, high intensity illumination, or 2) longer duration and low intensity illumination. The first option is usually achieved using shutters, illuminating the fluorescent specimen in very short pulses. The second option requires very sensitive photodetection devices. An interesting solution to the photobleaching problem called two-photon excitation microscopy has been developed. Since the advantages of the two-photon technique are particularly relevant to 3D microscopy, the topic is addressed more completely in Subsection 1.2.3.
The excited state relaxation process of quenching results in reduced fluorescence intensity through a variety of mechanisms involving non-radiative energy loss and frequently occurs as a result of oxidizing agents or the presence of salts or heavy metals or halogen compounds.
To summarize we can say that if the imaging experiment is carried on playing particular attention to avoid fading effects, emission signal will be proportional to the intensity excitation, or better to the fluence, and to the brightness of the fluorescent molecule.

1.2 Three-Dimensional Fluorescence Microscopy

1.2.1 Wide-field microscopy

The principles of operation for the fluorescence microscope are the same as for basic optical microscopy. The main difference is the incorporation of the appropriate optical filters at the excitation and emission wavelengths. The advantage of fluorescence microscopy is that an emission filter can be inserted in the viewing stage of the microscope to block out the illumination light. In this way only light emitted by the specimen (or specific dyed feature of the specimen) is collected by the detector. The image of the specimen which fluoresces is bright on a dark background. A schematic setup showing the basic components of a

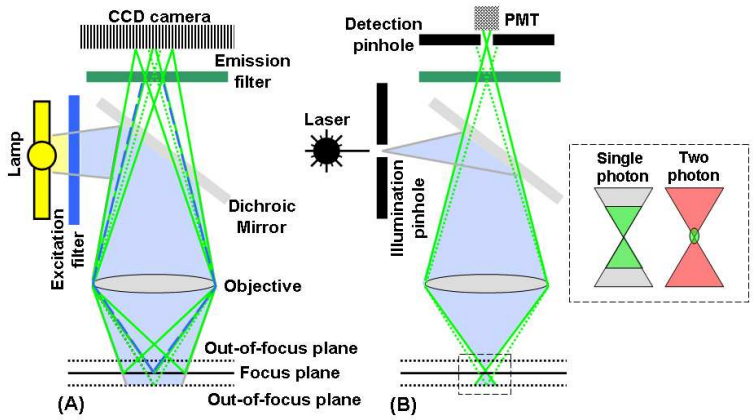

Figure 1.3: Schematic setups of wide-field (A) and confocal laser scanning microscopes (B). Inset represents the difference between the hourglass shape excitation volume (depicted in green) obtained with classical single photon excitation and the well localized excitation volume obtained with two-photon excitation (depicted in red).

fluorescence microscope is given in Figure 1.3 (A). This setup is identical to that of a standard optical microscope with the addition of a special source for illumination, and the filters required to exploit the excitation and emission characteristics described in the previous section. The most common fluorescence microscope is called wide-field microscope (WFM). It is equipped by a lamp, usually an arc lamp, able to emits a broad spectra. When coupled with a narrow-bandpass filter, wavelengths of light near the absorption peak of the fluorescent dye are incident on the specimen. As described in the previous section, the specimen then fluoresces at a longer wavelength. Typically, if the specimen contains only a single absorbance wavelength, a longpass filter is used as an emission filter to maximize the light collection efficiency. In this manner, all light emitted above the cutoff wavelength of the excitation filter is collected by a charge-coupled device (CCD) camera. However, if

multicolor dyes are being used simultaneously in a specimen, badpass filters must be used for both excitation and emission.
The most obvious approach to microscopy in three dimensions is to generate a set of 2D projections at specified intervals of focus along the focus axis [CL05]. One problem with wide-field microscope is that, regardless of where the microscope is focused vertically in a specimen, illumination causes the entire specimen thickness to fluoresce (see Fig. 1.3 (A)). Thus is not unusual that in a given bi-dimensional (2D) image most of the fluorescence light is out-of-focus that can completely obscure the in-focus detail and greatly reduce the contrast of what remains. A general approach to improve this problem is to use technique capable of optical section. Optical sectioning produces images of thin slices of a thick specimens by removing the contribution of out-of-focus light in each image plane. This removal of unwanted light provides greater contrast and permits three-dimensional (3D) reconstruction by computationally combining the image data from a stack of images. A first approach to avoid out-of-focus light is given by computational optical sectioning technique, also known as image restoration technique. A series of wide-field 2D images, optical slices, are collected at different plane thought the specimen, and a computational method derived from information about the process of image formation and recording is used to redistribute out-of-focus light to the original plane. Completely different approaches to reduce out-of-focus light are based on physical principles. Confocal scanning microscopy is presently the most widely used optical sectioning techniques among this category. Multiphoton fluorescence excitation microscopy is another very importation and powerful technique for optical sectioning microscopy. For completeness, even if we do not treat these techniques in this thesis, it is important to remember 3D interfering techniques [NH01] like the standing-wave microscope (SWM) [BFTL93], the 4Pi confocal microscope [HS92], and the incoherent illumination interference image interference microscope (I^5M) [GAS99].
We want to remark that image restoration technique has steadily gained acceptance as an alternative as well as a complement to the to confocal and two-photon excitation microscopy. It is demonstrated that the best performance of confocal and two-photon excitation techniques are attainable when they are combined with image restoration. Moreover, image restoration is essential to render unambiguous imaging improvement obtained by the interfering techniques.
In the following we briefly introduce the principles of confocal and two-photon excitation microscopy, for a complete review we suggest to read [Dia02].

1.2.2 Confocal microscopy

Whereas confocal microscopy can be implemented in many different ways, all the approaches are based on the same concepts. This idea was first described in patent applica-

tion[2] by Minsky [Min61] and subsequently described by him in delightful memoir [Min88]. All confocal techniques share the same fundamental attribute: they are scanning microscopes. "Scanning" means that the image of each optical section is build up by adding information from regions that are sampled in sequence. The central concept of confocal technique is to do two things simultaneously: scan the image by illuminating individual regions in sequence (scanning the illumination) and at the same time mask all but the illuminated regions from providing return light to the detector (scanning the detection). This concept is exploited in the following.
Figure 1.3 (B) shows a schematic setup of a confocal laser scanning microscope (CLSM). The CSLM is a variation of the confocal scanning optical microscope, while the first one uses for illumination a laser source the second one uses an arc lamp or similar broad band source (since laser sources are monochromatic excitation filter can be avoid from the setup of the CLSM). The excitation radiation passes through a pinhole and an objective lens which focuses it to a point at a chosen depth in the specimen. The benefit of this focusing action is readily apparent. Since the light is focused to a single diffraction limited point within the specimen, a much smaller amount of light passes through the specimen relative to conventional fluorescence microscopy with uniform illumination. Unwanted fluorescent emissions will only occur in an hourglass-shaped region (see Fig. 1.3 (B)) of the thick specimen, reducing stray emissions and reflections which obscure the image. In this manner, stray light emissions from points to either side of the point of interest are eliminated. However, there are still fluorescent emissions from in front of and behind the plane of interest to consider (out-of-focus light). This problem is controlled with a second pinhole aperture in the detection stage of the confocal microscope. The light emitted from the spot of interest is focused to a point at the second pinhole aperture, where it passes through in its entirety. The opaque material around the second aperture serves to block most of the remaining out-of-focus light which would otherwise degraded the image. The solid green lines represent the light fluorescing from the desired depth, which passes through the pinhole unhindered. However, unwanted out-of-focus light is not focused to a point at the aperture and is thus blocked from reaching the photodetector. Since the number of photons reached the photodetector is thereby reduced, sophisticate photodetectors, like photomultiplier tubes (PMTs) are required. The stray emissions are represented in the Figure by dashed green lines. Of course, the blocking filter is inserted in front of the detector as in wide-field microscopy to stop the illumination wavelengths from being collected. In this manner, a clear image of one single fluorescing point in the thick specimen is collected and recorded. The entire specimen is scanned in three dimensions to acquire a complete recording of it.

[2]Minsky, who is perhaps better known as the founder of the field known as artificial intelligence, as a young man built a confocal microscope to improve reflected-light images of brain in which the Golgi apparatus was stained in hopes of seeing more clearly the connections within a thick tissue block. Whereas his design and theoretical analysis was exactly correct, there was little interest in his idea at the time. He never published a paper using this technique and received no royalties over the 17-year life of the patent.

Confocal scanning microscopy has an added benefit too: the lateral or xy resolution of the system can be improved beyond what is possible with conventional WFM (see Sec. 3.4). According to Diaspro[Dia02], confocal techniques result in a spatial resolution 1.4 times better than that of a WFM. Where spatial resolution is the ability of the detection system to record details of the specimen under study, in imaging it is usually defined in terms of how close two features can be within an image and still be recorded as distinct (in the next Section we relate the resolution of a system to its point spread function and in Section 3.4 we will give a rigorous criteria to estimate it).
The use of CLSM brings with it several potential problems that do not apply to the WFM. Foremost among them is the increased risk of bleaching. The probability that a molecules bleaches depends on its exposure to the excitation light (fluence). This is the product of the irradiance a molecule receives and the time it receives it. Moreover since the number of photons emitted by the specimen is proportional to the fluence, it is obvious that there is a conflict between high signal-to-noise ratio (SNR) and low bleaching effects. SNR represents a topic point for image restoration, here we briefly summarized the different factors that can be impact on the SNR and the relative consequence for the bleaching.

- **Laser power.** Increasing the laser power is a simply way of increasing the apparent signal but greatly increase the risk of bleaching. Moreover fluorescence saturation must be taken into account: fluorescence molecules in the hourglass-shape region are cycling between excited state and ground state as fast that they can, and greater illumination irradiance provides no additional fluorescence.

- **Pinhole size.** Reducing the pinhole size rejects more out-of-focus light and thus increase the depth discrimination. Despite this improvement images from CLSM have low SNR because of the small number of photons passed trough the pinhole in the short time of scanning. In practice, larger pinholes must often be used when imaging thin specimen or fluorophores that bleach fast.

- **PMT gain.** Increasing the gain on the PMT will increase the signal but also results in higher noise. Such a trade-off may be necessary if bleaching is a problem.

- **Speed.** Noise reduction can be obtained increasing the dwell time of the laser by slowing the scan speed. Again, bleaching risk becomes higher.

- **Fluorophore concentration.** The fluorescent emission intensity is proportional to the concentration of fluorescent dye, but there are practical limitation to this concentration. A very high concentration could have toxic effects on the organism under study, interfere with the process being observed, leave a large amount non-specific fluorescence and lead to self-quenching effects [Lak99]. To avoid these pitfalls, the concentration of fluorescent dye is usually kept low with a resulting low fluorescence intensity that brings a low SNR.

- **Average.** If one keeps the laser power low one can still improve the SNR by collecting several images of the same plane and averaging them. This increases the SNR in direct proportion to the square root of the number of frames that are averages. Obviously repetition of scanning increases the probability of bleaching.

In summary fast scanning with low power and averaging is probably the best way to obtain a good image. However, this solution is inappropriate for studing of fast dynamic in living cell, where speed becomes a primary requisite. The obvious drawbacks of these "solutions" have held the door open for further research in this area. Very promising solutions to the photobleaching problem for 3D imaging microscopy is two-photon microscopy techniques which very recently have been developed.

1.2.3 Two-photon excitation microscopy

The two-photon approach to laser scanning fluorescence microscopy is similar to the confocal technique, but causes much less photo-damage. The two-photon excitation microscope (TPEM) was first demonstrated by Watt Webb and his coworkers [WJW90] in 1990, but the idea was originally proposed by Sheppard et al. [SK78] in 1978. As with the confocal technique, a laser beam is focused at the plane of interest within a labeled specimen and raster scanned across the focal plane. The fluorescence response is detected by a PMT to form the image. The advantage of TPEM lies in the fact that the fluorescent label is excited only at the focus of the beam. This is accomplished by illuminating the specimen with a high power laser which emits at twice the wavelength of the absorption peak of the fluorophore. For example, a fluorophore with peak absorption in the ultra-violet (UV) range of 350-400 nm can be excited in the two-photon mode with a high power laser emitting in the infra-red (IR) range of about 700-800 nm. The trick is to get two infrared photons to collide with a fluorophore near-simultaneously (in the scale of 10^{-18} seconds). The combined energy of two infrared photons excites the fluorophore in an identical way as a single photon of half the wavelength. After excitation, the fluorophore relaxes and fluorescent emission takes place as usual. The statistics involved in having two low energy photons excite a fluorescent molecule at precisely the same time have been calculated. Loosely speaking, since two-photon excitation require two statistical independent photons for excitation process, its rate depends on the square power of the instantaneous intensity. The power of the laser is adjusted so it is only probable that this occurs at the focus of the laser, where there is the highest density of photons. The process can result in a thereby reduction of the excitation volume (see inset Fig. 1.3(B)) leading to an intrinsic optical section capability. The advantages of this technique are multifold [Dia02]. Since there is no out-of-focus fluorescence, the dye is not bleached in the hourglass-shaped region shown in Figure 1.3 (B)), but only precisely at the diffraction-limited spot corresponding to the focus. Since only a single spot is excited, the need for an aperture at the detec-

tor is eliminated. In CSLM, the aperture is a source of reduced SNR since some of the light is necessarily blocked by the aperture. TPE can therefore have a SNR advantage over CSLM as well, since all the fluorescent emission can be collected. Imaging to greater depths is also possible with the two-photon excitation technique. The reason is that IR wavelengths penetrate deeper into most specimens than UV or visible light. There are a few disadvantages for TPEM, most deal with the large and expensive laser necessary to achieve the high power and short pulse (80-150 fs pulsewith and 80-100 MHz repetition rate). Also, the theoretical resolution of a TPEM system is worse than that of CSLM by a factor of about two, since the excitation wavelength is about twice as long (see Section 3.4).

1.3 Image Restoration Problem

In the previous Section we described in a free-and-easy manner the image process formation for various microscope systems and we saw that these systems fail in transmitting all information about the specimen to the image. Image restoration aims at reversing this degradation process. According to Roy Pike "the job of image restoration is to figure out what the instrument is actually trying to tell you". A first requisite to solve the image restoration problem is to obtain a properly mathematical modeling of the image formation process. In this Section we introduce the main components of this model which will be used in Chapter 2 for a very rigorous mathematical formulation of the image formation process.
Under particular assumptions (see Sec. 2.1) image formation process can be completely

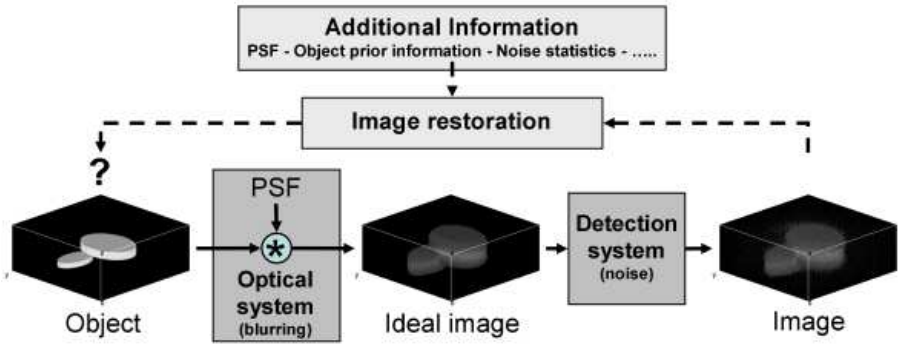

Figure 1.4: Block diagram of the flow of information during the imaging process formation and the restoration process. Left (Specimen): a synthetic three-dimensional representation of a tissue containing two cells . Center (Ideal image): the image, blurred by convolution with the PSF. Right (Image): the image as actually captured by the detection system.

described in terms of the point spread function (PSF) of the system. The point spread function determines how a single point in the sample is being imaged. Representing the sample as a collection of points, or better a cluster of fluorescent molecules, its image can be composed by replacing each point with the PSF weighted by the intensity emitted by the point. This operation is defined as a convolution. We model the formation of an image as a convolution of the sample with the point spread function of the microscope (see Fig. 1.4). For this reason image restoration associated to microscopy is often called image deconvolution

Figure 1.5 shows an example of a wide-field PSF (how to compute a PSF for the various

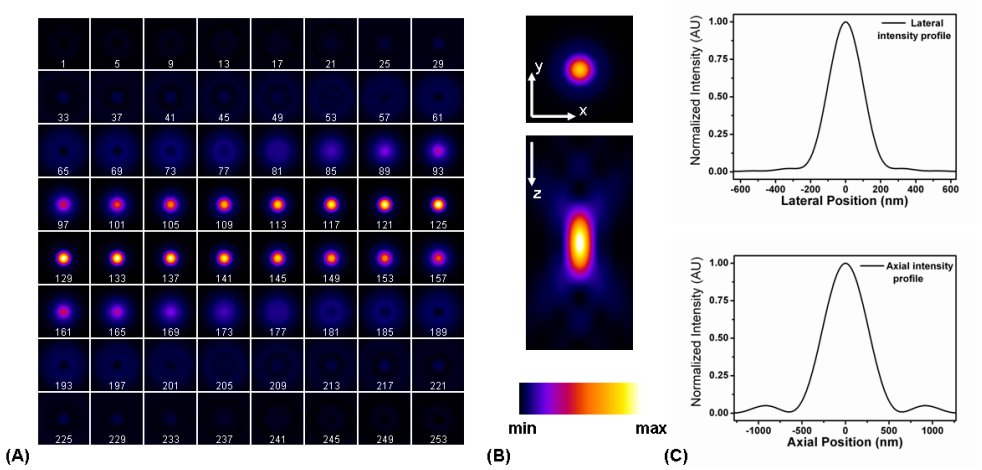

Figure 1.5: Point spread function of a wide-field microscope. Different representations can be used to show a three-dimensional PSF. (A) Montage of through-focus series. Sections at 200 nm z-intervals are shown. Section 129 represents the central plane, that coincides with the plane of sharp focus. (B) Lateral and axial views. If the origin of the Cartesian reference system is places on the center of max of the PSF, lateral view is the plane such that z=0, while axial view is the plane such that y=0. (C) Axial (z=0 and y=0) and lateral (x=0 and y=0) intensity profiles.

microscope systems will be exhaustively studied in Chapter 3) and different ways how to represent it, respectively, serial montage (A), lateral and axial views (B) and finally lateral and axial intensity profiles (C).

Resolution of the system strictly depends on its PSF. In particular a large PSF will impose significant blurring of the sample, hence a reduction of the resolution in the image. Therefore, a comparison between the resolution power of the various systems can be made in terms of their PSFs. The full-width-half maximum (FWHM) of the central peak of the axial and lateral intensity profiles are the most used criteria to compare different PSFs. From Figure 1.5 (C) one can conclude that axial resolution of WFM is worst than its lateral

resolution. We will see that this statement is true also for CLSM and TPEM (see Sec. 3.4). It is readily understood that the knowledge of the PSF represents an important information to solve the image restoration problem. However, another kind of method, known with the name of blind deconvolution, trys to produce both a reconstructed estimate of the PSF as well as the specimen [MC99, Hol92].
Besides blurring by the point spread function, two other factors principally influence the image formation in a fluorescence microscope, i.e. noise and background.
All acquired images are contaminated by noise from a variety of sources. Noise is a stochastic phenomenon that can neither be compensated for nor eliminated. Here we briefly resume the major sources of noise in fluorescence microscopy:

- **Poisson noise** results from the quantum nature of light. Light can also be considered as a series of particles called photons. Photon production by any light source is a statistical process governed by the laws of quantum physics. The source emits photons at random time intervals. The number of photons in a fixed observation interval will result in a number that obeys Poisson statistics. The probability distribution for counting p photons in an observation window of T seconds and given the photon flux in photons per seconds ρ is:

$$P(p|\rho T) = \frac{(\rho T)^p e^{-\rho T}}{p!} \tag{1.1}$$

  It is important to remark that photon noise is not independent to the signal, not additive and its SNR improves slowly with increasing counts, i.e. more intensity or longer integration time. This kind of noise do not depend on the kind of detector used

- **Dark noise** arises from statistical variation in the number to photoelectrons thermally generate in absence of light. Dark noise does not depend on the signal, and it is related to the observation time and temperature, therefore it can be greatly reduced by cooling systems. High performance CCD cameras are usually cooled to a temperature at which dark current is negligible. Similarly, the PMTs can be considered free-dark noise detector.

- **Read-out noise** originates in the process of reading the signal from the sensor. It is typical of CCD cameras and it is strongly depends on the read-out rate. For high read-out rate it increases and becomes a significant component of the overall noise. The read-out noise is additive, Gaussian distributed and independent of the signal. Read-out-noise can be reduced in well-designed electronics.

- **Quantization noise** arises in any kind of detector from the quantization of the recorded amplitudes into a finite number of discrete levels by the analog-to-digit converter (ADC). This noise is additive, is uniformly distributed, is independent of

the signal and dependent of the number of bits in ADC. Quantization noise is very small and usually ignored.

In summary some of these noises can be negligible by proper electronic design and careful operating system. One of them, photon noise, can never be eliminated and thus it forms the limiting case when all other noise have become negligible compared to this. Such situation is typical of CLSM and TPEM. While for WFM, where the number of photons collected is higher, and a CCD is used, it is usual to assume that read-out noise is crucial, and photon noise is neglected.
Background is a third factor that influences image formation in fluorescence microscope. It consists in the amount of intensity that arrives to the sensor due to the auto-fluorescence of the sample, reflection, extern light, etc. In other word all the intensity that does not come from the specimen emission. It is important to remark that background is independent from the signal.
Again a properly estimate of the background and a properly model of the statistical noise involved during the detection process represent two important parameters to address image restoration process in a properly direction.
Finally, all other prior information on the specimen, like non-negativity constraints, or statistical properties, help to find suitable solution to the image restoration problem.

Fluorescence microscopy has become a requisite tool for biologists, particularly for the study of living cells. The ability to link fluorescent dyes to separate biological structures has enabled researchers to study specific cell structures in living cells. Fluorescence microscopy takes advantage of the ability of some molecules to emit light of a characteristic wavelength when excited by another wavelength. Image contrast is improved by using filters to facilitate the fluorescence process and block interfering illumination radiation prior to recording. Moreover optical sectioning techniques, introduce a further spatial dimension on the study of biological structure improving resolution and depth discrimination. Optical sectioning technique can be mainly divided in two categories: computational-based, and optical-based. The Computational techniques obtain thin slice of the specimen inverting mathematically the image formation process. While optical-based techniques like confocal and two-photon excitation use properties of light and fluorescence to obtain the same results. However, the combination of the two approaches usually achieves a better resolution than possible with either method alone. For this reason in this thesis we present a general formulation of the image restoration problem that can be applied to any of the microscopy techniques presented in this Chapter.
In the next chapter we investigate again the image process formation but from a more rigorous point of view. Concepts like noise and blurring are formalized with the aim to obtain an equation that links the properties of the specimen to its image.

Chapter 2

Image Formation

> I have had my results for a long time: but I do not yet know how I am to arrive at them.
>
> Carl Friedrich Gauss

Since image restoration is a typical example of ill-posed problem (see Sec. 4.1) an accurate mathematical modeling of the image formation process is a necessary prerequisite for obtaining a correct formulation of the problem. In this Chapter we analyze the properties of the fluorescence microscope system, showing that under suitable physical assumptions, it can be modeled as a linear shift-invariant system. Moreover, we show that it is quite natural to look for a statistical modeling of the restoration problem.
Microscope imaging system consists, in general, of two parts:

- The first is an apparatus (formed by physical components such as sources, collimators, mirrors, lenses etc.) able to transform the fluorescence emitted by the sample to be imaged (in the following called the object) into a detectable radiation containing useful information about the spatial properties of the object.
- The second is a detector providing measured values of the incoming radiation; this is the part of the system introducing sampling and noise.

Therefore general features of the data are the following:

- Data are discrete and the discretization is not decided by the mathematician but by the physicist or engineer who designed the imaging system; in general, sampling theorems are taken into account in the design and, very often, data are oversampled.

- Data are realizations of random variables, as a consequence of the noise introduced by the detection system.

The final goal of this Chapter is to get a mapping that transforms the spatial distribution of fluorescent molecules inside the object into the radiation incoming on the detector.

In Section 2.1 we present a continuous formulation of the image formation model (infinite-dimensional functional space model), that will be used to explain ill-posedness of the image restoration problem. In Section 2.2 we discretize (finite-dimensional model) the original model formulated in the previous Section, in order to use it for a statistical formulation of the image restoration problem.

2.1 Continuous Model

We assume that images and objects are described by functions of 3 variables, $\mathbf{x} \in \mathbb{R}^3$. We denote by $f_0(\mathbf{x})$ the intensity emitted, at the point $\mathbf{x}$, of the object to be imaged, by $g_0(\mathbf{x})$ its image produced by the optical system before detection (ideal or noiseless image) and by $g(\mathbf{x})$ the final detected image.
At each location $\mathbf{x}$ of the object a certain number of fluorescence molecules is present, each cluster of molecules corresponds to a point source of light whose intensity is proportional to the number of molecules in the cluster (for a discussion about the relation between concentration and emission intensity see Sec. 3.3), therefore $f_0(\mathbf{x})$ can alternatively describe the distribution of fluorescence molecules inside the object at the point $\mathbf{x}$.
In fluorescence microscopy the ideal image is approximately a linear function of the object. Therefore the imaging system defines a linear operator A such that:

$$g_0 = Af_0. \tag{2.1}$$

Linearity means that:

$$A(\alpha f_0^1 + \beta f_0^2) = \alpha A f_0^1 + \beta A f_0^2 = g_0^1 + g_0^2, \tag{2.2}$$

weighted combination of inputs yields weighted combination of outputs. Since the detector measures an intensity the linearity assumption is guaranteed by spatial incoherence of the fluorescence light (no fixed phase relationship exist among the fluorescence molecules that compose the object).
Therefore, if the image formation can be modeled as a linear process, then $g_0(\mathbf{x})$ is a linear superimposition of the values of f_0 and it is given mathematically by:

$$g_0(\mathbf{x}) = \iiint k(\mathbf{x}, \mathbf{x}') f_0(\mathbf{x}') \mathrm{d}\mathbf{x}', \tag{2.3}$$

where $k(\mathbf{x}, \mathbf{x}')$ is the point spread function (PSF) of the linear image system. The term PSF for $k(\mathbf{x}, \mathbf{x}')$ derives from the fact that it is the image of a point like source located at the point $\mathbf{x}'$.
Moreover, we can assume that in fluorescence microscopy the PSF is invariant with respect to translations in the following sense: the image $k(\mathbf{x}, \mathbf{x}_0)$ of a point source located at $\mathbf{x}_0$ is translated by $\mathbf{x}_0$ of the image $k(\mathbf{x} - \mathbf{x}_0, \mathbf{0})$ of a point source located at the origin of the object plane, i.e. $k(\mathbf{x}, \mathbf{x}_0) = k(\mathbf{x} - \mathbf{x}_0, \mathbf{0})$. It follows that $k(\mathbf{x}, \mathbf{x}')$ is a function of the difference $\mathbf{x} - \mathbf{x}'$ and we can write $k(\mathbf{x} - \mathbf{x}')$ instead of $k(\mathbf{x}, \mathbf{x}')$. Such an imaging system is called space invariant and the corresponding PSF is also called space invariant. If the imaging system is not space invariant, its PSF is said to be space variant. In the case of space-variant system, Equation (2.3) becomes:

$$g_0(\mathbf{x}) = \iiint k(\mathbf{x} - \mathbf{x}') f_0(\mathbf{x}') \mathrm{d}\mathbf{x}', \tag{2.4}$$

or also

$$g_0 = k * f_0 = A f_0, \tag{2.5}$$

where the linear operator A becomes the convolution operator. Truthfulness of space-invariant assumption in fluorescence microscopy will be discuss in Chapter 3.
In principle, the PSF can be obtained by solving the direct problem associated with the imaging process or can eventually be measured by generating a point source. In Chapter 3 we will derived a full theory for the computation of the PSF for different kinds of microscope systems.
If we denote the Fourier transform (FT) of a function by the corresponding capital letter, then from the convolution theorem and Equation (2.4), we get:

$$G_0(\boldsymbol{\omega}) = K(\boldsymbol{\omega}) F_0(\boldsymbol{\omega}), \tag{2.6}$$

$\boldsymbol{\omega} \in \mathbb{R}^3$ being the coordinates in Fourier space, also called space frequencies.
The function $K(\boldsymbol{\omega})$ is the optical transfer function (OTF), which describes the behavior of the imaging system in the frequency domain.
The PSF of a microscope, and in general of most optical system, has the following properties:

$$i) \qquad k(\mathbf{x}) \geq 0; \tag{2.7}$$

$$ii) \quad \int k(\mathbf{x}) \mathrm{d}\mathbf{x} < +\infty. \tag{2.8}$$

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. Indeed, the Riemann-Lebesgue theorem implies that the OTF $K(\boldsymbol{\omega})$ is bounded and continuous and that it tends to zero when $|\boldsymbol{\omega}| \to +\infty$.
If the case of microscopy the OTF is zero outside a bounded domain $\boldsymbol{\Omega}$, the PSF is said

to be band-limited and the set $\mathbf{\Omega}$ is called the band of the imaging system. It is obvious from Equation (2.6) that the noiseless image g_0 is also band limited.
In many instances one must also take into account the existence of some background radiation, denoted by $b(\mathbf{x})$, so that we have

$$g_0 = Af_0 + b. \tag{2.9}$$

where, with abuse of notation, g_0 is the ideal image. Finally, the effect of the recording process is the additional of a noise contribution, so that the detected image $g(\mathbf{x})$ (also called noisy image), is given by:

$$g(\mathbf{x}) = (Af_0)(\mathbf{x}) + b(\mathbf{x}) + w(\mathbf{x}). \tag{2.10}$$

It is important to remark that 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 fluorescence microscopy the noise is a random process intimately related to photon emission and acquisition, thus the noise term w is a realization of a random process. This realization is not known in practice. One knows, at most, statistical properties of the random process, such as mean value, variance, etc. We can also know, of course, if it is additive or multiplicative, correlated or uncorrelated, Gaussian or Poisson, etc. These properties, when known, should be used in the treatment of the problem. As noise is intimately related to photon emission and acquisition, the best framework to model the noise corrupted data is provided by a discrete setting. In the next section we reformulate image formation process from a statistical and discrete point of view.

2.2 Discrete and Statistical Model

Because real images are discrete, image formation process must be reformulate from a discrete point of view. We assume that object is partitioned in N voxels, and each voxel can be characterized by a multi-index $\mathbf{n}$, which is the triple of indexes (n_1, n_2, n_3). Each triple defines the position of the voxel within the object volume and taking respectively N_1, N_2 and N_3 values (thus $N = N_1 \times N_2 \times N_3$). However, as usual, we assume that these voxels are ordered to form a vector with N components ($\mathbf{f} \in \mathbb{R}^N$). Since $\mathbf{f}(\mathbf{n})$ represents the amount of fluorescent molecules in the voxel $\mathbf{n}$, from a mathematical point of view $\mathbf{f}(\mathbf{n})$ is the integral of $f(\mathbf{x})$ over the voxel volume. If we assume linearity relation between fluorescence concentration and emission intensity, $\mathbf{f}(\mathbf{n})$ represents also the intensity generated by the fluorescence molecules contained in voxel $\mathbf{n}$.
In fluorescence microscopy, thanks to geometric optics, there is a one-to-one correspondence between the voxels in the image domain and the voxel in the object domain. Therefore,

we assume that the radiation emitted by the object is measured by means of N detectors (real implementation do not use N different detectors, for example in confocal microscopy, because the image is obtained by point-by-point scanning, only one detector is required) characterized again by a multi-index $\mathbf{n}$. We denote by $\mathbf{g}(\mathbf{n})$ the radiation recorded by the detector $\mathbf{n}$. Similarly both the PSF and the background are partitioned in N voxels, thus $k(\mathbf{x})$ and $b(\mathbf{x})$ are replaced by $\mathbf{k}(\mathbf{n})$ and $\mathbf{b}(\mathbf{n})$, respectively ($\mathbf{g}, \mathbf{k}, \mathbf{b} \in \mathbb{R}^N$).
Since a linear model was assumed for the image process formation, then the discretization of the model leads to a $N \times N$ matrix $\mathbf{A}$, relating the unknown object $\mathbf{f}_0$ and the ideal image $\mathbf{g}_0$:

$$\mathbf{g}_0 = \mathbf{A}\mathbf{f}_0. \tag{2.11}$$

Moreover, discretization of convolution integral of Equation 2.4 using cyclic convolution and periodic extension of the voxel values of $\mathbf{f}$ and $\mathbf{k}$ reduces $\mathbf{A}$ to a circulant matrix [BB98], hence the transformation

$$\mathbf{A}\mathbf{f}_0 = \mathbf{k} * \mathbf{f}_0 \tag{2.12}$$

can be easily computed by means of fast Fourier transform (FFT). $\mathbf{A}$ is also known as convolution matrix. The relation between the elements of the matrix $\mathbf{A}$ and the vector $\mathbf{k}$ can be found in [BB98]. We want to remark that, because we assume that each elements of $\mathbf{k}$ is non-negative (see Eq. (2.7)) also each elements of $\mathbf{A}$ is non-negative.
Signal-independent background can be included in the model as an additive term:

$$\mathbf{g} = \mathbf{A}\mathbf{f}_0 + \mathbf{b}. \tag{2.13}$$

In this thesis we assume that $\mathbf{b}$ is a constant vector.
Detection system introduces sampling and noise, noise is a random process, so that each detected value $\mathbf{g}(\mathbf{n})$ is a realization of a random variable (RV) $\mathbf{G}(\mathbf{n})$. Therefore a modeling of the image formation process requires a modeling of the noise, i.e. a model of its probability density distribution.
We denote by $\mathbf{G}$ the multi-valued RV consisting of the set of $\mathbf{G}(\mathbf{n})$ RVs and by $P_{\mathbf{G}}(\mathbf{g}; \mathbf{f}_0)$ its probability density. Note that the probability density depends on $\mathbf{f}_0$. The following assumptions are, in general, accepted as reasonable ones.

- The random variables $\mathbf{G}(\mathbf{n})$ and $\mathbf{G}(\mathbf{l})$ associated to different voxels are statistically independent, so that we can write:

$$P_{\mathbf{G}}(\mathbf{g}; \mathbf{f}_0) = \prod_{\mathbf{n}} P_{\mathbf{G}(\mathbf{n})}\big(\mathbf{g}(\mathbf{n}); \mathbf{f}_0\big). \tag{2.14}$$

- The expected value of $\mathbf{G}(\mathbf{n})$ is just given by the exact value of the incoming radiation so that we have:

$$E\{\mathbf{G}(\mathbf{n})\} = (\mathbf{A}\mathbf{f}_0)(\mathbf{n}) + \mathbf{b}(\mathbf{n}). \tag{2.15}$$

Mainly two different noise models are presented in literature, i.e. the additive Gaussian white noise and the Poisson noise. A more refined model, the so-called Gauss+Poisson noise is proposed by Snyder et. al [SHW93], but it is not studied in this thesis. In the following we present the Gaussian and Poisson case, and, on the basis of the discussion of Sec. 1.3, we state which model is more appropriate for each kind of microscope system.

2.2.1 Additive white Gaussian noise

If we assume that the number of photons collected by the sensor is high, we can consider photon counting noise negligible respect to read-out-noise. This assumption is safe in a wide-field microscope where specimen can be irradiate for long time without relevant photobleaching and all the photons emitted by the specimen are recorded by a CCD. While this assumption is not considered realistic for confocal and two-photon excitation microscopy. Read-out-noise follows a Gaussian distribution, therefore the multi-valued RV $\mathbf{G}$ is given by:

$$\mathbf{G} = \mathbf{A}\mathbf{f}_0 + \mathbf{b} + \mathbf{E}, \tag{2.16}$$

where $\mathbf{E}$ is a multi-valued RV with statistically independent components, all having the same Gaussian distribution, with expected value 0 and variance σ^2, so that:

$$P_{\mathbf{E}}(\mathbf{e}) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{e}\|_2^2\right), \tag{2.17}$$

where $\|\cdot\|_2$ denotes the usual 2-norm:

$$\|\mathbf{e}\|_2 = \Big(\sum_{\mathbf{n}} |\mathbf{e}(\mathbf{n})|^2\Big)^{1/2}. \tag{2.18}$$

From Equations (2.16) and (2.17) the probability distribution of the multi-valued RV $\mathbf{G}$ is given by:

$$P_{\mathbf{G}}(\mathbf{g};\mathbf{f}_0) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^N \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{f}_0 - (\mathbf{A}\mathbf{f}_0 + \mathbf{b})\|_2^2\right). \tag{2.19}$$

This probability distribution links the object function to the image, therefore it can be regarded as the modeling of the image formation process from a statistical point of view.

2.2.2 Poisson noise

If we assume that the number of photons collected is poor, we can considered read-out-noise negligible respect to photon counting noise. This situation is usual in confocal and two-photon excitation where a photomultiplier is used as detector. In confocal microsopy

the pinhole aperture rejects lot of photons emitted by the specimen, while in two-photon excitation microscopy only a small localize region of the sample emits photons. Photon counting noise follows a Poisson distribution, therefore each $\mathbf{G}(\mathbf{n})$ is a Poisson RV with expected value given by Equation (2.15):

$$P_{\mathbf{G}(\mathbf{n})}(\mathbf{g}(\mathbf{n}); \mathbf{f}_0) = \frac{e^{-\left((\mathbf{A}\mathbf{f}_0)(\mathbf{n})+b(\mathbf{n})\right)}\left((\mathbf{A}\mathbf{f}_0)(\mathbf{n}) + \mathbf{b}(\mathbf{n})\right)^{\mathbf{g}(\mathbf{n})}}{\mathbf{g}(\mathbf{n})!}. \tag{2.20}$$

From Equations (2.14) and (2.20) the statistical model for the detected data is given by:

$$P_{\mathbf{G}}(\mathbf{g}; \mathbf{f}_0) = \prod_{\mathbf{n}} \frac{e^{-\left((\mathbf{A}\mathbf{f}_0)(\mathbf{n})+\mathbf{b}(\mathbf{n})\right)}\left((\mathbf{A}\mathbf{f}_0)(\mathbf{n}) + \mathbf{b}(\mathbf{n})\right)^{\mathbf{g}(\mathbf{n})}}{\mathbf{g}(\mathbf{n})!}. \tag{2.21}$$

Different assumptions are made to obtain a computationally tractable image formation process. The final result is a statistical formulation of the image formation model able to include fundamental information like PSF, background and as well as crucial noise distribution.
The most important element of this model is represented by the PSF. For this reason in the next Chapter we present a very rigorous theory to estimate the PSFs for the various microscope systems studied in this thesis.

Chapter 3

Point Spread Function

> The laws of Nature are written in the language of mathematics ... the symbols are triangles, circles and other geometrical figures, without whose help it is impossible to comprehend a single word.
>
> Galileo Galilei

A very important component of the image formation model is given by the point spread function (PSF), which represents the image of the microscope system to an ideal point source. Therefore its estimation is a crucial information needed for image restoration. In general, three approaches have been used to obtain the PSF for deconvolution algorithms. In the experimental approach [dMLCU01, dMSCU03, TFUM06] the images of one or more pointlike objects are collected and used to obtain the PSF. In the theoretical approach [GL91, Hae03], a mathematical model is used to calculate the PSF. Finally, the blind-deconvolution [Hol92] approach is used to estimate the specimen and the PSF from the recorded image simultaneously. The experimental approach has the potential advantage that the PSF closely matches the experimental setup, but the pointlike objects used to record the PSF have to be small and thus are dim. Because of this, PSFs is contaminated by noise. The theoretical approach solves some of these problems. The theoretical PSF has the advantage that it is noise free and can be computed over the relatively large regions of support necessary for successful deconvolution of extended objects. In some cases, however, some of the necessary parameters for PSF computation are not known for a particular setup. In cases in which not all the PSF parameters are known and a PSF measurement is difficult, it may be preferable to estimate the PSF and the specimen function simultaneously, an approach called blind deconvolution. However, further dimensional

instability are introduced in the image restoration problem. Researchers have produced abundant literature for all these approaches, and, to our knowledge, until now no one of them overcomes the others. In this thesis we follow the theoretical approach.
Since the principal elements of a microscope are lenses, a physical model of the PSF is completely based on the diffraction theory. In particular, depending on the type of microscope, diffraction theory is necessary to model the intensity distribution generated by a laser beam focused by an objective lens, or/and to model the intensity distribution generated by the objective lens when it images a point source. Both cases can be seen as a transformation of a plane wave front in a convergent spherical wave front, or viceversa. For this reason the two problems are equivalent, and can be derived using the very same diffraction theory.
Different approximations can be considered to simplify the problem, a common approximation assumes that microscope system is a free-aberration system, however real life introduces different aberration sources, which must be taken in to account for an accurate estimation of the PSF.
The final goal of this Chapter is to use diffraction theory to derive an accurate physical model for the PSFs of the principal types of microscope studied in this thesis.

In Section 3.1 we study the electromagnetic field in the case of free-abberation system, from which we derive the intensity distribution generated by a lens into its focus. In section 3.2 we analyze the principal sources of abberation in fluorescence microscopy, and we show how to model such aberration. In Section 3.3 we formulate for each microscope types the expression of their PSF in terms of the diffraction theory previously introduced. Finally a comparison between the PSFs for the different microscopes is presented.

3.1 Abberation-Free System

The structure of the electromagnetic field near the focus of an aplanatic system which images a point source is extensively studied by Richardson and Wolf [RW59]. They proposed a formalism based on the angular spectrum of plane waves, from which an integral representation of the electromagnetic field is obtained. The very same representation is successively applied to studied the electromagnetic field in the region of focus of a coherent laser beam that emerges from an aplanatic system [BW65]. Here, for the sake of completeness, starting from the original work of Richardson and Wolf, we summarize the principal results that are necessary for a numerical PSF evaluation. Because we are interested to an intensity distribution we will analyze only the electric component of the electromagnetic field.
Figure 3.1 illustrates the notation used in this Section. The lens L represents an optical system of revolution with an optical axis z, which images a point source. The imaging will

be assumed to be aplanatic, i.e. axially stigmatic and obeying the Abbe sine condition. The source will be assumed to be at infinity in the direction of the axis, and to begin with it will be assumed that it gives rise to a linearly polarized monochromatic wave in the entrance pupil of the system. The aplanatic system converts the incident wave in a aberration-free convergent spherical wave in the image space. Moreover, it is assumed that the linear dimension of the exit pupil are large compared with the wavelength. The angle 2α which the diameter of the lens subtends at the Gaussian focus O, is the angular aperture on the image side. OX, OY and OZ are Cartesian rectangular axes, with origin at the focus, with OX in the direction of vibration of the incident electric field (since we assumed linearity polarization in the x-direction), and with OZ along the axis of revolution, pointing away from the lens. P is a generic point in the image space, which is not so close to the exit pupil. (x, y, z) and (r_P, θ_P, ϕ_P) are respectively the Cartesian and the spherical polar coordinates for P, with the polar axis $\theta = 0$ in the axis of revolution of the system and with the azimuth $\phi = 0$ in the OX direction:

$$r_P = \sqrt{x^2 + y^2 + z^2}, \quad \theta_P = \cos^{-1}(z/r_P), \quad \phi_P = \tan^{-1}(y/r_P), \tag{3.1}$$

$$x = r_P \sin\theta_P \cos\phi_P, \quad y = r_P \sin\theta_P \sin\phi_P, \quad z = r_P \cos\theta_P, \tag{3.2}$$

where r_P in $[0, \infty)$, ϕ_P in $[0, 2\pi)$, and θ_P in $[0, \pi)$, and the inverse tangent must be suitably defined to take the correct quadrant of (x,y) into account.
Let

$$\mathbf{E}(P, t) = \mathscr{R}\{\mathbf{e}(P)e^{-i\omega t}\} \tag{3.3}$$

represents the electric field, at the typical point P in the focal region, at time t, with $\mathscr{R}$ denoting the real part. Then it has been shown in [RW59] that the Cartesian components

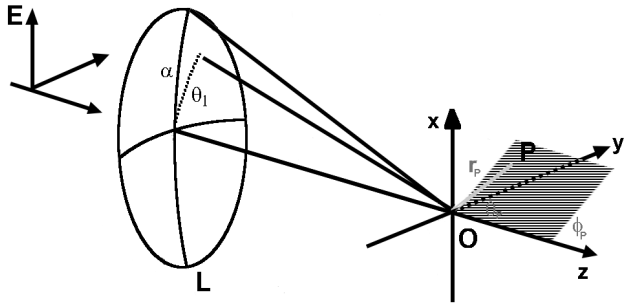

Figure 3.1: Diagram showing light focused by a lens into a single medium.

of the space-dependent part $\mathbf{e}$ of the electric field are given by:

$$\begin{aligned} e_x(P) &= -iA(I_0 + I_2 cos2\phi_P); \\ e_y(P) &= -iAI_2 sin2\phi_P; \\ e_z(P) &= -2AI_1 cos\phi_P; \end{aligned} \tag{3.4}$$

where

$$\begin{aligned} I_0 \equiv I_0(kr, \theta_P; \alpha) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 (1 + cos\theta_1) J_0(kr \sin\theta_1 \sin\theta_P) e^{ikr\cos\theta_1 \cos\theta_P} \mathrm{d}\theta_1, \\ I_1 \equiv I_1(kr, \theta_P; \alpha) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin^2\theta_1 J_1(kr \sin\theta_1 \sin\theta_P) e^{ikr\cos\theta_1 \cos\theta_P} \mathrm{d}\theta_1, \\ I_2 \equiv I_2(kr, \theta_P; \alpha) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 (1 - cos\theta_1) J_2(kr \sin\theta_1 \sin\theta_P) e^{ikr\cos\theta_1 \cos\theta_P} \mathrm{d}\theta_1, \end{aligned} \tag{3.5}$$

and J_0, J_1, J_2 are Bessel functions of the first kind. A is a normalization constant and is proportional to the total amplitude of the transmitted light [RW59]. λ is the vacuum wavelength and $k = 2\pi/\lambda$ is the wave number. We want to remark that if the lens L works, as usual, in a medium different from the vacuum, the wave number becomes $k = 2\pi n/\lambda$, where n is the refractive index in which the lens is embedded, for example $n = 1.515$ for a lens that works in oil.
It is convenient at this stage to introduce the following optical coordinates of the typical point P in the region of focus:

$$\begin{aligned} u &= kr_P \cos\theta_P \sin^2\alpha = kzsin^2\alpha, \\ v &= kr_P \sin\theta_P \sin\alpha = k\sqrt{x^2 + y^2} \sin\alpha. \end{aligned} \tag{3.6}$$

From now on we shall omit the subscript P in the symbol ϕ_P for the azimuthal angle and specify the point P by the three parameters u, v and ϕ. The three integrals defined by Equation (3.5) now become functions of u and v:

$$\begin{aligned} I_0(u, v) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 (1 + cos\theta_1) J_0\left(\frac{v \sin\theta_1}{\sin\alpha}\right) e^{iu\cos\theta_1/\sin^2\alpha} \mathrm{d}\theta_1, \\ I_2(u, v) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin^2\theta_1 J_1\left(\frac{v \sin\theta_1}{\sin\alpha}\right) e^{iu\cos\theta_1/\sin^2\alpha} \mathrm{d}\theta_1, \\ I_2(u, v) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 (1 - cos\theta_1) J_2\left(\frac{v \sin\theta_1}{\sin\alpha}\right) e^{iu\cos\theta_1/\sin^2\alpha} \mathrm{d}\theta_1. \end{aligned} \tag{3.7}$$

Since we want to compute the intensity distribution we have to evaluate the time-averaged electric density energy $\langle \mathbf{E}^2 \rangle$. Starting from Equations (3.3) and (3.4) by elementary computation one obtains:

$$\begin{aligned} k_{lp}(u,v,\phi) &= \langle \mathbf{E}(u,v,\phi,t)^2 \rangle = \mathbf{e}(u,v,\phi) \bullet \mathbf{e}^*(u,v,\phi) = \\ &= |I_0|^2 + 4|I_1|^2 \cos^2\phi + |I_2|^2 + 2\cos 2\phi \mathscr{R}(I_0 I_2^*), \end{aligned} \tag{3.8}$$

where constants of proportionality are omitted, and $\mathscr{R}$ denotes the real part.
So far, we have assumed that the wave entering the system is linearity polarized. In the case of unpolarized wave (or circular polarization) the distribution of the time-averaged electric energy is obtained by integration of the variable ϕ over all possible values:

$$\begin{aligned} k_{up}(u,v) &= \int_0^{2\pi} \langle \mathbf{E}(u,v,\phi,t)^2 \rangle \mathrm{d}\phi = \\ &= |I_0|^2 + 2|I_1|^2 + |I_2|^2, \end{aligned} \tag{3.9}$$

where, again constants of proportionality are omitted. It is very important to remark that for unpolarized wave the distribution intensity at a fixed plane z is radially symmetric. Therefore, in comparison to the polarized case (Eq. (3.8)), a numerical evaluation of Equation (3.9) in the three-dimensional space can be computationally reduced by a properly implementation.
Finally, Richardson and Wolf [RW59] showed that when α is small enough and u and v are not large compared with unity, the I_1 and I_2 integrals can be neglected in comparison to I_0, hence

$$e_x = -iAI_0, \quad e_y = 0, \quad e_z = 0, \tag{3.10}$$

Moreover, I_0 can be simplified as follows:

$$I_0 = e^{iu/\alpha^2} \int_0^1 \rho J_0(v\rho) e^{-\frac{1}{2}iu\rho^2} \mathrm{d}\rho, \tag{3.11}$$

where the new variable $\rho = \theta/\alpha$ is introduced.
Under these assumptions the expression of the time-averaged electric energy density (both in case of linearity polarized and unpolarized waves) becomes:

$$k_{lp/up}(u,v) = |I_0|^2, \tag{3.12}$$

where again constant terms are avoided. Equation (3.12) is of course a scalar approximation.
The parameter α of a microscope objective lens, usually is given in terms of the numerical aperture (NA):

$$NA = n \sin\alpha, \tag{3.13}$$

where n is the refractive index of the medium in which the objective lens works.
Figure 3.2 shows the axial (v=0) and lateral (u=0) profiles of the intensity distribution in the case of unpolarized wave computed using the vectorial theory (Eq. 3.9) and its scalar approximation (Eq. 3.12). In particular we compute the profiles for three different numerical apertures, respectively 0.9, 1.2 and 1.4, in order to study the error introduced by the approximation for different α angles. It is assumed that objective lens works on oil (n=1.514) and images point source of 488 nm wavelength. As expected, higher is the NA of the objective higher is the error introduced by the scalar approximation. Because, for image restoration of real data we are interested to evaluate PSFs for high-numerical aperture objective, in the following we will use always vectorial theory.

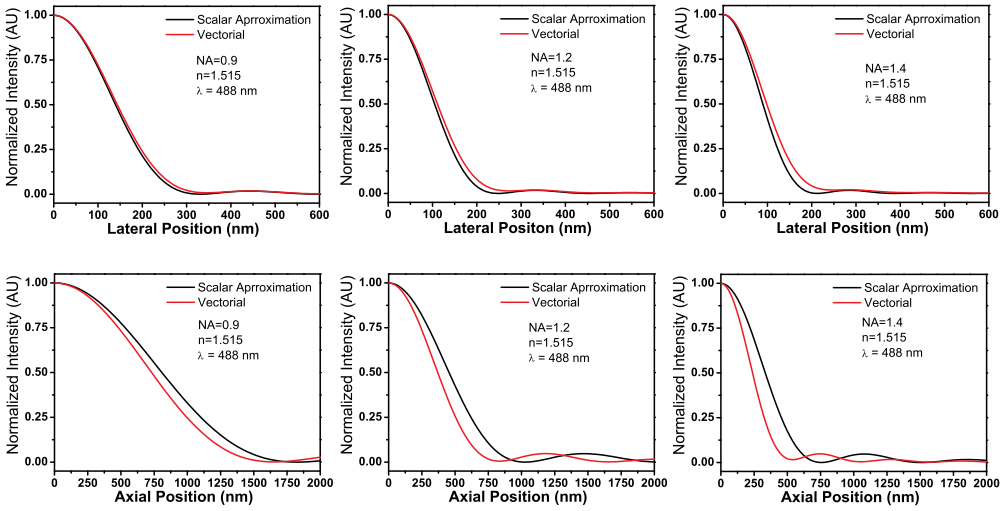

Figure 3.2: Lateral and axial intensity profiles in the case of rigorous vectorial theory and in the case of its scalar approximation. Scalar approximation error increases for high-numerical aperture objective. In particular differences in the axial profile become non negligible. Since we use unpolarized light intensity distribution is radially and axially symmetric, we report only intensity in positive positions.

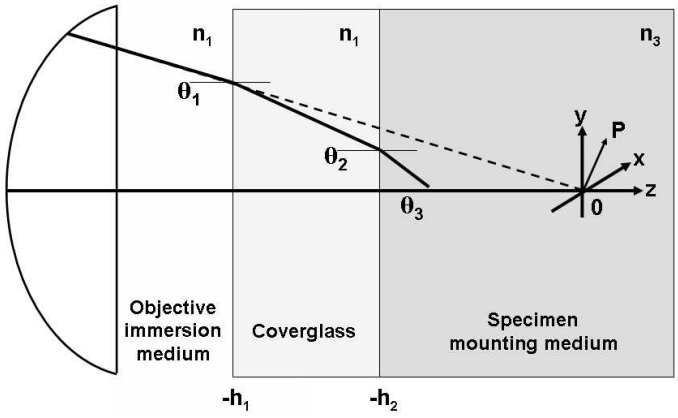

Figure 3.3: Focusing of an electromagnetic wave through a three-layer stratified medium. The origin O of the (x, y, z) reference frame is at the unaberrated Gaussian focal point.

3.2 Aberration System

In the previous Section we described the intensity distribution generated by a plane wave focused in a homogeneous medium. However, usually in fluorescence microscopy the plane wave is focused in a stratified medium, since the objective and the specimen are embedded in different mediums. In particular, the specimen is mounted between a coverslip and a glass slide, which in fact can be another coverslip, and it is immersed in a special mounting medium, such as an aqueous buffer or a more viscous solution based on glycerol. The disparity between the refractive indices of these mediums significantly aberrates the intensity distribution near the focus. Therefore, the model presented in the previous Section should lead to non-accurate estimation of the PSF. The importance of this problem has spurred many researchers to contribute to this field. For a more complete list of references see the volumes by Stamnes [Sta86] and Pawley [Paw06]. Here, we describe a model proposed by Torok and Varga [TV97] for high-numerical aperture objectives able to compute intensity distribution when focusing through a three-layer stratified medium, respectively objective immersion medium, coverglass and specimen mounting medium.
This model is a generalization of the integral representation of the electromagnetic field described in the previous Section, therefore we can use the same notations. In the following, we summarize this model, further details can be found in the original paper [TV97].
Figure 3.3 shows the configuration for illumination through a lens and the three-layer stratified medium. The origin O is again positioned at the Gaussian focus. The first interface, perpendicular to the optical z axis, is placed at $z = -h_1$, the second interface at $z = -h_2$. The wave numbers of the specimen mounting medium, coverglas, objective immersion medium and vacuum are $k_3 = 2\pi n_3/\lambda$, $k_2 = 2\pi n_2/\lambda$, $k_1 = 2\pi n_1/\lambda$ and $k_0 = 2\pi/\lambda$,

respectively. The angles in the mediums nth are denoted by θ_n, and their relation is given by the Snell's law.
In this situation the expressions of the three integrals I_0, I_1 and I_2 (Eq. (3.4)) change in the following way:

$$I_0(x,y,z) = \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 J_0\big(k_1\sqrt{x^2+y^2}\sin\theta_1\big)(T_{2s}+T_{2p}\cos\theta_3)e^{ik_0\Psi_i}e^{ik_3 z\cos\theta_3}\mathrm{d}\theta_1,$$

$$I_1(x,y,z) = \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 J_1\big(k_1\sqrt{x^2+y^2}\sin\theta_1\big)T_{2p}\sin\theta_3 e^{ik_0\Psi_i}e^{ik_3 z\cos\theta_3}\mathrm{d}\theta_1, \quad (3.14)$$

$$I_2(x,y,z) = \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 J_2\big(k_1\sqrt{x^2+y^2}\sin\theta_1\big)(T_{2s}-T_{2p}\cos\theta_3)e^{ik_0\Psi_i}e^{ik_3 z\cos\theta_3}\mathrm{d}\theta_1.$$

where the Cartesian coordinates (x,y,z) are used.
The so-called initial aberration function [TV97] is given by:

$$\Psi_i = h_2 n_3\cos\theta_3 - h_1 n_1\cos\theta_1. \quad (3.15)$$

The transmission coefficients for a three-layer medium are given by:

$$T_{2s,p} = \frac{t_{12s,p}\ t_{23s,p}\ e^{i\beta}}{1+r_{12s,p}\ r_{23s,p}\ e^{2i\beta}}, \quad (3.16)$$

with $\beta = k_2|h_2-h_1|cos\theta_2$ and the Fresnel coefficients for transmission and reflection being given by:

$$\begin{aligned}
t_{nn+1,s} &= \frac{2n_n\cos\theta_n}{n_n\cos\theta_n + n_{n+1}\cos\theta_{n+1}},\\
t_{nn+1,p} &= \frac{2n_n\cos\theta_n}{n_{n+1}\cos\theta_n + n_n\cos\theta_{n+1}},\\
r_{nn+1,s} &= \frac{n_n\cos\theta_n - n_{n+1}\cos\theta_{n+1}}{n_n\cos\theta_n + n_{n+1}\cos\theta_{n+1}},\\
r_{nn+1,p} &= \frac{n_{n+1}\cos\theta_n - n_n\cos\theta_{n+1}}{n_{n+1}\cos\theta_n + n_n\cos\theta_{n+1}}.
\end{aligned} \quad (3.17)$$

It is important to remark that transmission coefficient and Fresnel coefficients are computed assuming propagation of light from the objective to the focus.
Another model for high-numerical aperture objective based on the Huygens-Fresnel construction was proposed in 1993 by Hell at al.[HRCS93]. More recently [EH99], the two models have been shown to be equivalent in the case of a two-layer medium (when $n_2 = n_3$ for example). On the other hand, the Huygens-Fresnell method is computationally most intensive than the method proposed by Torok and Varga.

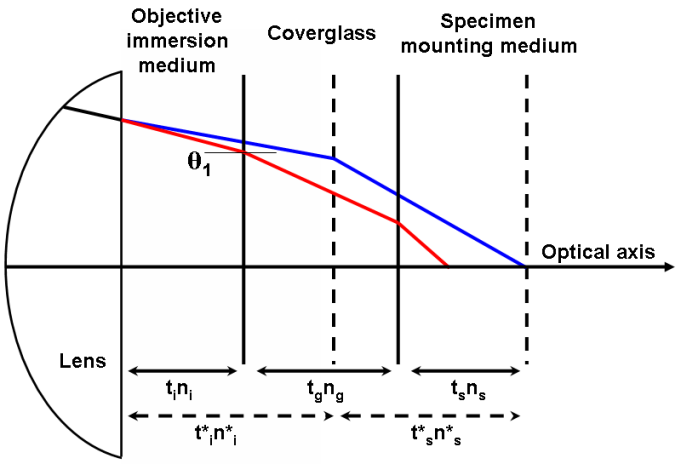

Figure 3.4: Paths of rays in a designed system (blue line) and non-designed system (red line). In the design system the focus is located immediately below the coverglass. Note that for designed conditions the refractive index of the lens is assumed equal to that of the immersion objective medium. This is of course a schematic simplification. The Gibson and Lanni model is valid not only for oil-immersion objective.

3.2.1 Gibson and Lanni notations

In a free-aberration system, we saw that the calculation of the intensity distribution is rather straightforward because it solely depends on the wavelength and the NA. A problem of the aberrated model proposed by Torok and Varga is that it is less directly usable by non-specialists, as practical acquisition conditions do not directly appear as computational parameters. An alternative model is that of Gibson and Lanni [GL91]. One advantageous feature of this model is that it specifically introduces as parameters to compute the intensity distribution the design conditions of use of the objective, as recommended by the manufacturer, and the actual acquisition conditions, when known by the user. However, this model is based on the scalar approximation of the electric field, and we demonstrated that this approximation is not safe for high-numerical aperture objectives.
Haeberle [Hae03] proposed a model, which combines the advantage of the Gibson and Lanni [1] namely the clear introduction of design and actual conditions, with the rigor of the vectorial calculus of Torok and Varga. The result is an easy method to compute the intensity distribution near the focus, that we summarize below. Further detail can be

[1]The original paper of Gibson and Lanni modeled the image of a point source, therefore a diverging spherical wave enter in the front element of the objective lens. While, in our derivation, we consider a plane wave entering in the back aperture of the objective lens, that transforms it in converging spherical wave.

Table 3.1: Relations between the parameters of the Gibson and Lanni model and Torok and Varga model

Parameters acquisition	Actual	Design	Torok and Varga
RI of the objective immersion medium	n_i	n_i^*	n_1
Thickness of the objective immersion medium layer	t_i	t_i^*	-
RI of the coverslip	n_g	n_g^*	n_2
Thickness of the coverslip	t_g	t_g^*	$h_1 - t_s$
RI of the specimen mounting medium	n_s	-	n_3
Depth of the specimen under the coverslip	t_s	-	h_2

found in [Hae03, GL91]
Any objective has only one surface such that when the focus lies upon this surface the intensity distribution is free of aberration. Under design condition for an oil-immersion objective this surface lies immediately below the coverslip, which must have a standard thickness t_g^* and a standard refractive index n_g^*. Also the coverslip must be separated from the objective lens by a layer of immersion oil of a specific thickness t_i^*, the so called working distance, and a specific refractive index n_i^*. However, in real life these design conditions can be violated, for example, when the microscope acquires deep inside a specimen it is obvious that the focus will be not placed immediately below the coverslip, in other words actual conditions usually are different to design conditions (see Fig. 3.4). In the following, the parameters with an asterisk * are values for the design conditions of use of the objective, those without an asterisk are the actual ones, moreover we will denote with t_s the depth of the specimen under the coverslip and with n_s the index of refraction of the mounting medium in which the specimen is embedded.
Table 3.1 resumes the design and the actual parameters that we introduced above, and when exist, the relations between the actual parameters with the parameters used in the Torok and Varga model.
When the designed conditions are violated, a non-negligible optical path difference (OPD) results between the designed (depicted by blu line in Fig. 3.4) and non-design rays (depicted by red line). The basic idea is to derive an aberration function, similar to Equation 3.15, able to take into account the phase changes due to the OPD. Using the Snell law of refraction and assuming that the microscope obeys the Abbe sine condition, Gibson and

Lanni [GL91] proposed an expression of the OPD as a function of the designed and actual conditions, and included it in the scalar model of the intensity distribution near the focus. More recently, Haeberle [Hae03] modified the expression of the OPD in order to include it in the more rigorous vectorial calculus of Torok and Varga:

$$\begin{aligned} OPD &= n_i z\sqrt{1-\left(\frac{NA\rho}{n_i}\right)^2} + \\ &+ n_g t_g\left(\sqrt{1-\left(\frac{NA\rho}{n_g}\right)^2}-\left(\frac{n_i}{n_g}\right)^2\sqrt{1-\left(\frac{NA\rho}{n_i}\right)^2}\right) - \\ &- n_g^* t_g^*\left(\sqrt{1-\left(\frac{NA\rho}{n_g^*}\right)^2}-\left(\frac{n_i}{n_g^*}\right)^2\sqrt{1-\left(\frac{NA\rho}{n_i}\right)^2}\right) - \\ &- n_i^* t_i^*\left(\sqrt{1-\left(\frac{NA\rho}{n_i^*}\right)^2}-\left(\frac{n_i}{n_i^*}\right)^2\sqrt{1-\left(\frac{NA\rho}{n_i}\right)^2}\right) + \\ &+ n_s t_s\left(\sqrt{1-\left(\frac{NA\rho}{n_s}\right)^2}-\left(\frac{n_i}{n_s}\right)^2\sqrt{1-\left(\frac{NA\rho}{n_i}\right)^2}\right). \end{aligned} \tag{3.18}$$

where $\rho = n_i \sin\theta_1/NA$ is the normalized radius, that under the Abbe sine condition is equivalent to the variable already introduced in Equation 3.11. As a consequence, combining the integral equations of Torok and Varga with the Gibson and Lanni method for computing the aberration function, the new expressions of the three integrals I_0, I_1 and I_2 become:

$$\begin{aligned} I_0(x,y,z) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 J_0\big(k_1\sqrt{x^2+y^2}\sin\theta_1\big)(t_{12s}t_{23s}+t_{12p}t_{23p}\cos\theta_3)e^{ik_0 OPD}\mathrm{d}\theta_1, \\ I_1(x,y,z) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 J_1\big(k_1\sqrt{x^2+y^2}\sin\theta_1\big)(t_{12p}t_{23p}\sin\theta_3)e^{ik_0 OPD}\mathrm{d}\theta_1, \\ I_2(x,y,z) &= \int_0^\alpha \cos^{1/2}\theta_1 \sin\theta_1 J_2\big(k_1\sqrt{x^2+y^2}\sin\theta_1\big)(t_{12s}t_{23s}-t_{12p}t_{23p}\cos\theta_3)e^{ik_0 OPD}\mathrm{d}\theta_1. \end{aligned} \tag{3.19}$$

where the Fresnel coefficients for transmission and reflection of Equations 3.17 are computed using the Gibson and Lanny notations ($n_3 = n_s$, $n_2 = n_g$ and $n_1 = n_i$) and θ_3 is related to θ_1 by the Snell's law. Note that the z-defocus term is included in the OPD expression.

3.3 Epi-Fluorescence Microscopy PSF

In the previous Sections we derived an integral representation of the intensity distribution generated by a lens which transforms a monochromatic plane wave in a converging spherical wave. Since a coherent laser can be considered a monochromatic plane wave, the Equations derived both in the case of free-aberrated and aberrated systems can be used to describe the intensity distribution generate by a laser beam in the focus of the objective lens. This distribution intensity is generally called illumination point spread function k_{ill}, since represents the illumination spot used in a scanning system to excite a portion of the specimen.
The same integral representation can be used to compute the image of a point source, since it can be modeled as a divergent spherical front wave collected by the lens and transformed in a plane wave. Modern microscopes use the so called infinity optics: light rays created by the point source in the focal plane of the lens objective are parallel after passing the lens, successively they are captured by a tube lens to form the image. The image distribution intensity generated by a point source is generally called detection point spread function k_{det}, since represents the image of the point in the detected plane of the system. We want to remark that to properly compute k_{det} for an abberation system the transmission (see Eq. (3.16)) and the Fresnel (see Eq. (3.17)) coefficients must be changed taking into account that the light passes through the layers in the opposite direction, i.e. from the specimen to the objective [Hae04]. However, we believe that if the difference between the refractive indexes of the different layers is very small, like for oil, glass and water, the error introduced is very small and we can assume the detection PSF may be computed as the illumination PSF.
In this Section we derive the PSFs for the different types of microscope studied in this thesis; i.e. wide-field, confocal and two-photon excitation microscopes.
Using linear system framework it is possible to formulate these PSFs simply in terms of illumination and detection PSFs. Most of the results reported here are based on [Kem99]. Figure 3.5 (A) shows the schematic setup of a general epi-fluorescence microscopy (GEFM) and its system diagram. Illumination and emission light path share the same objective which strongly reduces the amount of illumination light penetrating the emission light path. The fluorescent molecules inside the object (specimen) are excited by incident light of wavelength λ_{ex}. The excited molecules emit light of wavelength λ_{em} which is collected by the microscope forming the image. In the following derivation we will assume that the illumination and emission light are monochromatic. Although the light sources have broad spectra (except laser sources that can be considered with a good approximation monochromatic), and similarly most fluorescent molecules have broad emission spectra (see Subsec. 1.1.1). Monochromatic assumption can be fulfilled by placing narrow bandpass filters in the illumination and emission light paths. The illumination intensity i_{ill} is the amount of light that is projected from the source onto the sample. The illumination

intensity determines the probability that an excitation photon hits a fluorescent molecule at a certain point in the object. Similarly, the detection intensity quantifies the probability that a fluorescent photon emitted from a point in the object is able to propagate to the detector. Therefore, the detection intensity is dependent on the intensity of both the illumination and the emission light. Whenever the illumination light passes from a region

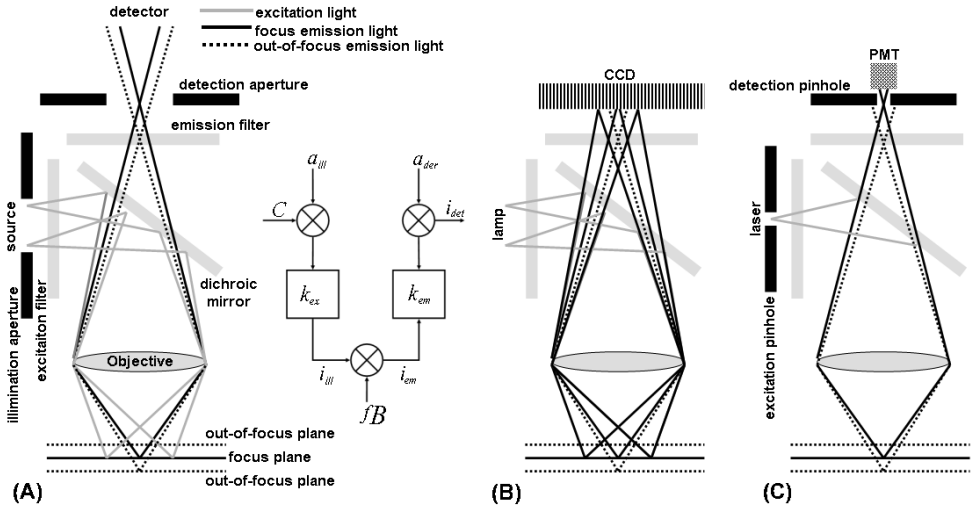

Figure 3.5: (A) Schematic setup and data-flow diagram of a general epi-fluorescence microscope. Schematic setups for wide-field (B) and confocal (C) microscopes.

having one RI to another region having a different RI, some light is scattered back towards the source. The typical case is the difference RI between the coverslip and medium in which the specimen is embedded. Therefore, a dichroic mirror is used to reflect back to the source the back scattered.

The system diagram of the general epi-fluorescence microscope makes the flow of data in the microscope clearer. This diagram also depicts the symbols we use to refer to the various components of the microscope. This model assumes that the illumination and the detection are limited by finite sized apertures, a_{ill} and a_{det} respectively. In a wide-field fluorescence microscope, the illumination aperture is determined by the size of diaphragm, in a confocal microscope by the illumination pinhole. The detection aperture is determined by the size of the detection pinhole in the case of a confocal microscope, and in the case of a wide-field microscope by the CCD camera.

In GEFM a point in the object is illuminated through the illumination aperture a_{ill} of the light source. The light source can be modeled as a collection of point sources spread over the aperture and with constant intensity C. Each point source is projected by the objective on the sample. The illumination intensity can therefore be written as the convolution between

the illumination aperture[2] and the diffraction point spread function:

$$i_{ill}(x) = \big(Ca_{ill}(x)\big) * k_{ill}(x), \tag{3.20}$$

where $*$ denotes the convolution product and k_{ill} the illumination PSF at the illumination wavelength λ_{ex}. For safe in complexity we use a single lateral coordinate x. At the end of this Section we will give the final results for the three-dimensional case.
If we assume that fading and saturation fluorescence effects [Lak99] can be neglected (see Subsec. 1.1.1), the emitted light i_{em} is proportional to the illumination intensity i_{ill}, to the fluorescence molecules concentration in the sample, represented by the function f, and to the brightness B (see Subsec. 1.1.1) of the fluorescence molecules:

$$i_{em}(x) = i_{ill}(x)f(x)B. \tag{3.21}$$

The emitted light is imaged by the same objective onto a photo-sensitive sensor. Behind the objective, the detection aperture limits the emission light with respect to the sensor. The intensity behind this aperture can be modeled as:

$$i_{det}(x) = a_{det}(x)\big(k_{det}(x) * i_{em}(x)\big). \tag{3.22}$$

Combining Equations (3.20)-(3.22) and substituting the object function f with the Dirac function δ, one obtains the PSF of a general epi-fluorescence microscopy:

$$\begin{aligned} k_{GEFM}(x) &= a_{det}(x)\left[\int_{-\infty}^{+\infty} k_{det}(x-\xi)i_{ill}(\xi)\delta(\xi)\mathrm{d}\xi\right] = \\ &= a_{det}(x)k_{det}(x)i_{ill}(0) = a_{det}(x)k_{det}(a_{ill} * k_{ill})(0) = \\ &= a_{det}(x)k_{det}(x)\int_{-\infty}^{+\infty} a_{ill}(\xi)k_{ill}(-\xi)\mathrm{d}\xi, \end{aligned} \tag{3.23}$$

where constant terms are neglected.
Choosing appropriate illumination and detection aperture functions it is possible to derive the PSFs for the major fluorescence microscopes, i.e. wide-field microscope (WFM), confocal laser scanning microscope (CLSM) and laser scanning microscope (LSM). We model both the illumination and the detection aperture using a block function able to represent the size and the shape of the aperture:

$$a_{ill}(x) = \begin{cases} 1, & x \text{ inside the aperture} \\ 0, & x \text{ outside the aperture} \end{cases} \tag{3.24}$$

$$a_{det}(x) = \begin{cases} 1, & x \text{ inside the aperture} \\ 0, & x \text{ outside the aperture} \end{cases} \tag{3.25}$$

[2]Structured illumination techniques [Paw06] use particular illumination mask to obtain suitable intensity pattern in the focus plane. These techniques obtain a resolution comparable to the confocal resolution, but with the main advantages of a wide-field system.

Table 3.2 shows the assumptions that are made on the illumination and detection aperture functions in order to obtain the PSFs of the principal microscopes. It is important to remark that the PSFs of WFM, CLSM and LSM are obtained as limiting cases of the general epi-fluorescence microscopy. In wide-field microscopy (Fig. 3.5 (B)) a lamp is used to

Table 3.2: Illumination and detection aperture functions.

Microscopes	illumination aperture	detection aperture
WFM	$\forall x\ a_{ill}(x) = 1$	$\forall x\ a_{det}(x) = 1$
ideal CLSM	$a_{ill}(x) = \delta(x)$	$a_{det}(x) = \delta(x)$
LSM	$a_{ill}(x) = \delta(x)$	$\forall x\ a_{det}(x) = 1$

excite the specimen. By a proper alignment of the lamp it is possible to obtain a homogeneous incoherent light on the entire field-of-view of the system (Kohler illumination). This situation can be mathematically modeled setting an illumination aperture infinitely large. Again, we can assume that the dimension of the CCD is enough to image all field-of-view. In this case also the detection aperture can be considered infinitely large. By elementary computations (see App. A) it is possible to show that the PSF of a WFM depends only on the detection PSF:

$$k_{WFM}(x, y, z) = k_{det}(x, y, z). \tag{3.26}$$

In confocal microscopy the object is excited through a pinhole aperture. This is often implemented by illuminating the focal plane of the sample with a focused laser beam. The emission light is collected with a light detector, typically a photomultiplier tube (PMT), placed behind a detection pinhole, which is able to blocks the out-of-focus light. Figure 3.5 (C) shows the schematic setup of a confocal microscope. A formulation for the signal i_{det} can be derived from the general epi-fluorescence microscope assuming that both excitation and detection pinhole are infinitely small. From a mathematical point of view this situation can be modeled using the Dirac function δ for both the apertures. It is important to remark that in a confocal microscope, the light that falls through the detection pinhole is collected by a single detector. Therefore, a bi-dimensional confocal image is obtained by scanning the object along the two dimension. In a CLSM the scanning is obtained moving the laser beam over the objective. Under these assumptions the following expression for the PSF is obtained (see App. A):

$$k_{CLSM}(x, y, z) = k_{det}(x, y, z)k_{ill}(-x, -y, -z). \tag{3.27}$$

However, Equation (3.27) represents the PSF of an ideal CLSM; in practice the detection pinhole is not infinitely small, but its size is a compromise between the resolution and the signal-to-noise ratio of the detected signal (see Sec. 1.2). In the real life the PSF of a CLSM is given by (see App. A):

$$k_{CLSM}(x,y,z) = \big(a_{det}(x,y) * k_{det}(x,y,z)\big)k_{ill}(-x,-y,-z), \tag{3.28}$$

where $*$ denotes the bi-dimensional convolution, and a_{det} describe shape and size of the detection pinhole. In this thesis we assume that the detection pinhole has a circular shape. Moreover, it is important to remark that a_{det} describes the shape of the pinhole in the object plane (backprojection pinhole), therefore, loosely speaking, the radius of the pinhole is given by its physical radius divided for the total magnification of the system. Usually, the size of the projection pinhole is reported in Airy disk unit; the radius r of the Airy disk is give by:

$$r = 0.61\frac{\lambda_{em}}{NA}, \tag{3.29}$$

where NA is the numerical aperture of the objective.
In practice, should be useful to completely open the pinhole of the CLSM, hence the detection aperture is infinitely large. This kind of system is called laser scanning microscope, and its PSF is given by (see App. A):

$$k_{LSM}(x,y,z) = k_{ill}(-x,-y,-z). \tag{3.30}$$

Although, laser scanning microscope coupled with classical one-photon excitation laser source is rarely used, while two-photon excitation scanning fluorescence microscope (TPEM) is becoming increasingly popular. In TPEM fluorescent molecules are excited by two photons of half the excitation energy (see Sec. 1.2). Only the presence of two photons at the same time and place enables the excitation of the fluorescent molecule by providing a sufficient amount of energy. Realizing that the illumination point spread function can be seen as a spatial probability function, the chance of a two photon excitation is dependent on the product of the PSFs of both photons, resulting in the following PSF for a TPEM:

$$k_{TPEM}(x,y,z) = k_{ill}^2(-x,-y,-z). \tag{3.31}$$

where it is important to remark that the illumination PSF k_{ill} is computed at a "double" wavelength respect to the classical one-photon excitation.

3.4 Conclusions and Remarks

In this Section we unify the results obtained in this Chapter in order to derive a representation for the PSF of each different microscope in terms of the integrals I_0, I_1 and

I_2. These representations are resumed below and numerically evaluated under the same conditions. The results are compared to derive the major properties of the different microscopes. However, in these derivations we do not treat completely the properties of the fluorescent molecules, but some assumptions are made in order to simplify the study. A rigorous study must model fluorescence molecule as a radiating electric dipole, that emits light in function of the illumination polarization and of its environment [HAF+03, TV97]. For the case when the fluorescent molecule can rotate freely between excitation and emission, the PSF of a wide-field microscope is given by:

$$k_{WFM} = |I_{0det}|^2 + 2|I_{1det}|^2 + |I_{2det}|^2, \tag{3.32}$$

where the notations I_{0det}, I_{1det} and I_{2det} denote that the integrals are computed using the detection wavelength λ_{em} and Equation (3.7) or Equation (3.19), respectively in the abberation-free case and abberation case
Tuning now to CLSM, for unpolarized illumination, or circularly-polarized illumination, and when the fluorescent molecule can rotate freely between excitation and emission, its PSF is given by:

$$k_{CLSM} = \left(|I_{0det}|^2 + 2|I_{1det}|^2 + |I_{2det}|^2\right) \times \left(|I_{0ill}|^2 + 2|I_{1ill}|^2 + |I_{2ill}|^2\right), \tag{3.33}$$

where the notations I_{0ill}, I_{1ill} and I_{2ill} denote that the integrals are computed using the illumination wavelength λ_{ex}. Finally, under the same assumption made for confocal microscopy the PSF of a two-photon excitation microscopy is given by:

$$k_{TPM} = \left(|I_{0ex}|^2 + 2|I_{1ex}|^2 + |I_{2ex}|^2\right)^2. \tag{3.34}$$

The theory described above does not result in an analytical description of the PSF. The PSF has to be calculated numerically as a function of the NA, illumination and detection wavelengths and design and actual acquisition parameters (see Tab. 3.1). In this thesis to evaluate the integrals I_0, I_1 and I_2 we use Gauss-Legendre method [PTV92].
Figure 3.6 (A-C) shows the computed PSFs for the various microscopes using Equations (3.32)-(3.34) and the Torok and Varga + Gibson and Lanni model. To compare the systems in the best conditions we assume that the design acquisition parameters are accomplished (see Figure legend 3.6), therefore aberration effects do not hampered the computed PSFs. From visual inspection it is evident the reduction of out-focus contribute in the case of CLSM and TPEM in comparison to WFM. However, the PSF in the case of two-photon excitation is larger due to the fact that a wavelength twice than in the wide-field and confocal case is responsible for the intensity distribution. For example in the typical case of specimen labeled with green fluorescence protein (GFP) one-photon excitation is performed at 488 nm while two-photon excitation is performed at 920 nm [SBHP+05].
Comparison of PSFs is reported also in terms of lateral and axial intensity profiles (Fig. 3.6 (A-C)). Table 3.3 reports the full-width-half maximum (FWHM) values of the main peaks

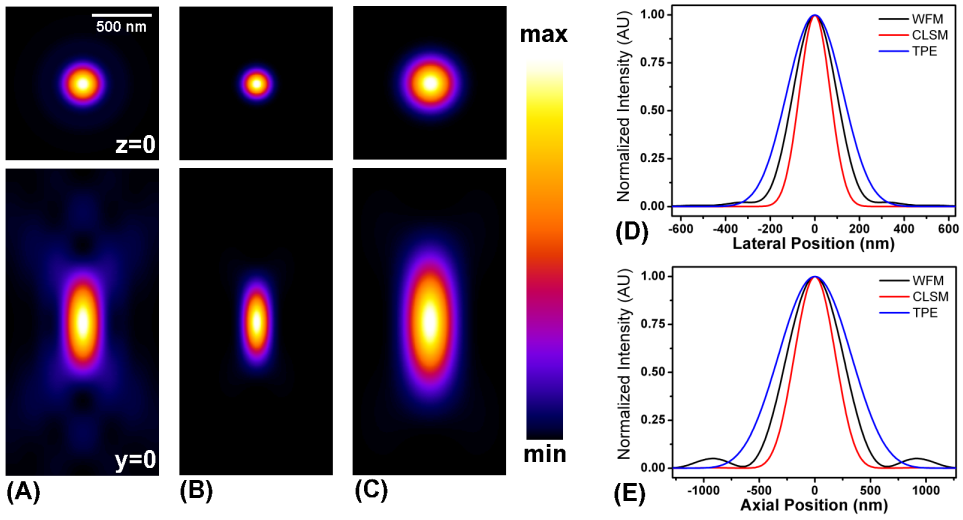

Figure 3.6: Computed PSFs for various fluorescence microscopes. (A) wide-field, (B) ideal confocal and (C) two-photon excitation. All PSFs are computed using: NA=1.4, $n_i = n_i^*$=1.515, $n_g = n_g^*$=1.518, n_s=1.33, $t_g = t_g^*$=0.17 mm, t_s=0 μm, t_i^*=100 μm, one-photon excitation wavelength λ_{ex}=488 nm, two-photon excitation wavelength λ_{ex}=920 nm and emission wavelength λ_{em}=510 nm. For each PSF transaxial (z=0) and coronal (y=0) planes are shown. Normalized lateral (D) and axial (E) intensity profiles for the different microscopes.

Table 3.3: Full-width-half-maximum comparison between wide-field, confocal and two-photon excitation microscopes.

Microscopes	Lateral FWHM (μm)	Axial FWHM (μm)
WFM	0.22	0.56
ideal CLSM	0.16	0.42
TPEM	0.28	0.74

of the lateral and axial profiles for the various microscopes. FWHM of the main peak is relatively simple to measure and hence has become the most common criteria to quantify lateral and axial resolution of a microscope. The same observation previously made from qualitative inspection of the PSFs can be confirmed by the quantitative FWHM analysis. It is important to remark that the resolution performance of a system can not be simply given in terms of point spread function FWHM. For example, from FWHM analysis one can

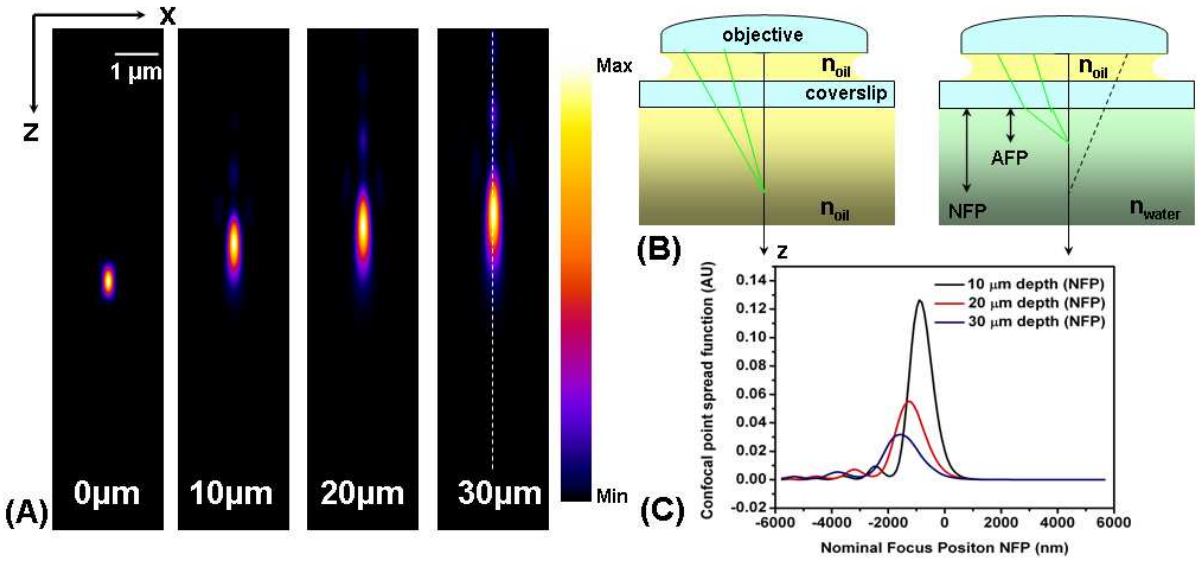

Figure 3.7: (A) Simulation of confocal PSF focused at different depth in water. All PSFs are computed using: NA=1.4, $n_i = n_i^*$=1.515 (oil), $n_g = n_g^*$=1.518, n_s=1.33 (water buffer), $t_g = t_g^*$=0.17 mm, t_s=0, 10, 20, 30 μm, t_i^*=100 μm, excitation wavelength λ_{ex}=543 nm, emission wavelength λ_{em}=565 nm, ideal pinhole size. The first image indicates the ideal situation encountered with the specimen mounted in immersion oil. (B) Schematic representation of the light ray paths in the case of free-aberrated system (left) and aberrated system (right). (C) Intensity axial profile for the aberrated PSFs along the dotted line. Intensity profile for the ideal free-aberrated system is not reported. All the profiles are normalized at the maximum of the ideal profile. This graph shows that the intensity profiles are smeared along the optical axis, and an additional peak appears below the main maximum.

conclude that TPEM resolution is worst than WFM resolution, however it is well-known that optical section performance of TPEM are greater than WFM [Dia02]. The analysis of the optical transfer functions of the different systems can show better optical section capability of confocal and two-photon excitation microscopes [NH01]. Further performance comparisons of the three systems is already given in Chapter 1.
Figure 3.6 shows the confocal PSF in the case of ideal conditions, however it is very interesting to analyzed confocal PSF for non-ideal conditions, where aberration effects take place. Here we present a typical CLSM imaging example in which designed conditions are violated.
While, it is usual that microscopy users do not violate conditions like RI and thickness of the coverslip ($n_g = n_g^*$ and $t_g = t_g^*$), and RI of the immersion medium ($n_i = n_i^*$), it

is usual that for the study of living biological cells and tissues, the imaging is carried on deep inside the specimen. Moreover, the specimen is embedded in medium with refractive index closer to that of the water ($n = 1.33$) than that of immersion oil ($n = 1.515$). Figure 3.7 shows the confocal PSFs for common situations encountered in fluorescence microscopy: the specimen is mounted using a properly glass coverslip ($n_g = n_g^* = 1.518$ and $t_g = t_g^* =$0.17 mm) and is immersed in an aqueous buffer ($n_s = 1.33$), an oil objective ($n_i = n_i^* = 1.515$) NA=1.4 is used to image at different depths inside the specimen, 0 μm, 10 μm, 20 μm and 30 μm respectively.
In this situation a light ray emerging from the oil-immersion objective lens that is coupled to the coverglass with the appropriate oil will not be refracted until it passes the interface from the coverglass into the mounting medium. The light ray is usually only slightly affected by the sample itself and is assumed to carry on straight towards the focal region once it has passed the interface between the mounting medium and the coverslip. In other words, the specimen itself will have a RI that is not much different from that of the mounting medium (for an accurate analysis of aberration introduced by the specimen it self see [SBW07] and therein). The discussion can therefore be restricted to the effects caused by the change in n at the coverslip-medium boundary and to the distance from this interface to the focus point somewhere inside the sample. What effects can be observed?

- A light ray is refracted at the coverslip-medium interface. The angle of the ray is changed; therefore, the different rays focus at different positions along the z-axis than they would in a perfectly matched optical system. In microscopy, n_g is usually larger than n_s, and the actual focus position (AFP) is, therefore, closer to the coverslip than the focus under ideal conditions, that is called nominal focus position (NFP) (see Fig. 3.7(B)). The position of an object will then appear to be further away from the coverslip. If n_g were smaller than n_s, the focus would be further from the coverslip than it should be and the object would then appear to be closer to the coverslip.

- Whenever light is refracted, some light is also reflected (Born and Wolf, 2002). As refraction occurs only when the angle of incidence is lower than the angle of total internal reflection, the NA of the immersion system is effectively reduced.

- Perfect imaging is only possible if the wavefront remains spherical. Any deviation from sphericity results in a larger spread of the focus and hence in a reduction in both spatial resolution and peak intensity.

- This spreading of the focus means that the image of the focal spot focused back towards the confocal pinhole is also spread. This second defocus effect means that less light penetrates the pinhole, and the observed intensity decreases still more.

The same observations like reduction of spatial resolution and peak intensity, as well shift between nominal and focus positions, can be extended also to the TPEM and WFM cases.

3.4.1 Shift-variant PSF

Before concluding we want to point out a possible drawback of the image formation model proposed in Chapter 2. This model assumes that the PSF is constant for all points in the object (shift-invariant model). Microscope optics generally meet this assumption; however, in this Chapter we showed that other issues such as mismatch of immersion and mounting media, cause shift-variations in the PSF, especially in thick specimens. At present, the limitation of a shift-variant model, is represented by computer power: i.e the computational evaluation of $\mathbf{Af}$, that is the base of any kind of algorithm, will become very time-expensive if the operator $\mathbf{A}$ is not a convolution operator, and thus evaluated by means of FFT. Increasing computer power may make it feasible to vary the PSF through the image in the near future.
In microscopy shift-variance is traditionally divided into radial and axial shift-variance. Radial shift-variance is the change of image of a point source as a function of the radial distance of the point source from the optical axis and it is caused by objective lens aberration such as curvature of field, coma and astigmatism. In high quality flat field lenses these off-axis aberrations are corrected, so the system could be considered radial shift-invariant on the most part of the imaging field [VSB$^+$07, VZB$^+$07]. Axial shift-variance is the change of PSF as a function of the position along the optical axis, and as previously demonstrated it depends by mismatch in the refractive indexes, and can not be neglected in thick specimen. To partially solve the problem of image restoration associated to very thick specimen, we can assume that the PSF does not change very rapidly with depth along the optical axis z. Therefore, it is possible to divide the total depth of the image domain in a finite number of sub-domain, in which the PSF can be approximately considered space-invariant. The restored objects obtained from each single process can be merge in a single results (piecewise image restoration). It is obvious that a satisfactory restoration of the complete object can be obtained only using a suitable decomposition of the image domain, for example, an overlapping sub-domain decomposition is high-recommended. However, in this thesis we will apply image restoration to thin sample, hence a linear shift-invariant model is a safe approximation. Shift-invariant image restoration is outside the scope of this thesis.

In this Chapter we has given a well developed framework for the computation of accurate theoretical PSFs of different microscopes. Therefore, combining the results of this Chapter with the results of Chapter 2 a complete model of the image formation process is given. Image restoration problem consists in inverting this model. In the next Chapter we will give the instruments to solve the image restoration problem.

Chapter 4

Image Restoration

We must know. We will know.

David Hilbert

Let assume that we have a complete model in the sense specified in the Section 2.2 and that we have a detected image $\mathbf{g}$. The problem of image restoration is to find an estimate $\hat{\mathbf{f}}$ of the unknown object $\mathbf{f}_0$ corresponding to the image $\mathbf{g}$. The trivial approach should be to look for a solution of the linear equation $\mathbf{Af}+\mathbf{b}=\mathbf{g}$ derived from Equation (2.13), but this approach, in general, is not successful since the matrix $\mathbf{A}$ is ill-conditioned. Indeed, the continuous version of this problem is ill-posed. The practical implication of ill-posedness is that the solution of the linear equation is completely corrupted by an excess of noise propagation. The solution to ill-posedness is that image formation model can not uniquely consist in establishing the equations relating the data to the solution, but it must to be include a model of the noise perturbing the data, and, as far as possible, a model of known properties of the solution.
The final goal of this chapter is to derive a method able to transform the image restoration problem from ill-posed to "well-posed".

In Section 4.1 we explain the difficulties generated by ill-posedness. The most simple way for understanding these difficulties is to discuss the image restoration problem in a continuous setting. A discussion about ill-posedness of image restoration in a discrete setting can be found in [BP06]. In Section 4.2 we present the maximum-likelihood method, this method is able to include the different noise models in the image restoration problem. In particular we derive the maximum-likelihood methods for the Gaussian and Poisson cases. For a study of the Gauss+Poisson case we suggest [BCT$^+$07]. In Section 4.3 we present the maximum *a-posteriori* approach, that reformulates the image restoration problem in-

cluding not only noise information but also available prior information of the solution.

4.1 Ill-Posedness of Image Restoration

In the continuous setting, the relation between the detected image g, the noiseless image g_0 and the unknown object function f_0 is given by:

$$g(\mathbf{x}) = g_0(\mathbf{x}) + w(\mathbf{x}) = (Af_0)(\mathbf{x}) + w(\mathbf{x}). \tag{4.1}$$

where to sake of compactness we neglect the background.
In terms of the Fourier transforms (FTs) we have:

$$G(\boldsymbol{\omega}) = G_0(\boldsymbol{\omega}) + W(\boldsymbol{\omega}) = K(\boldsymbol{\omega})F_0(\boldsymbol{\omega}) + W(\boldsymbol{\omega}); \tag{4.2}$$

and therefore, even if the noiseless image is band-limited, the detected image, in general, is not, because $W(\boldsymbol{\omega})$ may not be zero where $K(\boldsymbol{\omega}) = 0$ (out-of-band noise). Given the detected image g and the PSF k, the problem of image restoration is to determine a sound estimate $\hat{f}$ 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:

$$Af = g, \tag{4.3}$$

which, in Fourier space, becomes:

$$K(\boldsymbol{\omega})F(\boldsymbol{\omega}) = G(\boldsymbol{\omega}). \tag{4.4}$$

This elementary equation clarifies the difficulties of image restoration. 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 OTF is small can modify completely the solution, as we can understand by substituting Equation (4.2) into Equation (4.4) and solving for F the result is given by:

$$F(\boldsymbol{\omega}) = F_0(\boldsymbol{\omega}) + \frac{W(\boldsymbol{\omega})}{K(\boldsymbol{\omega})}. \tag{4.5}$$

All these remarks imply that the problem is ill-posed. The concept of being ill-posed was introduced by Hadamard [Had02] as a mathematical anomaly in the solution of a particular boundary problems for partial differential equations. However, an exhaustive definition of ill-posed is give by a negative characterization: ill-posed problems do not satisfy at least

one of the three conditions required for being well-posed, i.e., existence, uniqueness and continuous dependence of the solution on the data.
It is obvious that the true object f_0 is an approximate solution of Equation (4.3), 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 Equation (4.3), 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). Indeed, the model of the noise is needed in order to clarify in what sense one is looking for approximate solution, on the other the modeling of the object properties must be used for extracting meaningful solutions from the broad set of the approximate ones.

4.2 Maximum Likelihood Approach

Since we assume to know the probability density $P_{\mathbf{G}}(\mathbf{g}, \mathbf{f})$ of the data and since, in this density, the unknown object appears as a set of unknown parameters, at first glance the problem of image restoration appears as a classic problem of parameter estimation. Then the standard approach is the so-called maximum likelihood (ML) estimation. In our specific application, for a given detected image, it consists in introducing the likelihood function defined by:

$$\mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f}) = P_G(\mathbf{g}; \mathbf{f}); \tag{4.6}$$

clearly this is only a function of $\mathbf{f}$ since $\mathbf{g}$ is given and is just the detected image. Then the ML-estimate of the unknown object is any object $\mathbf{f}^*$ that maximizes the likelihood function:

$$\mathbf{f}^* = \arg\max_{\mathbf{f}} \mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f}). \tag{4.7}$$

In such a situation the ML estimator answers to the following question: "Which object $\mathbf{f}$ is most likely to produce the detected image $\mathbf{g}$?".
In our applications the likelihood function is the product of a very large number of factors, so that it is convenient to take the logarithm of this function; moreover, if we consider the negative logarithm (the so-called neglog) the maximization problem is transformed into a minimization one. Therefore we introduce the functional:

$$J_0(\mathbf{f}; \mathbf{g}) = -B \ln \mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f}) + C, \tag{4.8}$$

where B and C are suitable constants that can be introduced in order to simplify the expression of the functional. In the following, when necessary, we will call $J_0(\mathbf{f}; \mathbf{g})$ with the name, discrepancy functional, because it measures the discrepancy between the computed

image associated to $\mathbf{f}$ and the detected image $\mathbf{g}$.
Since the neglog function is strictly convex, the problem of Equation (4.7) is equivalent to the following one:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} J_0(\mathbf{f};\mathbf{g}). \tag{4.9}$$

We reconsider now the two noise models of Section 2.2.

4.2.1 Additive white Gaussian noise case

In the case of additive white Gaussian noise, by a suitable choice of the constants B, C (see Appendix B.2), we obtain:

$$J_0(\mathbf{f};\mathbf{g}) = \|\mathbf{A}\mathbf{f} + \mathbf{b} - \mathbf{g}\|_2^2, \tag{4.10}$$

and therefore the ML approach is just the discrete version of the well-known least-squares (LS) approach. Discussions about the ill-posedness of LS-problem and about the properties of its functional $J_0(\mathbf{f};\mathbf{g})$ can be found in [BLZ08, BP06]. We remark also that this approach is the starting point of the Tikhonov regularization theory of ill-posed problems [TA77]. Therefore, this theory is based on the tacit assumption that the noise affecting the data is additive and Gaussian.
Consequence of the ill-posedness of LS-problem is that their solutions are strongly perturbed by noise. To reduce the set of approximate solutions, since in our application $\mathbf{f}$ describe a concentration that can not be negative, we should consider the minimization of the functional of Equation (4.10) on the non-negative orthant. But, the solution of the new problem is also affected by numerical instability. However it is possible to introduce iterative algorithms converging to the non-negative LS-solutions and having the following semiconvergence property: first the iterates approach a sensible solution and then go away. The two principal iterative methods are provided by the projected Landweber method [BB98] and by iterative space reconstruction algorithm (ISRA). Here we directly present only the iterative scheme for ISRA:

- choose an initial guess $\mathbf{f}^{(0)}$ such that:

$$\mathbf{f}^{(0)}(\mathbf{n}) > 0; \tag{4.11}$$

- given $\mathbf{f}^{(i)}$ compute:

$$\mathbf{f}^{(i+1)} = \mathbf{f}^{(i)} \frac{\mathbf{A}^T\mathbf{g}}{\mathbf{A}^T\mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}}; \tag{4.12}$$

where $\mathbf{A}^T$ denotes the transposed of the matrix $\mathbf{A}$, $\mathbf{b}$ is a constant vector, and quotient and product of two vectors are defined element by element (see Eq. (4))
Discussion about the properties of ISRA can be found in [Pie87].

4.2.2 Poisson noise case

In the case of Poisson noise, if we introduce the so-called Kullback-Leibler (KL) divergence (also known as Csiszar divergence [Csi91]) of a vector $\mathbf{c}$ from a vector $\mathbf{d}$, defined by:

$$KL_D(\mathbf{d},\mathbf{c}) = \sum_{\mathbf{n}} \left\{ \mathbf{d}(\mathbf{n}) \ln \frac{\mathbf{d}(\mathbf{n})}{\mathbf{c}(\mathbf{n})} + \mathbf{c}(\mathbf{n}) - \mathbf{d}(\mathbf{n}) \right\}, \tag{4.13}$$

then, with a suitable choice of the constants B and C (see Appendix B.1), the functional $J_0(\mathbf{f};\mathbf{g})$ is given by:

$$\begin{aligned} J_0(\mathbf{f};\mathbf{g}) &= KL_D(\mathbf{g}, \mathbf{A}\mathbf{f}+\mathbf{b}) = \\ &= \sum_{\mathbf{n}} \left\{ \mathbf{g}(\mathbf{n}) \ln \frac{\mathbf{g}(\mathbf{n})}{(\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n})} + (\mathbf{A}\mathbf{f})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) - \mathbf{g}(\mathbf{n}) \right\}. \end{aligned} \tag{4.14}$$

It is quite natural to take the non-negative orthant as the domain of this functional. Moreover, it is well-known that it is convex, and strictly convex if the equation $\mathbf{A}\mathbf{f} = \mathbf{0}$, has only the solution $\mathbf{f} = \mathbf{0}$ [SV82], non-negative and locally bounded. Therefore it has global minima. The properties of the continuous version of this functional and its minimization are investigated in [MS87, MS89, Mul93]. In particular, in [Mul93], an example is given where the functional does not have a minimum in the classical sense, hence proving the ill-posedness of this minimization problem. As a consequence, we should expect that noise strongly affects the minima of the discrete problem (see Sec. 7.1.2).
An iterative method for the minimization of the functional $J_0(\mathbf{f};\mathbf{g})$ was proposed by several authors, in particular by Richardson [Ric72] and Lucy [Luc74] for spectra and image deconvolution, and by Shepp and Vardi [SV82] for emission tomography. As shown in [SV82], 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 both as Richardson-Lucy method and EM method. The very same method was adopted by Holmes and Liu [Hol88, HYH89] for 3D florescence microscopy. In this thesis we refer to it as a Richardson-Lucy method (RLM); it is as follows [BB98]:

- choose an initial guess $\mathbf{f}^{(0)}$ such that:

$$\mathbf{f}^{(0)}(\mathbf{n}) > 0; \tag{4.15}$$

- given $\mathbf{f}^{(i)}$ compute:

$$\mathbf{f}^{(i+1)} = \mathbf{f}^{(i)} \left(\mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}} \right). \tag{4.16}$$

Again quotient and product of two vectors are defined element by element (see Eq. (4)). We note an interesting property of this algorithm: since the element of the matrix $\mathbf{A}$ and the components of the image $\mathbf{g}$ are non-negative, then all the iterates are automatically non-negative. It must also pointed out that, in general, a positive initial guess is chosen, usually a constant vector, because, as a consequence of the multiplicative structure of the algorithm, if a component of the initial guess is zero, then the same component of all the iterates is zero. For $\mathbf{b} = \mathbf{0}$ convergence proof of the algorithm to a ML solution $\mathbf{f}^*$ is available [VSK85]. However, after a number of iterations, the iterates show the checkerboard degradation effect (see Sec. 7.1.2), indicating that the minima of the KL divergence are not reliable solution. The utility of the algorithm is due to the fact that it has the semiconvergence property[1].
Moreover, if the PSF $\mathbf{k}$ is normalized in such a way that the sum of all its voxel values is 1, for each iterate the following property is true:

$$\sum_{\mathbf{n}} \mathbf{f}^{(i)}(\mathbf{n}) = \sum_{\mathbf{n}} \mathbf{g}(\mathbf{n}). \tag{4.17}$$

This property is also called flux-conservation because it guarantees that the total number of counts of the reconstructed object coincides with the total number of counts of the detected image. It is not satisfied in the case $\mathbf{b} \neq \mathbf{0}$. Moreover the convergence of the algorithm seems not to be proved in such case.
Inclusion of flux-conservation constraints in the RLM for the case $\mathbf{b} \neq \mathbf{0}$ can be obtain by introducing a normalization step in the RLM iterative scheme:

- choose an initial guess $\mathbf{f}^{(0)}$ such that:

$$\mathbf{f}^{(0)}(\mathbf{n}) > 0, \tag{4.18}$$

$$\textstyle\sum_{\mathbf{n}} \mathbf{f}^{(0)}(\mathbf{n}) = \sum_{\mathbf{n}} \{\mathbf{g}(\mathbf{n}) - \mathbf{b}(\mathbf{n})\} \doteq c; \tag{4.19}$$

[1]We believe that it is necessary to clarify the difference between convergence and semi-convergence properties of an iterative algorithm. In our application convergence of an iterative algorithm means that it approaches to the solution $\mathbf{f}^*$ of the problem. Convergence should be demonstrated theoretically, however sometimes only numerically demonstrations are available. Semiconvergence of an algorithm means that it initially approaches to the unknown object $\mathbf{f}_0$ and then it start to diverge. It is obvious that such properties can be demonstrate only by numerical simulation since in real life $\mathbf{f}_0$ is not available. Therefore we want to remark that convergence of an algorithm does not mean that it approaches to the unknown object $\mathbf{f}_0$, that is the best solution of the image restoration problem. For example, RLM for $\mathbf{b}$=0 converges to the solution $\mathbf{f}^*$ of the ML problem, however $\mathbf{f}^*$ is not a suitable solution of the image restoration problem, loosely speaking $\mathbf{f}^*$ is to far from $\mathbf{f}_0$.

- given $\mathbf{f}^{(i)}$ compute:

$$\tilde{\mathbf{f}}^{(i+1)} = \mathbf{f}^{(i)}\Big(\mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}}\Big), \tag{4.20}$$

$$\tilde{c}^{(i+1)} = \sum_{\mathbf{n}} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n});$$

- set:

$$\mathbf{f}^{(i+1)} = \frac{c}{\tilde{c}^{(i+1)}} \tilde{\mathbf{f}}^{(i+1)}. \tag{4.21}$$

4.3 Maximum *a-posteriori* Approach

Section 4.2 shows that, in the case of image restoration, ML problems are ill-posed. This means that one is not interested in computing the minimum points $\mathbf{f}^*$ of the divergence functionals because they do not provide sensible estimates $\hat{\mathbf{f}}$ of the unknown object $\mathbf{f}_0$. In particular, as shown by numerical experience (see Sec. 7.1.2 for the Poisson case), the ML estimate is affected by noise artifacts especially at large number of iterations. A possible solution to obtain an acceptable (regularized) solution is to stop the iterations immediately before the noise amplification, however this solution is difficult in the real data case, therefore an alternative remedy must be found.
We showed that ML approach use only information about the noise with, possibly, the addition of the constraint of non-negativity. In situations where some prior knowledge about the object are available, we may utilize such information for improving the restoration process. In this thesis we adopt a completely probabilistic approach, called Bayesian approach, where the additional information are given in the form of statistical properties of the object. Bayesian approach is presented below.

It is assumed that $\mathbf{f}(\mathbf{n})$ is a realization of a RV $\mathbf{F}(\mathbf{n})$ (for each $\mathbf{n}$). We denote by $\mathbf{F}$ the multi-valued RV consisting of the set of $\mathbf{F}(\mathbf{n})$ and by $P_{\mathbf{F}}(\mathbf{f})$ its probability density, the so called prior. Then, a different interpretation of the probability distribution $P_{\mathbf{G}}(\mathbf{g};\mathbf{f})$ is introduced; this is considered as the conditional probability density of $\mathbf{G}$ when the random variable $\mathbf{F}$ assumes the value $\mathbf{f}$:

$$P_{\mathbf{G}}(\mathbf{g};\mathbf{f}) = P_{\mathbf{G}}(\mathbf{g}|\mathbf{F} = \mathbf{f}). \tag{4.22}$$

For simplicity we will write $P_{\mathbf{G}}(\mathbf{g}|\mathbf{f})$. We use the Bayesian framework to integrate the likelihood density with the prior density. From Bayes theorem we obtain the conditional probability density of $\mathbf{F}$ for a given value $\mathbf{g}$ of $\mathbf{G}$:

$$P_{\mathbf{F}}(\mathbf{f}|\mathbf{g}) = \frac{P_{\mathbf{G}}(\mathbf{g}|\mathbf{f})\ P_{\mathbf{F}}(\mathbf{f})}{P_{\mathbf{G}}(\mathbf{g})}, \tag{4.23}$$

where $P_{\mathbf{G}}(\mathbf{g})$ is the probability density of $\mathbf{G}$. The conditional probability density $P_{\mathbf{F}}(\mathbf{f}|\mathbf{g})$ is called posterior density as it gives the probability for an object distribution $\mathbf{f}$ given, that the observed data is $\mathbf{g}$. Thus, Bayes theorem relates the posterior density $P_{\mathbf{F}}(\mathbf{f}|\mathbf{g})$ to the likelihood density $P_{\mathbf{F}}(\mathbf{g}|\mathbf{f})$ and the prior density $P_{\mathbf{F}}(\mathbf{f})$.
Then, a maximum *a-posteriori* (MAP) estimate of the unknown object $\mathbf{f}_0$ is defined as any object $\mathbf{f}^*$ that maximizes the *a-posteriori* probability density:

$$\mathbf{f}^* = \arg\max_{\mathbf{f}} P_{\mathbf{F}}(\mathbf{f}|\mathbf{g}) = \arg\max_{\mathbf{f}} \mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f}) \frac{P_{\mathbf{F}}(\mathbf{f})}{P_{\mathbf{G}}(\mathbf{g})}. \tag{4.24}$$

As in the case of the likelihood it is convenient to consider the neglog function of $P_{\mathbf{F}}(\mathbf{f}|\mathbf{g})$. If we take into account the definition of Equation (4.8), we can introduce the following regularized functional:

$$J_\mu(\mathbf{f};\mathbf{g}) = J_0(\mathbf{f};\mathbf{g}) + \mu J_R(\mathbf{f}), \tag{4.25}$$

where:

$$J_R(\mathbf{f}) = -\ln P_{\mathbf{F}}(\mathbf{f}), \tag{4.26}$$

is a regularization functional (called also regularization term or penalization term); in the case of white Gaussian noise $J_0(\mathbf{f};\mathbf{g})$ is defined by Equation (4.10) and in the case of Poisson noise by Equation (6.16); μ is the regularization parameter. Note that in the case $\mu = 0$ MAP problem become the ML problem. Therefore the MAP estimates are also given by:

$$\mathbf{f}^* = \arg\min_{\mathbf{f}} J_\mu(\mathbf{f};\mathbf{g}), \tag{4.27}$$

and again one must look for the minimum points satisfying the non-negativity constraint. We conclude by remarking that it is not obvious that a minimum point $\mathbf{f}^*$ of $J_\mu(\mathbf{f};\mathbf{g})$ is a sensible estimate $\hat{\mathbf{f}}$ of the unknown object $\mathbf{f}_0$. In fact, in this formulation we have a free parameter μ. In the classical regularization theory, a wide literature exists on the problem of the optimal choice of this parameter [EHN96] but, as far as we know, this problem has not yet been thoroughly investigated in the more general framework provided by Bayesian regularization.

In this Chapter using the Bayesian statistical approach we transform the image restoration problem in a minimization problem of a suitable regularized functional. Moreover, Bayesian approach is able to include in the restoration process both noise information, both prior object information. These object prior information must be given in the form of statistical properties of the object. In the next Chapter we will show how Markov random field framework can be used to model different general statistical properties of the object simply using local properties. Local smoothness, edges, piece-wise constant, piece-wise quadratic properties of the object can be included in the image restoration to obtain more suitable solutions.

Chapter 5

Object Modeling by Prior Distribution

> One of the principal objects of my research in my department of knowledge is to find the point of view from which the subject appears in the greatest simplicity.
>
> Josiah Willard Gibbs

In the previous Chapter we have proposed the Bayesian approach to include prior information of the object in the image restoration process. This approach requires that the characteristics of the object, such as local smoothness, edges, piece-wise constant, piecewise quadratic, etc. are given in the form of statistical properties of the object, the so-called prior distribution $P_{\mathbf{F}}(\mathbf{f})$. In this thesis, in order to specify a suitable prior distribution, we model the object as a Markov random field (MRF). Markov random fields were first introduced into image processing by Geman and Geman [GG84]. This type of model is based on a similitude with the statistical physics. All the voxel of the object can be idealized as atoms, distributed on a regular lattice of sites. To each atom is associated a particular energetic level (the analogue of the intensity level for the voxel), the energetic level of each atom depends (from a statistical point of view) only on the energetic levels of their neighbors. Thus, in a similar way, we can suppose that the value of each voxel of the object $\mathbf{f}$ only depends on the values of their neighbors voxels. But the most important feature of the MRF is its equivalence with the Gibbs distribution. This equivalence allows to join the general prior distribution $P_{\mathbf{F}}(\mathbf{f})$ (a general characteristics of the object) to the local attributes of the object. In particular these characteristics can be modeled by suitable choices of the so-called potential function.

The final goal of this Chapter is to derive a suitable model able to include in the image restoration problem many different characteristics of the object to be restored.

In Section 5.1 we introduce the MRF and the associated concepts of neighbor system, clique and label. For further details we suggest to read [GG84]. In Section 5.2 we derive different regularization functionals based on the Gibbs prior distribution, in order to model different object characteristics. Moreover, we show that Gibbs prior distribution under particular circumstances can lead to classical Tikhonov regularization.
A simple and well-known regularization supposes that objects are globally smooth, and enforces a roughness penalty on the solution. This kind of regularization, obtained using a quadratic potential function, yields oversmooth solutions. A more realistic object model assumes that objects are made of smooth regions, separated by sharp edges. This is called edge-preserving regularization and requires a non-quadratic potential function. A wide class of non-quadratic potential function is studied. Moreover, in Section 5.3 we present alternative potential functions for edge-preserving regularization based on a the fuzzy-logic framework.

5.1 Markov Random Field

To obtain different prior distributions we can assume that the multi-valued RV $\mathbf{F}$ associated to the object is a MRF. Before understanding MRF, it is essential to introduce the concepts of neighborhood system, clique and label [Li95]. Let S the set of sites representing the three-dimensional voxels:

$$\big\{(n_1, n_2, n_3)|n_1 = 1, 2, ..., N_1, n_2 = 1, 2, ..., N_2, n_3 = 1, 2, ..., N_3\big\} = S. \tag{5.1}$$

The sites in S are related to one another via a neighborhood system H, defined as:

$$H = \{\mathscr{N}_{\mathbf{n}}|\forall \mathbf{n} \in S\}, \tag{5.2}$$

where $\mathscr{N}_{\mathbf{n}}$ is the set of sites neighboring $\mathbf{n}$. The neighboring relationship has the following properties: (1) $\mathbf{n} \notin \mathscr{N}_{\mathbf{n}}$ and (2) $\mathbf{n} \in \mathscr{N}_{\mathbf{n}'} \Leftrightarrow \mathbf{n}' \in \mathscr{N}_{\mathbf{n}}$. The pair (S,H) constitutes a graph in the usual sense where S contains the nodes and H determines the links between the nodes according to the neighboring relationship. Figure 5.1 (A,B) shows two example of neighborhood system for a 3D lattice, a first order neighborhood system (6-neighbors system) and a second order neighborhood system (10-neighbors system), respectively.
A clique c for (S,H) is defined as a subset of sites in S. In particular a clique is a single site or a set of sites such that each site in the clique is a neighbor of all other sites in the clique. For a general neighborhood system, a clique consists either of a single site $c = \{\mathbf{n}\}$, or of a pair of neighboring sites $c = \{\mathbf{n}, \mathbf{m}\}$, or of a triple of neighboring sites $c = \{\mathbf{n}, \mathbf{m}, \mathbf{l}\}$, or of a quadruple of neighboring sites $c = \{\mathbf{n}, \mathbf{m}, \mathbf{l}, \mathbf{o}\}$ and so on. Obviously, the maximum

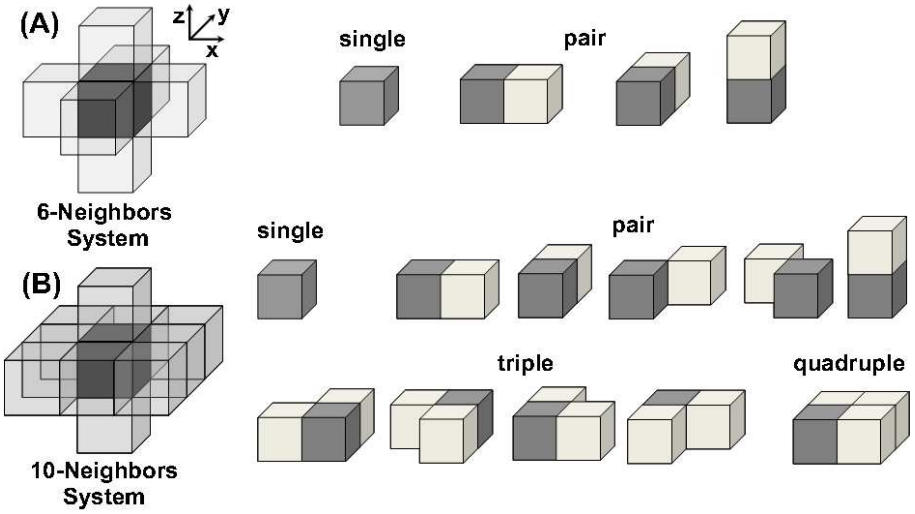

Figure 5.1: (A) Three-dimensional first-order (6-neighbor) system and associated cliques. (B) Three-dimensional second-order (10-neighbor) system and associated cliques.

size of a clique strictly depends on the type of neighborhood system considered (see Fig. 5.1). The collections of single-site, pair-site, triple-site and quadruple-site cliques will be denoted by $\mathscr{C}'$, $\mathscr{C}''$, $\mathscr{C}'''$ and $\mathscr{C}''''$, respectively, where

$$\begin{aligned}
\mathscr{C}' &= \{\mathbf{n}|\mathbf{n} \in S\};\\
\mathscr{C}'' &= \{\{\mathbf{n},\mathbf{m}\}|\mathbf{m} \in \mathscr{N}_{\mathbf{n}}, \mathbf{n} \in S\};\\
\mathscr{C}''' &= \{\{\mathbf{n},\mathbf{m},\mathbf{l}\}|\mathbf{n},\mathbf{m},\mathbf{l} \in S \text{ are neighbors to one another}\};\\
\mathscr{C}'''' &= \{\{\mathbf{n},\mathbf{m},\mathbf{l},\mathbf{o}\}|\mathbf{n},\mathbf{m},\mathbf{l},\mathbf{o} \in S \text{ are neighbors to one another}\}.
\end{aligned} \tag{5.3}$$

The notation and definition of cliques with higher size are obvious. Figure 5.1 shows for the two neighborhood systems their respectively types of clique. Finally we assume that each RV $\mathbf{F}(\mathbf{n})$ of the field $\mathbf{F}$ take a value $\mathbf{f}(\mathbf{n})$ in a set of label $\mathscr{L}$. In our case the set of label $\mathscr{L}$ corresponds to the set of possible intensities (or equivalently the set of possible concentrations).
If we denote by $\mathbf{F}(\mathscr{N}_{\mathbf{n}})$ the set of the RVs associated with the voxels of $\mathscr{N}_{\mathbf{n}}$, $\mathbf{F}(\mathscr{N}_{\mathbf{n}}) = \{\mathbf{F}(\mathbf{m})|\mathbf{m} \in \mathscr{N}_{\mathbf{n}}\}$. Then $\mathbf{F}$ is a MRF if for all $\mathbf{n} \in S$:

$$i) \qquad P_{\mathbf{F}(\mathbf{n})}(\mathbf{f}(\mathbf{n})) \geq 0 \text{ (Positivity)}; \tag{5.4}$$

$$ii) \quad P_{\mathbf{F}(\mathbf{n})}(\mathbf{f}(\mathbf{n})|\mathbf{f}(S-\{\mathbf{n}\})) = P_{\mathbf{F}(\mathbf{n})}(\mathbf{f}(\mathbf{n})|\mathbf{f}(\mathscr{N}_{\mathbf{n}})) \text{ (Markovianity)}, \tag{5.5}$$

where the first term of the Markovianity condition is the conditional probability of $\mathbf{F}(\mathbf{n})$ taking a value $\mathbf{f}(\mathbf{n})$, for given values of $\mathbf{f}$ in the other voxels except $\mathbf{n}$, and the second term

is the conditional probability of $\mathbf{F}(\mathbf{n})$ taking a value $\mathbf{f}(\mathbf{n})$, for given values of $\mathbf{f}$ in their neighbors. Thus Markovianity condition states that each random variable $\mathbf{F}(\mathbf{n})$ depends on the other random variables in $\mathbf{F}$ only through its neighbors in $\mathbf{F}(\mathscr{N}_{\mathbf{n}})$.

5.2 Gibbs Random Field

The characteristics of an object such as local smoothness, piecewise constant regions, etc. can be statistically modeled by the conditional probabilities. However, specifying such characteristics with conditional probability in the MRF framework is, in general, very difficult [Bes74]. This problem has been simplified in the case of MRF by the application of the Hammersley-Clifford theorem [HC71].

Theorem 1 *Hammersley-Clifford.* ***F*** *is an MRF on a set S with respect to H if and only if* ***F*** *is a Gibbs random field (GRF) on S with respect to H and vice-versa.*

A multi-value random variables $\mathbf{F}$ is said to be a Gibbs random field (GRF) on S with respect to H if and only if its configurations obey Gibbs distribution. Hence the resulting object distribution is a Gibbs distribution [Bes74], given by:

$$P_{\mathbf{F}}(\mathbf{f}) = \frac{1}{Z} \exp\Big\{ -\frac{1}{\beta} U(\mathbf{f}) \Big\} \tag{5.6}$$

where Z is a normalization constant, β is the Gibbs hyper-parameter and $U(\mathbf{f})$ is the energy function.
The energy function can be expressed as a sum of all clique potentials:

$$U(\mathbf{f}) = \sum_{c \in \mathscr{C}} \varphi_c(\mathbf{f}), \tag{5.7}$$

where, $\mathscr{C}$ is the collection of all possible cliques and $\varphi_c(\mathbf{f})$ is the potential function which depends only on the value of $\mathbf{f}$ at the sites into the clique c. It is obvious that higher is the energy function associated to an object $\mathbf{f}$, lower is its prior probability $P_{\mathbf{F}}(\mathbf{f})$, thus higher is the penalization for this object to be a suitable solution.
Thus, using the Gibbs prior function:

$$P_{\mathbf{F}}(\mathbf{f}) = \frac{1}{Z} \exp\Big\{ -\frac{1}{\beta} \sum_{c \in \mathscr{C}} \varphi_c(\mathbf{f}) \Big\}, \tag{5.8}$$

one is able to express general properties of $\mathbf{f}$ by local properties on the set of cliques $\mathscr{C}$ for obtaining a suitable regularization term. Combining Equations (4.26), (4.25) with Equation (5.8) one obtain the following regularization functional:

$$J_R(\mathbf{f}) = \sum_{c \in \mathscr{C}} \varphi_c(\mathbf{f}), \tag{5.9}$$

where the regularization parameter μ becomes equal to $1/\beta$.
In general clique potentials are functions of a derivative of the object [GR92]:

$$J_R(\mathbf{f}) = \sum_c \varphi\big(D_c^k(\mathbf{f})\big), \tag{5.10}$$

where D^k denotes a kth order derivative, k=0, 1, 2 or 3 and states also the subset of cliques in which the summation ranges. The order of the derivatives are chosen depending on the kind of object that is sought.
For $k = 0$ the summation ranges over single-site clique, D_c^0 denotes the identity and a constraint on the dynamic of the solution is imposed (see Subsec. 5.2.1). For k=1 the summation ranges over double-site cliques, D_c^1 denotes the discrete approximation of the first derivative and using appropriate potential functions promotes smoothing solution (see Sec. 5.2.1) or solution with piecewise constant areas (see Sec. 5.2.2). For k=2, 3 the terms D_c^k corresponds to discrete approximation to the differences between elements of the gradient vector (k=2) and Hessian matrix (k=3) at adjacent voxels. The summation in such cases ranges over cliques derived from system neighbors of order higher than two and are used to promote piecewise linear areas (k=2) or piecewise quadratic areas (k=3) in the solution [GR92].

In this thesis we are interested in the cases k=0 and $k = 1$ associated to a 3D second order neighbors system (see Fig. 5.1(B)). Not only the choose of the order of the derivative and the choose of the neighbors system have an important rule on the constraints imposed to the solution, but in the same way, also the form of the potential function φ is a very important parameter to control the characteristic of the object that is sought. In the following we present a wide range of potential functions able to impose different characteristics of the object. Moreover, we show how using a quadratic potential function it is possible to derive the classical Tikhonov regularization. Then different edge-preserving potential functions are proposed and discussed.

5.2.1 Quadratic potential functions

We start from Equation (5.10) in the case of k=0, thus the summation range over all the single-site cliques (all the voxels) and D^0 denotes the identity operator. Then, if we use a quadratic potential function $\varphi_{QP}(t) = t^2$, we obtain:

$$J_R(\mathbf{f}) = \sum_{c \in C'} (D_c^0(\mathbf{f}))^2 = \sum_{\mathbf{n}} \mathbf{f}(\mathbf{n})^2 = \|\mathbf{f}\|_2^2, \tag{5.11}$$

where $\|\cdot\|_2$ denotes the usual 2-norm (see Eq. (2.18). This term is exactly the Tikhonov [TA77] regularization term in the case of "default object" $\bar{\mathbf{f}}$ equal to the constant vector

zero:

$$J_R(\mathbf{f}) = \|\mathbf{f} - \bar{\mathbf{f}}\|_2^2. \tag{5.12}$$

This regularization term is extensively studied in tree-dimensional fluorescence microscopy. In the iterative constraints Tikhonov-Miller algorithm (ICTM) Tikhonov regularization term is combined with the Gaussian likelihood and the derived functional is minimized using the method of conjugate gradients [VJ97, LB91]. Non-negativity constraint in this algorithm is implemented simply setting the negative intensities to zero after each iteration. Like the ICTM algorithm, the Carrington [CLM+95] algorithm minimized the same functional under the non-negativity constraints. Finally, Conchello and McNally incorporated Tikhonov regularization term in the RLM, hence exploiting also the Poisson case [CM96]. For further discussion about different applications of the regularized term of Equation (5.11) see [VGJ99].
It is well known [CM96, MC97] that this kind of regularization simply constraints the dynamic of the solution, i.e. reduces the presence of isolated bright spots and biases the solution to low intensity values (see Sec. 7.1.3).
In the following we will refer to this kind of regularization with the name Tikhonov regularization.

Again, we start from Equation (5.10) but in the case of k=1, thus the summation range over all the pair-site cliques and D^1 denotes the first order derivative:

$$J_R(\mathbf{f}) = \sum_{c \in C''} \varphi\big(D_c^1(\mathbf{f})\big) = \sum_{\mathbf{n}} \sum_{\mathbf{m} \in \mathcal{N}_\mathbf{n}} \varphi\left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})\delta}\right), \tag{5.13}$$

where δ and $d(\mathbf{n},\mathbf{m})$ are two scaling parameters, the firs tuning the value of the gradient above which a discontinuity is detected, the second taking into account the different distances between the voxels of a clique. For example, in a real case in which sampling in the lateral direction ($delta_{xy}$) is different from sampling in the axial direction ($delta_z$), $d(\mathbf{n},\mathbf{m})$ are defined as follows:

$$d(\mathbf{n},\mathbf{m}) = \begin{cases} 1, & \mathbf{m} \in \big\{(n_1 \pm 1, n_2, n_3), (n_1, n_2 \pm 1, n_3)\big\} \\ \sqrt{2}, & \mathbf{m} \in \big\{(n_1 \pm 1, n_2 \pm 1, n_3)\big\} \\ delta_z/delta_{xy}, & \mathbf{m} \in \big\{(n_1, n_2, n_3 \pm 1)\big\}. \end{cases} \tag{5.14}$$

If we use again quadratic potential function $\varphi_{QP}(t) = t^2$ and we set δ equal to one we obtain:

$$J_R(\mathbf{f}) = \sum_{\mathbf{n}} \sum_{\mathbf{m} \in \mathcal{N}_\mathbf{n}} \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})}\right)^2 = \|L\mathbf{f}\|_2^2, \tag{5.15}$$

where L is the weighted Laplacian operator, which can be expressed in the form of Equation (5.12):

$$J_R(\mathbf{f}) = \|\mathbf{f} - \mathbf{C}\mathbf{f}\|_2^2, \tag{5.16}$$

where $\mathbf{C}$ is deduced from the $3 \times 3 \times 3$ matrix:

$$\begin{pmatrix} 0 & 0 & 0 \\ 0 & delta_{xy}/delta_z & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 1 & 1/\sqrt{2} \\ 1 & 0 & 1 \\ 1/\sqrt{2} & 1 & 1/\sqrt{2} \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & delta_{xy}/delta_z & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (5.17)$$

and $\mathbf{Cf}$ is the matrix form notation for the convolution operation between $\mathbf{f}$ and this mask. Thus also this case lies in the classical Tikhonov regularization. It is important to note that we set the scaling parameter δ to one without losing in generality since it is easy to demonstrate that δ can be isolated from the summation and included in the regularization parameter β. In the following we will refer to this kind of regularization with the name quadratic potential regularization [MVD07].
Quadratic potential regularization is extensively used in many reconstruction [Gre90a, GM85] and restoration [LRA02] application but, as far as we know, this does not happen in fluorescence microscopy. It is well known that quadratic potential regularization imposes smoothness constraint everywhere; as large gradients in the solution are penalized, the result is that noise artifacts (like checkerboard effect) are reduced but edges of the object are completely lost (see Subsec. 7.1.3).

5.2.2 Non-quadratic potential functions

In the previous Section we supposes that objects are globally smooth, but a more realistic object model assumes that objects are made of smooth regions, separated by sharp edges. Again, we start from Equation (5.10) in the case k=1:

$$J_R(\mathbf{f}) = \sum_{c \in C''} \varphi\big(D_c^1(\mathbf{f})\big) = \sum_{\mathbf{n}} \sum_{\mathbf{m} \in \mathcal{N}_\mathbf{n}} \varphi\left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})\delta}\right), \quad (5.18)$$

and we present a wide class of potential functions φ able to impose this model, or, in other words, able to preserve the edges in the restored object. This framework is extensively studied in positron emission tomography (PET) and single photon emission computerized Tomography (SPECT) image reconstruction [GM85, CBFAB94, TL89, CBFAB97], but to our knowledge, these studies lack in 3D fluorescence microscopy image restoration [VBDB08].
In recent years, the problem like "What properties must a potential function satisfy to ensure the preservation of edges?" has provided abundant literatures. Table 5.1 shows different edge-preserving potential functions proposed in literature and their corresponding weighting functions $\psi(t) = \varphi'(t)/2t$. φ' stands for the first derivative of φ. Since, quadratic potential function leads to overshooting restoration all edge-preserving potential functions presented in Table 5.1 are non-quadratic. Geman and Mc. Clure [GM85] advocated functions having a finite asymptotic behavior φ_{GM}. Hebert and Leahy [TL89]

Table 5.1: Edge-preserving potential functions and their associated weighting functions.

Regularization name	Potential function	Expression of $\varphi(t)$	Expression of $\psi(t) = \varphi'(t)/2t$	Convex
Quadratic-Potential	φ_{QP}	t^2	1	yes
Geman-McClure	φ_{GM}	$\frac{t^2}{1+t^2}$	$\frac{1}{(1+t^2)^2}$	no
Hebert-Leahy	φ_{HL}	$\log(1+t^2)$	$\frac{1}{1+t^2}$	no
Huber	φ_{HB}	$\begin{cases} t^2, & \lvert t\rvert \leq 1 \\ 2\lvert t\rvert - 1, & \lvert t\rvert > 1 \end{cases}$	$\begin{cases} 1, & \lvert t\rvert \leq 1 \\ 1/\lvert t\rvert, & \lvert t\rvert > 1 \end{cases}$	yes
Hyper-Surface	φ_{HS}	$2\sqrt{1+t^2} - 2$	$\frac{1}{\sqrt{1+t^2}}$	yes

proposed a compromise between quadratic and the asymptotic potential functions φ_{HL}. Lastly, other authors, as Schultz and Stevenson [Sch95], and Charbonnier et al.[CBFAB94] preferred using convex potential functions in order to ensure uniqueness of the solution, φ_{HB} and φ_{HS}. To unify all these approaches, Charbonnier et al. [CBFAB97] proposed a set of general conditions that the edge-preserving regularization potential function should satisfy, independently of the algorithm used to minimize the regularized functional (Eq. (4.25)). Such conditions can be divided in two groups, i.e. basic conditions and edge-preserving conditions. In this thesis we spent some words about the basic conditions and we completely refer to [CBFAB97] for details about the edge-preserving conditions. Basic conditions can be summarized as follows:

a) $\varphi(t) \geq 0 \ \forall t$ and $\varphi(0) = 0$;

b) $\varphi(t) = \varphi(-t)$;

c) φ continuously differentiable;

d) $\varphi'(t) \geq 0 \ \forall t \geq 0$.

Since we assume that a large value of the gradient corresponds to an edge while a small value of the gradient is an effect of noise, the basic idea is to build a potential function that assigns an high cost (penalize) to large gradients and low cost to small gradients. Thus potential function should have some obvious properties. First, it seems natural to give positive value to the potential. Therefore, $\varphi(t) \geq 0 \ \forall t$, and $\varphi(0) = 0$ for practical reason (a). Second, it is necessary to give the same importance to gradients of equal values

but opposite signs. Thus, φ is assumed to be an even function (b). Also, in designing a potential function, it is natural to assume that φ is an increasing function for $t \geq 0$ (c). Finally, in order to avoid introducing instability into the reconstruction process, we impose that φ is continuously differentiable (d). Note that condition (d) excludes function like $|t|$, which are not differentiable at 0. Moreover we remark that quadratic potential function satisfies all the basic conditions (a-d), but we know that it does not preserve the edges, for this reason further conditions must be imposed on the potential function (e-g). From an heuristic study of the first derivative of the potential function Charbonier et al. [CBFAB97] derived the following edge-preserving conditions:

e) $\psi(t) = \dfrac{\varphi'(t)}{2t}$ continuous and strictly decreasing on $[0, +\infty)$;

f) $\lim_{t \to +\infty} \psi(t) = 0$;

g) $\lim_{t \to 0} \psi(t) = M$, $0 < M < +\infty$.

As we expect φ_{QP} is not in agreement with edge-preserving conditions, in particular it violates condition (e) and (f). Moreover, also φ_{HB} violates condition (e), in fact, the weighting function of φ_{HB} is not strictly decreasing on $[0, +\infty)$. However, synthetic and real data tests, show that also φ_{HB} leads to well-preserved edge restorations. Therefore, we believe that condition (e) can be weakened and must be satisfied only for sufficiently large values of t (in the specific case for $t > 1$).
Figure 5.2 shows the plot of the four edge-preserving potential functions, suitably normalized, versus the difference between the two voxels in a clique. This normalization simplifies the interpretation of the effects of the relative regularization term and enables some speculation on the choose of the regulation parameter β for the different potential functions. For example φ_{GM}, φ_{HL}, φ_{HS} and φ_{HB} are normalized to 1 in $t = 5$ dividing for 1.04, 1.41, 8.19 and 9, respectively. Thus, we can suppose that higher regularization parameters should be necessary for φ_{HS} and φ_{HB} in comparison to φ_{GM} and φ_{HL} (see Sec. 7.1.4). Moreover, we want to remark the very similar behaviors between φ_{HS} and φ_{HB}.
We conclude with some important aspects. First, edge-preserving potential functions presenting in Table 5.1 have different behavior at infinity, some have an horizontal asymptote, as φ_{GM}, other do not. This means that the existence of an horizontal asymptote is not a condition necessary to ensure edge-preservation, as contrarily suggested by [GR92]. Second, no condition on the convexity of the potential function is required, edge-preserved potential function can be both convex, like φ_{HB} and φ_{HS}, and non-convex like φ_{GM} and φ_{HL}. However, we point out that φ_{GM} and φ_{HL} are quasi-convex, i.e. all their sublevels set are convex, and that they have a unique global minimum in $t = 0$.

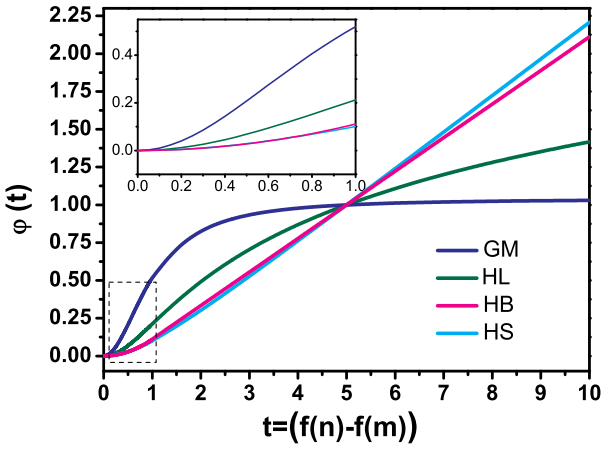

Figure 5.2: Plot of the normalized non-quadratic potential functions versus difference between two pixels in a clique.

5.3 Fuzzy-Logic Based Prior

In non-quadratic regularization, we have considered that small gradients must be smoothed, while large gradients must be preserved. Hence, we have implicitly made the following assumption: a large value of the gradient corresponds to an edge while a small value of the gradient is an effect of noise. This assumption is not necessarily satisfied in practice: small values of the gradient may as well correspond to an actual low amplitude discontinuity in the object and large gradients may be due to noise. Undoubtedly, this is a limitation of the model: especially in presence of strong noise, it may not always be possible to discriminate real features of the object from the effects of noise. On the other hand quadratic potential regularization, which smoothes both noise and edge equally, tends to produce an unfavorable overshooting effect. Such limitations can be ascribed to the fact that both non-quadratic and quadratic priors rely on voxel intensity differences information within a local neighborhood to determine the degrees of regularization on every voxels in the object. Therefore, none of them addresses global connectivity and continuity information in the object. Our aim in this Section is to find a new prior that is capable of exploiting more information of the object.
Recently, Van de Ville et. al [VNdW$^+$03] put forward a novel algorithm for image denoising based on fuzzy-logic. Inspired by their idea, we build [VMD06, VMD07, MVD07] a new fuzzy-logic based potential function for image restoration. The proposed approach is principally based on a fuzzy filtering step and on a smoothing step. Fuzzy filtering step helps in distinguishing the local variations due to object structure and due to noise. In particular,

fuzzy filtering is used to evaluate an array of weighting factors $\mathbf{w_f}$. Each weighting factor $\mathbf{w_f}(\mathbf{n}, \mathbf{m})$ with $\mathbf{m} \in \mathscr{N}_\mathbf{n}$ represents the possibility degree that an edge passes through the voxel $\mathbf{n}$ with a direction perpendicular to the direction of the neighbor voxel $\mathbf{m}$. We will se that it is safe to assume that $\mathbf{w_f}(\mathbf{n}, \mathbf{m}) \in [0, 1]$. The notation $\mathbf{w_f}$ denotes that the weight matrix is obtained by fuzzy filtering on the object $\mathbf{f}$. In the smoothing step the weighting factors are included in the regularization term in order to perform smoothing only on noisy voxel:

$$J_R(\mathbf{f}) = \sum_{\mathbf{n}} \sum_{\mathbf{m} \in \mathscr{N}_\mathbf{n}} \mathbf{w_f}(\mathbf{n}, \mathbf{m}) \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})} \right)^2. \tag{5.19}$$

It is important to note the similitude with the quadratic prior (Eq. 5.15) excepted for the scaling parameter δ that is set to one without lose of generality (see Subsec. 5.2.1) and the weighting factor $\mathbf{w_f}(\mathbf{n}, \mathbf{m})$. Therefore in the limit case $\mathbf{w_f}(\mathbf{n}, \mathbf{m}) = 0$ (this means that an edge passes through the voxel $\mathbf{n}$ with a direction perpendicular to the direction of the neighbor voxel $\mathbf{m}$) no regularization is applied in order to preserve the edge. While in the limit case $\mathbf{w_f}(\mathbf{n}, \mathbf{m}) = 1$ smoothing regularization is applied. It is obvious that such approach completely depends on the ability of the fuzzy filtering to find suitable weight matrix $\mathbf{w_f}$.
Here, we briefly define the fuzzy filter. For the sake of simplicity, we represent the fuzzy

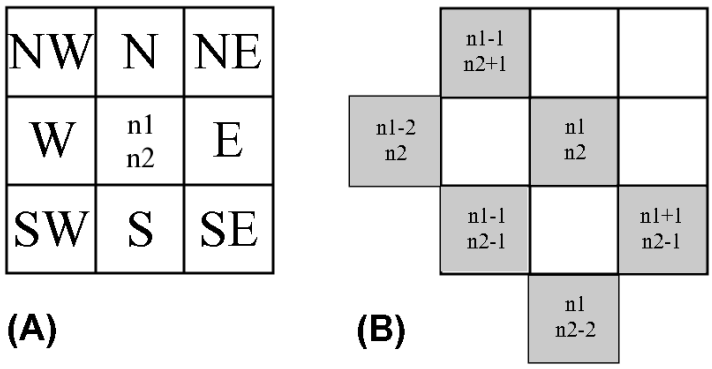

Figure 5.3: (A) 3×3 neighborhood window of a central pixel (n_1, n_2). (B) Pixel values indicated in gray are used to compute the fuzzy derivative of the central pixel (n_1, n_2) for the SW-direction.

filter for the 2D case, at the end the result will be extend to the 3D case. The main concern of the proposed filter is to distinguish between local variations due to noise and due to object structure. In order to accomplish this, for each pixel we derive a value that express the degree in which the derivative in a certain direction is small (fuzzy derivative). Such a value is derived for each direction corresponding to the neighboring pixels of the

processed pixel (see Fig. 5.3 (A)) by a fuzzy rule (Eq. (5.23)). The further construction of the weights is obtain by defuzzification of the fuzzy derivatives of each direction and is based on the observation that a small fuzzy derivative most likely is caused by noise, while a large fuzzy derivative most likely is caused by an edge in the object.
We consider the 3×3 neighborhood of a pixel (n_1, n_2) as displayed in Figure 5.3 (A).
A simple derivative of the pixel (n_1, n_2) in the direction direction $\hat{d}$, denoted by $\nabla(n_1, n_2)\hat{d}$ ($\hat{d} \in dir = \{\hat{E}, \hat{W}, \hat{N}, \hat{S}, \widehat{NE}, \widehat{NW}, \widehat{SE}, \widehat{SW}\}$), is defined as the difference between the pixel (n_1, n_2) and its neighbor in the direction $\hat{d}$.
For example $\nabla(n_1, n_2)\widehat{SW} = (\mathbf{f}(n_1, n_2) - \mathbf{f}(n_1 - 1, n_2 - 1))$ is the direction derivative of the pixel (n_1, n_2) in the direction $\widehat{SW}$. Next the principle of the fuzzy derivative is based on the following observation. Consider an edge passing through the neighborhood of a pixel (n_1, n_2) in the NW-SE direction (see Fig. 5.3 (A)). The direction derivative $\nabla(n_1, n_2)\widehat{SW}$ will be large and positive, or large and negative, but also the derivative values of neighboring pixels perpendicular to the edge's direction can expected to be large and positive, or large and negative. For example in the SW-direction we can calculate the values $\nabla(n_1, n_2)\widehat{SW}$, $\nabla(n_1 - 1, n_2 + 1)\widehat{SW}$ and $\nabla(n_1 + 1, n_2 - 1)\widehat{SW}$ (see Fig. 5.3 (B)). The idea is to cancel out the effect of one derivative value which turns out to be low due to noise. Therefore, if two out of three derivative values are large and positive, or large and negative, it is same to assume that an edge is present is the considered direction. This observation will be taken into account when we formulate the fuzzy rule to calculate the fuzzy derivative values.
To compute the value that expresses the degree to which the fuzzy derivative in a certain direction is large, we will make use of two fuzzy set *largepositive* and *largenegative*. The membership functions $mlp_{Th}(u)$ and $mln_{Th}(u)$ for the property *largepositive* and *largenegative* are as follows (see Fig. 5.4):

$$mlp_{Th}(u) = \begin{cases} 0, & u < 0 \\ |u|/Th, & 0 \leq u \leq Th \\ 1, & u > Th \end{cases} \tag{5.20}$$

$$mln_{Th}(u) = \begin{cases} 1, & u < -Th \\ |u|/Th, & -Th \leq u \leq 0 \\ 0, & u > 0 \end{cases} \tag{5.21}$$

where Th is an adaptive parameter. In this thesis the parameter Th is the mean of the absolute value of the directional derivative:

$$Th = \frac{1}{8} \sum_{\hat{d} \in dir} |\nabla(n_1, n_2)\hat{d}|. \tag{5.22}$$

Since we will use such fuzzy-filtering in a iterative scheme the threshold will be update for each iteration.

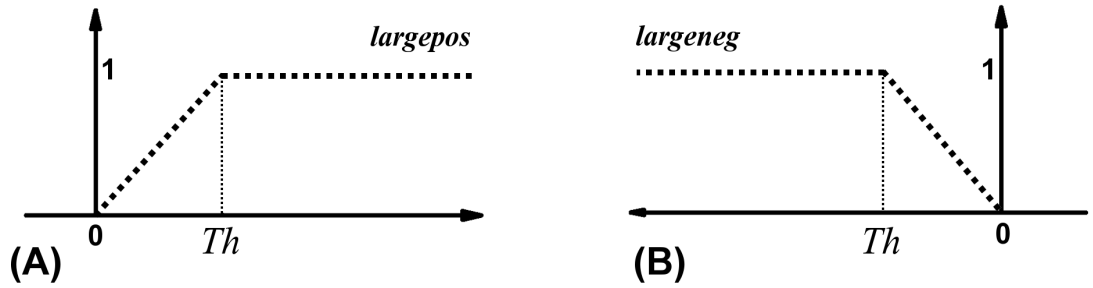

Figure 5.4: Linear membership functions (A) *largepos*, (B) *largeneg*.

For example, the value of the fuzzy derivative $\nabla_F(n_1,n_2)\widehat{SW}$ for the pixel (n_1,n_2) in the SW-direction is calculated by applying the following rule:

$$\texttt{if}\ \left\{\begin{array}{l}(\nabla(n_1,n_2)\widehat{SW}\ \texttt{is}\ \textit{largepos}\ \texttt{and}\ \nabla(n_1-1,n_2+1)\widehat{SW}\ \texttt{is}\ \textit{largepos})\ \texttt{or}\\(\nabla(n_1,n_2)\widehat{SW}\ \texttt{is}\ \textit{largepos}\ \texttt{and}\ \nabla(n_1+1,n_2-1)\widehat{SW}\ \texttt{is}\ \textit{largepos})\ \texttt{or}\\(\nabla(n_1-1,n_2+1)\widehat{SW}\ \texttt{is}\ \textit{largepos}\ \texttt{and}\ \nabla(n_1+1,n_2-1)\widehat{SW}\ \texttt{is}\ \textit{largepos})\ \texttt{or}\\(\nabla(n_1,n_2)\widehat{SW}\ \texttt{is}\ \textit{largeneg}\ \texttt{and}\ \nabla(n_1-1,n_2+1)\widehat{SW}\ \texttt{is}\ \textit{largeneg})\ \texttt{or}\\(\nabla(n_1,n_2)\widehat{SW}\ \texttt{is}\ \textit{largeneg}\ \texttt{and}\ \nabla(n_1+1,n_2-1)\widehat{SW}\ \texttt{is}\ \textit{largeneg})\ \texttt{or}\\(\nabla(n_1-1,n_2+1)\widehat{SW}\ \texttt{is}\ \textit{largeneg}\ \texttt{and}\ \nabla(n_1+1,n_2-1)\widehat{SW}\ \texttt{is}\ \textit{largeneg})\ \texttt{or}\end{array}\right.$$
$$\texttt{then}\ \nabla_F(n_1,n_2)\widehat{SW}\ \text{is}\ \textit{large} \tag{5.23}$$

A similar fuzzy rule is applied for each direction $\hat{d} \in dir$, to compute the degree of membership of the fuzzy derivatives $\nabla_F(n_1,n_2)\hat{d}$ to the fuzzy set *large*. These rules are implemented using the minimum to represent the `and`-operator, and the maximum for the `or`-operator. The final step is the defuzzification of the fuzzy derivative in order to compute the weighting factors for Equation (5.19):

$$\mathbf{w_f}(\mathbf{n},\mathbf{m}) = 1 - \nabla_F(\mathbf{n})\hat{d}, \tag{5.24}$$

where $\mathbf{m}$ is the neighbor of $\mathbf{n}$ in the direction $\hat{d}$. It is important to note that:

$$0 \leq \mathbf{w_f}(\mathbf{n},\mathbf{m}) \leq 1. \tag{5.25}$$

Extensions to the 3D case is obtained introducing two new fuzzy rules related to the axial directions: backward $\hat{B}$ and forward $\widehat{F}$ (see Fig.5.1(B)). Note that the evaluation of the adaptive parameter Th must take into account also the new directions:

$$Th = \frac{1}{10}\sum_{\hat{d}\in dir} |\nabla(n_1,n_2)\hat{d}|, \tag{5.26}$$

where $dir = \{\hat{E}, \hat{W}, \hat{N}, \hat{S}, \widehat{NE}, \widehat{NW}, \widehat{SE}, \widehat{SW}, \hat{B}, \hat{F}\}$.

The fuzzy rule described above can be extended to a 5×5 neighborhood window to study the effect of window size on the image quality. In the case of a 3×3 window, the fuzzy directional derivative is calculated using three derivatives per direction. The sensitivity of edge detection depends upon the number of derivatives used for edge detection. To enhance the detectivity of edges, five derivatives per direction are used for a 5×5 window. For example, the derivatives applied for edge detection in SW direction are $\nabla(n_1 - 2, n_2 + 2, n_3)\widehat{SW}$, $\nabla(n_1 - 1, n_2 + 1, n_3)\widehat{SW}$, $\nabla(n_1, n_2, n_3)\widehat{SW}$, $\nabla(n_1 + 1, n_2 - 1, n_3)\widehat{SW}$ and $\nabla(n_1 + 2, n_2 - 2, n_3)\widehat{SW}$ (see Fig. 5.5).

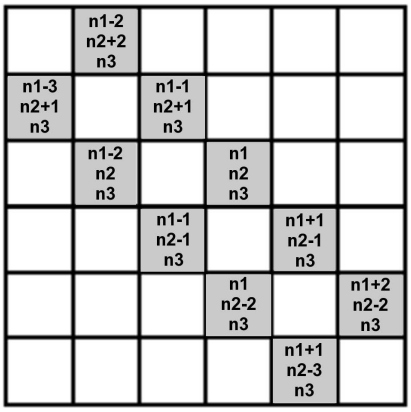

Figure 5.5: 5×5 neighborhood of a central pixel (n_1, n_2, n_3).

In this Chapter we use MRF to model statistical properties of the object to be restored. This model can be included in the restoration problem through a regularization functional, and impose particular constraints on the solution. Regularization functionals derived from quadratic, non-quadratic and fuzzy-logic based potential functions are presented. These regularization functionals can be used within the maximum *a-posteriori* approach in order to include in the restoration process prior information of the object. Maximum *a-posteriori* approach transforms the image restoration problem in a minimization problem. In the next Chapter we propose a general method, called split-gradient method, able to easily derive an iterative algorithm for each regularization functional obtained in this Chapter.

Chapter 6

Split-Gradient Method

> What is it indeed that gives us the feeling of elegance in a solution, in a demonstration? It is the harmony of the diverse parts, their symmetry, their happy balance; in a word it is all that introduces order, all that gives unity, that permits us to see clearly and to comprehend at once both the ensemble and the details.
>
> Henri Poincaré

In Chapter 4 we used Bayesian approach to reformulate the image restoration problem from a statistical point of view. This new formulation allows to include in the restoration problem both information about the recording process (noise model) and information about the object to be restored (object model), moreover the problem is transformed into a minimization problem of a suitable functional, the so called regularized functional. This functional is composed by two terms, the first one is the discrepancy functional and includes prior information about the noise, the second one is the regularized functional and includes prior information about the object. Two noise processes are generally considered: the additive Gaussian white process, that assumes that noise is independent by the signal, and the Poisson process that assumes perfect detection. On the other hand, a large number of regularized functionals are proposed in Chapter 5 to bias different properties on the solution of the problem. Since each combination of discrepancy and regularization term leads to a new functional, a wide class of regularized functionals is derived. These functionals have very different properties, for example some are convex, some other are non-convex, and some, like fuzzy-logic based functionals, have very complicate dependence

on the object function **f**. Therefore, we need a method to obtain minimization algorithms that does not require specific assumptions on noise model or regularization term.
Here, we present an approach, called split-gradient method (SGM) able to devise iterative algorithms for the minimization of a very general regularized functionals. Principal advantage of this method is that non explicit assumption must be made on noise model or regularization term. Moreover, it leads to simply multiplicative form algorithms and allows to include non-negativity and support constraints.
Since restored objects can be used for quantitative analysis, like for example co-localization [RPL05, Lan02] or volume-estimation [DMS+04], a useful support constraint is the so called flux-conservation: the total number of counts of the detected image has to coincide with the total number of counts of the restored object, in terms of intensity, this means that the total intensity of the restored object must be the same of the initial image.
The aim of this Chapter is to obtain an easy implementable algorithm for each regularization method proposed in Chapter 5.

In Section 6.1 we introduced the SGM in its general and original relaxed step descendent formulation, then we present its product formulation. In Section 6.2 we analyze the product formulation for the solution of the regularized problem. In particular, for each regularized functional proposed we devise the relative SGM based algorithm. Comparison between one-step-late method and split-gradient method is exploited.

6.1 General Formulation

In a series of recent papers Lantèri et al. [LRGA02, LRA02, LRCA01] proposed a general approach, denoted as split-gradient method (SGM), to devise iterative algorithms for the minimization of a wide class of regularized functionals of the following type (see Sect 4.3):

$$J_\mu(\mathbf{f};\mathbf{g}) = J_0(\mathbf{f};\mathbf{g}) + \mu J_R(\mathbf{f}), \tag{6.1}$$

with the additional constraints of non-negativity and flux-conservation:

$$\mathbf{f}(\mathbf{n}) \geq 0, \qquad \sum_{\mathbf{n}} \mathbf{f}(\mathbf{n}) = c. \tag{6.2}$$

c is the total intensity of the image as defined in Equation (4.19). In the expression of $J_\mu(\mathbf{f};\mathbf{g})$, the term $J_0(\mathbf{f};\mathbf{g})$ is the functional measuring the discrepancy between the computed images associated to **f** and the detected images **g**, that is the neglog likelihood, thus the Kulback-Leibler divergence in the case of Poisson noise (see Sec. 4.2.2), or the least-square distance in the case of white Gaussian noise (see Sec. 4.2.1). While $J_R(\mathbf{f})$ is a regularization term (see Ch. 5) allowing us to impose some properties of the solution, such

as smoothness, roughness, etc. μ is the regularization parameter (following MRF object modeling we have $\mu = 1/\beta$).
SGM is founded on the use of the basic gradient algorithm adapted to verify the Khun-Tucker first-order optimality conditions [LRCA01]. The basic idea in this method relies on the following decomposition of the gradient of the functional $J_\mu(\mathbf{f};\mathbf{g})$:

$$-\nabla_{\mathbf{f}} J_\mu(\mathbf{f};\mathbf{g}) = \mathbf{U}_\mu(\mathbf{f};\mathbf{g}) - \mathbf{V}_\mu(\mathbf{f};\mathbf{g}), \tag{6.3}$$

where $\mathbf{U}_\mu(\mathbf{f};\mathbf{g})$ and $\mathbf{V}_\mu(\mathbf{f};\mathbf{g})$ are positive functions (vectors) depending on $\mathbf{f}$. Obviously such a decomposition always exists and is not unique. The applicability of the method requires explicit expressions for the dependence of these functions on $\mathbf{f}$. Then the general structure of the iterative algorithm, as described in Lantéri et al. [LRA02, LRCA01], is presented as a descend method with a step-selection:

- (initialization) choose an initial guess $\mathbf{f}^{(0)}$, such that:

$$\mathbf{f}^{(0)}(\mathbf{n}) > 0, \tag{6.4}$$

$$\textstyle\sum_{\mathbf{n}} \mathbf{f}^{(0)}(\mathbf{n}) = c;$$

- (update) given $\mathbf{f}^{(i)}$ compute:

$$\tilde{\mathbf{f}}^{(i+1)} = \mathbf{f}^{(i)} + \alpha^{(i)}\mathbf{f}^{(i)}\left(\frac{\mathbf{U}_\mu(\mathbf{f}^{(i)};\mathbf{g}) - \mathbf{V}_\mu(\mathbf{f}^{(i)};\mathbf{g})}{\mathbf{V}_\mu(\mathbf{f}^{(i)};\mathbf{g})}\right), \tag{6.5}$$

$$\tilde{c}^{(i+1)} = \textstyle\sum_{\mathbf{n}} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n});$$

- (normalization) set:

$$\mathbf{f}^{(i+1)} = \frac{c}{\tilde{c}^{(i+1)}}\tilde{\mathbf{f}}^{(i+1)}, \tag{6.6}$$

where, $\alpha^{(i)}$ is a relaxation factor, and it can be computed in order to guarantee both the non-negativity of the iterates and the convergence of the algorithm [LRCA01, LRA02].
Note that Lantèri et al. [LRA02] demonstrated that flux-conservation constraint could be just taken into account simply by a normalization of the solution after each iteration.
The algorithm takes a very simple product-form (multiplicative form) if we use a unit step (i.e. $\alpha^{(i)} = 1, \forall i$); in such a case the initialization step does not change and we obtain:

- (update) given $\mathbf{f}^{(i)}$ compute:

$$\tilde{\mathbf{f}}^{(i+1)} = \mathbf{f}^{(i)}\left(\frac{\mathbf{U}_\mu(\mathbf{f}^{(i)};\mathbf{g})}{\mathbf{V}_\mu(\mathbf{f}^{(i)};\mathbf{g})}\right), \tag{6.7}$$

$$\tilde{c}^{(i+1)} = \textstyle\sum_{\mathbf{n}} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n});$$

- (normalization) set:

$$\mathbf{f}^{(i+1)} = \frac{c}{\tilde{c}^{(i+1)}}\tilde{\mathbf{f}}^{(i+1)}. \tag{6.8}$$

It is important to remark some interesting features. The first is that all the iterates are automatically non-negative if the initial guess $\mathbf{f}^{(0)}$ is non-negative, as one can easily verify. The second is that the algorithm in the simplified form of Equations (6.7) and (6.8) is a scaled-gradient method, with step size 1, since the update step can be written in the following form:

$$\tilde{\mathbf{f}}^{(i+1)} = \mathbf{f}^{(i)} - S^{(i)}\Big[\nabla_{\mathbf{f}} J_\mu(\mathbf{f}^{(i)};\mathbf{g})\Big] \tag{6.9}$$

where

$$S^{(i)} = \mathrm{diag}\left\{\frac{\mathbf{f}^{(i)}(\mathbf{n})}{\big(\mathbf{V}_\mu(\mathbf{f}^{(i)};\mathbf{g})\big)(\mathbf{n})}\right\}. \tag{6.10}$$

The convergence of the algorithm in the unrelaxed product-form (Eqs. (6.7), (6.8)) has not been proved in general, but only in some particular cases [LRCA01]; nevertheless it has been verified experimentally in all the applications we have considered (see Sec. 7.1.5). Obviously, the advantage of the product-form algorithm is that it does not require any relaxation factor $\alpha^{(i)}$ computation for each iteration, that sometimes can be very expensive.

6.2 Regularized Formulation

In the case of the product-form of the algorithm (Eqs. (6.7) and (6.8)) the dependence on the regularization parameter can be made explicit by introducing decompositions of the gradients of the discrepancy functional $J_0(\mathbf{f};\mathbf{g})$ and of the regularization functional $J_R(\mathbf{f})$ similar to Equation 6.1:

$$-\nabla_{\mathbf{f}} J_0(\mathbf{f};\mathbf{g}) = \mathbf{U}_0(\mathbf{f};\mathbf{g}) - \mathbf{V}_0(\mathbf{f};\mathbf{g}), \tag{6.11}$$

$$-\nabla_{\mathbf{f}} J_R(\mathbf{f}) = \mathbf{U}_R(\mathbf{f}) - \mathbf{V}_R(\mathbf{f}).$$

Then it turns out that the algorithm has the following structure:

- (update) given $\mathbf{f}^{(i)}$ compute:

$$\tilde{\mathbf{f}}^{(i+1)} = \mathbf{f}^{(i)}\left(\frac{\mathbf{U}_0(\mathbf{f}^{(i)};\mathbf{g}) + \mu\mathbf{U}_R(\mathbf{f}^{(i)})}{\mathbf{V}_0(\mathbf{f}^{(i)};\mathbf{g}) + \mu\mathbf{V}_R(\mathbf{f}^{(i)})}\right), \tag{6.12}$$

$$\tilde{c}^{(i+1)} = \textstyle\sum_{\mathbf{n}} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n});$$

- (normalization) set:

$$\mathbf{f}^{(i+1)} = \frac{c}{\tilde{c}^{(i+1)}}\tilde{\mathbf{f}}^{(i+1)}, \tag{6.13}$$

which shows the very simple dependence of the algorithm on the regularization parameter μ. Again the initialization step is the same of Equation 6.4.
Before proceeding with the study of the regularized formulation of the SGM we analyze some important features of the case $\mu = 0$, this means the case in which we are using SGM to solve the ML problem:

$$J_\mu(\mathbf{f};\mathbf{g}) = J_0(\mathbf{f};\mathbf{g}). \tag{6.14}$$

In Section 4.2 we introduce two kinds of noise statistic for the detection process, which lead to two different discrepancy functionals $J_0(\mathbf{g};\mathbf{f})$. Here, to help the reader, we rewrite the two functionals :

- white Gaussian case:

$$J_0(\mathbf{f};\mathbf{g}) = \sum_{\mathbf{n}} \left|(\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) - \mathbf{g}(\mathbf{n})\right|^2; \tag{6.15}$$

- Poisson case:

$$J_0(\mathbf{f};\mathbf{g}) = \sum_{\mathbf{n}} \left\{\mathbf{g}(\mathbf{n}) \ln \frac{\mathbf{g}(\mathbf{n})}{(\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n})} + (\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) - \mathbf{g}(\mathbf{n})\right\}. \tag{6.16}$$

In Appendix B we derive the gradient of the functionals in the two cases:

- white Gaussian case:

$$\nabla_{\mathbf{f}} J_0(\mathbf{g};\mathbf{f}) = \mathbf{1} - \mathbf{A}^T \frac{\mathbf{g}}{\mathbf{Af} + \mathbf{b}}, \tag{6.17}$$

- Poisson case:

$$\nabla_{\mathbf{f}} J_0(\mathbf{g};\mathbf{f}) = 2\mathbf{A}^T\mathbf{Af} + 2\mathbf{b} - 2\mathbf{A}^T\mathbf{g}. \tag{6.18}$$

Table 6.1: The functions $\mathbf{U}_0$ and $\mathbf{V}_0$ for the the Gaussian and Poisson noise.

	$\mathbf{U}_0(\mathbf{f}^{(i)};\mathbf{g})$	$\mathbf{V}_0(\mathbf{f}^{(i)};\mathbf{g})$
Gauss	$2\mathbf{A}^T\mathbf{g}$	$2(\mathbf{A}^T\mathbf{Af} + \mathbf{b})$
Poisson	$\mathbf{A}^T\frac{\mathbf{g}}{\mathbf{Af}+\mathbf{b}}$	$\mathbf{1}$

Since, we are considering the restriction of the discrepancy functional $J_0(\mathbf{f};\mathbf{g})$ to the non-negative vectors, and each elements of $\mathbf{A}$ is non-negative, the natural choices for the functions $\mathbf{U}_0(\mathbf{f}^{(i)};\mathbf{g})$ and $\mathbf{V}_0(\mathbf{f}^{(i)};\mathbf{g})$ for the two cases are given in Table 6.1. It is obvious that it is possible to obtain other acceptable choices by adding, for instance, a suitable constant both to $\mathbf{U}_0(\mathbf{f}^{(i)};\mathbf{g})$ and to $\mathbf{V}_0(\mathbf{f}^{(i)};\mathbf{g})$.
If we insert the expressions of Table 6.1 in the general scheme of Equations 6.12 and 6.13 with $\mu = 0$ we obtain for the Gaussian case the ISRA algorithm and for the Poisson case the RLM alghoritm, with the additional flux-conservation constraint. Therefore, if we particularize the general simplified version of the SGM algorithm to the two noise models, we obtain two well-known algorithms.
However, the most interesting feature of the proposed SGM is that one can easily derive different new iterative algorithms to solve the MAP problem, simply computing the functions $U_R(\mathbf{f})$ and $V_R(\mathbf{f})$ corresponding to the different regularization terms. Here we derive the functions $U_R(\mathbf{f})$ and $V_R(\mathbf{f})$ relating to the regularization terms presented in Chapter 5.

Tikhonov regularization

From the Tikhonov regularization term:

$$J_R(\mathbf{f}) = \sum_{\mathbf{n}} \mathbf{f}(\mathbf{n})^2, \tag{6.19}$$

it is easy to compute its gradient, and taking into account that we are considering the restriction of $J_R(\mathbf{f})$ to the cone of non-negative vectors, we can choose the following decomposition:

$$\mathbf{U}_R(\mathbf{f}) = \mathbf{0}, \qquad \mathbf{V}_R(\mathbf{f}) = \mathbf{2f} \tag{6.20}$$

Quadratic and non-quadratic regularization

Both for the quadratic and non-quadratic regularization we can start from the general formulation of Equation (5.18):

$$J_R(\mathbf{f}) = \sum_{\mathbf{n}} \sum_{\mathbf{m}\in\mathcal{N}_\mathbf{n}} \varphi\left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})\delta}\right). \tag{6.21}$$

By elementary computation (see App. C.1) we can derive its gradient:

$$(\nabla_\mathbf{f} J_R(\mathbf{f}))(\mathbf{n}) = \frac{4}{\delta^2} \sum_{\mathbf{m}\in\mathcal{N}_\mathbf{n}} \frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})^2} \psi\left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})\delta}\right). \tag{6.22}$$

Note that since φ is even and $\varphi'(t) \geq 0\ \forall t \geq 0$ it is easy to demonstrate that the weighting function $\psi(t) = \varphi'(t)/2t$ is non-negative $\forall t$. By taking into account that we are considering

the restriction of the regularization functional $J_R(\mathbf{f})$ to the cone of non-negative vectors, we can chose the following decomposition of its gradient:

$$(\mathbf{U}_R(\mathbf{f}))(\mathbf{n}) = \frac{4}{\delta^2} \sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \frac{\mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})^2} \psi\left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})\delta}\right), \tag{6.23}$$

$$(\mathbf{V}_R(\mathbf{f}))(\mathbf{n}) = \frac{4}{\delta^2} \sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \frac{\mathbf{f}(\mathbf{n})}{d(\mathbf{n},\mathbf{m})^2} \psi\left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})\delta}\right).$$

We want to remark that the non-quadratic potential functions φ used in this thesis are both convex and non-convex.

Fuzzy-logic based regularization

Starting from the fuzzy-logic regularization term:

$$J_R(\mathbf{f}) = \sum_{\mathbf{n}} \sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m}) \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})}\right)^2, \tag{6.24}$$

we have to derive its gradient. However, from Equations (5.23) we find that every weighting factor $\mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m})$ is determined by the image $\mathbf{f}$, which makes it difficult to obtain derivatives. Therefore we apply alternating two-step updating restoration algorithm in our study. The two steps can be described as follows:

- *Weight update*
  For every $\mathbf{n}$ and $\mathbf{m} \in \mathscr{N}_\mathbf{n}$ we compute $\mathbf{w}_\mathbf{f}$ using the current estimate of $\mathbf{f}$.

- *Solution update*
  For this step we use again SGM approach. Now we can assume that the weighting factors $\mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m})$ are constant terms, therefore by elementary computation (see App. C.2) we can derive its gradient:

  $$(\nabla_\mathbf{f} J_R(\mathbf{f}))(\mathbf{n}) = 2 \sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \left\{ \big(\mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m}) + \mathbf{w}_\mathbf{f}(\mathbf{m},\mathbf{n})\big) \frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})^2} \right\}. \tag{6.25}$$

  Since $w_\mathbf{f}$ are positive weights and we are considering the restriction of the regularization functional $J_R(\mathbf{f})$ to the non-negative vectors, we can chose the following

decomposition of its gradient:

$$(\mathbf{U}_R(\mathbf{f}))(\mathbf{n}) = 2\sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \left\{ \big(\mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m}) + \mathbf{w}_\mathbf{f}(\mathbf{m},\mathbf{n})\big) \frac{\mathbf{f}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})^2} \right\}, \tag{6.26}$$

$$(\mathbf{V}_R(\mathbf{f}))(\mathbf{n}) = 2\sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \left\{ \big(\mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m}) + \mathbf{w}_\mathbf{f}(\mathbf{m},\mathbf{n})\big) \frac{\mathbf{f}(\mathbf{n})}{d(\mathbf{n},\mathbf{m})^2} \right\}.$$

To help the reader in the compression of the SGM coupled with the fuzzy-logic based regularization we write the complete iterative scheme of the algorithm in the Poisson case:

- (initialization) choose an initial guess $\mathbf{f}^{(0)}$, such that:

$$\begin{aligned} \mathbf{f}^{(0)}(\mathbf{n}) &> 0, \\ \textstyle\sum_\mathbf{n} \mathbf{f}^{(0)}(\mathbf{n}) &= c; \end{aligned} \tag{6.27}$$

- (weight update) given $\mathbf{f}^{(i)}$ for each $\mathbf{n}$ and $\mathbf{m} \in \mathscr{N}_\mathbf{n}$ compute $\mathbf{w}_{\mathbf{f}^{(i)}}(\mathbf{n},\mathbf{m})$ using Equations (5.20-5.24);

- (solution update) given $\mathbf{f}^{(i)}$ compute:

$$\begin{aligned} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n}) &= \mathbf{f}^{(i)}(\mathbf{n}) \frac{\left(\mathbf{A}^T \dfrac{\mathbf{g}}{\mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}}\right)(\mathbf{n}) + \mu \displaystyle\sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \left\{ \big(\mathbf{w}_{\mathbf{f}^{(i)}}(\mathbf{n},\mathbf{m}) + \mathbf{w}_{\mathbf{f}^{(i)}}(\mathbf{m},\mathbf{n})\big) \dfrac{\mathbf{f}^{(i)}(\mathbf{m})}{d(\mathbf{n},\mathbf{m})^2} \right\}}{1 + \mu \displaystyle\sum_{\mathbf{m}\in\mathscr{N}_\mathbf{n}} \left\{ \big(\mathbf{w}_{\mathbf{f}^{(i)}}(\mathbf{n},\mathbf{m}) + \mathbf{w}_{\mathbf{f}^{(i)}}(\mathbf{m},\mathbf{n})\big) \dfrac{\mathbf{f}^{(i)}(\mathbf{n})}{d(\mathbf{n},\mathbf{m})^2} \right\}}, \\ \tilde{c}^{(i+1)} &= \textstyle\sum_\mathbf{n} \tilde{\mathbf{f}}^{(i+1)}(\mathbf{n}); \end{aligned} \tag{6.28}$$

- (solution normalization) set:

$$\mathbf{f}^{(i+1)} = \frac{c}{\tilde{c}^{(i+1)}} \tilde{\mathbf{f}}^{(i+1)}. \tag{6.29}$$

Note that since $\mathbf{w}_\mathbf{f}(\mathbf{n},\mathbf{m}) \neq \mathbf{w}_\mathbf{f}(\mathbf{m},\mathbf{n})$ it is necessary to store for each voxel $\mathbf{n}$ ten weights (one for each neighbor). This means that a matrix with dimension $N\times 10$ must be allocated during all the algorithm.

6.2.1 One step late algorithm

We demonstrated that the SGM is a suitable approach to derive iterative algorithms for the solution of the MAP problem. It may be interesting to compare SGM in the Poisson case with the one-step-late algorithm proposed in 1990 by Green [Gre90a, Gre90b] that, in our notation, is given by:

$$\tilde{\mathbf{f}}^{(i+1)} = \frac{\mathbf{f}^{(i)}}{\mathbf{1} + \mu \nabla_{\mathbf{f}} \mathbf{J}_R(\mathbf{f}^{(i)})} \mathbf{A}^T \frac{\mathbf{g}}{\mathbf{A}\mathbf{f} + \mathbf{b}} \quad . \tag{6.30}$$

As noted in [PY01], the OSL algorithm may fail to converge and may not produce objects with non-negative intensities. This is mainly due to possible oscillations in the derivative of the regularization terms, which can be amplified or diminished by the amount of the regularization parameter μ. In particular for high regularization parameter μ (this means low regularization parameter β) the denominator in Equation (6.30) may become very small or even negative, causing numerical problems. Later that same year, Lange [Lan90] added a line search and proved that if the regularization parameter μ was sufficiently small, the modified OSL algorithm was convergent and produced positive iterates that monotically decrease the regularized functional. However, this condition restricts the choice of the penalty parameter. To solve this problem, the original algorithm was further modified [GY05, MZ06] to ensure non-negativity of the estimates, and (numerical) convergence of the regularized functional; however these methods lead to cumbersome steps in the algorithm.
Since we are looking for a general approach able to work also at very low signal to noise ratio, where high regularization parameters μ are necessary, we believe that the SGM in the product and regularized form can address better than OSL method to these requirements.

In this Chapter we have proposed the SGM general approach to solve the MAP problems both in the case of Poisson and Gaussian noise. This method provides iterative algorithms that, in general, are not very efficient (they require a large number of iterations) but have a very nice feature: for very broad classes of regularization terms they can be obtained by means of a very simply modification of RLM or ISRA and therefore can be easily implemented by everybody. In such a way it is easy to compare the results obtainable by means of different regularization terms. In particular we derive the SGM formulation for Tikhonov, quadratic, non-quadratic and fuzzy-logic based regularization. Split-gradient method formulation for maximum entropy regularization can be found in [Anc06]. Further drawbacks of the SGM in the multiplicative form is that convergence is not demonstrated, even if numerical experiments suggest it. However, convergence is ensured in the relaxed descent form of the SGM.
In the next Chapter we will use simulated and real confocal microscopy data to test the performance of the different regularized algorithms derived in this Chapter using SGM.

Chapter 7

Application to Confocal Microscopy

> I didn't fail the test, I just found 100 ways to do it wrong
>
> Benjamin Franklin

All the derivations obtained in this thesis are made without loss of generality in terms of the microscope system analyzed. This means that all the algorithms proposed can be applied indiscriminately to images obtained from different systems, simply choosing the appropriate PSF and the appropriate noise model. In particular we saw that wide-field images are principally hampered by white Gaussian noise while confocal and two-photon excitation images are distorted by Poisson noise. Therefore, to validate the algorithms proposed we have to take into account these knowledge, especially when the validation is performed on real data. The tests carried out in this Chapter are made only on confocal images, however we believe that similar results can be obtained on two-photon and wide-field images.
The goal of this Chapter is to validate by syntectic simulated data and by real data all the algorithms that we developed by means of the split-gradient method.

In the first Section we use a syntectic phantom and the Kullback-Leibler divergence analysis to compare the performance of the different algorithms at various level of noise. Discussions, remarks and conclusions are the main contents of this Section. The observations carried out on numerical experiments are confirmed by real data experiments in Section 7.2. Both biological sample and calibration sample are used in this Section.

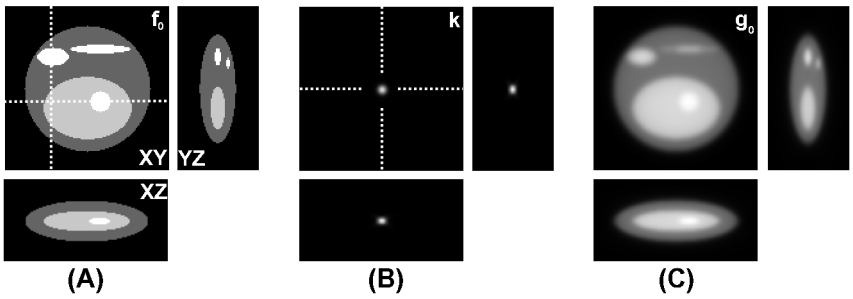

Figure 7.1: Simulate confocal image. All images are composed by a slice view through the middle of the stack (XY), and two axial views along the dotted lines (YZ and XZ). (A) Original three-dimensional phantom. (B) Confocal point spread function (see text for details). (C) Ideal image plus background. Note that the sphere look oblate in axial views because the pixel distance in the axial direction is larger than in lateral direction.

7.1 Results on Numerical Simulation

7.1.1 Syntectic phantom and figures of merit

We design and elaborate a 3D test phantom (Fig. 7.1(A)) consisting of a sphere containing 5 ellipsoids with different sizes and intensities (intensity of the larger sphere: 100 units; larger ellipsoid: 200 units; other structures: 255 units). The dimension of the test phantom is $128 \times 128 \times 64$ voxels, in the x, y and z direction, respectively. These phantom can be considered the idealization of a cell containing different subcellular structures, i.e. nucleus, nucleulus, endoplasmatic reticulum, mitocondria and cytoplasm. A similar test phantom was proposed by [LB02] for denoising of microscopy images.
To simulate CLSM image formation, the object is convolved with a confocal PSF (Fig. 7.1(B)), a constant background of 5 units is added to the ideal image (Fig. 7.2(C)), and finally the result is corrupted by Poisson noise (Fig. 7.2). Confocal free-aberration PSF is computed (see Sec. 3.4) with the following acquisition parameters: excitation wavelength λ_{ex}=488 nm; emission wavelength λ_{em}=520 nm; NA of the objective 1.4; refractive index in which the objective is embedded 1.515 (oil); diameter of the circular pinhole 1 Airy unit. Because we test the algorithms by an inverse crime experiment, namely the PSF used for generating the images is also used for restored the object, it is not necessary to use a more accurate PSF, it is equivalent to choose an aberrate PSF or a free-aberrated PSF.
Nyquist samplings for a confocal system under the acquisition parameters described above are ∼150 nm along the optical axis and ∼50 nm along the lateral axis. In this simulation we over-sample the object with respect to the Nyquist criteria, assuming $delta_{xy}$=35 nm

in the lateral direction and $delta_z$=105 nm in the axial direction. To change the signal-to-noise ratio (SNR) of the simulated image we change τ, which is the reciprocal of the photon-conversion factor. We assume that the noise is described by a Poisson process, implying that the signal-to-noise ratio in fluorescence imaging depends solely on the total number of detected photons. For this reason we choose the mean of the Poisson processes equal to $\tau(\mathbf{A}\mathbf{f}_0 + \mathbf{b})$, with fixed intensities of $\mathbf{A}\mathbf{f}_0 + \mathbf{b}$. Thus by increasing τ, the average number of detected photons increases and hence the noise level decreases. In the real life photon-conversion factor is determined by several multiplicative factors that depend both on system properties and specimen properties [JAJ89]. The relation between SNR and τ is given by:

$$SNR = 10 \log \left[\tau \max_{\mathbf{n}} (\mathbf{A}\mathbf{f}_0 + \mathbf{b})(\mathbf{n}) \right]. \tag{7.1}$$

Figure 7.2 shows confocal simulated images of the proposed phantom for four different SNRs, 10, 15, 20 and 25, respectively. Moreover Table 7.1 reports the τ parameters and

Table 7.1: Photon conversion factors τ associated to different signal-to-noise ratios and their global maximum of detected photons.

SNR	τ	Photons
10	0.046	10
15	0.145	32
20	0.461	100
25	1.457	316

the global maximum of detected photons (see Eq. (7.1)) corresponding to the different SNRs. We want to remark that we choose very low SNR because we want to simulate real confocal situation, where the use of the detection pinhole thereby reduced the number of photons collected and where the power of the laser must be controlled to not damage the sample (see Subsec. 1.2.2) Higher SNR are more appropriate for WFM.

In order to measure the quality of the different restoration we must introduce a figure of merit. We use the Kullback-Leiber divergence (already introduced in Sec. 4.2.2), which is the best measure in the presence of a non-negativity constraint [Csi91]. For numerical comparison the Kullback-Leiber divergence of the vector $\mathbf{q}$ from the vector $\mathbf{r}$ is defined as follows:

$$KL_D(\mathbf{r}, \mathbf{q}) = \frac{1}{N} \sum_{\mathbf{n}} \left(\mathbf{r}(\mathbf{n}) \ln \frac{\mathbf{r}(\mathbf{n})}{\mathbf{q}(\mathbf{n})} - (\mathbf{r}(\mathbf{n}) - \mathbf{q}(\mathbf{n})) \right). \tag{7.2}$$

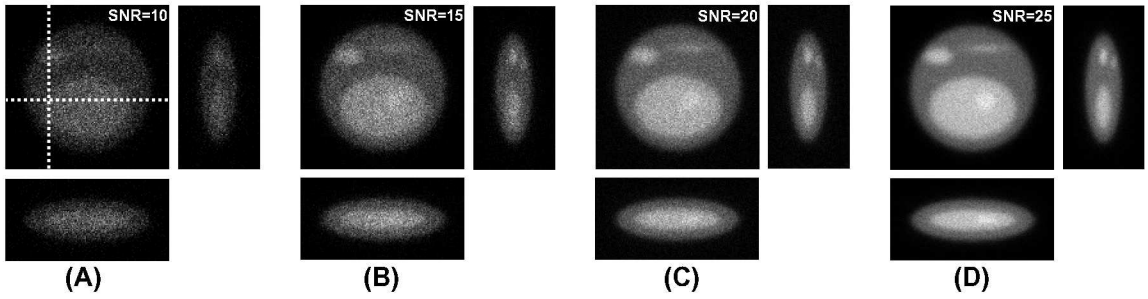

Figure 7.2: Simulated confocal images corrupted by different levelx of Poisson noise. SNR = 10, 15, 20 and 25 for (A), (B), (C) and (D), respectively.

Note that the KL_D is non-symmetric, i.e., $KL_D(\mathbf{r}, \mathbf{q}) \neq KL_D(\mathbf{q}, \mathbf{r})$. Therefore, it is important to remark that we monitor the divergence of the restored solution $\hat{\mathbf{f}}$, from the true object $\mathbf{f}_0$, $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$.
The Kullback-Leiber divergence is also used to find a suitable criterion for stopping the iterations of the devised algorithms, especially for the algorithms that converge to a suitable solution. When the difference between the values of the $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ corresponding to two consecutive iterations is smaller that a certain threshold Th_s we stop the iterations:

$$KL_D(\mathbf{f}_0, \mathbf{f}^{(i)}) - KL_D(\mathbf{f}_0, \mathbf{f}^{(i+1)}) \leq Th_s. \tag{7.3}$$

For algorithms with semi-convergent behavior one can simply think to stop the iterations when the KL divergence starts to grow up.
A simple rule for choosing Th_s does not exist, and its choice follows from our experience. In this thesis, we used a threshold Th_s equal to 10^{-6} as a reliable value for obtaining satisfactory reconstruction. For the methods that converge to a suitable solution, this means that the reconstructed solution $\hat{\mathbf{f}}$, and the associated $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ value do not change much after the threshold has been reached.
For the methods that include a regularization term, the choice of the regularization parameter β is also crucial for a good restoration result. Moreover, non-quadratic regularizations require also the choice of an optimal scaling parameter δ. Unfortunately, a general analytical method, suitable for all the different methods tested in this thesis, is not available for selecting these parameters. A study about different methods for the choice of the regularization parameter in the case of Thikhonov regularization can be found in [KV00b]. Since the original object was available, we use again the KL_D criterion, i.e. the value of β and δ providing the minimum $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ are considered as the appropriate values for obtaining the best restoration. Only, to find reliable β and δ parameters the number of iterations for each restoration is reduced using a stopping threshold Th_s equal to 10^{-4}.
It is obvious that the KL based criterion proposed for finding the optimal number of iter-

ations, the regularization and scaling parameters does not constitute a feasible approach in the case of real data, where the true object is unknown. For real data restoration a different solution must be used (see Sec. 7.2).
Although the KL_D is a good choice as a global measure of quality of a restoration, it may not be able to detect small artifacts which may be important in a given application. Thus, the following comparisons are only the first step in determining which algorithm is most suitable for a given application. For this reason, future work should perform comparisons using criteria that are specifically aimed at features of interest. For example, a segmentation step could be performed to extract regions of interest before computing the KL_D.

All the proposed methods imply the use of an iterative algorithm with an initialization step. A constant object is used as the first estimate f^0. A study about the influence of the choice of the first estimate for a large class of algorithms, among which RLM, can be found in [KV00b].

All the derived algorithms are implemented in Visual C++ environment, where FFT computation to evaluate convolution product is obtained by means of FFT in the west (FFTW) version 2.1.5 [FJ05] compiled in float precision. All experiments were performed on a HP workstation XW4200 Pentium(R)4 CPU 3.40Ghz equipped with 2.00 GB of RAM.

7.1.2 Richardson-Lucy method results

Image restoration is a notorious ill-posed problem, thus to obtain suitable solution it is necessary to introduce all available prior information. Richardson-Lucy method (RLM) is able to integrate in the image restoration problem statistical information about the noise introduced in the recording process (Poisson process). Here we present the results obtained using RLM on the phantom previously described.
In Figure 7.3 we plot for different SNRs the behavior of the KL divergence of the RLM iterates $\mathbf{f}^{(i)}$ from the original object $\mathbf{f}_0$, $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$, as a function of the number of the iterations i. As follows from Figure 7.3, $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ first decreases, goes through a minimum and then increases up to very large values (semiconvergence property of the method). Therefore the algorithm must not be pushed to convergence and an early stopping is required. It turns out that the estimate object $\hat{\mathbf{f}}$ provided by the iteration corresponding to the minimum of the $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ (critical iteration) is, in general, a reliable solution of the restoration problem.
Table 7.2 reports the critical iterations number for different SNRs, and the KL divergence of the associated restored image $\hat{\mathbf{f}}$ from the real object $\mathbf{f}_0$, $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$. Numerical instability of RLM can be observed also by visual inspection of the restored images. It is evident that RLM-restored images suffer severely from noise effect, known as checkerboard effect,

Table 7.2: $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ and critical iteration analysis for RLM at different noise levels

SNR	Critical Iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$
10	19	1.49117
15	38	1.17855
20	73	0.95485
25	165	0.79864

namely many component of the solution are zero. Figure 7.4 shows RLM-restored object at different iterations (SNR=20). If the algorithm is stopped too early, the blurring in the solution is still predominantly (Fig. 7.4(A)). On the other hand, checkerboard effects in the solution become more and more evident at higher iterations (Fig. 7.4(C,D)). Figure 7.4(B) is the solution associated at the critical iterations.
It is obvious that the optimal number of iterations, corresponding to the minimum of the $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$, can be computed only in the case of numerical simulations. Its estimation

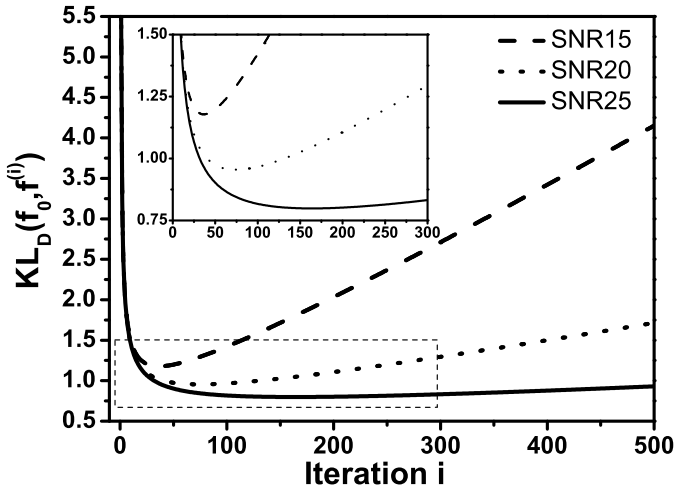

Figure 7.3: Behavior of the KL_D of the iterates $\mathbf{f}^{(i)}$ from the original object $\mathbf{f}_0$ as a function of the number of iterations for RLM. The behaviors for different SNRs are reported. Noise amplification and semiconvergence properties of RLM are evident.

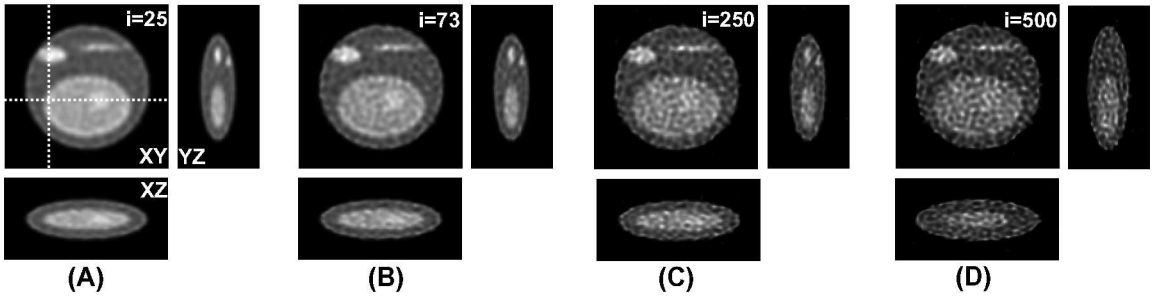

Figure 7.4: RLM-restored objects at different iterations i. Respectively, i=25, 73, 250 and 500 for (A), (B), (C) and (D). Restored object (B) represents the iterates associated at the critical iteration (SNR=20).

in the case of real data is a difficult problem and must be investigated for any particular application of image restoration. As a general rule one can say that the optimal number of iterations depends on the SNR and decreases for decreasing values of the SNR (see Fig. 7.3 and Tab. 7.2).

7.1.3 Quadratic potential function results

Richardson-Lucy method and more in general ML methods require an early stopping of the algorithm in order to regularize the solution. An alternative approach is given by the MAP methods. These methods are able to include not only information about the noise involved in the recording process, but also prior information about the object to be restored. In particular statistical properties of the object can be described by means of prior distributions, and integrated in the minimization problem associated to the restoration problem using penalizing terms (regularization term). The performance of MAP methods are strictly correlated to the choice of the penalizing term. Here, we present the results obtained using two very common penalizing terms, the Tikhonov and the quadratic potential regularization terms.
In Figure 7.5 (A) the $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ is plotted for MAP Tikhonov (MAPTK) and MAP quadratic potential (MAPQP) algorithms as a function of the regularization parameter β (SNR=20). We want to remark again that to find suitable regularization parameters we use a stopping threshold Th_s equal to 10^{-4}. We believe that the very same behavior should be obtained also with a lower threshold. One important observation can be made from this graph: underestimating the regularization parameter leads to larger errors than overestimating it by the same amount. This is due to the fact that for $\beta \to \infty$ any MAP

algorithms become the ML algorithm, therefore $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ has a vertical asymptote equal to $KL_D(\mathbf{f}_0, \hat{\mathbf{f}}_{RLM})$, where $\hat{\mathbf{f}}_{RLM}$ is the solution of the RLM (see Tab. 7.2). The values for β that yield minimal values in these plots, are considered as appropriate values for generating quality restorations, and are used in the further restorations. Figure 7.5(B) shows the behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of the iterations i (SNR=20). It is important to note that, using the appropriate regularization parameters previously found, both MAPTK and MAPQP converge to a suitable solution.
Table 7.3 resumes the optimal regularization parameter β for different SNRs. Moreover, it reports the $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ associated to the restored object $\hat{\mathbf{f}}$, obtained using a stopping threshold Th_s equal to 10^{-6}, and the relative iteration in which the threshold is reached. In the cases marked with an asterisk the threshold is not reached within the maximum number of iterations allowed by the code. It should further be noted that the regularization parameter needs to be modified at a different SNR. This is because of the fact that noisy images comparatively need more smoothing. The regularization parameter controls the amount of smoothness to be introduced in the restoration process. Hence, for a low SNR, the regularization parameter should decrease so that large prior contribution can be introduced for smoothing out the noise and vice-versa.
A better quality in terms of KL divergence of MAPQP respect to MAPTK can be ob-

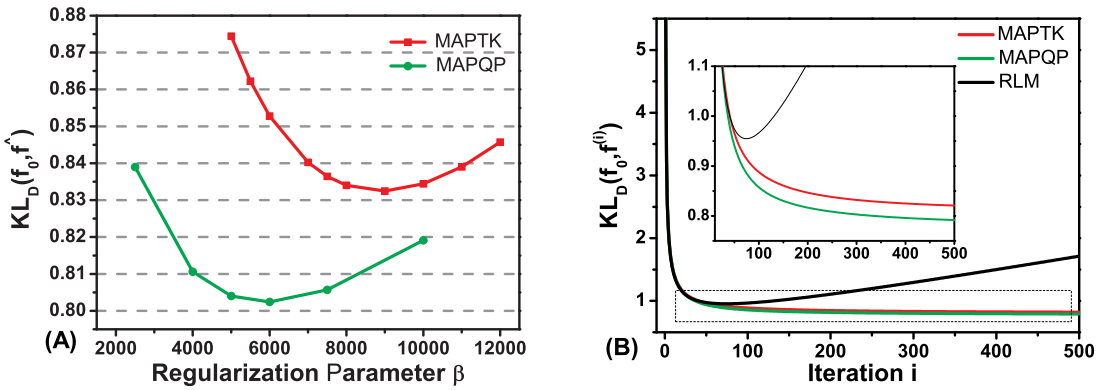

Figure 7.5: (A) Behavior of $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ as a function of the regularization parameter β for Tikhonov and quadratic potential regularizations (SNR=20). (B) Behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of iterations i for Tikhonov and quadratic potential regularizations (SNR=20).

Table 7.3: Values of $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ corresponding to the optimal restorations provided by Tikhonov and quadratic potential regularization for different values of SNR. Values of the regularization parameter β and numbers of iterations used for the optimal restorations are also reported.

SNR	MAPTK			MAPQP		
	β	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	β	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$
10	2000	357	1.40145	750	1000*	1.18181
15	4000	1000*	1.05379	2250	1000*	0.95455
20	9000	1000*	0.81269	6000	1000*	0.78388
25	15000	1000*	0.66614	15000	1000*	0.66186

served especially at low SNR, while for high SNR this difference starts to disappear. Same observation can be made by visual inspection. Figure 7.6 (A,B) shows the restored object obtained using MAPTK and MAPQP, respectively (SNR=20). Checkerboard artifacts are partially reduced in MAPTK and further reduced in MAPQP. However, it is well known that quadratic potential regularization imposes smoothness constraint everywhere; as large gradients in the solution are penalized, the results is that in the MAPQP solution edges are completely lost. Therefore simultaneous suppression of noise amplification and edge-preservation can not be obtained by means of Tikhonov or quadratic potential regularization.

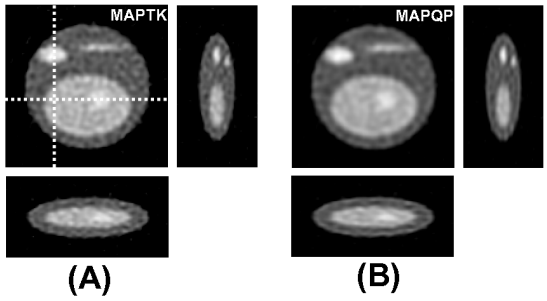

Figure 7.6: Restored objects using Tikhonov (A) and quadratic potential regularization (B) (SNR=20).

One-step-late method versus split-gradient method

To demonstrate the capability of SGM to overcame the limits of the original OSL method we compare the results obtained from the two derived algorithms when coupled with the quadratic potential regularization. Figure 7.7 (A) plots the $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ as function of the regularization parameter β for the two proposed methods (SNR=10). It is very important to note that with OSL method is not possible to use the appropriate β parameter since for values below a certain threshold the algorithm diverges. Therefore, we are obliged to use a sub-optimal regularization parameter, that leads to a worst reconstruction. Figure 7.7 (B) compares the $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ obtained from SGM using an optimal regularization parameter and by OSL using a sub-optimal regularization parameter. The inset shows that SGM leads to a better result than OSL method. The same problem is amplified when OSL is used to solve the MAP problem associate to non-quadratic regularization, especially for the potential function that required low regularization parameter. We want to remark that this problem tends to disappear for high SNR, since the optimal regularization parameter required increases, however we are interested to general algorithms, which are robust also at very low SNR, typical of CLSM.

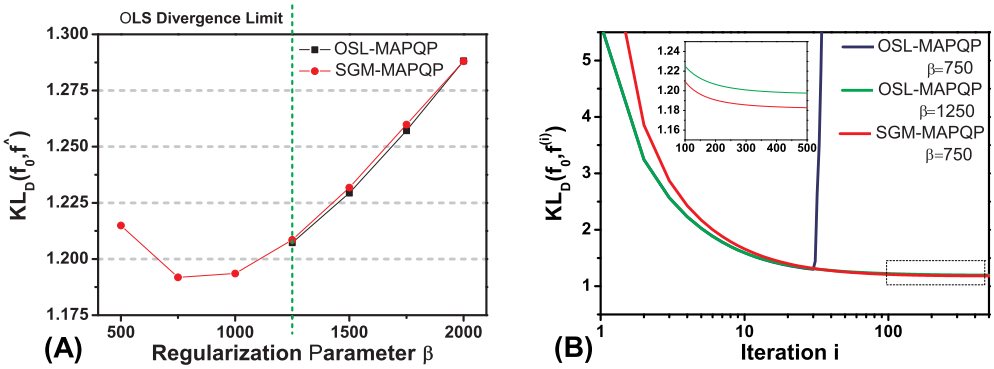

Figure 7.7: (A) Behavior of $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ as a function of the regularization parameter β for quadratic potential regularization in the case of OSL and SGM algorithms (SNR=20). Green dotted line indicate the divergence limit for OSL algorithm. A β value below this limit leads to a divergence for $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$. (B) Behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of iterations i for quadratic potential regularization in the case of OSL and SGM algorithms (SNR=20).

7.1.4 Non-quadratic potential function results

Quadratic potential functions reduce the amplification of noise but introduce overshooting in the solution, with the consequence that the edge of the sub-structure of the object are lost. Here we present the results obtained using the SGM algorithm incorporating the non-quadratic penalization terms derived in Section 5.2.2. In the following we will denote by MAPGM, MAPHL, MAPHB and MAPHS the SGM algorithms derived from the respectively potential functions presented in Table 5.1.
All these non-quadratic potential functions contain a scaling parameter δ. Thus, for an accurate quantitative evaluation of the derived regularized algorithms it is necessary to compare their results not only using an appropriate value of the regularization parameter

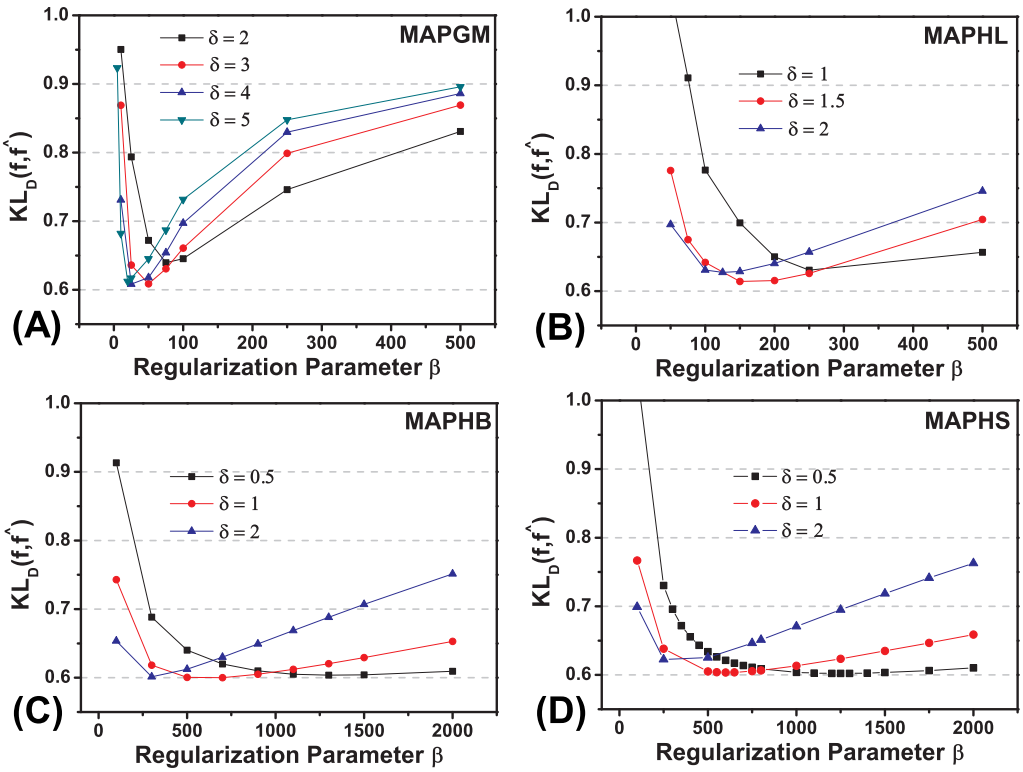

Figure 7.8: Relationship between the parameters β and δ (SNR=20).

β, but also an appropriate value of the scaling parameter δ. Figure 7.8 shows for each potential function the behavior of the $KL_D(\mathbf{f}, \hat{\mathbf{f}})$ as a function of the regularization parameter β for different values of the scaling parameter δ (SNR=20). As expected, there is a relation between the β and the δ parameters: higher is the δ parameter, lower is the optimal β parameter. Moreover, it is easy to check that all the non-quadratic potential functions used in this thesis (see Tab. 5.1) behave as t^2 for small t, thus their limit for $\delta \to +\infty$ is just the quadratic potential. Since we show that MAPQP leads to worst results in terms of KL divergence, a too large scaling parameter can lead to higher KL_D.
Table 7.4 reports the most appropriate combination δ and β values, for different levels

Table 7.4: Optimal values of the parameters δ and β for the different non-quadratic regularization and different values of SNR.

SNR	MAPGM		MAPHL		MAPHB		MAPHS	
	δ	β	δ	β	δ	β	δ	β
15	4	20	1.5	125	0.5	500	0.5	600
20	4	25	1.5	150	1	700	0.5	1200
25	5	20	2	125	1.5	600	0.5	1800

of noise. These values are used for the further restorations. It should be noted that the parameter combination depends on the SNR. The motivation is similar to the motivation already discussed for quadratic potential functions. For a small SNR, combination of regularization and scaling parameters should decrease so that large prior contribution can be introduced for smoothing out the noise. Moreover, we want to remark that MAPGM and MAPHL require smaller regularization parameters respect to MAPHB and MAPHS. This results was already predicted by the analysis of the relative potential functions (see Subsec. 5.2.2).
Figure 7.9 shows the behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of the iterations i (SNR=20). It is important to note that MAPGM and MAPHL have a minimum of the KL_D (semiconvergence of the method), while, MAPHB, MAPHS converge to a suitable solution.
Table 7.5 reports $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ associated to the restored object $\hat{\mathbf{f}}$, obtained using the parameters values reported in Table 7.4. Moreover we report for each restoration processes the iteration in which the stopping threshold is reached. In the cases marked with an asterisk the threshold is not reached within the maximum number of iterations allowed by the code. The first idea this Table illustrate is the superiority in terms of KL divergence of convex potential functions ($\varphi(GM)$ and $\varphi(HL)$) in comparison to non-convex ones ($\varphi(HB)$ and $\varphi(HS)$). The second idea is the similarity between MAPHB and MAPHS results. This is

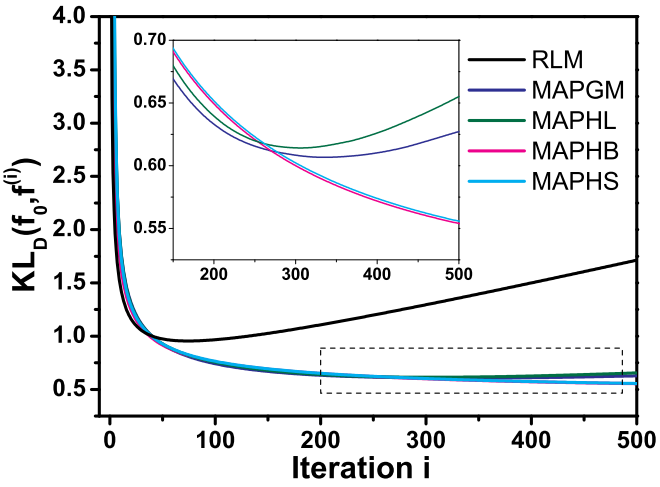

Figure 7.9: Behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of iterations i for non-quadratic regularizations (SNR=20).

Table 7.5: Values of $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ corresponding to the optimal restorations provided by non-quadratic regularizations for different values of SNR. Numbers of iterations used for the optimal restorations are also reported.

SNR	MAPGM		MAPHL		MAPHB		MAPHS	
	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$
15	182	0.82228	176	0.84093	770	0.71077	674	0.71554
20	336	0.60669	307	0.61416	1000*	0.5248	1000*	0.52263
25	416	0.47145	388	0.48353	1000*	0.40938	1000*	0.4113

the consequence of the similarity of the respective potential functions, as shown in Figure 5.2. Finally, all non-quadratic regularizations yield better results than quadratic potential regularization and Tikhonov regularization (see Tab. 7.3).
Figure 7.10 shows the restored objects associated to the results reported in Table 7.5 (SNR = 20). All the restorations using non-quadratic potential functions exhibit excellent performance in both suppressing noise effect and preserving edges: the corresponding restored objects are free from the unfavorable oversmoothing effect and the checkerboard effect. It is important to note as the MAPGM- (Fig. 7.10 (A)) and MAPHL-restored

(Fig. 7.10 (B)) objects are very similar. The same observation can be made comparing MAHB- and MAPHS-restored objects (Fig. 7.10 (C,D), respectively). These results are in complete agreement with the KL_D results. It is also interesting to remark that MAPGM and MAPHS restorations yield very sharp edges in comparison to MAPHB and MAPHS restorations. This observation becomes more evident from line plot investigation. Figure 7.11 shows the intensities profile along predefined lines in the original object and restored objects, both for a lateral profile (Fig. 7.11 (A-C)) and for an axial profile (Fig. 7.11(D-E)). To better appreciate the results obtained with non-quadratic potential functions, we report also the intensity profile for RLM and MAPQP restorations (Fig. 7.11(A,D)). It is possible to see that the non-quadratic regularization restorations have a profiles that agree much better with the real profiles of the synthetic object. While oscillations due to noise amplification are clearly visible in RLM profiles, and partially reduced in MAPQP profiles. Insets of Figure 7.11 (B,C,E,F) show in detail the edge-preserving capability of the non-quadratic potential functions studied, and reveal clearly the very sharp effect obtained by MAPGM and MAPHL, in comparison to MAPHB and MAPHS.

7.1.5 Fuzzy-logic based prior results

Non-quadratic regularization assumes that a large value of the gradient corresponds to an edge while a small value of the gradient is an effect of noise. A different approach consists in discriminate intensity variations due to edge structure or due to noise, using not only the values of the gradient but also global connectivity and continuity information of the object. To explore such properties of the object we use the fuzzy logic (FL) framework. The fuzzy-logic based algorithm derived using SGM, is denoted by MAPFL. In particular two different algorithms are derived, the first explores connectivity and continuity infor-

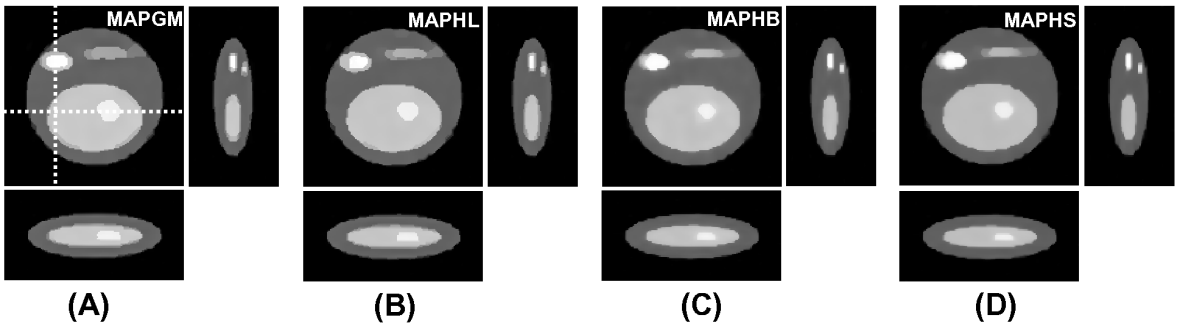

Figure 7.10: Restored objects using Geman & McClure regularization (A), Hebert & Leahy regularization (B), Huber regularization (C) and hyper-surface regularization (D) (SNR=20).

mation of the object using a fuzzy filtering on a 3×3 window (MAPFL3), the second using a 5×5 window (MAPFL5).
In Figure 7.12 (A) the $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ is plotted for MAPFL3 and MAPFL5 algorithms as a function of the regularization parameter β (SNR=20). Also in this case underestimating the regularization parameter leads to larger errors than overestimating it by the same amount. The values for β that yield minimal values in these plots, are considered as appropriate values for generating quality solutions, and are used in the further restorations. Table 7.6 resumes the optimal regularization parameters for different SNRs. It is important to remark that in the FL approach it is not necessary to estimate further parameters, like the scaling parameter for the non-quadratic regularization. The choice of the threshold parameter Th of Equation (5.26) is automatically estimated by the algorithm. This helps enormously the user, that must play attention only in the choice of the regularization pa-

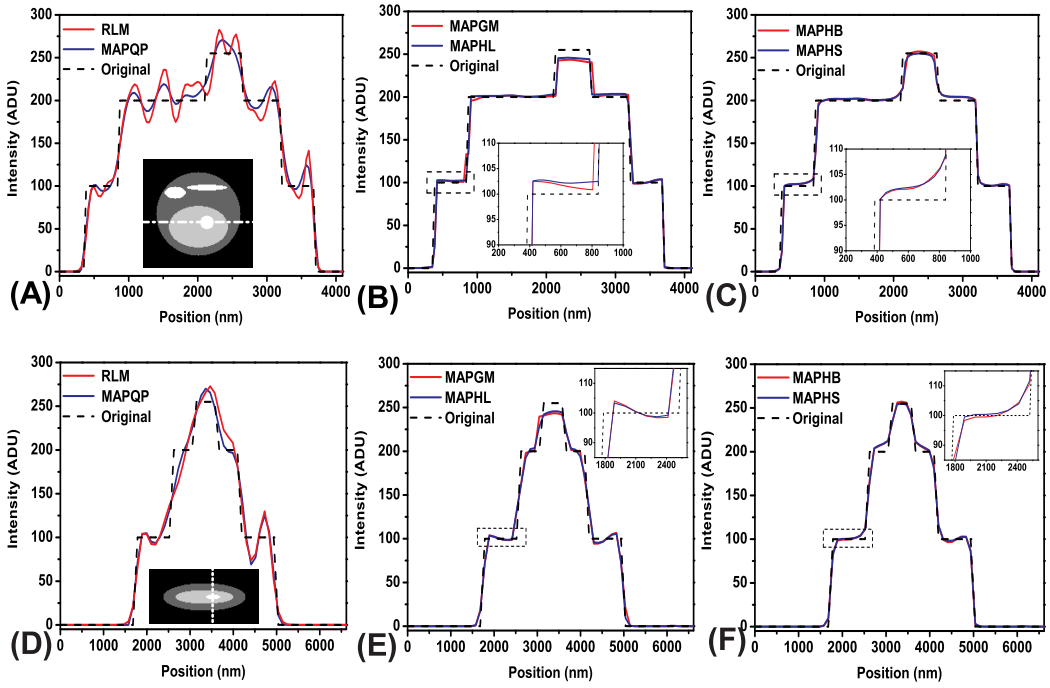

Figure 7.11: Intensity profiles along a lateral line (A,B,C) and an axial line (D,E,F) in the original object and in the non-quadratic restored objects. Insets of (A) and (D) represent respectively the regions where the intensity profile are obtained (SNR=20)

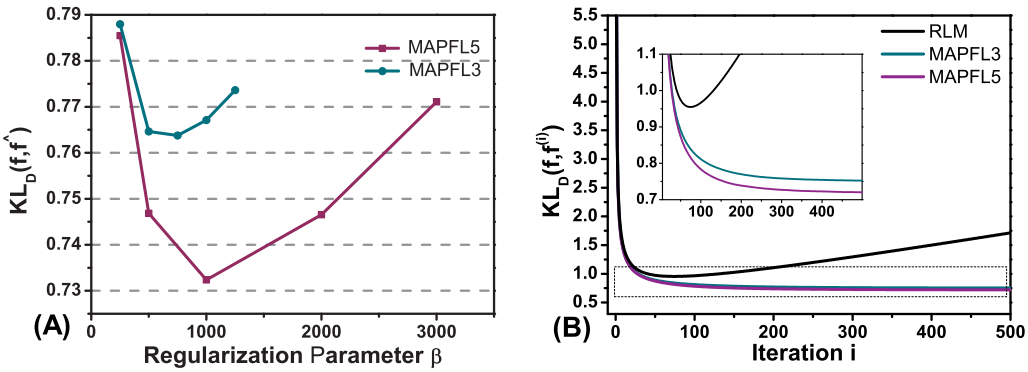

Figure 7.12: (A) Behavior of $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ as a function of the regularization parameter β for fuzzy-logic based regularizations (SNR=20). (B) Behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of iterations i for for fuzzy-logic based regularizations (SNR=20).

rameter β.
Figure 7.12 (B) shows the behavior of $KL_D(\mathbf{f}_0, \mathbf{f}^{(i)})$ as a function of the number of the iterations i (SNR=20). Convergence both of MAPFL3 and MAPFL5 to suitable solutions can be deduced from the graph. Moreover Table 7.6 reports $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ associated to

Table 7.6: Values of $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ corresponding to the optimal restorations provided by fuzzy-logic based regularizations for different values of SNR. Values of the regularization parameter β and numbers of iterations used for the optimal restorations are also reported.

SNR	MAPFL3			MAPFL5		
	β	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	β	iter.	$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$
15	250	225	0.91497	250	212	0.89086
20	750	338	0.75653	1000	409	0.72258
25	3000	1000*	0.63712	3000	1000*	0.60544

the restored objects $\hat{\mathbf{f}}$ and the number of the iteration in which the stopping threshold is reached. Basically two consideration can be made: first the rate of convergence of the two algorithms is the same, second MAPFL5 leads to a better restoration than MAPFL3 in

terms of KL divergence. However, no evident differences between the two methods can be infer by visual inspection (Fig. 7.13) and line plots investigation (Fig. 7.14) of the restored objects associated to the KL divergences reported in the Table (SNR=20).
Simultaneous noise suppression and structure preservation are the advantages of the MAPFL algorithms.
Finally, we want to remark that in this thesis we briefly present the fuzzy-logic approach, but we believe that further investigation can improve thereby its results. In particular we believe that mainly two directions can be followed. The first perspective of work should be to test more sophisticate fuzzy membership functions. The second perspective should be to improve the choice of the threshold Th of Equation (5.26), for example using a more robust statistic and than the average.

7.1.6 Conclusions

In the previous Section we demonstrated that MAP approach is essential to obtain suitable solutions. We want to emphasize the generality of the SGM to devise a large class of algorithms to solve the minimization problem derived from the MAP approach. In particular its simplified multiplicative form helps thereby the user in the implementations of the different regularized algorithms. Therefore, thanks to SGM we started a qualitative and quantitative comparison for a large class of regularization methods. Regularization in the sense of Thikhonov, by non-quadratic potential functions and new fuzzy-logic based potential functions are performed. Even if the convergence of the simplified multiplicative algorithm is not mathematically demonstrated, we did not observe divergent cases for all the regularization methods tested. It is important to remark that we obtain experimental convergence both with convex potential functions, that generally make the minimization problem well posed, and non-convex potential functions. For each regularization function,

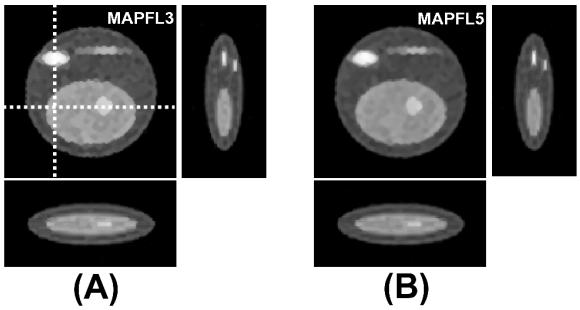

Figure 7.13: Restored objects using 3×3 fuzzy-logic filtering regularization (A) and 5×5 fuzzy-logic filtering regularization (B) (SNR=20).

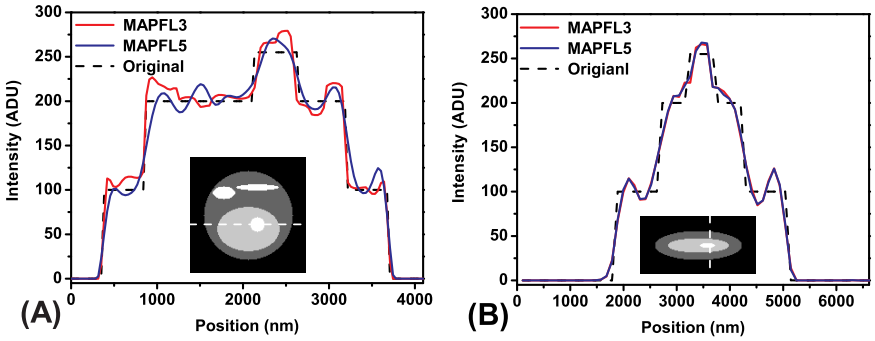

Figure 7.14: Intensity profiles along a lateral line (A) and an axial line (B) in the original object and in the fuzzy-logic based restored objects. Insets of (A) and (B) represent respectively the regions where the intensity profile are obtained (SNR=20)

an algorithm is given and tested on simulated images blurred by a realistic PSF and corrupted by different levels of Poisson noise. The precise determination of the regularization and scaling parameters are outside the scope of this thesis, so that only empirical values of this parameter are used to check the ability and flexibility of the algorithms to give regularized solutions. A figure of merit based on KL divergence is introduced for a quantitative comparison of the proposed regularization methods. This figure of merit shows that non-quadratic convex potential functions yield the best results. To help the reader in the comparison of the proposed regularization methods in terms of KL divergence we resume their relative values in Table 7.7 (SNR=20).

However, we believe that a rigorous comparison between the proposed regularized methods requires more experiments and the definition of more quality criteria. For example here, we use a phantom composed of constant regions with sharp boundaries. We think that this kind of phantom brings out better the quality of non-quadratic regularization, while pushes the quality of quadratic potential regularization and fuzzy-logic based regularization to low levels. A comparison with a phantom composed by smoothing regions, may lead to different results. Further discussions about this point will be carry out during the analysis of real data results.

Table 7.7: $KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$ analysis for all the regularization studied in this thesis. For each regularization the computational time per iteration is reported.

	MAPTK	MAPQP	MAPGM	MAPHL
$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	0.81269	0.78388	0.60669	0.61416
Iter. time (s)	2.1	2.35	2.57	2.61
	MAPHB	MAPHS	MAPFL3	MAPFL5
$KL_D(\mathbf{f}_0, \hat{\mathbf{f}})$	0.5248	0.52263	0.75653	0.72258
Iter. time (s)	2.65	2.75	3.01	3.95

Since, we works in three dimensional image restoration, a very important point to explore is the computational costs for the different regularization methods. Table 7.7 shows the times required for each iteration for the different regularized algorithms. Quadratic and non-quadratic regularizations are comparable in terms of computational costs. While, due to fuzzy-filtering MAPFL3 and MAPFL5 are comparably higher.
Moreover, in general we observe that split-gradient method presents a slow convergence rate. For this reason should be important to exploited the accelerate version proposed by Lanteri et al. [LRCA01].

7.2 Results on Real Data

All the experimental observation of this Section are carry out using Leica TCS (true confocal scanning system) SP5 (Leica Microsystems, Heidelberg, Germany) spectral confocal and multiphoton microscope system. This system is equipped by the SP® spectral detection module. This module allows the tuning of emission bands to any specification within the range of detection. The principle of spectral detection is shown in Figure 7.15. The emission light is spread spectrally by a prism and then guided to a spectrometer slit that allows any emission band to be selected. To operate with multiple bands simultaneously, the slit barriers consist of high-reflectance mirrors. This opens up the possibility of using the residual fraction of the spectrum to apply to further photometer slit devices. Leica TCS SP5 implement the dichroic-emission-excitation filters scheme of Figure 1.3 using not only the spectral module but combining it with a series of other modules, represented by the acousto optical tunable filters (AOTF), for fast selection and stepless attenuation of laser intensity lines, and by the acousto optical beam splitter (AOBS®), for excitation-emission separation.

System is equipped with an Ar 100mW (457-476-488-514 nm wavelength) laser, an HeNe (543 nm wavelength) 1.5 mW laser and is couplet with a Ti:sapphire tunable ultrafast pulsed laser source Chameleon XR (Choherent, Santa Clara) for two-photon excitation (the pulse width of the laser beam is in the range of the 100 fs and the repetition frequency is 90 MHz).
The system mounts a variable numerical aperture objective, immersion oil (n=1.515) HCX PL APO 63×/NA=0.6-1.4 (Leica Microsystems, Heidelberg, Germany) with a working distance of 100 μm. The objective is equipped with an internal iris diaphragm; opening or closing the iris diaphragm determines the size of the objective rear aperture yielding a variable numerical aperture range between 0.6 (close iris diaphragm) and the objective's upper limit 1.4 (open iris diaphragm).

7.2.1 Nano-beads restoration

To show the advance to use a properly model of the PSF we investigate the results obtained by applying image restoration to well-known patterns, such as fluorescent nano-beads of approximately 170 nm of diameter (Ps-Speak kit, Molecular Probes, Eugene, OR). The excitation and emission peaks of these beads are 505 nm and 515 nm respectively. Therefore, we use 488 nm laser line to excite them and we collected the signal in the 500-560 nm spectral window. To image them at different depth into a watery medium (refractive index of approximately 1.33) the nano-beads are imbedded in a 1% agarose gel. Agarose

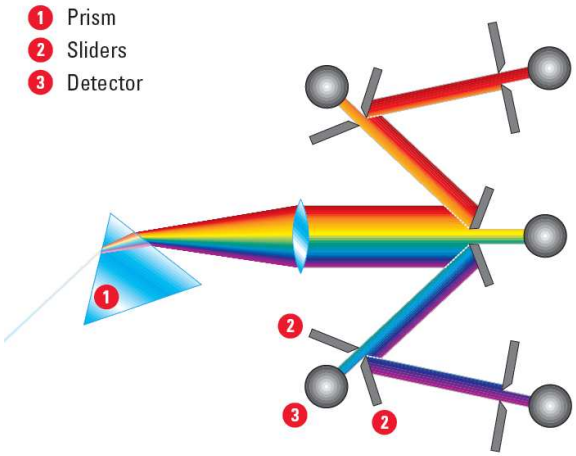

Figure 7.15: Emission from fluorophores in the sample is dispersed by a prism (1) into a spectrum. Detectors (3) receive section of the spectrum defined by movable mirrors, which form spectrometer slits (2).

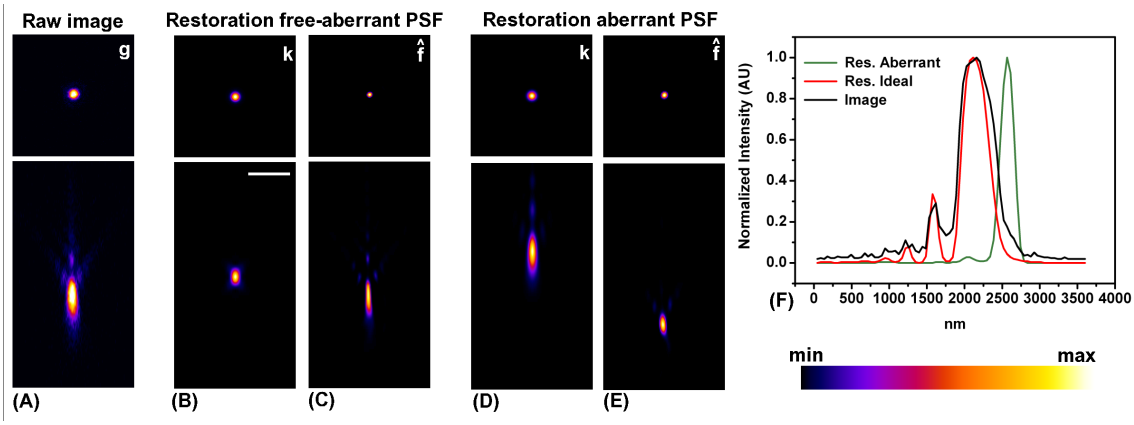

Figure 7.16: Comparison between restoration using a free-aberrated PSF and a more accurate aberrated PSF. (A) Lateral and axial views of a single nano-bead image at approximately 15 μm deep inside agarose gel (imaging condition detail in the text). Aberration effects are evident. (B) Computed free-aberrated (ideal) PSF. (C) Restored bead using free-aberrated PSF. (D) Computed aberrated PSF. (E) Restored bead using aberrated PSF. (F) Intensity axial profiles in the center of the bead for raw data, restored data with aberrated PSF and restored data with ideal PSF. Improvement using aberrated PSF is evident both by visual inspection and by line plots analysis.

is sealed between a properly coverslip (glass 0.17 mm thikness) and a microscope glass. For this measurement pinhole is set to 1 Airy unit (backprojected radius of 208 nm) and iris diaphragm is completely open in order to accomplish the maxima numerical aperture. Sampling is 30 nm in the lateral direction and 90 nm in the axial direction. Figure 7.16 (A) shows the image of a nano-beads at a depth of approximately 15 μm. Figure 7.16(B) represent the PSF computed using the free-aberration model. When this PSF is used to process the raw data, not enough light is removed from above the bead, leaving residual V-shaped tails (see 7.16 (C)). A more accurate restoration is obtained using an appropriate aberrant PSF (see Fig. 7.16 (C)) that takes into account mismatch refractive index between agarose mounting medium and oil immersion objective. In this case asymmetry in the restored object starts to disappear (see Fig. 7.16 (D)). To better understand the improvement obtained using an aberrant PSF, the line intensity plots along the axial profile of the bead for raw image and restored objects are shown in Figure 7.16 (E). Elongation of the bead is remains following restoration with free-aberrated PSF, and interfering secondary peak are not removed. Moreover shift of the bead in the nominal focus position can be obtained using only the aberrant PSF. Quadratic potential regularization is used.

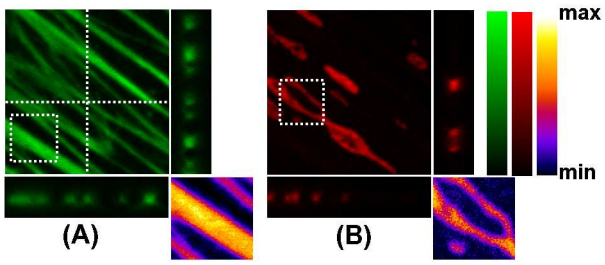

Figure 7.17: (A) Raw image of actin structures in a BPAE cell. (B) Raw image of mitochondria structures in the very same region of (A). For each raw image a lateral view and two axial views along the dotted line are reported. Insets represent the regions of interest in the white dot boxes.

7.2.2 Bovine pulmonary artery endothelial cells restoration

To test the biological validity of any restoration algorithm is always challenging, because there is no so-called ground truth about the cell features of dimensions near the resolution limit of the microscopes. Therefore, we mainly considered visual inspection as direct evidence for quantifying the quality of the 3D restoration. Bovine pulmonary artery endothelial (BPAE) cells are used as model system to test the performance of the proposed algorithms.
The bovine pulmonary artery endothelial cell line was initiated in January 1978 by P. Del Vecchio from the main stem of a pulmonary artery belonging to a young cow. Pulmonary arteries, which extend from the heart to the lungs, are the only arteries in the mammalian body that carry dark, unoxygenate blood. The BPAE line of endothelial cells is positive for bovine diarrhea virus, one of the most important known bovine viral pathogens, which causes a broad array of clinical syndromes that results in significant losses in the beef industry each year. BPAE cells are also positive for angiotensin converting enzyme, an enzyme that is intrinsically involved in the maintenance of bloody pressure and volume. Due to this fact, BPAE cells are often utilized in hypertension research as well as studies of atherosclerosis and coronary hear disease.
An adherent monolayer culture of BPAE cells was labeled for the cytoskeletal filamentous actin and intracellular mitochondrial networks with green-fluorescent BODYPY FL phallacidin and MitoTracker Red CMXRos, respectively. Figure 7.17 shows the images of portion of a BPAE cell. Individual 8 bit image channels are pseudocolored with RGB values corresponding to each of the fluorophore emission spectral profiles, i.e. actin filamentous sub-cellular structures are colored in green (see Fig. 7.17(A)) and mitochondrial network is colored in red (see Fig. 7.17(B)). The insets represent a digital zoom of particular regions of interest contained in the withe dash box.

MitoTracker Red CMXRos and green-fluorescent BODYPY FL phallacidin are excited using 543 nm and 488 nm laser source, respectively. Mitochondria and actin substructure emission were collected "simultaneously" within 500-540 nm and 570-700 nm spectral windows, respectively. The imaging is performed within the ideal condition parameters, the pinhole size is set to 1 Airy unit (backprojected radius of 208 nm) and the Voxel sampling is 45 nm in the lateral direction and 135 nm in the axial direction.
Since we do not know the exact shape of the object, it is not easy to carry out an unambiguous comparison of the different algorithms, particularly because the regularization parameter, the scaling parameter and the optimal number of iterations can not be determined in the way used for synthetic images. To partially overcome this problem, we try several values of the regularization and scaling parameters and inspect the resulting image visually to obtain a restoration with optimal sharpness of the edges without introducing artifacts due to noise amplification. Moreover we use the following criterion based on Kullback-Leiber distance for stopping the iterations of regularized and unregularized algorithms:

$$KL_D(\mathbf{g}, \mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}) - KL_D(\mathbf{g}, \mathbf{A}\mathbf{f}^{(i+1)} + \mathbf{b}) \leq Th_s, \tag{7.4}$$

where $\mathbf{A}\mathbf{f}^{(i)} + \mathbf{b}$ is the computed image, associated to $\mathbf{f}^{(i)}$ from $\mathbf{g}$ and Th_s is set to 10^{-4}. Both for mitochondria and actin images a background of 7 units is estimated using the histogram-based methods [KV00a].
As a consequence of the limited field of view (FOV) of the microscope, it may happen that an extended object is not completely contained within the image domain and an abrupt boundary truncation of an image is introudeuced; in other words the boundaries of the image do not correspond to the free-space (see 7.18 (A) and (B)). In such a case the standard fast Fourier transform (FFT) based methods can not be successfully used. Indeed, the use of the FFT implicitly assumes a periodic continuation of the image outside the original domain; as a consequence, discontinuities appear at the boundaries and, in the restored object, these discontinuities generate Gibbs oscillations (sometimes called ripples), which can propagate inside the object domain and degrade completely the quality of the restoration. These effects are also known as boundary artifacts. A simply methods to overcome this

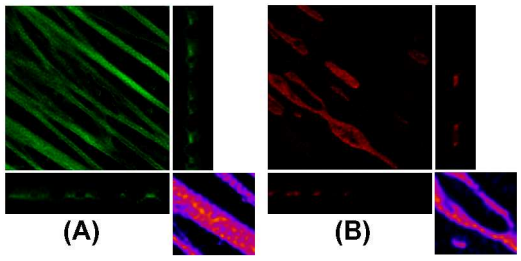

Figure 7.18: Restored objects associated to actin (A) and mitochondria (B) images using RLM.

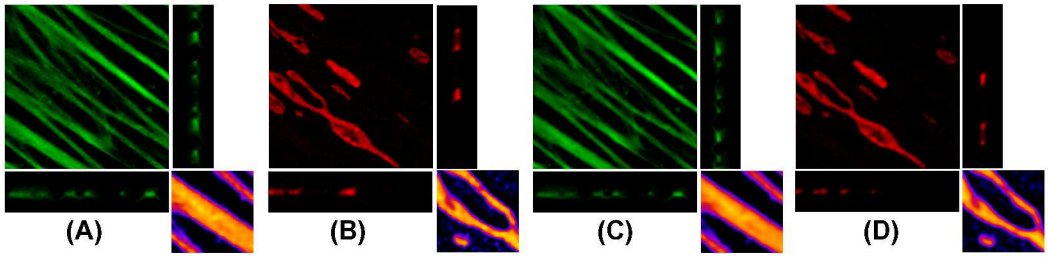

Figure 7.19: Restored objects associated to actin (A,C) and mitochondria (B,D) images using Tikhonov (A,B) and quadratic potential (C,D).

problems is that of symmetrically extending the image [AW96]. This methods consisted of adding a number of layers to the six faces of the original three-dimensional stack of images, using symmetry with respect to the original boundary layers to set the values of the added layers. The number of layers added on each face was determined in such a way that a PSF centered somewhere on the original boundary layers would not appear truncated in the padded image.
Figure 7.18 shows RLM-restored objects; noise amplifications in the form of checkerboard effects appear both in actin (Fig. 7.18 (A)) and in mitochondria restoration (Fig. 7.18 (B)). The solutions are distorted by artifacts originating from an incomplete suppression of noise and by small structures appearing in the filamentous actin for which no indication in the original data is present.
These artifacts partially disappear on MAPTK-restored objects (see Fig. 7.19 (A) and (B)) and are completely suppressed on MAPQP-restored objects (see Fig. 7.19 (C) and (D)). However, smoothing nature of MAPQP is evident.
Figure 7.20 shows the results derived from non-quadratic regularizations. Appreciable restoration of the localized sub-cellular structures can be seen in all the different regularization methods. Noise amplification is suppressed and structure edges are well preserved. Small localized features such as mitochondria are well-restored using edge-preserving algorithms as compared to their counterparts provided by RLM and MAPQP. Similarly to what observed on syntectic data results GM (see Fig. 7.20 (A),(B)) and MAPHL (see Fig. 7.20 (C),(D)) methods lead to similar restored objects. Moreover it is important to remark that staircase effects, do not observed on synthetic data simulation, appears using non-convex potential functions. We believe that this is strictly correlate to the object model assumption made by these potential functions, which enforce constant regions in the solution. Staircase effects are comparatively reduced by using convex potential functions, i.e. MAPHB (see Fig. 7.20 (E),(F)) and MAPHS (see Fig. 7.20 (G),(H)).
Figure 7.21 shows FL-restored objects. Visual inspection brings out that MAPFL-restored object are better resolved compared to RLM- and MAPQP-restored objects. Simultaneous noise suppression, structure preservation, and minimal staircase effects are the advantages

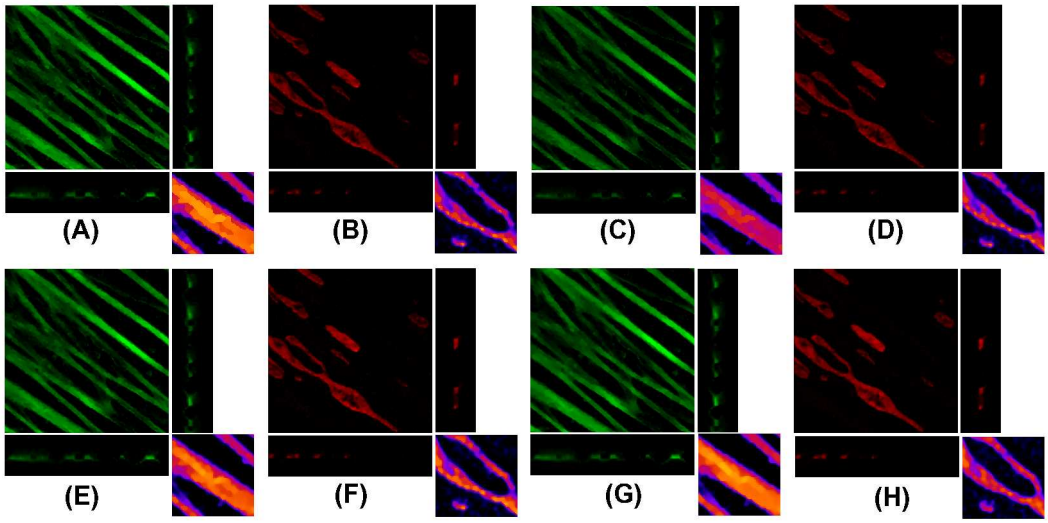

Figure 7.20: Restored objects associated to actin (A,C,E,G) and mitochondria (B,D,F,H) images using Geman & McClure regularization (A,B) Hebert & Leahy regularization (C,D), Huber regularization (E,F) and hyper-surface regularization (G,H).

of the MAPFL algorithm. No evident difference are attainable by comparison between MAPFL3- (see Fig. 7.21 (A)) and MAPFL5-restored objects.
These comparison on real data suggest that convex non-quadratic and fuzzy-logic based regularizations are the suitable methods to obtain free-artifacts restored objects. This remark should be taken into account when deciding what kind of edge-preserving potential it is convenient to use in a practical application. Moreover, we observe that higher regularization and higher scaling parameters respect to that used on simulated data for SNR equal to 25 must be used to obtain suitable restorations.

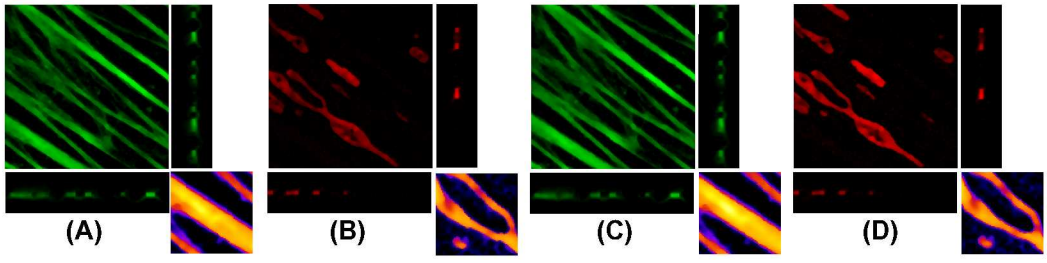

Figure 7.21: Restored objects associated to actin (A,C) and mitochondria (B,D) images using 3×3 fuzzy-logic filtering (A,B) and 5×5 fuzzy-logic filtering (C,D).

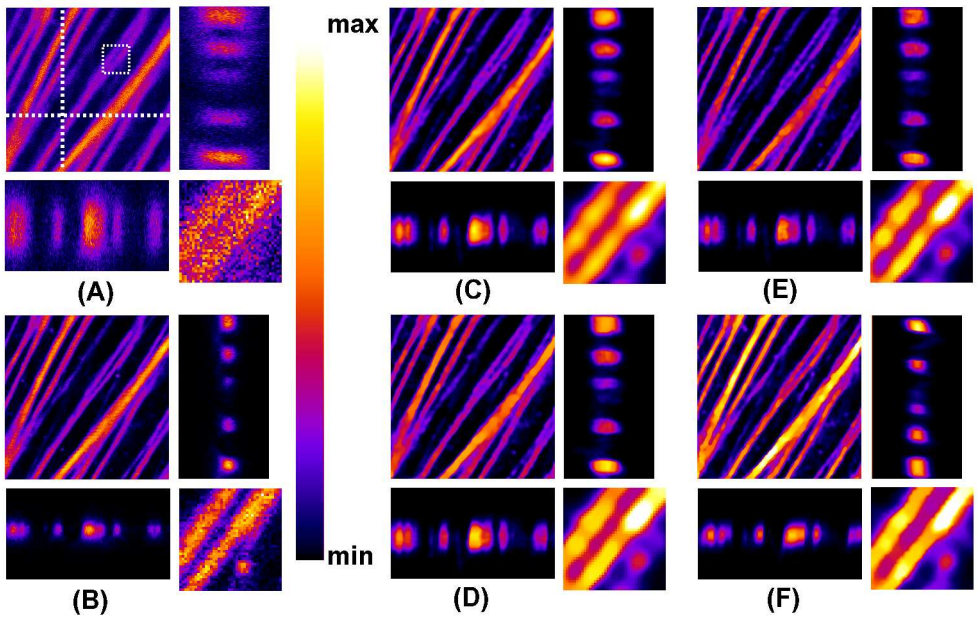

Figure 7.22: Comparison between restored low-numerical aperture images and raw high numerical aperture images. (A,B) Raw actin images obtained using low-numerical aperture objective configuration (NA=0.6) and a high-numerical aperture objective configuration (NA=1.4), respectively. Note that both images represent the same structure. Restored objects obtained from low-numerical aperture image using regularized methods with quadratic potential (C), hyper-surface potential (D), Geman & McClure potential (E), fuzzy logic-based potential (F). For each raw image and restored object two axial views along the dotted lines are reported. The insets represent for each images and restored objects the region of interest contained in the white dash box. Actin structures previously unresolved by lower numerical aperture imaging can be fully resolved by imaging restoration.

A nice trick to qualitatively compare the performance of the proposed algorithms is to acquire the very same sample at different resolution level, using variable numerical aperture objective. Figure 7.22(A) shows actin structures acquired using the objective under the minimal numerical aperture (NA=0.6). The resolution improvement is evident using the objective under the maximal numerical aperture (NA=1.4) (Fig. 7.22 (B)). Restored images (Fig 7.22 (C-F) show thereby the increase of information provided by the reconstruction; actin filament structure previously unresolved, after image restoration are fully resolved (see insets), while artifacts are not generated.

Conclusions

> If I have been able to see further, it was only because I stood on the shoulder of giants.
>
> Isaac Newton

In this thesis we investigated the problem of image restoration associated to multidimensional fluorescence microscopy. Image restoration is a post-processing method that uses known properties of the microscope to mathematically reverse the image formation process and obtain high-resolved and sharper images. The principal results accomplished by this thesis can be resumed as follows.

- A proper mathematical modeling of the image formation process represents a crucial component for a good restoration algorithm. Because a very accurate model for the process of image formation would be mathematically and computationally intractable, different models rely on different simplification assumptions. Principally components of the image formation process are the blurring introduced by band-limited nature of the optical microscope, that can be described in term of its point spread function, and the noise introduced during the recording process.
  An important approximation usually made is to consider the fluorescence microscope as a free-aberration system. However, for example in live-cell imaging, where the most important thing is to put cells in a reliable environment, refractive indexes mismatching represents a non-negligible source of abberations. We showed that the use of a point spread function able to model these aberration effects is a crucial information needed for image restoration. Therefore in this thesis we approached the modeling of the point spread function using the rigorous theory of vectorial light. From this theory a very accurate point spread function model for microscope systems mounting high-numerical aperture objectives and working in abberation conditions is derived. Moreover, this model is parameterized by a well-known notation used by biologist to describe imaging conditions of their experiments, this helps enormously biologist

to compute a properly point spread function for their image restoration experiments [BBS+04].

- Images in fluorescence microscopy are quantum limited, each pixel receives only a limited and random number of photons for unite time. Especially in confocal microscopy where most of the out-of-focus light being rejected by the small confocal aperture. Moreover, electronic introduce noise during the detection process. As a consequence, noise problems are usually important in image restoration and must be treated with particular case. However, the inclusion in the image restoration problem of available noise information, by means of maximum likelihood method, can not be enough to obtain suitable solutions. This is a consequence of the band-limited property of the system, so that inversion of the imaging process is an ill-posed problem and a huge amplification of the noise is introduced. Therefore some constraints (regularization) on the restored object must be applied. Bayesian approach appears the most appropriate methods to applied stochastic constrains on the object during the restoration process. In this thesis we have demonstrated the capability of Markov random field to model suitable statistical a-priori information of the object. In particular different regularizations can be applied simply modeling the potential function associated to the Markov random field. A large class of potential functions are designed to states different three-dimensional properties of the object. Smoothness constraints can be applied using classical quadratic potential function, however edges of the object are completely lost, for this reason edge-preserving potential functions are studied deep in this thesis. Non-quadratic convex and non-convex potential functions are able to produce restored objects with very sharp edges. However, from real biological data application, we observed that non-convex potential functions can introduce non-desiderate staircase effects. A solution to this problem is locked for in the fuzzy-logic framework, and a new-class of fuzzy logic-based potential functions are developed. Very promising results are obtained by their applications on three-dimensional microscope images.

- Bayesian approach transforms the image restoration problem into a constraint minimization problem of a suitable functional. Bayesian approach combined with Markov random field open the possibility to impose a huge amount of different constraints on the restored object simply modeling an appropriate potential function. Therefore each potential function leads to a different functional to minimize. In order to compare their results it is natural to look for a general method able to easily devise a minimization algorithm for any of the functionals that we propose. Split-gradient method represents a perfect tool to address this request. Effective iteration algorithms including also non-negativity and flux-conservation constraints are devised and used to test the different regularized potential functions. Split-gradient-method is based on a decomposition of the gradient of the functional to minimize. In the case of fuzzy-logic base potential function this decomposition becomes hard to compute.

Therefore we developed a new two-step updating algorithm, the first step is devoting to simplify the evaluation of the gradient, the second step is again derived from the split-gradient method. Numerical convergence of the spit-gradient method is observed for all case implemented, and also for different signal-to-noise ratio. This last property represents a key feature of this method when compared with another famous method called one-step-late, since we showed that one-step-late is not convergent for the typical signal level of fluorescence microscopy.

In this thesis we putted more emphases on confocal application, and in particular we showed syntectic and real results only related to such technique, however, we stress that all the tools that we described in this thesis, i.e. point spread function modeling, object modeling, split-gradient method are not restricted to this confocal microscopy image restoration. But they can be used also in two-photon excitation microscopy as well as wide-field microscopy. Moreover, we believe that Markov random field object modeling can be applied with great success to image restoration associated also to other microscopy techniques, like 4PI microscopy [MVAD08], or multiple-image microscopy [CRV$^+$08, BBDV06].

Future Scope

Research will never come to the end. In the same way that we present the results also the future works can be divided in the same three areas, i.e. image formation modeling, object modeling and minimization algorithm.

- A reasonably precise estimation of the point spread function is essential in microscopy image restoration. The physical model that we proposed is able to model aberration induced by refractive mismatching indexes, however also this model presents some limitation. For example, refractive indexes inside a biological structure are very heterogeneous and a simplified three-layer stratified medium model can not be enough, moreover a precise values of such refractive indexes can not be easy to determine. An alternative approach is given by blind and parametric-blind deconvolution. The basic idea of blind deconvolution is to estimate "simultaneously" the point spread function and the object function, however, loosely speaking the new image restoration problem becomes more ill-posed, with the natural consequence that its solution is not-unique. Most existing methods avoid the non-uniqueness by enforcing constraints on the PSF, like non-negativity and bound-limiting. The alternative is given by parametric-blind deconvolution, in which one assumes that the point spread function follows a mathematical model that depends on a small number of parameters. In this case the estimation of the unknown parameters of the point spread function do not require an enforcement of constraints, because the mathematical model automatically satisfies all the constraints.

To keep the computational load practical, researchers make simplifying assumptions in the derivation of the image formation model. Linearity and space-invariance are the main. However, we saw that for thick biological specimens, space-invariance does not necessarily hold. Therefore estimating the object function assuming space-invariance produces artifactual estimated images. Unfortunately, the space-varying model of image formation results in a forward model with a very high computational complexity in terms of CPU time and memory. A compromise between these two models represents an attractive research field.

- In this thesis we assume that objects are piecewise constant. However, we believe that a quadratic piecewise assumption can further improve the results. Therefore, properly potential functions based on higher-order neighbor systems that describe such constraints should be developed.
  Fuzzy-logic based approach has showed great promising results in distinguishing noisy voxels from edge voxel. We have modeled the membership function for the property small and large for the presence of noise in a voxel. The edge transitions in the restored objects are found to be sharp. It is hoped that using more appropriate membership functions and an accurate choice of their thresholds can lead to much better and highly resolved restored objects.

- We believe that most of the open research line described above can be again combined with the split-gradient method. However, an increasing of the computational time will represent an inevitable drawback for all these future advances. Because slow convergence of the split-gradient method is observed an increasing of the rate of convergence is mandatory. Two different approaches are proposed to increase the convergence rate of the split-gradient method, the first one is again proposed by Lanteri et al. [LRCA01] and require a relaxed forms of the algorithm, the second one, known as scaled projection method, has been recently proposed [BLZ08]. By scaled project method one can obtain both a theoretical and a practical improvement of the method. Indeed, it is possible to prove convergence of the modified algorithm and first numerical results in the case of Richardson-Lucy method applied to the image restoration of 2D images indicate a reduction in the computational time by at least an order of magnitude [BZL07].

Appendix A

Epi-Fluorecence Microscopy PSF

In this appendix, starting from intensity detection i_{det} model of a general epi-fluorescence microscope (Eqs. (3.20)-(3.25)), we derive the PSFs for the wide-field microscope, the confocal laser scanning microscope and the laser scanning microscope. All the derivations are performed in one-dimension, the extensions to the three-dimensional case are directly reported in the main text.

A.1 Wide-Field Microscopy

In a wide-field microscope (WFM) we can assume that the illumination aperture a_{ill} is infinitely large, hence the intensity illumination (Eq. (3.20)) becomes a constant:

$$i_{ill}(x) = \int_{-\infty}^{+\infty} a_{ill}(x-\xi)k_{ill}(\xi)\mathrm{d}\xi = \int_{-\infty}^{+\infty} k_{ill}(\xi)\mathrm{d}\xi = C_{ill}. \tag{A.1}$$

In the same way assuming the detection aperture a_{det} infinitely large the detection intensity (Eq. 3.22) becomes:

$$i_{det}(x) = k_{det}(x) * f(x). \tag{A.2}$$

Simply, substituting the object function f with the Dirac function δ, we obtain the PSF of a WFM:

$$k_{WFM}(x) = k_{det}(x). \tag{A.3}$$

We want to remark that Equation (A.3) assumes that the sampling introduced by the CCD is enough to not lose signal frequency during the detection process, in other words if it is possible to neglect the finite dimensions of the pixel of the CCD [You89].

A.2 Confocal laser scanning microscopy

In a confocal laser scanning microscope (CLSM) both illumination and detection aperture are assumed to be infinitesimally small, hence a_{ill} and a_{det} can be substituted by the Dirac function δ, and the detection intensity (Eq. (3.22)) becomes:

$$i_{det}(x') = \delta(x')\Big(k_{det}(x') * \Big(\big(k_{ill}(x') * \delta(x')\big)f(x')\Big)\Big). \tag{A.4}$$

The detection intensity is recorded by a single photo-multiplier (PMT), therefore i_{PMT} yields at a single point and its value is obtain by integrating over the surface of the sensor (that we assume infinitely large):

$$\begin{aligned} i_{PMT}(0) &= \int_{\infty}^{+\infty} \delta(x')\Big(k_{det}(x') * \Big(\big(k_{ill}(x') * \delta(x')\big)f(x')\Big)\Big) = \\ &= \int_{\infty}^{+\infty} \delta(x')\Big(k_{det}(x') * \Big(k_{ill}(x')f(x')\Big)\Big) = \\ &= \big(k_{det} * (k_{ill}f)\big)(0) = \int_{\infty}^{+\infty} k_{det}(-\xi)k_{ill}(\xi)f(\xi)\mathrm{d}\xi. \end{aligned} \tag{A.5}$$

Since the bi-dimensional image in a CLSM is obtained by scanning the specimen, the intensity recorded by the PMT for each point can be derived from Equation (A.5) by shifting both the illumination and detection over the object:

$$\begin{aligned} i_{PMT}(x) &= \int_{\infty}^{+\infty} \delta(x'-x)\Big(k_{det}(x') * \big((k_{ill}(x') * \delta(x'-x))f(x')\big)\Big)\mathrm{d}x' = \\ &= \int_{\infty}^{+\infty} \delta(x'-x)\Big(k_{det}(x') * \big(k_{ill}(x'-x)f(x')\big)\Big)(x')\mathrm{d}x' = \\ &= \int_{\infty}^{+\infty} \delta(x'-x)\Big(\int_{-\infty}^{+\infty} k_{det}(x'-\xi)k_{ill}(\xi-x)f(\xi)\mathrm{d}\xi\Big)\mathrm{d}x' = \\ &= \int_{-\infty}^{+\infty} k_{det}(x-\xi)k_{ill}(\xi-x)f(\xi)\mathrm{d}\xi = \big(k_{det}(x)k_{ill}(-x)\big) * f(x). \end{aligned} \tag{A.6}$$

Substituting the object function with the Dirac function one obtains the PSF of the CLSM:

$$k_{CLSM}(x) = k_{det}(x)k_{ill}(-x) \tag{A.7}$$

In practice the detection pinhole is not infinitely small, hence we can not use the Dirac function to model the detection aperture, but we have to use a finite size aperture:

$$\begin{aligned} i_{PMT}(x) &= \int_{\infty}^{+\infty} a_{det}(x-x')\Big(k_{det}(x') * \big((k_{ill}(x') * \delta(x'-x))f(x')\big)\Big)\mathrm{d}x' = \\ &= \int_{\infty}^{+\infty} a_{det}(x-x')\bigg(\int_{-\infty}^{+\infty} k_{det}(x'-\xi)k_{ill}(\xi-x)f(\xi)\mathrm{d}\xi\bigg)\mathrm{d}x' = \\ &= \int_{-\infty}^{+\infty}\bigg(\int_{-\infty}^{+\infty} a_{det}(x-x')k_{det}(x'-\xi)\mathrm{d}x'\bigg)k_{ill}(\xi-x)f(\xi)\mathrm{d}\xi = \\ &= \int_{-\infty}^{+\infty}(a_{det} * k_{det})(x-\xi)k_{ill}(\xi-x)f(\xi)\mathrm{d}\xi = \big((a_{det} * k_{det})(x)h_{ex}(-x)\big) * f(x). \end{aligned} \tag{A.8}$$

Therefore a more accurate PSF for the CLSM is given by:

$$k_{CLSM}(x) = \big(a_{det} * k_{det}(x)\big)k_{ill}(-x) \tag{A.9}$$

Finally, in the case of laser scanning microscope (LSM) the detection aperture a_{det} is enlarged to infinite. Therefore the convolution in Equation (A.9) becomes a constant that we can avoid from the expression of the PSF:

$$k_{LSM}(x) = k_{ill}(-x) \tag{A.10}$$

Appendix B

Discrepancy Functionals

B.1 Poisson Case

Starting from Equation 2.21 that describes the probability density distribution $P_{\mathbf{G}}(\mathbf{g};\mathbf{f})$ in the Poisson case we apply the neglog to $\mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f})$ in order to derive the relative discrepancy functional $J_0(\mathbf{g};\mathbf{f})$:

$$-\ln \mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f}) = \sum_{\mathbf{n}} \Big\{ (\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) - \mathbf{g}(\mathbf{n}) \ln \big((\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) \big) + \ln \big(\mathbf{g}(\mathbf{n})! \big) \Big\}. \tag{B.1}$$

Using the formula of Stiriling for the factorial term:

$$\ln \big(\mathbf{g}(\mathbf{n})! \big) = \Big\{ \mathbf{g}(\mathbf{n}) \ln \big(\mathbf{g}(\mathbf{n}) \big) - \mathbf{g}(\mathbf{n}) + \ln \big(\sqrt{2\pi \mathbf{g}(\mathbf{n})} \big) \Big\}, \tag{B.2}$$

one obtains:

$$-\ln \mathcal{L}_{\mathbf{g}}^{\mathbf{G}}(\mathbf{f}) = \sum_{\mathbf{n}} \Big\{ \mathbf{g}(\mathbf{n}) \ln \frac{\mathbf{g}(\mathbf{n})}{(\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n})} + (\mathbf{Af})(\mathbf{n}) + \mathbf{b}(\mathbf{n}) - \mathbf{g}(\mathbf{n}) + \ln \big(\sqrt{2\pi \mathbf{g}(\mathbf{n})} \big) \Big\}. \tag{B.3}$$

Thus, using the definition of KL divergence (Eq. (4.13)), one obtains:

$$J_0(\mathbf{g};\mathbf{f}) = KL_D(\mathbf{g}, \mathbf{Af} + \mathbf{b}) = -\ln P_{\mathbf{G}}(\mathbf{g};\mathbf{f}) - \sum_{\mathbf{n}} \ln \big(\sqrt{2\pi \mathbf{g}(\mathbf{n})} \big), \tag{B.4}$$

and taking in to account that we know $\mathbf{g}$, the choice of the constants B and C in Equation (4.8) are obvious.

To easily derive the gradient of the discrepancy functional $J_0(\mathbf{g};\mathbf{f})$ we use the lexicographic ordering vector notation for $\mathbf{g}$, $\mathbf{b}$ and $\mathbf{f}$. Thus, expanding the matrix-vector multiplication $\mathbf{Af}$, the discrepancy functional $J_0(\mathbf{g};\mathbf{f})$ can be written as:

$$J_0(\mathbf{g};\mathbf{f}) = \sum_m \left\{ \mathbf{g}(m) \ln \frac{\mathbf{g}(m)}{\sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m)} + \sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m) - \mathbf{g}(m) \right\}. \quad \text{(B.5)}$$

By elementary computation:

$$\begin{aligned}
\big(\nabla_{\mathbf{f}} J_0(\mathbf{g};\mathbf{f})\big)(n) &= \sum_m \left\{ \mathbf{g}(m) \frac{\sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m)}{\mathbf{g}(m)} \times \right. && \text{(B.6)} \\
&\times \left. \frac{-\mathbf{g}(m)}{\big(\sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m)\big)^2} \mathbf{A}(m,n) \right\} + && \text{(B.7)} \\
&+ \sum_m \mathbf{A}(m,n) = && \text{(B.8)} \\
&= -\left(\mathbf{A}^T \frac{\mathbf{g}}{\mathbf{Af}+\mathbf{b}} \right)(n) + \alpha_n, && \text{(B.9)}
\end{aligned}$$

where: α_n is the sum of the elements of the nth column of the matrix $\mathbf{A}$; the quotient of the two vectors $\mathbf{g}$ and $\mathbf{Af}+\mathbf{b}$ is defined components by components (see Eq. (4); $\mathbf{A}^T$ denotes the transposed of the matrix $\mathbf{A}$.
Finally, we want to remark that $\mathbf{A}$ is a cyclic matrix thus all coefficients α_n are equal. Moreover, if the PSF $\mathbf{k}$ is normalized in such a way that the sum of all its voxel values is 1, then α_n is equal to 1. In conclusion, in the Poisson case the gradient of the discrepancy functional is given by:

$$\nabla_{\mathbf{f}} J_0(\mathbf{g};\mathbf{f}) = \mathbf{1} - \mathbf{A}^T \frac{\mathbf{g}}{\mathbf{Af}+\mathbf{b}}, \quad \text{(B.10)}$$

where $\mathbf{1}$ is the vector whose entries are equal to 1.

B.2 White Gaussian Case

Starting from Equation 2.19 that describes the probability density distribution $P_{\mathbf{G}}(\mathbf{g};\mathbf{f})$ in the white Gaussian case we apply the neglog to $\mathcal{L}^{\mathbf{G}}_{\mathbf{g}}(\mathbf{f})$ in order to derive the relative discrepancy functional $J_0(\mathbf{g};\mathbf{f})$:

$$-\ln \mathcal{L}^{\mathbf{G}}_{\mathbf{g}}(\mathbf{f}) = -N \ln \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) + \left(\frac{1}{2\sigma^2} \|\mathbf{f}_0 - (\mathbf{Af}_0 + \mathbf{b})\|_2^2 \right). \quad \text{(B.11)}$$

Thus:

$$J_0(\mathbf{g};\mathbf{f}) = -2\sigma^2 \ln P_{\mathbf{G}}(\mathbf{g};\mathbf{f}) + 2\sigma^2 N \ln\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right), \tag{B.12}$$

the choice of the constants B and C in Equation (4.8) are obvious.

To easily derive the gradient of the discrepancy functional $J_0(\mathbf{g};\mathbf{f})$ we use the lexicographic ordering vector notation for $\mathbf{g}$, $\mathbf{b}$ and $\mathbf{f}$. Thus, expanding the matrix-vector multiplication $\mathbf{Af}$, the discrepancy functional $J_0(\mathbf{g};\mathbf{f})$ can be written as:

$$J_0(\mathbf{g};\mathbf{f}) = \sum_m \Big| \sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m) - \mathbf{g}(m)\Big|^2. \tag{B.13}$$

By elementary computation:

$$\begin{aligned}
\big(\nabla_{\mathbf{f}} J_0(\mathbf{g};\mathbf{f})\big)(n) &= \sum_m \Big\{ 2\Big| \sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m) - \mathbf{g}(m)\Big| \times & \text{(B.14)}\\
&\times \operatorname{sgn}\Big(\sum_n \mathbf{A}(m,n)\mathbf{f}(n) + \mathbf{b}(m) - \mathbf{g}(m)\Big)\mathbf{A}(m,n)\Big\} = & \text{(B.15)}\\
&= \sum_m \Big\{ 2\Big((\mathbf{Af})(m) + \mathbf{b}(m) - \mathbf{g}(m)\Big)\mathbf{A}(m,n)\Big\} = & \text{(B.16)}\\
&= 2(\mathbf{A}^T\mathbf{Af})(\mathbf{n}) + 2(\mathbf{A}^T\mathbf{b})(\mathbf{n}) - 2(\mathbf{A}^T\mathbf{g})(\mathbf{n}). & \text{(B.17)}
\end{aligned}$$

Finally, if the PSF $\mathbf{k}$ is normalized in such a way that the sum of all its voxel values is 1, and $\mathbf{b}$ is a constant vector, the final form of the gradient of the discrepancy functional in the Gaussian case is:

$$\nabla_{\mathbf{f}} J_0(\mathbf{g};\mathbf{f}) = 2\mathbf{A}^T\mathbf{Af} + 2\mathbf{b} - 2\mathbf{A}^T\mathbf{g}. \tag{B.18}$$

Appendix C

Gradient of the Regularization Term

C.1 Quadratic and non-quadratic cases

Both quadratic and non-quadratic case can be derived using a general potential function on the first derivative $\varphi\big(D^1(\mathbf{f})\big)$. Starting from Equation (6.21) let us calculate the local expression of $\nabla_{\mathbf{f}} J_R(\mathbf{f})$ at site $(n1, n2, n3)$. In Equation 6.21 twenty different terms contain the value of $\mathbf{f}$ at site $(n1, n2, n3)$, in particular four terms for each direction. In order to simplify our presentation, we forget about vertical, axial and diagonals differences for the moment. Their case will be added directly in the final formulation as obvious extension. This leads to the following expression for the gradient of the regularization term:

$$\begin{aligned}
&(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(n_1, n_2, n_3) = \\
&= \varphi'\left(\frac{\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 + 1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} - \varphi'\left(\frac{\mathbf{f}(n_1 - 1, n_2, n_3) - \mathbf{f}(n_1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} + \\
&+ \varphi'\left(\frac{\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 - 1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} - \varphi'\left(\frac{\mathbf{f}(n_1 + 1, n_2, n_3) - \mathbf{f}(n_1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} +
\end{aligned} \tag{C.1}$$

Since we assume that φ is even we obtain that φ' is odd, therefore:

$$\begin{aligned}
&(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(n_1, n_2, n_3) = \\
&= \varphi'\left(\frac{\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 + 1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} + \varphi'\left(\frac{\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 - 1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} + \\
&+ \varphi'\left(\frac{\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 - 1, n_2, n_3)}{\delta}\right)\frac{1}{\delta} + \varphi'\left(\frac{\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 + 1, n_2, n_3)}{\delta}\right)\frac{1}{\delta},
\end{aligned} \tag{C.2}$$

Using $\mathscr{N}_{\mathbf{n}}$ to indicate the set of neighbors and adding also the vertical, axial and diagonals terms the gradient can be written in the following compact form:

$$(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(\mathbf{n}) = \frac{2}{\delta} \sum_{\mathbf{m} \in \mathscr{N}_{\mathbf{n}}} \varphi' \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})\delta} \right) \frac{1}{d(\mathbf{n}, \mathbf{m})}. \tag{C.3}$$

Moreover, Equation (C.3) can be rewritten in terms of the weighting function ψ:

$$(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(\mathbf{n}) = \frac{4}{\delta^2} \sum_{\mathbf{m} \in \mathscr{N}_{\mathbf{n}}} \frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})^2} \psi \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})\delta} \right). \tag{C.4}$$

C.2 Fuzzy-logic based case

Starting from Equation (6.24) let us calculate the local expression of $\nabla_{\mathbf{f}} J_R(\mathbf{f})$ at site $(n1, n2, n3)$. In Equation (6.24) twenty different terms contain the value of $\mathbf{f}$ at site $(n1, n2, n3)$, in particular four terms for each direction. In order to simplify our presentation, we forget about vertical, axial and diagonals differences for the moment. Their case will be added directly in the final formulation as obvious extension. This leads to the following expression for the gradient of the regularization term:

$$\begin{aligned}
&(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(n_1, n_2, n_3) = \\
&= 2w_{\mathbf{f}}\big((n_1, n_2, n_3), (n_1 + 1, n_2, n_3)\big)\big(\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 + 1, n_2, n_3)\big) - \\
&- 2w_{\mathbf{f}}\big((n_1 - 1, n_2, n_3), (n_1, n_2, n_3)\big)\big(\mathbf{f}(n_1 - 1, n_2, n_3) - \mathbf{f}(n_1, n_2, n_3)\big) + \\
&+ 2w_{\mathbf{f}}\big((n_1, n_2, n_3), (n_1 - 1, n_2, n_3)\big)\big(\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 - 1, n_2, n_3)\big) - \\
&- 2w_{\mathbf{f}}\big((n_1 + 1, n_2, n_3), (n_1, n_2, n_3)\big)\big(\mathbf{f}(n_1 + 1, n_2, n_3) - \mathbf{f}(n_1, n_2, n_3)\big) = \\
&= 2\big\{w_{\mathbf{f}}\big((n_1, n_2, n_3), (n_1 + 1, n_2, n_3)\big) + w_{\mathbf{f}}\big((n_1 + 1, n_2, n_3), (n_1, n_2, n_3)\big)\big\} \times \\
&\times \big(\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 + 1, n_2, n_3)\big) + \\
&+ 2\big\{w_{\mathbf{f}}\big((n_1, n_2, n_3), (n_1 - 1, n_2, n_3)\big) + w_{\mathbf{f}}\big((n_1 - 1, n_2, n_3), (n_1, n_2, n_3)\big)\big\} \times \\
&\times \big(\mathbf{f}(n_1, n_2, n_3) - \mathbf{f}(n_1 - 1, n_2, n_3)\big)
\end{aligned} \tag{C.5}$$

Using $\mathscr{N}_{\mathbf{n}}$ to indicate the set of neighbors and adding also the vertical, axial and diagonals terms the gradient can be written in the following compact form:

$$(\nabla_{\mathbf{f}} J_R(\mathbf{f}))(\mathbf{n}) = 2 \sum_{\mathbf{m} \in \mathscr{N}_{\mathbf{n}}} \big(w_{\mathbf{f}}(\mathbf{n}, \mathbf{m}) + w_{\mathbf{f}}(\mathbf{m}, \mathbf{n})\big) \left(\frac{\mathbf{f}(\mathbf{n}) - \mathbf{f}(\mathbf{m})}{d(\mathbf{n}, \mathbf{m})^2} \right). \tag{C.6}$$

Bibliography

[Anc06] B. Anconelli. *Multiple image deconvolution: methods, software and application to the Large Binocular Telescope*. PhD thesis, Ph.D. Thesis in Computer Science, 2006.

[AW96] F. Aghdasi and R.K. Ward. Reduction of boundary artifacts in image restoration. *Image Processing, IEEE Transactions on*, 5(4):611–618, 1996.

[BB98] M. Bertero and P. Boccacci. *Introduction to Inverse Problem in Imaging*. IoP Publishing, Bristol, 1998.

[BBDV06] M. Bertero, P. Boccacci, G. Desidera, and G. Vicidomini. High-resolution imaging by multiple-image deconvolution. In Gabriel Cristabal, Bahram Javidi, and Santiago Vallmitjana, editors, *Information optics: 5th International Workshop on Information Optics (WIO'06)*, volume 860, pages 3–14. API, 2006.

[BBS+04] P. Bonetto, P. Boccacci, M. Scarito, M. Davolio, M. Epifani, G. Vicidomini, C. Tacchetti, P. Ramoino, C. Usai, and A. Diaspro. Three-dimensional microscopy migrates to the web with "powerup your microscope". *Microsc Res Tech*, 64(2):196–203, 2004.

[BCT+07] F. Benvenuto, A. La Camera, C. Theys, A. Ferrari, H. Lanteri, and M. Bertero. Study of an iterative method for the reconstruction of images corrupted by poisson and gauss noise. *submitted to Inverse Problems*, 2007.

[Bes74] J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36:192–236, 1974.

[BFTL93] B. Bailey, D. L. Farkas, D. L. Taylor, and F. Lanni. Enhancement of axial resolution in fluorescence microscopy by standing-wave excitation. *Nature*, 336:44–48, 1993.

[BLZ08] M. Bertero, H. Lantéri, and L. Zanni. Iterative image reconstruction: a point of view. In Y. Censor, M. Jiang, and A.K. Louis, editors, *Mathematical Methods in Biomedical Imaging and Intensity-Modulated Radiation Therapy (IMRT)*, volume (to appear) of *Edizioni della Normale, Pisa, Italy*, 2008.

[BP06] M. Bertero and M. Piana. Inverse problems in biomedical imaging: modeling and methods of solution. In L. Formaggia A. Quarteroni and A. Veneziani, editors, *Complex Systems in Biomedicine*, pages 1–33. Springer, Berlin, 2006.

[BW65] A. Boivin and E. Wolf. Electromagnetic field in the neighborhood of the focus of a coherent beam. *Phys. Rev.*, 138(6B):B1561–B1565, 1965.

[BZL07] S. Bonettini, R. Zanella, and L.Zanni. A scaled gradient projection method for constrained image deblurring. Technical Report 78, Dept. of Mathematics, University of Modena and Reggio Emilia, 2007.

[CBFAB94] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Two deterministic half-quadratic regularization algorithms for computed imaging. In *Image Processing, Proceedings. ICIP-94, IEEE International Conference*, volume 2, pages 168–172. IEEE, 1994.

[CBFAB97] P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *Image Processing, IEEE Transactions on*, 6(2):298–311, 1997.

[CL05] J.A. Conchello and J.W. Lichtman. Optical sectioning microscopy. *Nature Methods*, 2(12):920–931, 2005.

[CLM+95] W.A. Carrington, R.M. Lynch, E.D. Moore, G. Isenberg, K.E. Fogarty, and F.S. Fay. Superresolution three-dimensional images of fluorescence in cells with minimal light exposure. *Science*, 268(5216):1483–1487, 1995.

[CM96] J.A. Conchello and J.G. McNally. Fast regularization technique for expectation maximization algorithm for optical sectioning microscopy. In C. J. Cogswell, G. S. Kino, and T. Wilson, editors, *Three-Dimensional Microscopy: Image Acquisition and Processing III*, volume 2655, pages 199–208. SPIE, 1996.

[CRV+08] F. Cella, E. Ronzitti, G. Vicidomini, P. P. Mondal, and A. Diaspro. Studying the illumination puzzle towards an isotropic increase of optical resolution. In Jose-Angel Conchello, Carol J. Cogswell, and Tony Wilson, editors, *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XV*, volume 6861, page 686112. SPIE, 2008.

[Csi91] I. Csiszar. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2032–2066, 1991.

[Dia02] A. Diaspro. *Confocal and Two-Photon Microscopy: Foundations, Applications and Advances*. John Wiley & sons, New York, 2002.

[dMLCU01] J. B. de Monvel, S. Le Calvez, and M. Ulfendahl. Image restoration for confocal microscopy: Improving the limits of deconvolution, with application to the visualization of the mammalian hearing organ. *Biophys. J.*, 80(5):2455–2470, 2001.

[DMS+04] F. Difato, F. Mazzone, S. Scaglione, M. Fato, F. Beltrame, L. Kubnov, J. Jancek, P. Ramoino, G. Vicidomini, and A. Diaspro. Improvement in volume estimation from confocal sections after image deconvolution. *Microsc Res Tech*, 64(2):151–155, 2004.

[dMSCU03] J. B. de Monvel, E. Scarfone, S. Le Calvez, and M. Ulfendahl. Image-adaptive deconvolution for three-dimensional deep biological imaging. *Biophys J*, 85(6):3991–4001, 2003.

[EH99] A. Egner and S. W. Hell. Equivalence of the huygens-fresnel and debye approach for the calculation of high aperture point-spread functions in the presence of refractive index mismatch. *Journal of Microscopy*, 193(3):244–249, 1999.

[EHN96] H.W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Dordrecht: Kluwer Academic Publishers, 1996.

[FJ05] M. Frigo and S.G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. special issue on "Program Generation, Optimization, and Platform Adaptation".

[GAS99] M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat. I5m:3d widefield light microscopy with better than 100 nm axial resolution. *J.Microsc.*, 195:1016, 1999.

[GG84] S. Geman and D. Geman. Stochastic relaxation, gibbs distribution and the bayesian restoration of images. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6(6):721–741, 1984.

[GL91] S.F. Gibson and F. Lanni. Experimental test of an analytic model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *J. Opt. Soc. Am. A*, 8:16011613, 1991.

[GM85] S. Geman and D.E. McClure. Bayesian image analysis: An application to single photon emission tomography. In *Proc. Statistical Computation Section*, pages 12–18. Amer. Statistical Assoc., Washington, DC, 1985.

[GR92] D. Geman and C. Reynolds. Constraint restoration and the recovery of discontinuities. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 14(3):367–383, 1992.

[Gre90a] P.J. Green. Bayesian reconstructions from emission tomography data using a modified em algorithm. *Medical Imaging, IEEE Transactions on*, 9(1):84–93, 1990.

[Gre90b] P.J. Green. On the use of the em algorithm for penalized likelihood estimation. *J. Roy. Statist. Soc. B*, 52(3):443–452, 1990.

[GY05] E. Gravier and Y. Yang. Motion-compensated reconstruction of tomographic image sequences. *IEEE Trans. Nucl. Sci.*, 52(1):51–56, 2005.

[Had02] J. Hadamard. Sur les problemes aux derivees partielles et leur signification physique. *Princeton Univ. Bull.*, 13:49–52, 1902.

[Hae03] O. Haeberlé. Focusing of light through a stratified medium: a practical approach for computing microscope point spread functions. part i: Conventional microscopy. *Optics Communications*, 216:55–63, 2003.

[Hae04] O. Haeberlé. Focusing of light through a stratified medium: a practical approach for computing microscope point spread functions. part ii: confocal and multiphoton microscopy. *Optics Communications*, 235(1–3):1–10, 2004.

[HAF+03] O. Haeberlé, M. Ammar, H. Furukawa, K. Tenjimbayashi, and P. Trk. Point spread function of optical microscopes imaging through stratified media. *Optics Express*, 11(22):2964, 2003.

[HC71] J.M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. *Unpublished*, 1971.

[Hol88] T.J. Holmes. Maximum-likelihood image restoration adapted for noncoherent optical imaging. *J Opt Soc Am A*, 5(5):666–673, 1988.

[Hol92] T.J. Holmes. Blind deconvolution of quantum-limited incoherent imagery. *J. Opt. Soc. Am. A*, 9:10521061, 1992.

[HRCS93] S.W. Hell, G. Reiner, C. Cremer, and E.H.K. Stelzer. Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *J. Microsc.*, 169:391405., 1993.

[HS92] S. Hell and E. H. K. Stelzer. Properties of a 4pi confocal fluorescence microscope. *J. Opt. Soc. Am. A*, 9(12):2159–2166, 1992.

[HYH89] T.J. Holmes and L. Yi-Hwa. Richardson-lucy/maximum likelihood image restoration algorithm for fluorescence microscopy: further testing. *Applied Optics*, 28(22):4930–4938, 1989.

[JAJ89] T. M. Jovin and D. J. Arndt-Jovin. Luminescence digital imaging microscopy. *Annu Rev Biophys Biophys Chem*, 18:271–308, 1989.

[Kem99] G.M.P. Van Kempen. *Image Restoration in Fluorescence Microscopy.* PhD thesis, 1999.

[KV00a] G. M. Van Kempen and L. J. Van Vliet. Background estimation in nonlinear image restoration. *J Opt Soc Am A Opt Image Sci Vis*, 17(3):425–433, 2000.

[KV00b] G.M.P. Van Kempen and L.J. Van Vliet. The influence of the regularization parameter and the first estimate on the performance of tikhonov regularized non-linear image restoration algorithms. *J Microsc*, 198 (Pt 1):63–75, 2000.

[Lak99] J.R. Lakowicz. *Principles of fluorescence spectroscopy.* New York: Kluwer Accademic/Pleum Publisher, 2nd edition, 1999.

[Lan90] K. Lange. Convergence of em image reconstruction algorithms with gibbs smoothing. *IEEE Trans. Med. Imag.*, 9(4):439446, 1990.

[Lan02] L. Landmann. Deconvolution improves colocalization analysis of multiple fluorochromes in 3d confocal data sets more than filtering techniques. *Journal of Microscopy*, 208(2):134–147, 2002.

[LB91] R.L. Lagendijk and J. Biemond. *Iterative Identification and Restoration of Images.* Kluwer Academic Publishers, Dordrecht, 1991.

[LB02] J. Ling and A. C. Bovik. Smoothing low-snr molecular images via anisotropic median-diffusion. *IEEE Trans. Med. Imaging*, 21(4):377–384, 2002.

[Li95] S.Z. Li. *Markov random field modeling in computer vision.* Springer-Verlag, London, UK, 1995.

[LRA02] H. Lantéri, M. Roche, and C. Aime. Penalized maximum likelihood image restoration with positivity constraints: multiplicative algorithms. *Inverse Problems*, 18:1397–1419, 2002.

[LRCA01] H. Lantéri, M. Roche, O. Cuevas, and C. Aime. A general method to devise maximum-likelihood signal restoration multiplicative algorithms with non-negativity constraints. *Signal Processing*, 81:945–974, 2001.

[LRGA02] H. Lantéri, M. Roche, P. Gaucherel, and C. Aime. Ringing reduction in image restoration algorithms using a constraint on the inferior bound of the solution. *Signal Processing*, 82:1481–1504, 2002.

[Luc74] L.B. Lucy. An iterative technique for the rectification of observed distributions. *Astron. J*, 79:745–754, 1974.

[MC97] J. Markham and J.A. Conchello. Tradeoffs in regularized maximum-likelihood image restoration. In C. J. Cogswell, G. S. Kino, and T. Wilson, editors, *Three-Dimensional Microscopy: Image Acquisition and Processing III*, volume 2984, pages 136–145. SPIE, 1997.

[MC99] J. Markham and J. A. Conchello. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *J Opt Soc Am A Opt Image Sci Vis*, 16(10):2377–2391, 1999.

[Min61] M. Minsky. Us patent 3013467, 1961.

[Min88] M. Minsky. Memoir on inventing the confocal scanning microscope. *Scanning*, 10:128–139, 1988.

[MS87] H. N. Multhei and B. Schorr. On an iterative method for a class of integral equations of the first kind. *Math. Methods Appl. Sci.*, 9:137–168, 1987.

[MS89] H. N. Multhei and B. Schorr. On properties of the iterative maximum likelihood reconstruction method. *Math. Methods Appl. Sci.*, 11:331–342, 1989.

[Mul93] H. N. Multhei. Iterative continuous maximum likelihood reconstruction methods. *Math. Methods Appl. Sci.*, 15:275–286, 1993.

[MVAD08] P. P. Mondal, G. Vicidomini, and " Applied Physics Letters 2008 (in press). A. Diaspro, ˙Image reconstruction for multi-photon fluorescence microscopy. *Applied Physics Letters*, In press 2008.

[MVD07] P. P. Mondal, G. Vicidomini, and A. Diaspro. Markov random field aided bayesian approach for image reconstruction in confocal microscopy. *Journal of Applied Physics*, 102(4):044701, 2007.

[MZ06] B.A. Mair and J. Zahnen. A generalization of green's one-step-late algorithm for penalized ml reconstruction of pet images. *Nuclear Science Symposium Conference Record, 2006. IEEE*, 5:2775–2777, 2006.

[NH01] M. Nagorni and S. W. Hell. Coherent use of opposing lenses for axial resolution increase in fluorescence microscopy. i. comparative study of concepts. *J. Opt. Soc. Am. A*, 18(1):36–48, 2001.

[Paw06] J. B. Pawley. *Handobook of biological confocal microscopy.* Springer, New York, 3th edition, 2006.

[Pie87] A. R. De Pierro. On the convergence of the iterative space reconstruction algorithm for volume ect. *IEEE Trans. Med. Imaging*, 6:124–125, 1987.

[PTV92] W. H. Press, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C.* Cambridge University Press, Cambridge, 2nd edition edition, 1992.

[PY01] A. R. De Pierro and M. E. B. Yamagishi. Fast em-like methods for maximum a posteriori estimates in emission tomography. *IEEE Trans. Med. Imag.*, 20(4):280–288, 2001.

[Ric72] W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62(1):55, 1972.

[RPL05] F. Rooms, W. Philips, and D. S. Lidke. Simultaneous degradation estimation and restoration of confocal images and performance evaluation by colocalization analysis. *J Microsc*, 218(1):22–36, 2005.

[RW59] B. Richards and E. Wolf. Electromagnetic diffraction in optical systems. ii. structure of the image field in an aplanatic system. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 253(1274):358–379, 1959.

[SBHP+05] E. Spiess, F. Bestvater, A. Heckel-Pompey, K. Toth, M. Hacker, G. Stobrawa, T. Feurer, C. Wotzlaw, U. Berchner-Pfannschmidt, T. Porwol, and H. Acker. Two-photon excitation and emission spectra of the green fluorescent protein variants ecfp, egfp and eyfp. *Journal of Microscopy*, 217(3):200–204, 2005.

[SBW07] M. Schwertner, M. J. Booth, and T. Wilson. Specimen-induced distortions in light microscopy. *J Microsc*, 228(Pt 1):97–102, 2007.

[Sch95] R.L. Schultz, R.R.; Stevenson. Stochastic modeling and estimation of multispectral image data. *Image Processing, IEEE Transactions on*, 4(8):1109–1119, 1995.

[SHW93] D.L. Snyder, A.M. Hammound, and R.L. White. Image recovery from data acquired with a charge-coupled-device camera. *J. Opt. Soc. Am. A*, 10:1014–1023, 1993.

[SK78] C. J. R. Sheppard and R. Kompfner. Resonant scanning optical microscope. *Applied Optics*, 17:2879–2883, 1978.

[Sta86] J.J. Stamnes. *Waves in Focal Regions.* Adam Hilger, Bristol, 1986.

[SV82] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Trans. Med. Imaging*, 1:113–122, 1982.

[TA77] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill-posed problems*. New York: Wiley, 1977.

[TFUM06] M. Von Tiedemann, A. Fridberger, M. Ulfendahl, and J. Boutet De Monvel. Image adaptive point-spread function estimation and deconvolution for in vivo confocal microscopy. *Microscopy Research and Technique*, 69(1):10–20, 2006.

[TL89] T.Hebert and R. Lehay. A generalized em algorithm for 3-d bayesian reconstruction from poisson data using gibbs priors. *IEEE Transaction on Medical Imaging*, 8(2):194–202, 1989.

[TV97] P. Török and P. Varga. Electromagnetic diffraction of light focused through a stratified medium. *Applied Optics*, 36(11):2305–2312, 1997.

[VBDB08] G. Vicidomini, P. Boccacci, A. Diaspro, and M. Bertero. A novel iterative method for edge-preserving image deconvolution in confocal microscopy. *submitted to Journal of Microscopy*, 2008.

[VGJ99] P. J Verveer, M. J Gemkow, and T. M Jovin. A comparison of image restoration approaches applied to three-dimensional confocal and wide-field fluorescence microscopy. *J Microsc*, 193(1):50–61, Jan 1999.

[VJ97] P.J. Verveer and T.M. Jovin. Acceleration of the ictm image restoration algorithm. *Journal of Microscopy*, 188:191195., 1997.

[VMD06] G. Vicidomini, P.P. Mondal, and A. Diaspro. Fuzzy logic and maximum a posteriori-based image restoration for confocal microscopy. *Optics Letters*, 31(24):3582–3584, 2006.

[VMD07] G. Vicidomini, P. P. Mondal, and A. Diaspro. Soft computing approach to confocal and two-photon excitation microscopy. In Jose-Angel Conchello, Carol J. Cogswell, and Tony Wilson, editors, *Three-Dimensional and Multidimensional Microscopy: Image Acquisition and Processing XIV*, volume 6443, page 644319. SPIE, 2007.

[VNdW+03] D. Van De Ville, M. Nachtegael, D. Van der Weken, E.E. Kerre, W. Philips, and I. Lemahieu. Noise reduction by fuzzy image filtering. *Fuzzy Systems, IEEE Transactions on*, 11(4):429–436, 2003.

[VSB+07] G. Vicidomini, M. Schneider, P. Bianchini, S. Krol, T. Szellas, and A. Diaspro. Characterization of uniform ultrathin layer for z-response measurements in three-dimensional section fluorescence microscopy. *J Microsc*, 225(1):88–95, 2007.

[VSK85] Y. Vardi, L. A. Shepp, and L. Kaufman. A statistical model for positron emission tomography, , 80, 8-20. *J. Am. Stat. Association*, 80:8–20, 1985.

[VZB+07] G. Vicidomini, J.M. Zwier, P. Bianchini, F. Cella, E. Ronzitti, S. Krol, T. Szellas, G.F. Brakenhoff, and A. Diaspro. Sipcharts using uniform ultra-thin and thin layers for z-response measurements in two-photon excitation fluorescence microscopy. In Ammasi Periasamy and Peter T.C. So, editors, *Multiphoton Microscopy in the Biomedical Sciences VII*, volume 6442, page 644224. SPIE, 2007.

[WJW90] W.Denk, J.Strickler, and W.Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248:1990, 1990.

[You89] I.T. Young. Image fidelity: Characterizing the imaging transfer function. In D. L. Taylor, editor, *Methods in Cell Biology*, pages 1–45. Academic Press, San Diego, CA, 1989.