Fast and Automated Wavelet-Regularized Image Restoration in Fluorescence Microscopy

Cédric Vonesch

Thèse N° 4306 (Mai 2009)

Thèse présentée à la faculté des sciences et techniques de l'ingénieur
pour l'obtention du grade de docteur ès sciences
et acceptée sur proposition du jury

Prof. Martin Hasler, *président*
Prof. Michael Unser, *directeur de thèse*
Dr. Rainer Heintzmann, *rapporteur*
Prof. Olivier Rioul, *rapporteur*
Prof. Martin Vetterli, *rapporteur*

École polytechnique fédérale de Lausanne—2009

Abstract

The subject of this thesis is image restoration, that is, deconvolution and denoising. Our work is motivated by applications in the rapidly expanding field of biological imaging and in particular fluorescence microscopy. Inverse problems in this area typically involve regularly structured but high-dimensional data sets, thus requiring a specific tradeoff between restoration quality, computational complexity and user interaction. Our interest lies in novel algorithmic solutions for addressing these challenges. The common point of our developments is that they rely centrally on wavelets and multilevel concepts.

The thesis is organized in three main parts. In the first part, we provide a general overview of fluorescence microscopy in the context of biological research. We then review the current state of the art in deconvolution microscopy, within a unifying variational framework.

The second part is concerned with deconvolution using a non-quadratic (typically ℓ_1) wavelet-domain regularization. In the context of deconvolution microscopy, this approach is novel. It amounts to imposing a sparsity constraint on the object to be recovered, a concept that has recently attracted considerable interest in signal and image processing. This type of regularization was shown to belong to the most effective restoration methods, but it involves a computationally challenging optimization problem. We propose a multilevel version of the standard "thresholded Landweber" algorithm that yields a substantial acceleration. This allows us to apply wavelet regularization to large-scale deconvolution problems encountered in deconvolution microscopy.

In the third part, we present some applications of risk estimation techniques for automating the process of image restoration. We first propose a denoising method that is specifically designed for very low light intensities. The algorithm is designed from the beginning for complete self-adjustment of its parameters and for low computational complexity. It also performs competitively with most existing approaches for Poisson intensity estimation. Our second contribution is a general

method for estimating the signal-to-noise ratio during the execution of iterative image-restoration procedures; it is applicable to most algorithms emanating from variational formulations, in particular those involving wavelet regularization.

Résumé

Le sujet de cette thèse est la restauration d'images, c'est à dire la déconvolution et le débruitage. Les applications en microscopie par fluorescence pour l'imagerie biologique, qui connaît actuellement un développement considérable, constituent l'une des principales motivations de notre travail. Les problèmes inverses dans ce domaine ont typiquement une structure régulière mais une dimension algébrique très grande, ce qui nécessite des compromis spécifiques entre qualité de restauration, complexité calculatoire et intervention de l'utilisateur. Nous nous intéressons donc à de nouvelles solutions algorithmiques permettant de répondre à ces contraintes. Les fils conducteurs de nos développements sont les ondelettes et des concepts dits "multi-niveaux" ou "multi-grilles".

Cette thèse est organisée en trois grandes parties. Dans la première, nous donnons une vue d'ensemble de la microscopie par fluorescence dans le contexte de la recherche en biologie. Nous faisons ensuite l'état de l'art en microscopie par déconvolution, dans un cadre variationnel unificateur.

La seconde partie concerne la déconvolution basée sur une régularisation par ondelettes non-quadratique (typiquement une norme ℓ_1). Dans le cadre de la microscopie par déconvolution, cette approche est nouvelle. Elle revient à appliquer une contrainte de "parcimonie" à l'objet que l'on cherche à reconstituer. Ce concept a récemment attiré un intérêt considérable en traitement du signal et des images. Il a été montré que ce type de régularisation fait partie des méthodes de restauration les plus performantes, mais il implique un problème d'optimisation relativement difficile. Nous proposons une version multi-niveaux nettement plus rapide d'un algorithme devenu standard, le "Landweber seuillé". Ceci nous permet d'appliquer la régularisation par ondelettes à des problèmes à grande dimension rencontrés en microscopie par déconvolution.

Dans la troisième partie, nous présentons des applications de techniques d'estimation de risque à l'automatisation du processus de restauration d'images. Nous proposons premièrement une méthode de débruitage conçue spécifiquement pour

les très faibles intensités de lumière. L'algorithme est conçu dès le départ de sorte qu'il puisse auto-ajuster entièrement ses paramètres. Notre seconde contribution est une méthode générale pour estimer le rapport signal à bruit durant l'exécution de procédures de restauration d'image itératives ; elle est applicable à la plupart des algorithmes émanant d'une formulation variationnelle, en particulier ceux utilisant une régularisation par ondelettes.

Mots clés : problèmes inverses, variationnel, reconstruction, restauration, déconvolution, débruitage, non-linéaire, parcimonie, régularisation ℓ_1, ondelettes, multi-résolution, multi-échelle, multi-niveaux, multi-grille, rapide, convergence, accélération, préconditionnement, optimisation de majorant, majoration-minimisation, optimisation par substitution, 3D, microscope, champs large, confocale, fluorescence, estimation de risque, estimateur de risque non-biaisé de Stein, combinaison linéaire de seuils.

Acknowledgements

of my thesis. Special thanks also to my dear friend, I. Khalidov.

A ma famille.

Contents

Chapter 1

Motivation and overview

1.1 Image restoration in fluorescence microscopy

The widespread availability of fluorescent protein markers and digital microscopy in the 1990's has revolutionized biological research. Today, fluorescence microscopy is the primary modality for biological imaging, and experimental requirements—such as high-resolution, small-animal or live-specimen imaging—are continuously stimulating new developments.

Specifically, three disciplines can be identified as playing a key role for modern microscopy, namely, biochemistry, optics and digital image processing. Of fundamental importance is the engineering of *novel fluorescent probes* with improved characteristics. This topic is intimately connected with the design of *advanced optical-microscopy techniques*; indeed, the latter increasingly depend on non-linear characteristics of fluorescence, such as stimulated emission depletion or photoactivation [104, 101, 20]. This evolution calls for *sophisticated computational methods* that are able to efficiently synthesize the images of interest from the measured data.

This thesis is primarily concerned with *image restoration*, that is, *deconvolution* and (to a lesser extent) *denoising*. Deconvolution represents the archetype of an ill-posed inverse problem and naturally arises in the context of structured imaging schemes; this includes some of the most recent types of microscopes (4Pi, STED...) as well as conventional widefield and confocal systems. Denoising differs from deconvolution in that it is a better-conditioned problem.

Image restoration can be distinguished from *image reconstruction* as well as from *image enhancement*. From our standpoint, image enhancement refers to an

operation that improves the subjective quality of an image, possibly to facilitate its interpretation. Image enhancement is a generic procedure in the sense that it relies on image-processing primitives—such as morphological processing or feature enhancement—that do not take into account the physical acquisition process. This is in contrast with image restoration, which aims at a quantitative improvement of the image quality based on physical insights. An extreme situation is when the interpretation of the data is *only* possible after a nontrivial numerical transformation; one then deals with image reconstruction. This is typically a less-structured inverse problem where the measurement domain can differ significantly from the image domain. These considerations are summarized in Table 1.1.

For a given light budget, the benefits of image restoration are typically a higher signal-to-noise ratio and an improved contrast. In the context of bioimaging, these advantages can be exploited in a variety of ways. On the post-acquisition side, image restoration facilitates both the visual and the computational analysis of the data; in fact, it can be considered as indispensable within a quantitative imaging approach (e.g., for colocalization studies). Image restoration can also help to optimize the acquisition process itself: indeed, for a target image quality, it typically allows for weaker excitation light and/or shorter exposure times. These are critical parameters when imaging live specimens.

1.2   Algorithmic challenges

Image restoration is a challenging task because it depends on many parameters; ideally, it should be part of an integrated approach that takes into account the complete image-acquisition chain (see Fig. 1.1). The first task is to *obtain a forward*
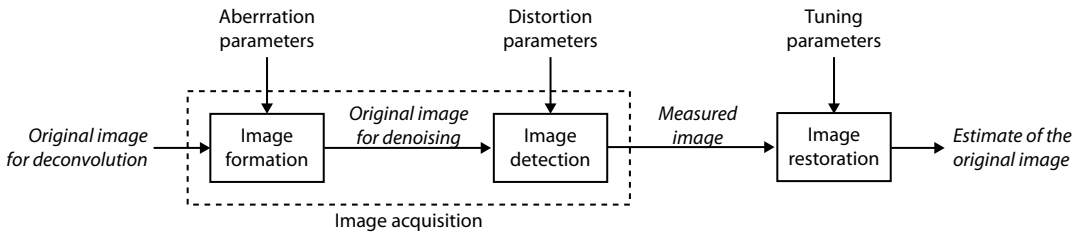
Figure 1.1: The image acquisition and restoration chain.

Image-processing task		Structure of the forward model	Nature of the inverse problem	Relation between original and measurement domain	Examples of relevant imaging modalities
Image reconstruction		Weak (e.g. non-cartesian, shift-variant)	Ill-conditioned	Different geometry and/or dimensionality	Structured illumination, PALM (PSF-based source localization), optical tomography
Image restoration	Deconvolution	Strong (shift-invariant)	Ill-conditioned	Qualitatively the same	Widefield, confocal, 4Pi, STED
	Denoising	Ideal (identity operator)	Well-conditioned	Exactly the same	
Image enhancement		Not taken into account	N/A	Assumed to be the same	Widefield, confocal, 4Pi, STED, PALM (Gaussian fitting)

Table 1.1: Comparison of image restoration with related image-processing tasks.

model that accurately describes the physics of the image-acquisition process. Such a model plays an important role for optimizing the acquisition process (that is, for generating the best-possible data); it is also a fundamental component of almost any image-restoration procedure.

The forward model usually depends on various elements (e.g., point spread function, characteristic function of the detector) and parameters: in Fig. 1.1, we distinguish between distortion parameters (e.g., gain, quantization step) and aberration parameters (e.g., bandwidth), the latter being mostly relevant to deconvolution. The next step is thus to characterize the imaging system, that is, to *fit the model* to the actual experimental conditions. This may require separate calibration data (e.g., fluorescent-microbead or dark-count images); sometimes the model identification can also be based directly on the measured data (blind approach).

The task of image-restoration itself can be divided in two parts. First, one must *design a good restoration procedure*, that is, choose its degrees of freedom for optimal restoration quality and processing complexity. Approximation theory and estimation theory are important tools in this context. Second, one must provide a rationale for adjusting these degrees of freedom in practice.

In summary, an algorithm developer must find a compromise between three usability criteria:

- restoration quality;

- computational complexity[1];

- automation[2].

This compromise eventually determines the attractivity of the image-restoration procedure for the end-user.

1.3   Contributions and organization of the thesis

The thesis is organized as follows.

A general introduction to bioimaging and fluorescence microscopy is given in Part I. Chapter 2 represents our contribution to the promotion of life sciences as an exciting application field for image processing. It is followed by a mathematical description of the image-acquisition process and a review of existing deconvolution methods in Chapter 3.

---

[1]This includes both CPU and memory requirements.
[2]In the sense of minimizing the user input.

The core of the thesis (Part II) is concerned with non-linear wavelet-based deconvolution. This approach can be justified using approximation-theory arguments [40, 58], sparsity principles [205] or a more standard Bayesian formalism [78]; it is also known to yield very good results for 2D image deconvolution [78]. Our main contribution is a multilevel version of the so-called thresholded Landweber algorithm (Chapter 6). Compared to the standard version, the multilevel version yields an acceleration of one order of magnitude in experimental situations, which allows us to address large deconvolution problems; in particular, we present one of the first applications of wavelet-regularized deconvolution to 3D fluorescence microscopy.

Our approach is strongly inspired from multigrid techniques; in fact, it represents one of the first successful applications of multigrid ideas to image deconvolution. The wavelet representation is an important component of our "divide-and-conquer" strategy, because it naturally leads to effective multilevel schemes. Indeed, its structure is inherently multiscale and invariant to dyadic shifts; when dealing with a convolution operator, this allows for inter-level transfer operators that are hardly more expensive than Mallat's fast wavelet-transform algorithm. In addition, it provides a decomposition of the spectral domain that is well adapted to the low-pass operators that are typically encountered in deconvolution problems. An important step towards this general observation is to consider the particular case of Shannon wavelets (Chapter 5).

The other main topic of our work (Part III) is the application of risk-estimation techniques to image restoration. We first propose a restoration method that is specifically designed for very low light levels (Chapter 8). In this situation the effect of shot noise predominates over the convolutive nature of the imaging system. Here too, we build upon previous variational and Bayesian approaches that have established wavelet-domain thresholding as one of the most effective denoising methods. We also borrow ideas from a recent class of data-driven restoration methods [140, 24] where the image priors (such as interscale dependencies) are directly built into the restoration algorithm and formulated so as to minimize its computational complexity. The result is a completely automatic denoising procedure that provides state-of-the-art results, while its computational cost is essentially equivalent to a non-redundant Haar wavelet transform. Again, this represents an attractive solution for processing high-dimensional data sets typically produced in fluorescence microscopy.

In Chapter 9, we also present a recursive risk-estimation approach for iterative restoration algorithms. The method is generic and applicable to ill-conditioned deconvolution problems. It essentially allows for monitoring the (estimated) signal-to-noise ratio during the execution of any iterative algorithm.

At the beginning of Part II and Part III, we provide introductory chapters that

aim at a higher-level discussion of our work.

Part I

An introduction to fluorescence microscopy and deconvolution

Tell me, and I will forget.
Show me, and I will remember.
Involve me, and I will understand.

Chapter 2

The colored revolution of bioimaging
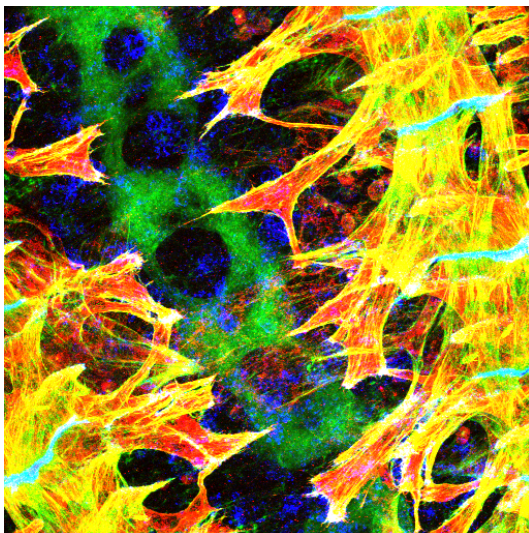
With the recent development of fluorescent probes and new high-resolution microscopes, biological imaging has entered a new era and is presently having a profound impact on the way research is being conducted in the life sciences. Biologists have come to depend more and more on imaging. They can now visualize sub-cellular components and processes in vivo, both structurally and functionally. Observations can be made in two or three dimensions, at different wavelengths (spectroscopy), possibly with time-lapse imaging to investigate cellular dynamics.

The observation of many biological processes relies on the ability to identify and locate specific proteins within their cellular environment. Cells are mostly transparent in their natural state, and the immense number of molecules that constitute them are optically indistinguishable from one another. This makes the identification of a particular protein a very complex task — akin to finding a needle in a haystack. However, if a bright marker were attached to the protein of interest, it could very precisely indicate its position. Much effort has gone into finding suitable markers for this purpose, but it is only over the course of the past decade, with the advent of fluorescent proteins, that this concept has been revolutionized. These biological markers have the crucial properties necessary for dynamic observations of living cells: they are essentially harmless to the organism and can be attached to other proteins.

Fluorescence microscopy was invented almost a century ago, when microscopists were experimenting with ultraviolet light to achieve higher resolutions. In the very beginning, observations were limited to specimens that naturally fluoresce[1]. Rapidly, fluorescent dyes for staining tissues and cells were investigated. But it was not until the 1940s that fluorescence microscopy became popular, when A. Coons and M. Kaplan introduced a technique to label antibodies with a fluorescent dye to study antibody-antigen interactions, which profoundly changed the field of immuno-histochemistry [55]. The discovery that really brought fluorescence microscopy to the forefront came in 1994, when M. Chalfie et al. succeeded in expressing a naturally fluorescent protein, the now famous green fluorescent protein (GFP), in living organisms [39]. This was a landmark evolution in the field[2], fostering a whole new class of tagging methods.

While genetic engineering is at the origin of this new methodology, a number of innovations from the fields of physics, optics, mechanical and electrical engineering have been combined to provide the necessary instrumentation. Impressive enhancements in classical microscopy have been achieved, and new imaging systems are actively being developed. A key element for the evolution of microscopy in gen-

---

[1]This property is called autofluorescence or primary fluorescence.
[2]The 2008 Nobel Prize in Chemistry went to O. Shimomura, M. Chalfie and R. Y. Tsien, "for the discovery and development of the green fluorescent protein, GFP".

eral was the shift to digital imaging in the 1990s, with the availability of affordable high-sensitivity acquisition devices and powerful computer hardware.

The capabilities of today's systems often lead to enormous data sets that, in most cases, require post-processing for their interpretation. Signal processing methods for biological research are only at their prelude; the needs are considerable and most probably not even clearly formulated yet. It is thus predictable that signal processing will be one of the main challenges of fluorescence microscopy in the forthcoming years.

The goal of this chapter is to provide an overview of the main aspects of modern fluorescence microscopy. We first cover the principles of fluorescence and highlight the key discoveries in the history of fluorescence microscopy. In subsequent sections, we present the optics of fluorescence microscopes and examine various types of detectors. Finally, we discuss the signal and image processing challenges in fluorescence microscopy and highlight some of the present developments and future trends in the field.

2.1 Fluorescence in molecular and cellular biology

2.1.1 The physical principles of fluorescence

Definition

Fluorescence is a phenomenon by which a molecule, upon illumination at a specific wavelength, reemits light at another (typically longer) wavelength. A molecule that has the ability to fluoresce is called a *fluorophore* or *fluorochrome*[3]. It has distinctive *excitation* and *emission* spectra (see Fig. 2.1), although in practice, it is often characterized by the two wavelengths corresponding to the respective peak intensities of these spectra.

A molecule can exist in a variety of energetic states, which, for the most part, are determined by the configuration of its electrons and the vibrational agitation of its atomic nuclei. If a photon with sufficient energy is absorbed by a fluorophore, the latter moves from its ground state to an excited electronic state (see Fig. 2.2(a)). Fluorescence occurs when the excited molecule returns to the ground state by releasing energy through emission of a photon. Because some of the energy gained during excitation is converted to heat, the emitted photon has a lower energy than the absorbed one. This explains the difference in wavelength mentioned earlier

---

[3]Specifically, the former describes an atomic compound responsible for fluorescence, while the latter is a more general term for a dye that renders a body fluorescent.
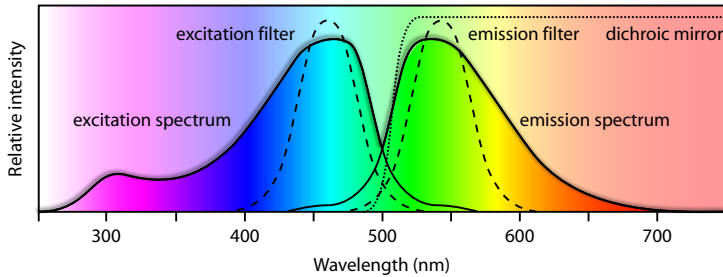
Figure 2.1: Representation of typical excitation/emission spectra of a fluorophore (in relative intensities). The excitation spectrum shows the emission intensity as a function of excitation wavelength, and the emission spectrum shows the relative emission intensities as a function of emission wavelengths for an excitation at the peak absorption wavelength. Explanations on the filters are given in Section 2.2.

(since $E = h\nu = hc/\lambda$), which is also known as the Stokes shift. Fluorophores whose spectra present a large Stokes shift are usually preferred since their emitted light can be separated from the excitation light more easily by the means of optical filters (see Fig. 2.1).

Related phenomena

When in an excited state, a fluorophore can be forced to the ground state in a process called *stimulated emission*: in the presence of an incident photon, the molecule emits a new photon that has the same wavelength, direction, polarization and phase. This phenomenon is relevant to some of the concepts discussed later, but is best known as the light-amplification principle behind lasers.

Another important concept is that of *multiphoton excitation*. A fluorophore can also be excited by the simultaneous absorption of two or more photons, given that the combined energy of the photons corresponds to the energy required for single-photon excitation (see Fig. 2.2(b)). In this particular situation the excitation wavelength is longer — in the case of two-photon excitation, twice as long as the single-photon excitation wavelength.
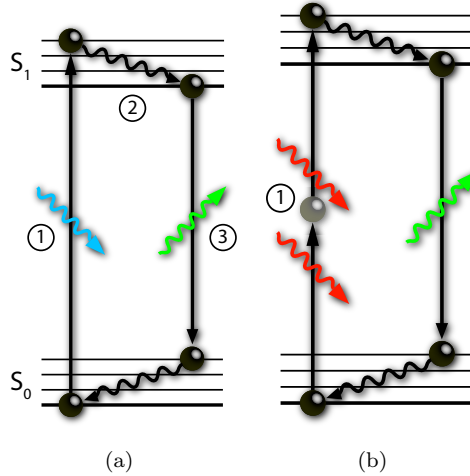
Figure 2.2: *Jablonski diagrams* representing the energy-level transitions involved in the fluorescence of GFP. Thick lines represent electronic energy levels; thin ones are associated vibrational energy levels. (a) Upon absorption of a photon at a specific wavelength (blue), the molecule moves from the ground state S_0 to the excited state S_1 (1). Vibrational energies are immediately converted into heat in a process called vibrational relaxation (2). When the molecule returns to the ground state, the remaining energy is released via emission of a new photon at a longer wavelength (green). (b) In the case of two-photon excitation, the excitation wavelength (red) is longer than the emission wavelength (green). The intermediate virtual state is indicated by (1).

2.1.2   The green revolution

The developments that had the biggest impact on biological research and made fluorescence microscopy ubiquitous took place during the past two decades. Consequently, we shall focus on this period for the remainder of this section. However, these recent developments could not have occurred without previous discoveries and inventions in a variety of fields, starting in the 16th century. An overview of events that played an essential role in contributing to the development of modern fluorescence microscopy is given in the form of a timeline in Figures 2.3 and 2.4.

1565	Nicolás Monardes observes luminescence in the extract of the wood *lignum nephriticum*, a substance used for treating kidney ailments.
1646	Athanasius Kircher relates that an infusion of *lignum nephriticum* reflects blue and transmits yellow light.
1800	Frederick William Herschel discovers infrared radiation.
1801	Johann Wilhelm Ritter discovers the ultraviolet region of the spectrum.
1833	David Brewster observes red radiation in a solution containing chlorophyll upon illuminating it with white light.
1845	John F. W. Herschel (the son of F. W. Herschel) discovers the phenomenon of fluorescence in a quinine solution.
1852	George Gabriel Stokes describes the luminescence observed in the fluorspar mineral as fluorescence, and formulates the Stokes law.
1871	Adolf von Baeyer discovers and synthesizes *fluorescein*, still widely used as a dye.
1878	Ernst Abbe formulates the theory that links resolution to the wavelength of light, and proposes the use of ultraviolet light to increase the resolution of microscopes.
1904	August Köhler and Moritz von Rohr develop the ultraviolet microscope. Autofluorescence observed in the visible domain starts the era of fluorescence microscopy.
1905	Albert Einstein describes the photoelectric effect.
1908	August Köhler and Henry Friedrich Wilhelm Siedentopf build and demonstrate the first fluorescence microscope.
1911	Max Haitinger coins the term *fluorochrome* to describe dyes that render non-fluorescent objects fluorescent. Numerous investigations into fluorescent dyes begin.
1911 1913	The first commercialized fluorescence microscopes, using carbon arc lamps, are produced by Reichert (Otto Heimstädt and Carl F. W. Reichert, 1911) and Carl Zeiss (Heinrich Lehmann, 1913). Observations are limited to specimens that present autofluorescence.
1929	Philipp Ellinger and August Hirt propose a fluorescence microscope with epi-illumination for the observation of living organisms. They use the fluorochromes *fryptaflavine*, which stains cell nuclei, and *fluorescein*, to study kidney function.

Figure 2.3: Early history of fluorescence microscopy

1935	Alexander Jablonski presents his model to explain fluorescence.
1935	The first photomultiplier tube is built by Harley E. Iams and Bernard Salzberg [3].
1941	Albert Coons and Melvin Kaplan introduce immunofluorescence, a labeling method still in use today [1].
1948	Theodor Förster describes resonance energy transfer.
1957	Marvin Minsky submits his patent for the confocal microscope [4].
1958	Arthur Schawlow and Charles Townes invent the laser.
1960	The first functioning laser is built by Theodore Maiman [5].
1969	George Smith and Willard Boyle invent the CCD.
1978	Colin J. R. Sheppard and Rudolf K. Kompfner propose the concept of two-photon microscopy [6].
1980	John G. J. Bauman et al. introduce fluorescence *in situ* hybridization (FISH), a method for localizing specific genetic sequences by labeling RNA molecules with a fluorochrome [7].
1983	David A. Agard and John W. Sedat reconstruct chromosomes from *Drosophila* larvae using computer deconvolution [8].

The Colored Revolution

1990s	Affordable digital imaging hardware and commercial confocal fluorescence microscopes become widely available, launching a new era in microscopy.
1990	Winfried Denk, James H. Strickler, and Watt W. Webb experimentally demonstrate two-photon microscopy [9].
1991	Joseph R. Lakowicz et al. describe fluorescence lifetime imaging (FLIM) [10].
1994	Martin Chalfie et al., and Satoshi Inouye and Frederick Tsuji express GFP in living organisms (*e. coli* and *c. elegans*) [2] [11].
1994	Stefan Hell and Jan Wichmann invent STED microscopy [12].
2006	Localization-based fluorescence techniques devised by Eric Betzig et al. (PALM [13]) and Michael Rust et al. (STORM [14]) yield super-resolved images that rival the resolution of electron microscopy.

0   500   1000   1500   2000   2500

ISI Web of Science citations for *Fluorescence Microscopy* from 1985-2006

Figure 2.4: Modern history of fluorescence microscopy

In the early 1990s, fluorescent labeling techniques such as immunofluorescence[4] and covalent marking[5] were already widely in use for imaging. However, a straightforward means for selectively labeling a given protein with a non-perturbing fluorescent marker was not yet available. Only such a tag would make the *in vivo* observation of interactions between a specific protein with other proteins and the environment feasible.

The breakthrough came in 1994, when Chalfie et al. [39] succeeded in expressing a fluorescent protein that naturally occurs in a jellyfish species in other organisms by modifying their genome to code for this protein. At the origin of this innovation, accordingly dubbed "the green revolution" [208], was the discovery of GFP by Shimomura et al. in 1961 [199]. During their studies of the jellyfish *aequorea victoria*, whose fluorescing nature was described for the first time in 1955, they discovered that the source of the fluorescence was a naturally produced protein. Its chemical structure was reported by Shimomura in 1979, and in 1992, Prasher et al. cloned and determined its genetic sequence [177], paving the way for the work of Chalfie et al.

Since the first experiments with GFP, many variants have been engineered and discovered. From the naturally occurring GFP, called wtGFP for wild-type GFP, and from similar fluorescent proteins occurring in other marine organisms, new, more powerful mutants have been derived. Their properties range from different excitation and emission spectra[6] to stronger fluorescence and higher resistance to photobleaching [258]. Two widespread examples are cyan fluorescent protein (CFP) and yellow fluorescent protein (YFP), named for their characteristic emission spectra.

Biologists can label virtually any desired protein with a fluorescent protein by means of straightforward procedures. The first step leading to the creation of a labeled protein is to append the marker protein's sequence to that of the target. The resulting sequence is then introduced into cells, where its transcription results in the synthesis of a *fusion protein*. A common means for doing this is by placing the gene onto a plasmid[7], which can then be taken up by a cell. Such plasmids

---

[4]A technique (also called immunostaining) for detecting an antigen (protein) with a fluorochrome-labeled antibody.

[5]Proteins are purified, covalently labeled with a fluorescent molecule, and then introduced into cells.

[6]The currently available fluorescent protein tags offer a wide choice of wavelengths within the visible spectrum.

[7]Plasmids are small, circular, double-stranded sequences of DNA that naturally occur in bacteria and are part of their genome. They can easily be introduced into cells, where they are expressed in the same fashion as chromosomal DNA. Plasmids are not replicated upon cellular division; however, in some cases they are integrated into the cell's chromosomal DNA.
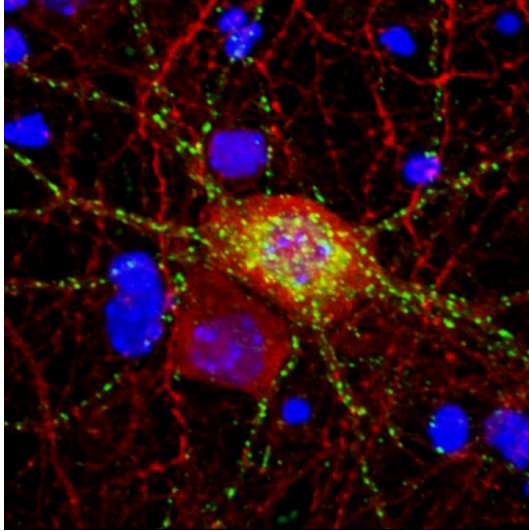
Figure 2.5: Image of a neuron where specific receptor proteins (delta opioid) have been fused with eGFP and appear in green. The red dots (GABA immunostaining) correspond to intracellular proteins located inside the neuron and its extensions. The nuclei of surrounding cells are stained in blue with DAPI, a fluorochrome that specifically binds to DNA. Courtesy of G. Scherrer, P. Tryoen-Toth and B. L. Kieffer, IGBMC, Illkirch, France.

exist for a wide range of fluorescent proteins and are available from specialized companies.

The fusion protein (Fig. 2.5) is expressed throughout the lifetime of the cell, as long as its sequence is present in the cell's nucleus. Note that this procedure typically results in the expression of both the fusion and natural versions of the protein, since the genetic sequence of the former does not replace that of the latter. Although the function and localization of the two variants are in most cases identical, it is necessary to verify that the label has no influence on cellular functions.

The availability of fluorescent protein tagging techniques led to a fundamental change in the way biological research is conducted and to an explosion of experimental possibilities. For further details, we refer to Section 2.5 and [134, 156].

2.2   Microscopes and image formation

We now turn our attention to the instrumentation for fluorescence imaging. This section provides a brief description of the two main types of image forming systems: widefield and confocal microscopes. While the former are usually less expensive than the latter (depending on the configuration), their optical resolution is intrinsically more limited, especially in the axial (i.e., z) direction. Both systems can yield volume images of the sample under inspection, possibly with the help of deconvolution. However, in widefield microscopes the volume is acquired plane by plane (as opposed to point by point in standard confocal systems), which allows for faster acquisitions.

2.2.1   The widefield microscope

Principle

Widefield microscopy is based on the paradigm of Köhler illumination, according to which the sample is observed under a uniform light beam. Fig. 2.6(a) shows how this is obtained in a simplified epi-illumination microscope: the light source (an arc or filament lamp) is magnified by the collector lens and projected onto the iris diaphragm. This aperture is located in a conjugate plane of the objective's back focal plane. Therefore the latter acts as a condenser lens and the intensity from the iris is uniformly dispatched on the sample.

Let us now consider a single point of the sample. It will reemit light by reflection and possibly by fluorescence. If located in the focal plane, this will generate a beam of parallel light rays through the microscope tube. The image is formed by integrating the effect of all secondary point sources within the specimen; it can be observed through the eyepiece, or recorded by placing a CCD sensor in the image plane.

One of the critical parameters in this setting is the *numerical aperture* (NA); that is, the angular opening of the light cone emerging from the object and collected by the objective. The magnification effect results from the combination of the objective, tube and ocular lenses.

Components for fluorescence imaging

Fluorescence imaging requires specific additional components for controlling the spectrum of the light (see also Figure 2.1). While usual lamps produce "white light" (covering the whole visible spectra, with some peaks at characteristic wavelengths), the fluorescent sample has to be illuminated with a specific excitation wavelength.
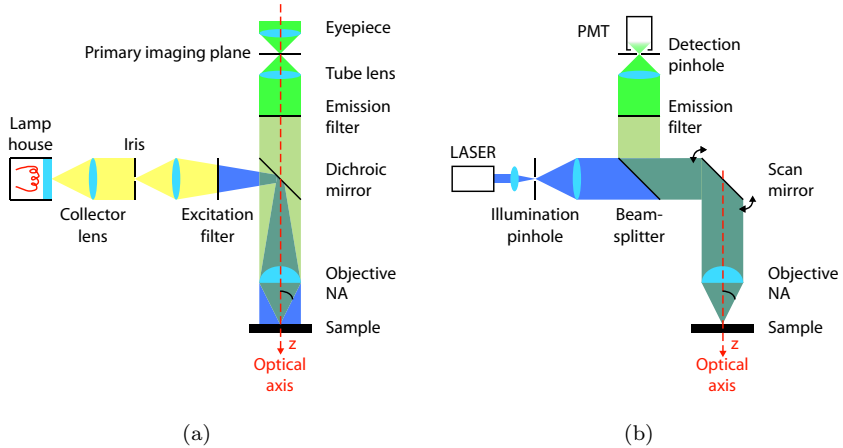
Figure 2.6: Schematics of widefield (a) and confocal (b) fluorescence microscopes, showing their main components. The illumination path is shown in yellow and/or blue (excitation at 395 nm) and the image-forming path in green (emission at 509 nm), to suggest the spectral composition of the different light beams in the case of a GFP-tagged sample.

This is ensured by inserting an *excitation filter* on the illumination path. The *emission filter*, on the other hand, ensures that only the wavelength corresponding to fluorescence reemission gets transmitted to the sensor or to the eyepiece, whereas reflected light (at the excitation wavelength) is discarded. A *dichroic mirror* helps achieving this by reflecting light below a certain *transition wavelength* (which is chosen to be between the excitation and emission wavelengths of the fluorophore) and transmitting light above that wavelength.

Incoherent point spread function

Because of the random nature of photon reemission, fluorescence microscopy is an incoherent imaging process. This means that each point of the sample contributes independently (without interference) to the light intensity distribution in the image space. Moreover, in the paraxial approximation, moving the object does not influence its image, except for a shift. From a signal-processing standpoint, a widefield microscope can thus be modeled as a linear space-invariant system in intensity. In
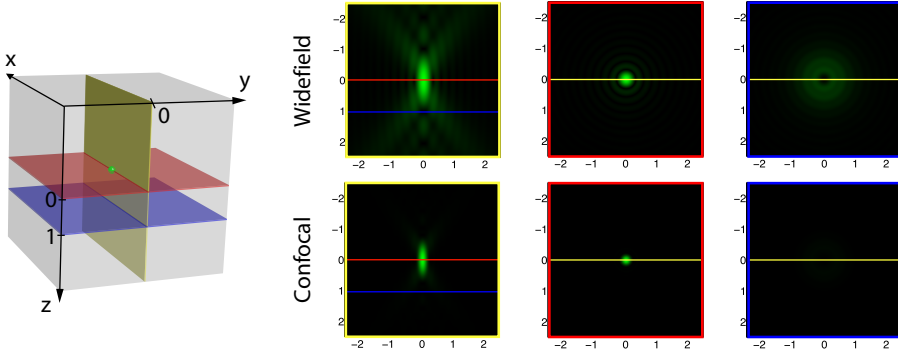
Figure 2.7: Cross-sections of the image of a fluorescent bead with a 1-NA objective (ideal 3D intensity PSF, or 3D Airy function). The upper row corresponds to a widefield microscope ($|h_{\lambda_{\mathrm{em}}}|^2$), while the lower one is for a confocal system ($|h_{\lambda_{\mathrm{ex}}}|^2|h_{\lambda_{\mathrm{em}}}|^2$). Here $\lambda_{\mathrm{ex}} = 395$ nm and $\lambda_{\mathrm{em}} = 509$ nm represent the respective excitation and emission wavelengths of GFP. Left sides: x-z section; right sides: x-y sections at two different depths. All units are in μm and refer to the object space.

other words, the light intensity (which is the physical value measured by a photodetector) in the neighborhood of the primary imaging plane ($z = 0$) is given by a convolutive expression:

$$I(x,y,z) \propto \int_{\mathbb{R}^3} \left| h_{\lambda_{\mathrm{em}}} \left(\frac{x}{M} - u, \frac{y}{M} - v, \frac{z}{M^2} - w \right) \right|^2 \chi(u,v,w) \, du \, dv \, dw \qquad (2.1)$$

where M is the magnification of the objective (notice that the axial magnification is M^2). Here, χ is the characteristic function of the object; it describes its ability to convert incident light into fluorescence intensity at the emission wavelength λ_{em} and is thus mostly related to the fluorophore concentration. The impulse response $|h_{\lambda_{\mathrm{em}}}|^2$ is called the incoherent (or intensity) *point spread function* (PSF), since it defines the image of an ideal point object ($\chi(x,y,z) = \delta(x,y,z)$). For a given wavelength λ it is defined by a 2D Fourier transform:

$$h_\lambda(x,y,z) = \int_{\mathbb{R}^2} P(u,v) \exp\left(i2\pi z \frac{u^2 + v^2}{2\lambda f^2} \right) \exp\left(-i2\pi \frac{xu + yv}{\lambda f} \right) \, du \, dv. \qquad (2.2)$$

In this expression f is the focal length of the objective. P represents the *pupil function*, which is an indicator function that corresponds to the circular aperture of the objective. Its radius r is related to the focal length by $\mathrm{NA} \simeq r/f$. Notice the presence of the depth coordinate z in the phase factor — it accounts for the defocusing effect illustrated in Fig. 2.7.

2.2.2 The confocal scanning microscope

Principle

In a confocal microscope ([154], Fig. 2.6(b)) the illuminating point source is usually obtained from a laser. The latter illuminates a pinhole located in a plane conjugate to the sample. In this way, the light is focused onto a very small volume of the sample, and the returning fluorescence radiation is collected by a photomultiplier tube (PMT [164]). The essential difference with a widefield microscope is the detection pinhole, which drastically reduces the proportion of light coming from out-of-focus points, especially in the axial direction.

Since only one point is observed at a time, the object has to be *scanned*. In the x and y dimensions, this is achieved by using a scan mirror, which deflects the illumination beam, hence moving the illumination spot in the same plane. In the z direction, the sample is usually moved mechanically by the means of a motorized stage. The ability to resolve different planes within the object is called *optical sectioning* and leads to a complete volumetric representation (a *stack* of 2D images).

A critical parameter in this setting is the pinhole diameter, which is usually expressed in *Airy units* (AU)[8]. One Airy unit corresponds to the size of the central disc of the PSF of the system (Fig. 2.7, second image in the bottom row). The smaller the pinhole, the better the resolution; however, this also means that less light is collected, implying a higher noise level.

Incoherent point spread function

The imaging process of fluorescent material can be modeled as follows: first, we have to take into account the effect of illumination, which consists in multiplying the fluorescence strength of the object by the PSF of the objective (taking into account the scan coordinate (x_0, y_0, z_0)). The reemitted light intensity is then given by

$$|h_{\lambda_\mathrm{ex}}(x - x_0, y - y_0, z)|^2 \chi(x, y, z - z_0) \tag{2.3}$$

---

[8]After back-projection in the object space, i.e. dividing the effective diameter by the magnification factor.

where λ_{ex} denotes the excitation wavelength. The intensity at the detector is the convolution of this expression with the objective PSF, evaluated at the origin (the position of the detection pinhole, which is assumed to be ideal):

$$\int_{\mathbb{R}^3} |h_{\lambda_{\text{em}}}(x_0 - x, y_0 - y, -z)|^2 |h_{\lambda_{\text{ex}}}(x - x_0, y - y_0, z)|^2 \chi(x, y, z - z_0) \, dx \, dy \, dz \quad (2.4)$$

where λ_{em} denotes the reemission wavelength. Notice that we did not indicate the magnification factors here, which is equivalent to back-projecting the image into object space; also, the returning light beams are descanned when they hit back on the scanning mirror. Since $h_{\lambda_{\text{ex}}}$ and $h_{\lambda_{\text{em}}}$ are symmetric in x, y and z, the final intensity PSF of the system is $h_{\text{total}} = |h_{\lambda_{\text{ex}}} h_{\lambda_{\text{em}}}|^2$, illustrated in the lower part of Fig. 2.7. It shows that a confocal microscope has a PSF that is more concentrated in space than a widefield one; i.e., a better resolution, especially in the axial direction.

2.2.3   Sample setup and aberrations

In an ideal optical system, wavefronts propagate without undergoing phase distortions, also called aberrations. Modern microscope optics are highly sophisticated and are corrected to high levels of precision to avoid such distortions. The optical properties of the sample play an important role in the formation and correction of aberrations. Samples are placed onto a glass slide and need to be covered with a glass coverslip for use with most objectives. As shown in Fig. 2.8, there is an *immersion* layer between the objective and the sample (to increase resolution, an immersion medium with a high refractive index, such as oil, is used). To minimize aberrations, each objective is designed for a specific setup, corresponding to parameters such as the refractive index of the immersion medium, the coverslip thickness, and the imaging depth. Small deviations from these optimal values (e.g., due to temperature changes or incorrect sample preparation) can introduce aberrations. A common and often unavoidable source of aberrations is the imaging depth in situations where the refractive indices of the specimen and immersion layers are mismatched. In the case where this mismatch is significant, it may result in the PSF becoming non-stationary, especially along the axial direction z [89].

2.3   Detectors

Fluorescence imaging can sometimes be a real challenge due to very low light conditions. Especially for live samples undergoing fast biological changes, it may not
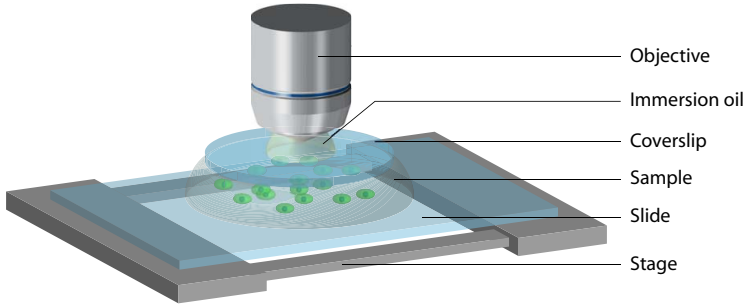
Figure 2.8: Schematic representation of a typical sample setup.

be possible to integrate more than a few tens of photons at each sampling position. Such conditions call for very sensitive detection devices with very accurate synchronization and control.

2.3.1 Characteristic parameters of detection systems

Most detectors are actually designed for specific applications and provide increased accuracy along the corresponding dimensions. In what follows, we briefly review the main characteristics of a detection device with respect to different parameters.

Wavelength

One of the critical parameters of a detector is its *quantum efficiency*; i.e., the average rate of incoming photons that are converted into an electronic charge. This rate strongly depends on the wavelength, and is therefore also called *spectral sensitivity*.

Intensity

Some detectors operate by internal amplification of the light signal they receive, which leads to the notion of *gain*. Detectors with gain adjustment offer increased *interscene dynamic range*, that is, the range of intensity levels that they can adapt to, for different imaging situations. The *intrascene dynamic range* characterizes the range of intensities to which the sensor can respond linearly, for a given imaging

situation. The maximum value, divided by the noise level, defines the peak *signal-to-noise ratio* (SNR). The *quantization precision* (number of bits per sample) must be chosen accordingly.

Spatial resolution

For array/line (resp. point) detectors, the *pixel size* (resp. detection aperture) represents a tradeoff between resolution and noise level. Another parameter is the effective *photosensitive area*, which may not cover the whole detector.

Temporal resolution

Long integration times will reduce noise but slow down the acquisition process. High *readout speeds* will allow faster frame rates/scan frequencies. For fluorescence lifetime measurements, one needs to precisely know when a given event occurred, which requires high synchronization and timing accuracy.

Operating temperature

Cooling is often mandatory for noise reduction.

2.3.2   Detection technologies

We can distinguish between two main types of detectors for fluorescence microscopy:

Semiconductor detectors

They are based on an internal photoelectric effect and are most often encountered as 2D or 1D array detectors. Typical examples are CCD (Charge Coupled Device, [29]) cameras for widefield microscopy.

Figure 2.9(a) presents the structure of a CCD sensor. Its elementary building block (pixel) is a MOS (Metal-Oxide-Semiconductor) photocapacitor, whose role is to convert photons into electric charge. While the internal conversion process can achieve rates close to 100 %, the transmission coefficients of the electrodes and the insulating layer limit the overall quantum efficiency of the detection to 40 % at most. To improve upon this value, the photocapacitor can be illuminated from the silicon substrate side, or *back-illuminated*. However, this requires a complex (and expensive) etching process to reduce the substrate thickness.

The accumulated charge is essentially a linear function of the number of incoming photons, until a saturation level is reached. Above that level, additional
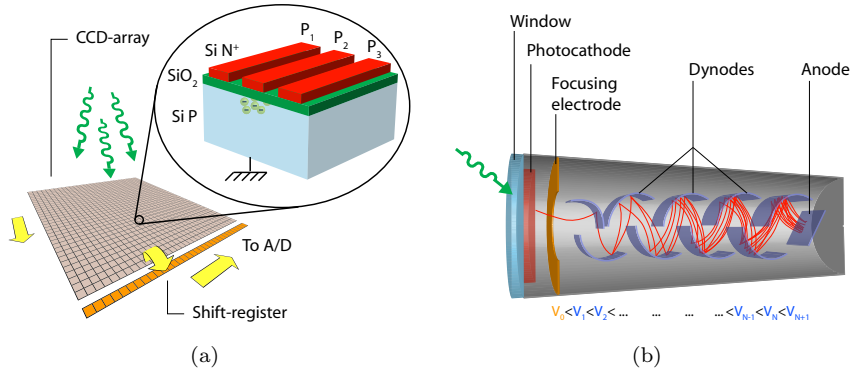
Figure 2.9: Examples of detectors for fluorescence imaging. (a) MOS photocapacitor in a full-frame-transfer CCD array. Two layers of a semiconductor silicon crystal with (P-type or N-type) impurities are separated by insulating silicon dioxide (SiO_2). When a photon penetrates into the substrate, it can excite electrons that are part of the crystalline structure. By applying suitable voltages at the electrodes (or gates) P_1, P_2 and P_3, these are "trapped" under the SiO_2 layer. (b) Transmission-mode PMT, composed of a vacuum tube and several electrodes that are applied a voltage gradient. The first one is protected from outside oxidation by a window that is transparent to light.

charges may diffuse to neighbouring pixels, an effect known as *blooming*. The maximum number of electrons, divided by the average number of electrons generated by noise (see below), gives the peak SNR. High sensitivity cameras can reach ratios of 30000:1 or better.

In CCD detectors, the charges are read-out using an analog shift-register[9]. In the full-frame transfer scheme illustrated in Fig. 2.9(a), the charges of each pixel row are sequentially moved towards the shift register by adjusting the gate voltages P_1, P_2 and P_3 in a periodic pattern. At the register, each row is again sequentially "emptied" and the charges are converted to voltages which are amplified and digitized.

---

[9]As opposed to digital shift-registers used in CMOS detectors, which allow each pixel to be accessed individually, at the expense of a reduced photosensitive area.

Photomultiplier tubes

PMTs ([164, 243]) are based on the photoelectric effect. They have no intrinsic spatial resolving power and are thus mostly used in combination with a scanning system and a detection aperture, typically in confocal microscopes. Figure 2.9(b) shows the schematic diagram of a PMT.

When photons with sufficient energy hit the *photocathode*, they may excite electrons and induce their release inside the tube. These are directed towards the first dynode by means of a focusing electrode. There, some are reflected, while others are absorbed and can excite secondary electrons. The number of reflected and secondary electrons divided by the incoming electrons defines a single dynode gain g. In total, a series of n dynode stages (typically a dozen or more) is traversed before the anode finally collects the electrons that have been produced. This principle can lead to very high gains, g^n being of the order of 10^6 to 10^8.

The photocathode is a key element as it determines the quantum efficiency of the system for the most part. Typically, less than 30 % of the incoming photons are effectively "converted" into electrons, depending on the wavelength.

2.4   Limiting factors of fluorescence imaging

Two sources act as the principal limiting factors in fluorescence imaging: 1) the instrumentation, which, apart from its inherent resolution limitation, introduces measurement noise, and 2) the sample itself, whose optical properties and emission characteristics are often non-ideal.

2.4.1   Noise sources

Photon shot noise

The fundamental limitation of any photodetector resides in the random nature of photon emission. The arrival of photons at the detector is well described by a Poisson process whose (statistical) intensity is proportional to the (physical) intensity of the fluorescence signal.

Background noise

The ambient radiation, especially in the infra-red domain, can also be a significant source of noise; it often requires the use of additional filters at the detection stage.

Dark current

Among the numerous internal noise sources of the detector, thermal agitation is the most important. The higher the temperature, the higher the kinetic energy of the electrons. For semiconductor detectors, this results in so-called *dark currents* (that exist even in the absence of light), which tend to charge the photocapacitors when the integration time and/or the temperature are too high. For point detectors such as PMTs, thermal energy can trigger spontaneous electron emissions. Consequently, high-sensitivity detectors are very often cooled down to reduce thermal noise.

Auxiliary noise sources

For semiconductor devices, additional noise is generated at read-out time. In particular, the charge transfer in CMOS sensors is less efficient than in CCD chips. Both technologies are subject to amplifier noise. For PMTs, there can be fluctuations in the internal gain of the unit, which also result in noise. Finally, any detector with digital output produces quantization noise (i.e., roundoff errors).

2.4.2 Sample-dependent limitations

Photobleaching

An important property of fluorophores is that they become more chemically reactive as they are being excited. Depending on the environment, they can undergo reactions that lead to permanent changes, by which the molecule loses its capability to fluoresce altogether, or becomes non-absorbent for the specified excitation wavelength. This effect, called *photobleaching*, limits the total intensity of light, and accordingly, the exposure time, until loss of fluorescence occurs. As a result, the observation time of a fluorescence-tagged specimen is limited. Photobleaching is a cumulative effect; this means that reducing the exposure time or excitation intensity will not prevent it, but merely reduce the rate at which it occurs.

Autofluorescence

Many organic molecules are naturally fluorescent, and thus even unstained biological samples can emit fluorescence in the visible domain. This autofluorescence is an important source of noise when it overlaps with the emission of a selected fluorophore, especially when the latter is sparsely expressed or exhibits weak fluorescence. This interference can render the detection of a signal very difficult.

Absorption and scattering of the medium

In a biological specimen, the intensity of the fluorescence signal decreases as the fluorophore's depth within the specimen increases. This attenuation is due to the absorption and scattering[10] of light; it strongly limits both the depth at which a fluorophore can be excited and the depth at which a fluorescence signal can be detected[11]. These effects are not always negligible. Therefore, to obtain truly quantitative measurements, it may be necessary to develop reconstruction algorithms that take into account the space-varying and complex nature of the refractive index.

2.5 Advanced experimental techniques

Besides standard imaging that involves the quantitative analysis of local fluorophore concentrations, there exist more sophisticated experimental techniques for studying protein-protein interactions and for investigating biological processes at the molecular scale. Among the techniques presented below, FLIM and FRET can be performed on both widefield and confocal microscopes. The photobleaching techniques, however, are usually performed with lasers and often require the ability to precisely define the region to be bleached; they are therefore mostly implemented on confocal microscopes.

2.5.1 FRET

Protein-protein interactions take place at scales that are too small to be resolved by optical microscopy; however, they can be detected by exploiting a mechanism called fluorescence resonance energy transfer (FRET). This process is a direct transfer of energy (i.e., it does not involve the emission or absorption of a photon) between a suitable *donor* and an *acceptor*, as illustrated in Fig. 2.10. FRET is only possible between two fluorophores if the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor. An example of a suitable pair of fluorescent proteins is the aforementioned CFP/YFP couple.

The efficiency of FRET strongly depends on the distance that separates the two molecules (the rate is inversely proportional to the 6th power of the distance) and on the relative orientation of their dipole moments. This means that FRET

---

[10]Scattering is the phenomenon by which particles with a refractive index different from the medium's index partially diffuse electromagnetic radiation in all directions. It commonly occurs when the particle sizes are comparable to the wavelength.

[11]Typically in the 100 μm range for one-photon confocal microscopy.

can be used to study the optical subresolution colocalization of a labeled protein pair of interest. FRET can also serve as an indicator of conformational changes in a protein: if complementary markers are placed at the extremities of the protein, then an energy transfer can occur when the protein folds [133]. Fig. 2.11 illustrates a FRET experiment.

2.5.2 FRAP

Although photobleaching has already been mentioned as a limitation, it can be exploited to study the intracellular dynamics of proteins. FRAP (fluorescence recovery after photobleaching) consists in intentionally bleaching a small region of a cell using high-intensity light, thereby rendering it non-fluorescent. The region then regains its fluorescence as fluorophores from the surroundings enter and pass through it, which yields information about the diffusion and mobility of the labeled protein [197].
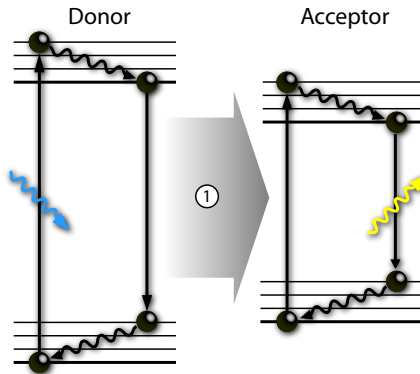
Figure 2.10: The principle of FRET between a suitable donor-acceptor pair: the energy of the excited donor molecule is transferred (without emission of a photon) to the acceptor after vibrational relaxation (1). For FRET to occur, the distance between the two molecules must typically be in the range of 1-10 nm [258].

(a)   (b)   (c)
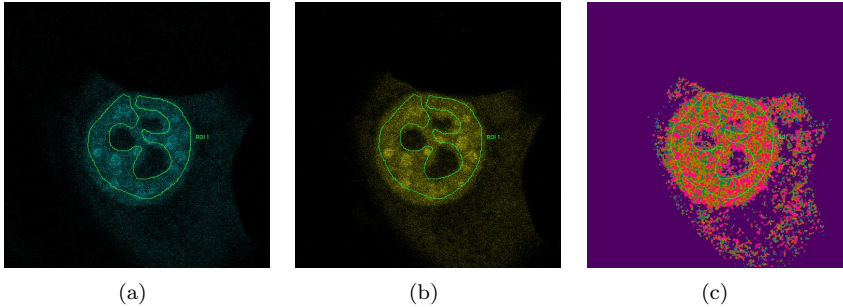
Figure 2.11: Images from a FRET experiment, showing the (normalized) donor (a) and acceptor (b) channels. From these images, a computer-generated FRET-efficiency image is obtained (c). In this case, the scientists were interested in the average FRET efficiency inside a region of interest corresponding to the cell nucleus. Courtesy of M. C. Rio, A. Baguet, and P. Kessler, IGBMC, Illkirch, France.

2.5.3   FLIM

All of the techniques discussed up to this point rely on intensity-based measurements. In the presence of autofluorescence, or when multiple fluorophores with similar emission spectra are used, it can be difficult to discriminate among the different signals. Intensity-based imaging is also highly dependent on fluorophore concentration.

In fluorescence lifetime imaging microscopy (FLIM), image contrast is generated based on the lifetime of fluorophores, which is the average time a fluorophore remains in the excited electronic state. The key point is that every fluorophore has a unique lifetime. A common method for measuring fluorescence lifetimes consists in exciting fluorophores with a picosecond pulsed laser source and recording the arrival times of the emitted photons with a high speed photodetector.

The lifetime of a fluorophore is sensitive to many environmental factors such as oxygen concentration, pH, and calcium ion concentration. Thus, FLIM can be used to obtain information about the local environment of a particular fluorophore. FLIM can also serve as an experimental indication that FRET occurs, because FRET induces a change in fluorescence lifetime (see Fig. 2.12).
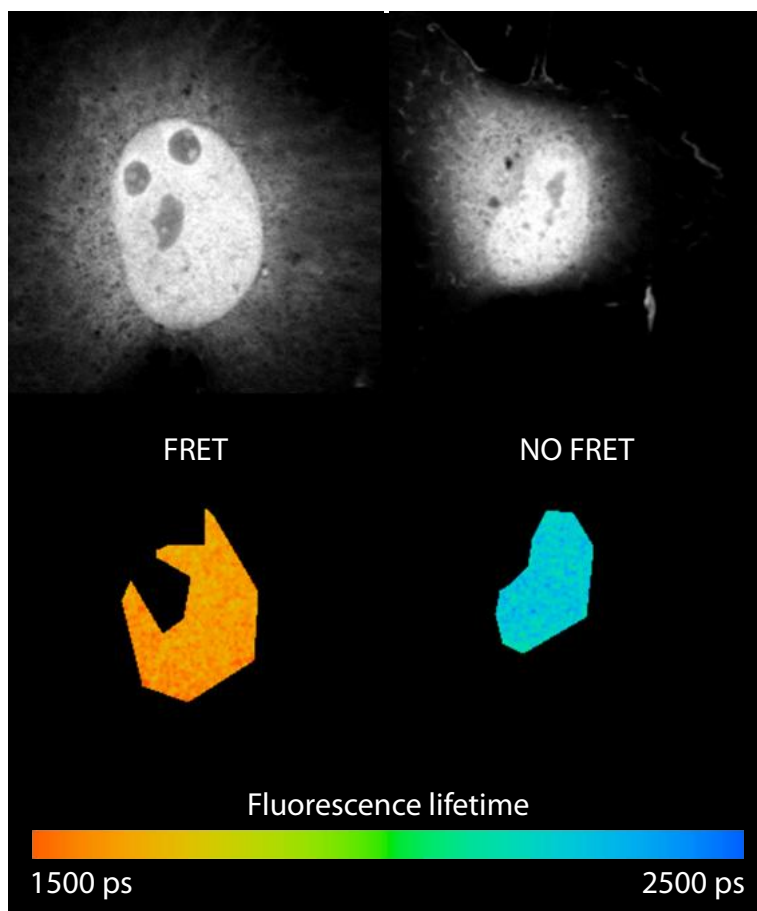
Figure 2.12: FLIM experiment for confirming the occurence of FRET. The images show the fluorescence lifetime of the donor over two regions of interest corresponding to different cell nuclei. In the presence of FRET (left-hand side), the fluorescence lifetime is significantly reduced, due to the energy transfer to the acceptor. Courtesy of C. Rochette-Egly, S. Lalevée and P. Kessler, IGBMC, Illkirch, France.
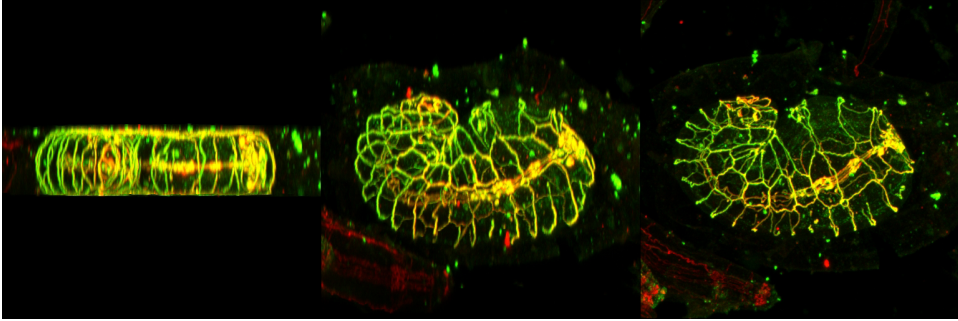
Figure 2.13: 3D projections of a nematode worm embryo (*C. elegans*) [147]. Courtesy of L. McMahon, J.-L. Vonesch and M. Labouesse, IGBMC, Illkirch, France.

2.6 Signal and image processing challenges

2.6.1 Data size and dimensionality

Modern research in biology requires quantitative experimental data. As a consequence, microscopes have developed into sophisticated digital image acquisition workstations that are capable of acquiring very large data sets of high dimensionality.

To get a better feeling of what is involved, consider an experiment monitored with a confocal microscope that requires the periodic (*time-lapse*) 3-D acquisition of a sample labeled with two fluorophores. This yields a 5D data set indexed by the space coordinates x, y and z, the time t, and the wavelength parameter λ. Assuming that each image has a resolution of 1024×1024 and that 32 slices are acquired per volume every 20 minutes over 24 hours with a 12-bit quantizer, the whole experiment results in nearly 7 gigabytes of data. If a comparative analysis is performed, this figure must be multiplied by the total number of samples.

Studies involving comparable or even larger amounts of data are becoming commonplace. Even with today's performance level of computer hardware, the storage, extraction, manipulation and representation of such data sets remain complex. One major challenge lies in the design of database systems and compression formats allowing for efficient retrieval and visualization (projections, 3D rendering – see the example in Fig. 2.13).

But most importantly, signal processing is becoming an indispensable expertise

for the analysis and understanding of quantitative biological experiments; in fact, it is increasingly considered part of the experimental protocol itself, as a way to infer the validity of a biological model.

Without claiming exhaustiveness, we give examples of current image processing problems in biology among five main categories: image preparation, image restoration, image registration, image segmentation, and quantitative image analysis.

2.6.2 Image preparation

Image calibration

Calibration is an important step both for image analysis and visualization. It can involve various preprocessing tasks such as histogram equalization, inhomogeneous illumination compensation, background correction, or image rescaling. While these tasks may appear relatively simple, some of them can rely on advanced signal processing.

Image simplification

In some cases, biological structures are too complex to be processed directly and an image simplification step is required. To preserve the objects of interest, the operator can choose among the wide range of available tools (morphological operations, filtering, multiresolution structures, diffusion equations, ...); application-specific solutions can also be envisaged.

Feature detection

Biological images often present characteristic elements such as particles and filaments. The detection of these features may require the development of optimized filters [112, 187], as well as multiresolution methods [168]. Here, a challenging aspect is the shape variability observed in live-cell imaging.

Experimentalists should at least be aware of the aforementioned preparation operations; otherwise, they run the risk of a significant loss of information, thereby leading to questionable results at publication time. Algorithm designers, on the other hand, should put more efforts into education and the development of user-friendly imaging software.

2.6.3 Restoration

Restoration encompasses the classical problems of denoising and deconvolution.

Denoising

Simple methods such as median filtering often need adaptation; for example, a 3-D stack may exhibit lower SNR levels as deeper portions of the object are imaged, due to absorption and/or autofluorescence. More generally, restoration methods should be based on physically realistic noise models (e.g., Poisson statistics) and take into account various noise sources (see Section 2.4). Advanced algorithms, relying on wavelet-domain thresholding strategies [27], PDE and variational formulations, or statistical frameworks, are just starting to be used in the field of bioimaging, and deserve more exploration.

Deconvolution

This operation requires an adequate characterization of the underlying imaging system, which can be either theoretical (involving a PSF model) or, frequently, experimental. In the latter case, the PSF is obtained by imaging subresolution fluorescent beads, under conditions as close as possible to the actual biological preparation. By averaging, possibly with simplifying assumptions (e.g., symmetry), a relatively noise-free PSF can be obtained. For further details concerning algorithmic deconvolution methods, we refer to [200, 242].

One of the main challenges is the design of computationally tractable methods that take into account the nonstationarity of the PSF, especially in the axial direction (see section 2.2.3). A recent attempt is the EM-algorithm proposed by Preza and Conchello [178]. In their image-formation model, the object is divided into several layers that are associated with a series of (depth-dependent) PSFs. The image of each layer is then obtained from classical (stationary) convolutions.

Other inverse problems

Restoration can also be considered in the wider framework of inverse problems.

One example concerns relatively thick objects with surface labeling, observed under a widefield microscope. Because of the 3-D conical extension of its PSF (Fig. 2.7, top row on the left), such a system has a limited depth of field; that is, only a small slice of the object around the focal plane appears sharp. To compensate for this, images at different focal depths can be taken and fused together so as to obtain a single, entirely sharp image[12]. A state-of-the-art (so-called *extended-depth-of-field*) algorithm is decribed in [82]. Such a method can also be used to extract 3-D maps of the object's surface.

---

[12]This process should not be confused with deconvolution, which in particular yields a 3D stack instead of a single image.

Another problem of interest is related to the detection and localization of sub-resolution particles [215]. New methods have been developed that take into account the 3-D nonstationarity of the PSF to achieve precision in the nanometer range [2].

2.6.4   Registration

Registration is a frequently-needed post-acquisition step. Here, researchers can take advantage of the availability of high-quality registration algorithms that were initially developed for medical imaging [34].

Mosaicing

Because of the limited field of view of high-magnification objectives, it can be necessary to acquire multiple images of a sample; for example, in a mosaic scheme. Despite the high accuracy that sample stages can achieve, perfect alignment is never possible. Rigid-body registration algorithms can correct this, provided the acquired images or volumes slightly overlap [214]. Within a given stack, it might also be necessary to compensate for pixel shifts between successive images. In addition, refraction indices – and thus focusing depths – are wavelength-dependent, which can necessitate the realignment of the different fluorescence channels.

Imaging live samples

During time-lapse acquisitions, spatial drifts can occur due to thermal processes; the sample itself might also be subject to motion. Therefore, even if the parts of interest lie in a relatively thin slice, they may not be observable in a unique focal plane over the whole experiment. This implies that either a stack of neighboring planes must be acquired and the planes of interest must be extracted (or fused); or a real-time focusing algorithm must be used to control the stage.

More sophisticated elastic registration may be required for compensating the deformation of living tissues, or for matching specimens of comparable shape [203].

If a periodic biological process is too fast to be imaged with sufficient time-resolution (such as the repetitive 3-D flow of blood cells in a heart), a registration approach may also be applied: in [129], images over several periods are recorded and reassembled so as to obtain a single period at a high frame-rate.

2.6.5   Segmentation

Segmentation is a mandatory step for image analysis. User interaction for the manual delineation of regions of interest is time-consuming and lacks reproducibility.

The need for automated segmentation methods is therefore important, e.g., for local intensity measures, object and event counting, as well as tracking.

While simple approaches such as prefiltering and thresholding are available in commercial software packages, advanced techniques – for instance active contours [261] – have not yet been much exploited in the context of biological image analysis.

The most accurate segmentation methods are often application-dependent and typically require specific developments. For example, the tracing of neuronal dendrites can be improved using graph-optimization techniques [149].

In the context of live microscopy, it also makes good sense to adapt the segmentation methods so that they exploit temporal coherence, e.g., for the labeling of cells.

2.6.6   Quantitative analysis

Data preprocessing

In multispectral imaging, each pixel consists of (possibly a large number of) intensity measures at different wavelengths (obtained using different filter sets, an interferometer, or a diffractive system). If several fluorophores are used, their spectra are likely to overlap, and channel-crosstalk must be expected. This gives rise to *unmixing problems* [263] that can be solved by taking separate reference images of each fluorophore (to measure its contribution to each channel) and using, for example, a singular-value decomposition [222]. Blind separation methods may also be applicable. As a general observation, the correct normalization of spectral data is critical for the interpretation of fluorescence images. Quantitative assessments using FRET or ratio imaging (comparing the relative intensities of different wavelengths) require careful preprocessing based on physical parameters such as spectrum overlap or fluorophore concentration.

Model fitting

Other advanced fluorescence techniques are based on the fitting of parametric models: in FLIM, the fluorescence lifetimes are obtained by fitting (possibly multiple) exponential trends to the photon arrival densities; in FRAP, the diffusion coefficients characterize fluorescence recovery curves [36]. Generally speaking, quantitative research often relies on the mapping of physical or biochemical models in the image and/or time domains, especially for dynamic processes.

Motion assessment and tracking

The diffusion of fluorescent proteins can be characterized by estimating motion fields. In many instances, it is interesting to track individual objects, which can also be a challenging task. We refer the reader to other sources that cover the broad field of movement analysis [150, 262].

Pattern recognition and classification; screening

Screening experiments consist in a systematic, automated study of a large number of samples (up to several hundreds of thousands), e.g., for the study of gene function or for drug discovery. This can involve terabytes of data and several weeks of computerized analysis. Pattern recognition and classification algorithms play a major role in this analysis. In particular, one must identify how the biological characteristics of interest translate into measurable image features. Computational complexity is a strong limiting factor, while the reliability of the methods must be thoroughly validated [260].

2.7 Current and future trends

In addition to the signal processing tools that have been discussed in the previous section, both the probes [258] and the instrumentation are being refined constantly. We therefore close our discussion with a description of current trends and future directions in the field.

2.7.1 Fluorescent labels

Quantum dots

Among the most recent developments are *quantum dots* [153], labels composed of a core nanometer-sized semiconductor crystal and an external protective shell. Their main advantages with respect to earlier fluorophores are their broader absorption and narrower emission spectra, resulting in brighter fluorescence. Also, they are more stable chemically and thus less subject to bleaching. These inorganic structures can be used for *in vivo* imaging, although they cannot be expressed by cells [69].

Labeling of recombinant proteins

The principle of this technique is to create fusion proteins that are not fluorescent by themselves, but which express a receptor to which a specific label can be added at a later time [93]. The label can be chosen from a wide range of fluorophores, with properties that GFP-type proteins may not be able to provide (such as higher resistance to photobleaching and stronger fluorescence). The receptor continues to be expressed in newly synthesized proteins, but only the stained proteins exhibit fluorescence, which allows for the selective labeling of a protein population at a given point in time.

Enhanced fluorescent proteins

New fluorescent proteins are being developed that provide increased quantum efficiency (e.g., enhanced GFP, or eGFP, with a 35-fold increase in brightness with respect to the original GFP) or whose emission spectra are closer to infra-red wavelengths (700 nm and above). These wavelengths are generally less absorbed by biological samples, hence allowing deeper observation. They are also less masked by cell autofluorescence occuring in the visible spectrum.

Photocontrolable proteins

Recent research has also been devoted to the design of *photoactivatable* [173] and *photoswitchable* [143, 6] proteins. The former exhibit little fluorescence in their initial, quiescent state. When exposed to a strong irradiation at a specific wavelength (usually lower than the fluorescence excitation wavelength), a 100-fold or higher increase in fluorescence brightness can be observed. For switchable proteins, strong irradiation changes both the excitation and emission spectrum. For example PS-CFP [46] is sensitive to irradiation at 405 nm, which produces a 1500-fold increase in its green-to-cyan ratio. Both types of labels can be used to activate and observe proteins in a specific region of a cell, without the interference of newly synthesized proteins or proteins outside of the selected region. This property is useful for protein lifetime and tracking as well as cell lineage studies.

2.7.2   Advanced microscopy systems

We conclude this section with some of the more advanced developments in the field of optics.

Faster scanning – slit detectors and Nipkow-disks

To cope with the high speed of some biological processes, the traditional confocal scanning microscope equipped with a point detector is often not sufficient. To accelerate the scanning process, a whole line can be imaged simultaneously by replacing the pinholes and the PMT by slit apertures and a linear camera. More generally, using e.g., a Nipkow-disk system [175], a 2-D illumination pattern can be shifted accross the sample, allowing time lapse imaging at up to 120 frames per second. This comes with a significant tradeoff in terms of resolution, due to crosstalk between the different detection apertures.

Deeper imaging – multiphoton microscopy

In a multiphoton microscope [60], optical sectioning is achieved by properties of the illumination; as a consequence, there is no need for a detection pinhole in such a system. Very short laser pulses (in the pico- to femtosecond range) are sent to the sample in brief intervals (of the order of nanoseconds). The probability that two photons encounter the same molecule, hence bringing it to its excited state and making it fluoresce, is significant only in the very central region of the illumination spot. A key advantage is that the corresponding infrared wavelengths are less absorbed by biological tissues so that samples can be imaged much deeper than with traditional confocal systems (at a comparable resolution). Photobleaching and toxicity are also reduced because the excitation intensity is effectively concentrated at the focal spot.

Increased axial resolution – multiple objective imaging

Since the numerical aperture has such a fundamental influence on resolution, Hell et al. proposed to insert the sample between two objectives, so as to send and collect light from both sides; accordingly, they called the method 4π microscopy [103]. Using computational methods, an improvement in axial resolution by a factor of six can be achieved. Such systems are commercially available, but they suffer from limitations on the sample thickness and sensitivity to differences in the length of the two optical paths. More recently, these ideas have been applied to widefield microscopy (I^5M, [96]).

Increased resolution – STED microscopy

One of the most advanced microscopy techniques to emerge from the quest for increased resolution is called STED, which stands for stimulated emission depletion

[104] (see also Section 2.1.1). The principle is to prevent fluorescent molecules outside the very central region of the illumination spot from emitting light by forcing them back to their fundamental state. This is achieved by dividing the excitation into two brief successive laser pulses, where the second pulse is redshifted and doughnut-shaped, having zero intensity at its center. Superimposed on the focal spot of the initial pulse, it induces stimulated emission, dramatically reducing the excitation volume, hence augmenting the resolution.

2.7.3 Super-resolution – photoactivated localization-based techniques

Photoactivated localization microscopy (PALM [20]) and stochastic optical reconstruction microscopy (STORM [15]) are novel techniques based on imaging sparse subsets of photoactivable, respectively photoswitchable, fluorescent proteins (see Section 2.7.1). A sparse subset of activated molecules is obtained by illuminating the sample with a short laser pulse. The sample is then imaged until the molecules are either bleached or switched off, after which the process is repeated until the pool of activable proteins in the sample is depleted, or until a sufficient amount of switchable proteins has been imaged. In a post-processing step, the individual molecules from each subset image are localized to nanometer-scale accuracy. The resulting estimated positions and corresponding intensities are then used to render a composite super-resolved image. Limited only by the number of photons collected from each fluorophore, resolutions of 10-20 nm have been reached in practice, which rivals the performance of electron microscopes.

Other developments

Another promising technique to improve the resolution of widefield systems is structured illumination. Illuminating the object with sinusoidal patterns, combined with adequate processing, can result in a two-fold or higher [101] improvement of the microscope's spatial bandwidth. It also yields optical sectioning properties and the processing can be done on specific hardware for real-time observation [155].

Another approach proposed by Stelzer et al. (SPIM, [110]) consists in projecting the light onto the object perpendicularly to the optical axis, in a diffraction-limited plane; then only fluorescence from molecules within this excitation plane is collected using a traditional CCD sensor. This system provides true optical sectioning for widefield systems.

2.8   Conclusion

Although this panorama is necessarily incomplete, we hope to have convinced the reader of the invaluable role of fluorescence microscopy in modern biology. It owes its current popularity to the GFP-like fluorescent proteins that are the key ingredient for *in vivo* studies of molecular processes in cells. These are currently opening up a plethora of experimental possibilities that only begin to be explored.

This colored revolution could clearly not have happened without numerous technological advances. In particular, progress in optics and instrumentation has been considerable in recent years; there is now a consistent trend towards non-linear techniques, such as multiphoton and saturated illumination imaging, which, with the help of computational methods, are contributing to overcoming Abbe's resolution barrier.

Signal processing is also at the heart of these developments and is expected to play an ever-increasing role in the field. It is already an integral part of optics and is becoming an essential tool for biologists, who rely more and more on imaging software to quantitate their data.

Therefore, a crucial aspect of the research lies in the successful collaboration between signal processing engineers and biologists. In modern research institutes, imaging core facilities are expected to play an important mediating role in this interaction. Our advice to colleagues that want to be part of this effort is that they try to understand the physics and, to some extent, the biology in order to design better and more useful algorithms. We believe that it is truly worth the effort.

Chapter 3

Deconvolution in fluorescence microscopy

The goal of this chapter is to give some background information on the problem of deconvolution in the context of fluorescence microscopy. Most of the material will be part of a forthcoming article [193] on the "DeconvolutionLab" plugin for ImageJ. This plugin implements some of the most common methods for deconvolution microscopy. It was the subject of a Master's project [192] that was supervised during this thesis[1].

As is usual for inverse problems, we adopt a variational formulation; this allows for a unified presentation of most existing approaches. Our purpose is to provide a self-contained description that can serve as a technical reference, in particular for end-users of the plugin. Alternative presentations of this topic can be found in review papers [233, 158, 207, 179, 188] or popular textbooks [17, 235].

3.1   Status quo of the field

The combination of three-dimensional (3D) deconvolution with optical-sectioning fluorescence microscopy was demonstrated more than twenty-five years ago[2] in the

---

[1]Note that a companion plugin for fitting and generating 3D PSF models was also developed around the same time in relation with another Master's project [83].

[2]Let us also mention the earlier work of Weinstein and Castleman [247], who introduced the so-called nearest-neighbor deconvolution method. However, according to our classification in Table 1.1, this approach rather belongs to the category of image enhancement.

visionary work of Agard and Sedat [1]. Following the discovery of novel fluorescent marker proteins and the rapid development of digital microscopy during the 1990's [3], this technique became widely popular in biological imaging. It is now commonly termed *deconvolution microscopy* [148]. We refer to [200, 188] for recent reviews of the subject, as well as [242] for a discussion of practical aspects.

Since the work of Agard and Sedat, several comparative studies of deconvolution algorithms have been performed in the context of fluorescence microscopy. Initially, Conchello and Hansen [52] used synthetic data to compare a constrained Van-Cittert algorithm with maximum-likelihood deconvolution. Van Kempen et al. [227, 230] consider the Richardson-Lucy algorithm, the iterative constrained Tikhonov-Miller (ICTM) algorithm and Carrington's algorithm [37]. Several deconvolution methods are compared within a Bayesian framework in [233, 146]. More theoretical studies related to the recovery of out-of-band information by maximum (penalized) likelihood algorithms are also available [50, 100].

Alternative approaches where the denoising and deblurring steps are decoupled have also been proposed. For example, [107] uses the Anscombe transform and Wiener filtering for the denoising part; the deblurring part is based on the method of projections onto convex sets (POCS). [27] uses wavelet-domain denoising before a maximum-a-posteriori (MAP) deconvolution algorithm.

More recently, the Richardson-Lucy algorithm was combined with wavelet-domain denoising [185] (note however that this work is not based on a variational formulation comparable to ours) and total-variation regularization [63].

A number of studies on deconvolution are specifically concerned with confocal microscopy, e.g., [53, 225, 228, 229]. Applications to two-photon, 4-Pi, 3D FLIM and STED microscopy are described in [117], [194], [204] and [250] respectively. Another variation on the topic is deconvolution for extended-depth-of-focus imaging [51].

The previous works can be distinguished from more generic approaches such as blind deconvolution [11, 144, 28, 185, 216] or shift-variant deconvolution [145, 178, 198].

The current state-of-the-art in practical deconvolution microscopy can be summarized as follows:

- Most works are based on a Bayesian formulation of the deconvolution problem. This typically leads to (non-linear) iterative algorithms with a good level of flexibility (additional constraints such as positivity can be easily incorporated). However, the acceleration of such methods was rapidly recognized as an important issue [106]. This has for example motivated the use of conjugate-gradient algorithms [234, 191].
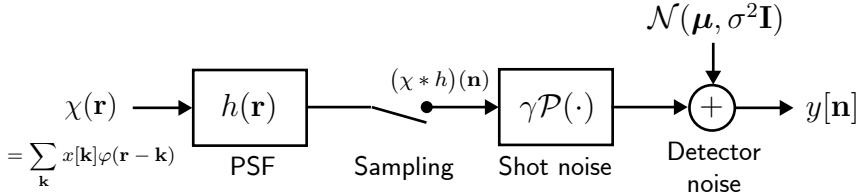
Figure 3.1: Image-formation and measurement model.

- Irrespective of the deconvolution method and its underlying paradigm, the mean squared error (MSE) is the most widespread measure of restoration quality [227, 230, 233, 228, 100], together with the I-divergence [227, 230, 63].

- Both Poissonian and Gaussian noise models have been considered (see e.g. [233, 191]). While the former model is better matched for low light intensities, the latter model is more generic. A Gaussian approximation (with spatially-varying variance) of the Poisson statistics was proposed in [146]. To our knowledge a mixture of both models has not been used for the restoration of fluorescence micrographs so far.

- Generalized Cross Validation (GCV) is the prefered method for parameter estimation (see e.g. [204, 229, 191]).

3.2 General image-acquisition model

In fluorescence microscopy one aims at measuring $\chi(\mathbf{r})$, the spatial distribution of fluorophores inside a prepared sample. Following upon our discussion in Chapter 2, we will now describe the image-acquisition chain—which is depicted in Fig. 3.1—in some more details.

When operating the microscope under design conditions, image formation is essentially shift-invariant (if one sets magnification aside). It can be modeled by a D-dimensional convolution[3] with an optical point spread function that models the illumination, the diffraction-limited objective and possibly the pixel geometry of the sensor [94, 102].

The measurement of the image intensity is affected by several noise sources (see Chapter 2). Here we propose a relatively flexible statistical model that can

---

[3]$D = 2$ or $D = 3$ depending on the thickness of the object.

account for a broad range of imaging situations (from photon-limited to sensor-limited imaging).

The first component of the model is *signal-dependent* and models the fluctuation of the number of photons arriving at a given pixel. This so-called shot noise follows a Poisson distribution whose mean depends on the incoming light intensity. We also introduce a parameter γ that represents the gain of the measurement device. The second component accounts for various other distortions such as a background signal, read-out noise or quantization noise. It is modeled as a vector of independent Gaussian random variables of mean $\boldsymbol{\mu}$ and covariance matrix $\sigma^2 \mathbf{I}$, where $\mathbf{I}$ denotes the identity matrix.

In the sequel, we use a discretized description of the problem to simplify the presentation. We assume that the characteristic function of the object can be represented with sufficient accuracy in a shift-invariant space[4] spanned by a function $\varphi(\mathbf{r})$:

$$\chi(\mathbf{r}) = \sum_{\boldsymbol{k} \in S} x[\boldsymbol{k}] \, \varphi(\mathbf{r} - \boldsymbol{k}).$$

We further assume periodic boundary conditions, in the sense that φ has integer periodicity along all dimensions. Then the signal of interest is defined by a finite number of coefficients $x[\boldsymbol{k}]$ indexed by the set $S = [\![0, N_1 - 1]\!] \times \ldots \times [\![0, N_D - 1]\!]$.

To summarize our model, the measurement is related to the original signal through

$$\mathbf{y} \sim \gamma \mathcal{P}(\mathbf{Hx}) + \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I}). \tag{3.1}$$

Here $\mathbf{y}$ and $\mathbf{x}$ are vectors in $\mathbb{R}^N$, where $N = N_1 \times \ldots \times N_D$; they contain the lexicographically ordered samples of the measurement and of the original signal, respectively. $\mathbf{H}$ is a block-circulant matrix corresponding to the discrete convolution kernel $(\varphi * h)(\mathbf{n})$.

Note that simplifications of (3.1) may be possible or required depending on the application. For example, a purely Poissonian model $\mathbf{y} \sim \gamma \mathcal{P}(\mathbf{Hx})$ is obtained when $\boldsymbol{\mu} = \mathbf{0}$ and $\sigma^2 = 0$. Setting $\mathbf{x} = \tilde{\mathbf{x}}/\gamma$ and $\boldsymbol{\mu} = \mathbf{0}$, one obtains the simple Gaussian model $\mathbf{y} \sim \mathcal{N}(\mathbf{H\tilde{x}}, \sigma^2 \mathbf{I})$ as a limit case when $\gamma \to 0$.

Our goal is now to estimate $\mathbf{x}$ from a realization of $\mathbf{y}$.

---

[4]For example, a space of bandlimited functions. We consider unit shifts and a unit sampling period along all dimensions for simplicity.

3.3 Variational deconvolution

3.3.1 General formulation

Most deconvolution approaches can be formulated as an optimization problem. One seeks to construct an estimate $\mathbf{x}^*$ of the original signal that minimizes a cost functional $\mathcal{C}(\mathbf{x})$ over a preassigned set of admissible solutions Q, i.e.,

$$\forall\,\mathbf{x} \in Q, \mathcal{C}(\mathbf{x}^*) \leq \mathcal{C}(\mathbf{x}). \tag{3.2}$$

In the present discussion, both the constraint set Q and the cost functional $\mathcal{C}(\mathbf{x})$ are assumed to be convex; the latter typically takes the form

$$\mathcal{C}(\mathbf{x}) = \mathcal{D}(\mathbf{x}) + \lambda \mathcal{R}(\mathbf{x}). \tag{3.3}$$

This formulation is often obtained in a Bayesian framework (see below). A more generic interpretation can be stated as follows.

- The *data term* $\mathcal{D}(\mathbf{x})$ enforces a certain level of consistency between the estimate and the measurement.

- The *regularization term* $\mathcal{R}(\mathbf{x})$ is constructed using prior knowledge on the original signal and typically favors estimates with particular smoothness properties.

- The *regularization parameter* λ determines the relative influence of both terms.

3.3.2 Connection with the Bayesian formalism

The cost functional defined above can be motivated from a statistical standpoint. Here we briefly establish the connection between the above variational framework and the Bayesian methodology. A relatively accessible presentation of Bayesian deconvolution can be found in the review papers [158, 207], which primarily deal with applications in astronomy.

The basic assumption is that the signal to be estimated is a random variable with known probability density function $p(\mathbf{x})$; this function is termed the *prior*. The Bayesian paradigm consists in estimating $\mathbf{x}$ by maximizing the posterior probability $p(\mathbf{x}|\mathbf{y})$, that is, the probability of $\mathbf{x}$ given that we have observed $\mathbf{y}$. The standard manipulation consists in applying Bayes' rule twice so as to obtain

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} = \frac{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{y})}.$$

Data term $\mathcal{D}(\mathbf{x})$	Noise model
$\|\mathbf{y} - \mathbf{Hx}\|_2^2$	Gaussian: $\mathbf{y} \sim \mathcal{N}(\mathbf{Hx}, \sigma^2 \mathbf{I})$
$\langle \mathbf{1}, \mathbf{Hx} \rangle - \langle \mathbf{y}, \log(\mathbf{Hx}) \rangle$	Poisson: $\mathbf{y} \sim \mathcal{P}(\mathbf{Hx})$

Table 3.1: Standard choices for the data term and corresponding noise models (see text for notations).

The function $p(\mathbf{y}|\mathbf{x})$ is known as the *likelihood* and is determined by the choice of a noise model (see Table 3.1).

Thus, maximizing $p(\mathbf{x}|\mathbf{y})$ with respect to $\mathbf{x}$ is equivalent to maximizing $p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})$. This in turn is equivalent to minimizing

$$-\log\left(p(\mathbf{y}|\mathbf{x})\, p(\mathbf{x})\right) = -\log\left(p(\mathbf{y}|\mathbf{x})\right) - \log\left(p(\mathbf{x})\right).$$

If we compare the right-hand side with (3.3), the log-likelihood $\log\left(p(\mathbf{y}|\mathbf{x})\right)$ acts as the data term, while the log-prior $\log\left(p(\mathbf{x})\right)$ acts as the the regularization term (up to the negative signs).

Examples of commonly-used prior functions are given in Table 3.2. From a statistical perspective, there are three main estimation approaches:

- *Maximum likelihood (ML)*: $p(\mathbf{x})$ is assumed to be the uniform distribution over a (bounded) constraint set Q; thus the solution only depends on the likelihood function (no regularization).

- *Maximum a posteriori (MAP)*: $p(\mathbf{x})$ is a non-uniform (e.g., Gaussian) probability-density function.

- *Penalized likelihood*: $-\log\left(p(\mathbf{x})\right)$ is replaced by an appropriate semi-norm that penalizes non-smooth (or non-regular) solutions.

3.4 Mathematical results for nonsmooth convex optimization

In order to set the framework of the next section, we introduce some tools from convex analysis. We refer to [231, 18, 4] for a detailed coverage and proofs.

Regularization term $\mathcal{R}(\mathbf{x})$	Notations	Statistical interpretation
Constant		Maximum likelihood
Tikhonov: $\|\mathbf{Lx}\|_2^2$	$\mathbf{L}$: "roughness-measuring" operator (typically high-pass filter); $\|\cdot\|_2$: Euclidian norm	Maximum a posteriori with a Gaussian signal prior ($\mathbf{L}^T\mathbf{L}$: signal autocorrelation matrix)
Wavelet-domain ℓ_1 regularization: $\|\tilde{\mathbf{W}}^T\mathbf{x}\|_1$	$\tilde{\mathbf{W}}^T$: wavelet-decomposition operator $\|\cdot\|_1$: ℓ_1 norm	Maximum a posteriori with a wavelet-domain Laplacian prior
Total variation regularization: $\|\mathbf{x}\|_{\mathrm{TV}}$	$\|\cdot\|_{\mathrm{TV}}$: total variation semi-norm	Penalized likelihood

Table 3.2: Examples of regularization terms and corresponding statistical approach.

In general, convexity assumptions ensure the existence of minimizers and allow for their concise characterization. In addition there are standard numerical procedures and convergence results for convex optimization (see e.g. the aforementioned references and the discussion in Chapter 4).

3.4.1 Existence and unicity of a minimizer

We first state some general conditions under which the problem of minimizing $\mathcal{C}$ is "well-posed".

Property 1. *Let Q be a non-empty closed subset of $\mathbb{R}^N$. Assume that $\mathcal{C}: \mathbb{R}^N \to \mathbb{R}$ is a convex function such that $\mathcal{C}(\mathbf{x}) \to +\infty$ whenever $\|\mathbf{x}\| \to +\infty$. Then $\mathcal{C}$ admits a minimizer over Q, i.e., there exists a vector $\mathbf{x}^* \in Q$ such that (3.2) holds. If in addition $\mathcal{C}$ is strictly convex, then this minimizer is unique.*

3.4.2 Projection onto a closed convex set

The previous property allows us to define $\mathcal{P}_Q(\mathbf{x})$, the projection of $\mathbf{x}$ onto a closed convex set Q.

Definition 1. *Let Q be a non-empty closed convex subset of $\mathbb{R}^N$. For every $\mathbf{x} \in \mathbb{R}^N$, the function $\mathbf{x}' \mapsto \|\mathbf{x}' - \mathbf{x}\|_2^2$ admits a unique minimizer in Q, which is denoted*

$\mathcal{P}_Q(\mathbf{x})$. *Equivalently, $\mathcal{P}_Q(\mathbf{x})$ is characterized by the property*

$$\forall\ \mathbf{x}' \in Q,\ \langle \mathcal{P}_Q(\mathbf{x}) - \mathbf{x}, \mathcal{P}_Q(\mathbf{x}) - \mathbf{x}' \rangle \le 0. \tag{3.4}$$

Note that when Q is a sub vector space[5] of $\mathbb{R}^N$, the above characterization coincides with the algebraic notion of projection: indeed, in this case, (3.4) is equivalent to the fact that the "projection error" $\mathcal{P}_Q(\mathbf{x}) - \mathbf{x}$ is orthogonal to Q.

3.4.3   The notion of subgradient

Geometrically, a subgradient corresponds to a hyperplane that is "subtangent" to a convex function at a given point.

Definition 2. *For a function $\mathcal{C} : \mathbb{R}^N \to \mathbb{R}$, a* subgradient *at $\mathbf{x} \in \mathbb{R}^N$ is a vector $\mathbf{g} \in \mathbb{R}^N$ such that*

$$\forall\ \mathbf{x}' \in \mathbb{R}^N,\ \mathcal{C}(\mathbf{x}') \ge \mathcal{C}(\mathbf{x}) + \langle \mathbf{g}, \mathbf{x}' - \mathbf{x} \rangle.$$

The set of all such vectors is denoted $\partial\mathcal{C}(\mathbf{x})$ and is called the subdifferential *of $\mathcal{C}$ at $\mathbf{x}$.*

The subdifferential generalizes the notion of gradient for a differentiable function, as shown by the following property.

Property 2. *Let $\mathcal{C} : \mathbb{R}^N \to \mathbb{R}$ be a convex function.*

- *At any given point $\mathbf{x} \in \mathbb{R}^N$, the subdifferential $\partial\mathcal{C}(\mathbf{x})$ is non-empty.*

- *If $\mathcal{C}$ is differentiable at $\mathbf{x}$, then the subdifferential of $\mathcal{C}$ at $\mathbf{x}$ contains exactly one element, namely the gradient of $\mathcal{C}$ at $\mathbf{x}$: $\partial\mathcal{C}(\mathbf{x}) = \{\nabla\mathcal{C}(\mathbf{x})\}$.*

3.4.4   A general optimality criterion

From the definition of a subgradient, one can easily show that $\mathbf{x}^*$ is a minimizer of $\mathcal{C}$ over $\mathbb{R}^N$ if and only if $\mathbf{0} \in \partial\mathcal{C}(\mathbf{x}^*)$. More generally, the notion of subgradient allows us to formulate an optimality criterion for *constrained* optimization problems.

Property 3. *Let $\mathcal{C} : \mathbb{R}^N \to \mathbb{R}$ be a convex function and let Q be a convex subset of $\mathbb{R}^N$. $\mathbf{x}^*$ is a minimizer of $\mathcal{C}$ over Q if and only if*

$$\exists\ \mathbf{g}^* \in \partial\mathcal{C}(\mathbf{x}^*) \mid \forall\ \mathbf{x} \in Q,\ \langle \mathbf{g}^*, \mathbf{x} - \mathbf{x}^* \rangle \ge 0. \tag{3.5}$$

---

[5]In particular it is a closed and convex subset of $\mathbb{R}^N$.

In particular, if $Q = \mathbb{R}^N$, this condition is equivalent to

$$\mathbf{0} \in \partial\mathcal{C}(\mathbf{x}^*). \tag{3.6}$$

Property 3 gives the basis for deriving the algorithms presented in the next section. In general, either the optimality criterion can be solved directly, or it is used for designing a *fixed-point iteration*. In particular, given a a strictly positive parameter τ, we can rewrite (3.5) as

$$\exists\, \mathbf{g}^* \in \partial\mathcal{C}(\mathbf{x}^*) \mid \forall\, \mathbf{x} \in Q, \ \langle \mathbf{x}^* - (\mathbf{x}^* - \tau\mathbf{g}^*), \mathbf{x}^* - \mathbf{x} \rangle \leq 0,$$

In view of (3.4), the optimality criterion can thus be interpreted as follows[6]: $\mathbf{x}^*$ minimizes $\mathcal{C}(\mathbf{x})$ over Q if and only if

$$\exists\, \mathbf{g}^* \in \partial\mathcal{C}(\mathbf{x}^*) \mid \mathbf{x}^* = \mathcal{P}_Q(\mathbf{x}^* - \tau\mathbf{g}^*). \tag{3.7}$$

3.5 Standard deconvolution methods

3.5.1 Direct methods

When $Q = \mathbb{R}^N$ and the functional is differentiable over $\mathbb{R}^N$, Property 3 reduces to the following well-known criterion: $\mathbf{x}$ minimizes the cost functional over $\mathbb{R}^N$ if and only if its gradient vanishes at $\mathbf{x}$, i.e.,

$$\nabla\mathcal{C}(\mathbf{x}) = \mathbf{0}.$$

For the cases considered below, this equation can be solved directly.

Inverse filtering

The simplest approach to deconvolution is the *least squares* method. The idea is that the "reblurred estimate" $\mathbf{Hx}$ should be as close as possible to the measurement (at least in the noiseless case). This can be enforced using a quadratic data term alone:

$$\mathcal{C}(\mathbf{x}) = \|\mathbf{y} - \mathbf{Hx}\|_2^2. \tag{3.8}$$

The gradient is given (up to a multiplicative constant) by

$$\nabla\mathcal{C}(\mathbf{x}) \propto \mathbf{H}^T\mathbf{Hx} - \mathbf{H}^T\mathbf{y}. \tag{3.9}$$

---

[6]Provided that the convex set Q is closed.

Note that if $\mathbf{H}$ has a non-trivial kernel, then $\mathcal{C}(\mathbf{x})$ does not have a unique minimizer; usually one selects the solution of $\mathbf{H}^T\mathbf{H}\mathbf{x} = \mathbf{H}^T\mathbf{y}$ with the smallest Euclidian norm.

Since $\mathbf{H}$ is block-circulant, it is diagonalized in the Fourier (DFT) domain, which is accessible via the FFT algorithm. The numerical computation of the solution can thus be performed directly. Denoting $\hat{x}[\boldsymbol{\nu}]$, $\hat{y}[\boldsymbol{\nu}]$ and $\hat{h}[\boldsymbol{\nu}]$ the DFT coefficients of $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{H}$ respectively, this *inverse filtering* amounts to setting

$$\hat{x}[\boldsymbol{\nu}] = \begin{cases} \hat{y}[\boldsymbol{\nu}]/\hat{h}[\boldsymbol{\nu}] & \text{if } \hat{h}[\boldsymbol{\nu}] \neq 0; \\ 0 & \text{otherwise.} \end{cases}$$

In practice, for numerical stability, the condition $\hat{h}[\boldsymbol{\nu}] \neq 0$ is replaced by $|\hat{h}[\boldsymbol{\nu}]| > \epsilon$ for some positive ϵ. Unfortunately even this *stabilized inverse filtering* may considerably amplify measurement noise, resulting in spurious high-frequency oscillations (for low-pass $\mathbf{H}$).

Regularized inverse filtering

To avoid this effect the estimate should not be allowed to vary too rapidly. This is usually achieved by penalizing the energy of the "derivative" of the signal; this approach is related to *Tikhonov-Phillips regularization* [217, 176] and to the method of *ridge regression* in statistics [105]. The cost functional then takes the form

$$\mathcal{C}(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{L}\mathbf{x}\|_2^2, \tag{3.10}$$

where $\mathbf{L}$ is a block-circulant matrix corresponding to the discretization of a differential (e.g. Laplacian) operator. The gradient satisfies

$$\nabla\mathcal{C}(\mathbf{x}) \propto (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})\mathbf{x} - \mathbf{H}^T\mathbf{y}.$$

The solution

$$\mathbf{x} = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})^{-1}\mathbf{H}^T\mathbf{y} \tag{3.11}$$

can again be computed in the DFT domain according to

$$\hat{x}[\boldsymbol{\nu}] = \frac{\hat{h}[\boldsymbol{\nu}]^*\hat{y}[\boldsymbol{\nu}]}{|\hat{h}[\boldsymbol{\nu}]|^2 + \lambda|\hat{\ell}[\boldsymbol{\nu}]|^2},$$

where the DFT coefficients of $\mathbf{L}$ are denoted $\hat{\ell}[\boldsymbol{\nu}]$. This equation also describes *Wiener filtering* [248]; the regularization filter $\hat{\ell}[\boldsymbol{\nu}]$ then depends on the relative power spectral densities of the noise and of the input signal.

3.5.2 Iterative methods

When the optimization problem cannot be solved directly, the general approach is to construct a sequence of estimates $\mathbf{x}^{(k)}$ (starting from an initial guess $\mathbf{x}^{(0)}$) that converges to a minimizer of $\mathcal{C}(\mathbf{x})$. The first two methods presented here can be seen as *(projected) gradient-descent* methods, which are both simple and memory-efficient; this is particularly important in the context of multi-dimensional microscopy, where the data sets can be extremely large.

At every iteration step, the estimate is corrected in the direction of steepest decrease of $\mathcal{C}(\mathbf{x})$, that is, the direction opposite to its gradient. Additionally, one can enforce constraints at every iteration [71]. The general update equation is

$$\mathbf{x}^{(k+1)} = \mathcal{P}_Q\{\mathbf{x}^{(k)} - \tau\nabla\mathcal{C}(\mathbf{x}^{(k)})\}, \tag{3.12}$$

where τ is a positive step-size parameter and Q is a closed convex set. For example, one may use the fact that the signal of interest is known to be positive. Q is then the set of positive vectors $(\mathbb{R}^+)^N$ and $\mathcal{P}_Q$ is a "clipping" operator that sets all negative components to zero.

Note that a fixed point of (3.12) is guaranteed to be a minimizer of $\mathcal{C}(\mathbf{x})$ over Q, according to the characterization given in (3.7). The convergence of iterative algorithms will be discussed in the next chapter.

Projected Landweber algorithm

For the least squares functional (3.8) whose gradient was given in (3.9), the iteration (3.12) can be written as

$$\mathbf{x}^{(k+1)} = \mathcal{P}_Q\{\mathbf{x}^{(k)} + \tau\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(k)})\}.$$

When no constraint is used (that is, when $\mathcal{P}_Q\{\mathbf{x}\} = \mathbf{x}$), this algorithm is known as the Landweber iteration [124]. This procedure will eventually exhibit the same noise amplification problems as with inverse filtering. However, one may obtain a satisfactory compromise between deconvolution and noise amplification if the algorithm is stopped on the way. In other words, the number of iterations can act as a pseudo regularization parameter [235].

Iterative constrained Tikhonov-Miller algorithm

Similarly, for the Tikhonov-regularized functional (3.10) we obtain

$$\mathbf{x}^{(k+1)} = \mathcal{P}_Q\{\mathbf{x}^{(k)} + \tau[\mathbf{H}^T\mathbf{y} - (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L})\mathbf{x}^{(k)}]\}.$$

When the projection operator is expressed in the space domain and $\mathbf{H}^T\mathbf{H} + \lambda\mathbf{L}^T\mathbf{L}$ is a convolution matrix, this algorithm (as well as the projected Landweber algorithm above) can be efficiently implemented using two FFTs per iteration.

Note that in the fluorescence-microscopy community, the minimization of (3.10) is often performed using a non-linear conjugate-gradient (CG) algorithm. It has been argued that this procedure exhibits better convergence properties; however, it is also more memory-intensive than a simple gradient descent. There are essentially two approaches that differ in how the constraints are incorporated. The one of Van Kempen et al. (see e.g. [229]) is built upon a linear CG iteration for the optimality equation corresponding to (3.10); the constraints are applied in a heuristic manner at every iteration of the algorithm. The approach of Carrington [37] is based on the Fletcher-Reeves CG method. It uses a more rigorous dual formulation to implement the constraints, but it is only applicable when $\mathbf{L}$ is the identity operator.

Richardson-Lucy algorithm

Maximum-likelihood estimation with a Poisson noise model leads to the cost functional[7]

$$\mathcal{C}(\mathbf{x}) = \mathbf{1}^T\mathbf{H}\mathbf{x} - \mathbf{y}^T\log(\mathbf{H}\mathbf{x}).$$

Here we look for a strictly positive solution and we can thus perform the change of variable $\mathbf{x} = \exp(\mathbf{z})$. If we compute the gradient of the cost functional $\mathcal{C}'(\mathbf{z}) = \mathcal{C}\left(\exp(\mathbf{z})\right)$ and re-express it using $\mathbf{x}$, we obtain

$$\nabla\mathcal{C}'(\mathbf{z}) = \mathbf{x} \times \left(\mathbf{H}^T\left[\mathbf{1} - \frac{\mathbf{y}}{\mathbf{H}\mathbf{x}}\right]\right).$$

Here the multiplication $\times$ and the division $\mathbf{y}/\mathbf{H}\mathbf{x}$ are understood pointwise. Assuming that $\mathbf{H}$ is normalized such that $\mathbf{H}^T\mathbf{1} = \mathbf{1}$, the constraint that the gradient should vanish leads to the following fixed-point iteration:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} \times \mathbf{H}^T\left[\frac{\mathbf{y}}{\mathbf{H}\mathbf{x}^{(k)}}\right]. \tag{3.13}$$

This is the standard Richardson-Lucy algorithm [184, 138]. Its implementation requires four FFTs per iteration (two per convolution).

An accelerated version of the Richardson-Lucy algorithm is for example used in [106]. It takes the form

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \tau_k\mathbf{x}^{(k)} \times \left[\mathbf{H}^T\left(\frac{\mathbf{y}}{\mathbf{H}\mathbf{x}^{(k)}} - \mathbf{1}\right)\right],$$

---

[7]Our convention is that $\log(\mathbf{v})$ and $\exp(\mathbf{v})$ stand respectively for the componentwise application of the logarithm and exponential functions to the vector $\mathbf{v}$.

where the step sizes τ_k are optimized using a line search that takes into account the positivity constraint. When $\tau_k = 1$ and $\mathbf{H}$ is normalized such that $\mathbf{H}^T\mathbf{1} = \mathbf{1}$, one retrieves the standard algorithm.

Richardson-Lucy algorithm with total-variation regularization

The problem of the above variational formulation is the same as for least squares: it is not regularized and thus iteration (3.13) will eventually lead to significant noise amplification.

To counterbalance this effect, Conchello and Hansen [54] proposed to use Tikhonov regularization. More recently, Dey et al. [62, 63] considered total-variation regularization [186], leading to the functional

$$\mathcal{C}(\mathbf{x}) = \mathbf{1}^T\mathbf{H}\mathbf{x} - \mathbf{y}^T\log(\mathbf{H}\mathbf{x}) + \lambda\|\mathbf{x}\|_{\mathrm{TV}}.$$

Depending on the definition of $\|\cdot\|_{\mathrm{TV}}$ (see e.g. [235]), this functional may not be differentiable everywhere. Using the same change of variable and the same normalization assumption as before, a subgradient of $\mathcal{C}$ at $\mathbf{x}$ has the form

$$\mathbf{x} \times \left(\mathbf{1} - \mathbf{H}^T\left[\frac{\mathbf{y}}{\mathbf{H}\mathbf{x}}\right] + \lambda\mathbf{g}\right),$$

where $\mathbf{g}$ is a subgradient of $\|\cdot\|_{\mathrm{TV}}$ at $\mathbf{x}$.

For every estimate $\mathbf{x}^{(k)}$, the authors consider a specific choice of subgradient $\mathbf{g}^{(k)}$ and apply the update formula

$$\mathbf{x}^{(k+1)} = \frac{\mathbf{1}}{\mathbf{1} + \lambda\mathbf{g}^{(k)}} \times \mathbf{x}^{(k)} \times \mathbf{H}^T\left[\frac{\mathbf{y}}{\mathbf{H}\mathbf{x}^{(k)}}\right].$$

They also use a regularization parameter λ that is sufficiently small to maintain the positivity constraint. Note that according to (3.6) a fixed point of this iteration is guaranteed to be a minimizer, but the converse is not necessarily true for the specific choice of subgradient made by the authors.

Part II

Fast wavelet-based deconvolution

Scale is the most important element of beauty.
Esa Piironen, *On Architecture.*

Chapter 4

Design strategies for image-reconstruction algorithms: a multilevel tutorial

The main subject of the remainder of this thesis will be the design and optimization of wavelet-based image-restoration algorithms. There are two common threads in our work: exploiting the structure of the restoration problem and preserving this structure accross scales. The combination of these ideas with the properties of wavelet representations can lead to efficient multilevel procedures.

The purpose of this introductory chapter is to give an overview of the design principles that we have used, paving the way for Chapter 5 and Chapter 6. Our work originates from multigrid concepts. Although "multigrid design guides" are available in various levels of generality [98, 32, 253], one is forced to admit that there is no systematic way to derive a multigrid algorithm. The term "multigrid" refers much more to a general methodology—not to a specific algorithm. It is a framework that can require substantial adaptation to the particular problem under consideration.

Our main application of interest will be the acceleration of ℓ_1-regularized deconvolution using a blend of wavelet and multigrid concepts. Nevertheless the principles described here could be (and have been) applied to more general (linear

or even non-linear, possibly non-convex) inverse problems. Our hope is that this summary may be useful to other researchers and fellow PhD students making their first steps in the world of algorithm design.

We emphasize some key aspects that constitute the foundation of our work. However, we voluntarily take a step back implementation issues, which are addressed in detail in the next two chapters. In addition to the references mentioned there, we provide pointers to works in the areas of numerical analysis and image processing that we found helpful. It is impossible to be exhaustive, though, as the literature on multilevel methods is overwhelmingly large[1]. In fact, many of the principles discussed here date back to the early days of numerical analysis; they have been "iteratively refined" during two centuries and are often associated with prestigious names such as Gauss, Jacobi, Seidel, Schwarz or Lions.

4.1 The model problem: image reconstruction using wavelet-domain ℓ_1 regularization

4.1.1 Definition and approximation-theoretic motivation

Our starting point is the definition of a "fine-level problem" using the variational form considered in Section 3.3.1:

$$\mathcal{C}(\mathbf{x}) = \mathcal{D}(\mathbf{x}) + \lambda \mathcal{R}(\mathbf{x}). \tag{4.1}$$

In general, it is obtained from the discretization of a continuous-domain functional (see below).

Specifically, we will be interested in the minimization of

$$\mathcal{C}(\mathbf{x}) = \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}^T \mathbf{x}\|_1. \tag{4.2}$$

where ^{T} denotes transposition and $\mathbf{W}^T$ is the matrix corresponding to a discrete wavelet decomposition. For simplicity, we will assume in this chapter that $\mathbf{W}$ is orthonormal, implying that $\mathbf{W}$ is the associated reconstruction matrix (such that $\mathbf{W}\mathbf{W}^T = \mathbf{I}$).

This will be our "model problem" throughout the second part of this thesis: it will support most of our analysis and derivations. To benefit from the intuition of Fourier analysis, it is a good idea to think of $\mathbf{H}$ as a convolution matrix (which is diagonalized in the DFT domain), but our intention is to cover more general image-formation operators (cf. the considerations of Section 6.1.4).

---

[1]Often with the frustrating side-effect of a widely redundant terminology.

As an alternative to the statistical framework mentioned in Chapter 3, we can motivate the functional (4.2) using an approximation-theoretic standpoint [40, 58]. It is a finite dimensional version of the following continuous-domain problem: given a set of measurements $(y[\boldsymbol{k}])_{\boldsymbol{k}\in S}$, minimize the quadratic data discrepancy

$$\sum_{k\in S} (y[k] - (h * \chi)[k])^2$$

under the constraint that the function χ is piecewise smooth.

To assess the smoothness of a function, a relatively flexible (though involved) tool in approximation theory is the family of Besov spaces and their associated norms. This family provides three degrees of freedom for measuring the smoothness of a function: essentially, if $f \in L_p(\mathbb{R}^d)$ has a finite Besov norm $\|f\|_{s,p,q}$, then its s-th derivative is in $L_p(\mathbb{R}^d)$; the parameter q is related to finer properties. An important result in approximation theory is that the Besov norm of a function f is equivalent to a suitable norm on the wavelet coefficients of f [152, 61]. In the particular case where $p = q = 1$, and for a specific choice of the parameter s (depending on the dimension d), this norm is simply the ℓ_1 norm. Hence the ℓ_1 regularization term in (4.2).

Although we will not use it later in our case, it is worth emphasizing that the continuous-domain formulation of a problem can provide valuable insights for the design of multilevel algorithms; see e.g. [10].

4.1.2 Derivation of the fine-level thresholded Landweber algorithm

Let us briefly go back to the general cost functional (4.1). Here we will assume that the data term $\mathcal{D}$ is differentiable.

According to the criterion formulated in Section 3.4, $\mathbf{x}^*$ is a minimizer of the functional (4.2) functional if and only if

$$\mathbf{0} \in \nabla\mathcal{D}(\mathbf{x}) + \lambda\partial\mathcal{R}(\mathbf{x}), \tag{4.3}$$

where $\partial\mathcal{R}$ denotes the subgradient of $\mathcal{R}$ (cf. Section 3.4.3). In particular, for (4.2), this is equivalent to

$$
\begin{aligned}
-\nabla\mathcal{D}(\mathbf{x}) &\in& \lambda\partial\mathcal{R}(\mathbf{x}) \\
\Leftrightarrow \quad \mathbf{x} - \frac{\tau}{2}\nabla\mathcal{D}(\mathbf{x}) &\in& \mathbf{x} + \frac{\lambda\tau}{2}\partial\mathcal{R}(\mathbf{x}),
\end{aligned}
$$

where τ is a non-zero parameter that will be positive in the sequel. Although the mapping $\mathbf{x} \mapsto \mathbf{x} + \lambda\tau/2\ \partial\mathcal{R}(\mathbf{x})$ is multivalued (set-valued) at points where $\mathcal{R}(\mathbf{x})$ is non-differentiable, it is actually invertible[2].

When $\mathcal{R}(\mathbf{x}) = \|\mathbf{W}^T\mathbf{x}\|_1$, the inverse (in the orthonormal-wavelet domain) is known as the soft-thresholding function $\mathcal{T}_{\lambda\tau/2}$, which is defined e.g. in Section 6.1.3. The gradient of $\|\mathbf{y} - \mathbf{Hx}\|_2^2$ is given by $\nabla\mathcal{D}(\mathbf{x}) = -2\mathbf{H}^T(\mathbf{y} - \mathbf{Hx})$. This establishes that the minimizers of (4.2) correspond exactly to the fixed points of the mapping

$$\mathbf{u}(\mathbf{x}) = \mathbf{W}\mathcal{T}_{\lambda\tau/2}\big\{\mathbf{W}^T\big(\mathbf{x} + \tau\mathbf{H}^T[\mathbf{y} - \mathbf{Hx}]\big)\big\}, \qquad (4.4)$$

where τ acts as a step-size parameter whose choice is discussed below. In view of this characterization, a minimizer is often approximated numerically using the sequence $\mathbf{x}^{(k)}$ defined by some initial estimate $\mathbf{x}^{(0)}$ and the update rule

$$\mathbf{x}^{(k+1)} = \mathbf{u}(\mathbf{x}^{(k)}). \qquad (4.5)$$

Algorithms that are derived using decompositions of the form (4.3) are known under the general name of *operator-splitting* algorithms. They have been studied, among others, by Passty [171], Lions and Mercier [132] as well as Eckstein [70].

The above algorithm can also be derived using an alternative functional interpretation [40, 58]; this is the approach we use in Chapter 5 and Chapter 6. It is related to the concept of *proximal operator* [49].

4.1.3   Convergence considerations

We will now review some basic facts related to the convergence of iterative procedures of the form (4.5). More precisely, we are interested in the decay properties of $\mathbf{x}^* - \mathbf{x}^{(k)}$, the error with respect to a fixed point $\mathbf{x}^*$ of $\mathbf{u}$. While explicit decay-rate estimates can only be obtained in particular cases, these examples will motivate most of our subsequent developments.

We start with the notion of Lipschitz continuity. As an example, one can verify that the pointwise soft-thresholding operator $\mathcal{T}_{\lambda\tau/2}$ is 1-Lipschitz over $\mathbb{R}^N$.

Definition 3. *Let L be a positive constant and let Ω be a subset of $\mathbb{R}^N$. A mapping $\mathbf{u} : \mathbb{R}^N \to \mathbb{R}^N$ is said to be L-Lipschitz over Ω if, for every $\mathbf{x}, \mathbf{x}' \in \Omega$,*

$$\|\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}')\| \le L\|\mathbf{x} - \mathbf{x}'\|. \qquad (4.6)$$

---

[2]This is a general result for so-called maximal monotone operators [70].

A mapping for which (4.6) holds with $L < 1$ is a *contraction*. A fundamental result for iterations of the form (4.5) is the so-called *contraction-mapping principle*[3] [169, 220].

Property 4. *Assume that* $\mathbf{u}$ *maps* Ω *into itself and that it is* L-*Lipschitz over* Ω, *with* $L < 1$. *Then, in* Ω, $\mathbf{u}$ *has a unique fixed point* $\mathbf{x}^*$ *and the sequence defined by (4.5) converges to* $\mathbf{x}^*$ *for any initial estimate* $\mathbf{x}^{(0)} \in \Omega$. *Moreover one has the error estimate*

$$\|\mathbf{x}^* - \mathbf{x}^{(k)}\| \leq CL^k, \tag{4.7}$$

where $C = \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|/(1 - L)$.

Algorithms for which such a bound holds are said to converge geometrically or linearly (because the logarithm of the error decays linearly); the tightest possible constant L is called the *convergence rate*. While this quantity only describes the asymptotic behavior of the error, it is a common measure of convergence speed [98, 32].

As a simple example, consider the case where $\lambda = 0$ in (4.4); one then obtains the classical Landweber update formula, which is linear. In particular, for any $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^N$,

$$\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}') = (\mathbf{I} - \tau \mathbf{H}^T \mathbf{H})(\mathbf{x} - \mathbf{x}'). \tag{4.8}$$

Via a diagonalization of $\mathbf{I} - \tau \mathbf{H}^T \mathbf{H}$, this shows that the linear Landweber iteration is L-Lipschitz over $\mathbb{R}^N$, where $L = \lambda_{\max}(\mathbf{I} - \tau \mathbf{H}^T \mathbf{H})$ is the largest eigenvalue[4] of $\mathbf{I} - \tau \mathbf{H}^T \mathbf{H}$ (in absolute value).

In particular, when $\lambda_{\min}(\mathbf{H}^T \mathbf{H}) > 0$, a standard analysis shows that the linear Landweber iteration is a contraction mappig ($L < 1$) for any value of τ in the open interval $]0, 2/\lambda_{\max}(\mathbf{H}^T \mathbf{H})[$. Unfortunately this does not leave a lot of freedom when $\mathbf{H}^T \mathbf{H}$ is poorly conditioned; that is, when the ratio $\lambda_{\min}(\mathbf{H}^T \mathbf{H})/\lambda_{\max}(\mathbf{H}^T \mathbf{H})$ is close to zero. Indeed the eigenvalue $1 - \tau \lambda_{\min}(\mathbf{H}^T \mathbf{H})$ of $\mathbf{I} - \tau \mathbf{H}^T \mathbf{H}$ is then always close to 1, even if one could set τ to the upper bound $2/\lambda_{\max}(\mathbf{H}^T \mathbf{H})$. In terms of the error $\mathbf{x}^* - \mathbf{x}^{(k)}$, this means that the components spanned by eigenvectors corresponding to small eigenvalues of $\mathbf{H}^T \mathbf{H}$ will always decay slowly; indeed, from (4.8),

$$\mathbf{x}^* - \mathbf{x}^{(k+1)} = (\mathbf{I} - \tau \mathbf{H}^T \mathbf{H})(\mathbf{x}^* - \mathbf{x}^{(k)}).$$

The general thresholded Landweber algorithm has the same Lipschitz regularity over $\mathbb{R}^N$ as the linear Landweber iteration. This essentially follows from the fact

---

[3] Also known as the Banach fixed-point theorem.
[4] Hereafter we use $\lambda_{\min}$ to refer to the smallest eigenvalue; these notations should not be confused with the regularization parameter λ.

that the soft-thresholding operation is 1-Lipschitz (and from our assumption that $\mathbf{W}$ is orthonormal).

$$
\begin{aligned}
\|\mathbf{u}(\mathbf{x}) - \mathbf{u}(\mathbf{x}')\| &= \left\|\mathcal{T}_{\lambda\tau/2}\left\{\mathbf{W}^T\left(\mathbf{x} + \tau\mathbf{H}^T[\mathbf{y} - \mathbf{Hx}]\right)\right\}\right. \\
&\quad \left. - \mathcal{T}_{\lambda\tau/2}\left\{\mathbf{W}^T\left(\mathbf{x}' + \tau\mathbf{H}^T[\mathbf{y} - \mathbf{Hx}']\right)\right\}\right\| \\
&\leq \left\|\mathbf{W}^T(\mathbf{I} - \tau\mathbf{H}^T\mathbf{H})(\mathbf{x} - \mathbf{x}')\right\| \\
&= \left\|(\mathbf{I} - \tau\mathbf{H}^T\mathbf{H})(\mathbf{x} - \mathbf{x}')\right\| \\
&\leq \lambda_{\max}(\mathbf{I} - \tau\mathbf{H}^T\mathbf{H})\left\|\mathbf{x} - \mathbf{x}'\right\| \qquad (4.9)
\end{aligned}
$$

However, for many mappings of interest it can be difficult or even impossible to establish a strict contraction property. To prove convergence, one may then have to resort to weaker forms of contractions, e.g., iterated contractions [169] or specific classes of nonexpansive mappings [231]. Another approach is to prove the contraction property only over a subdomain Ω of $\mathbb{R}^N$ (see e.g. [99]). However, obtaining an explicit and tractable characterization of a suitable subdomain can be difficult. As a result, Lipschitz constants that are derived analytically are often too conservative because they apply to an overly large domain.

For example, our numerical experiments (see column 1 of Table 6.1) suggest that the convergence rate of the thresholded Landweber algorithm (for $\lambda > 0$) is systematically better than the convergence rate of the corresponding linear Landweber algorithm (for $\lambda = 0$). Note that we are probably the first to provide a quantitative experimental demonstration of this property. Our results support the idea that the estimate (4.9) is usually too pessimistic. The thresholding operation seems to compensate for the poor conditioning of $\mathbf{H}^T\mathbf{H}$: for higher thresholds, the convergence becomes faster.

Another illustration can be given when $\lambda_{\min}(\mathbf{H}^T\mathbf{H}) = 0$. In this case the thresholded Landweber algorithm is at best 1-Lipschitz over $\mathbb{R}^N$, but several theoretical investigations support the fact that it can still converge linearly [43, 99, 30]. Our numerical experiments tend to confirm this prediction (at least for a sufficiently large regularization parameter).

4.2 Derivation of a multilevel method

4.2.1 The subspace-correction principle

The previous analysis suggests that certain components of the estimate $\mathbf{x}^{(k)}$ are subject to slower convergence than others (at least in the linear case). Thus we

would like to design an algorithm that acts more effectively on these components. This leads almost naturally to the principle of *subspace correction*, which was introduced by Xu [252] as a general framework for the numerical resolution of linear systems of equations. This principle was recently transposed by Fornasier [80] to more general non-linear problems defined in a variational setting, with emphasis on ℓ_1 regularization.

The idea is to concentrate each update on a particular subspace E_0 of $\mathbb{R}^N$, whose choice essentially depends on the nature of the aforementioned components. In other words, we introduce a new update mapping $\mathbf{u}_0$ that performs a correction only in the subspace E_0: for any given estimate $\mathbf{x}$, $\mathbf{u}_0(\mathbf{x}) = \mathbf{x} + \mathbf{e}_0$ for some $\mathbf{e}_0 \in E_0$. Some ideas for deriving such a mapping will be discussed in the next subsection.

Potentially, this strategy has two advantages. The clearest advantage, which was put forward in [80], is *dimension reduction*: by saving computation time from components that are "easy" to recover, one can spend more iterations on "hard" ones. An additional advantage is *better conditioning*: the efficiency of each individual update can be improved by adapting to the components. This is an implicit though central aspect of our work in Chapter 5 and Chapter 6.

Of course, this strategy must be applied to a collection of subspaces $E_0, E_1, \ldots, E_J$ that span the entire signal space, i.e.,

$$\mathbb{R}^N = E_0 + E_1 + \ldots + E_J.$$

In the sequel, we will denote by $\mathbf{u}_j$ the corresponding update mappings. There are essentially two ways to combine these mappings. The first one consists in a *parallel* update strategy, of the form

$$\mathbf{u}_{\mathrm{Par}}(\mathbf{x}) = c_0 \mathbf{u}_0(\mathbf{x}) + c_2 \mathbf{u}_2(\mathbf{x}) + \ldots + c_J \mathbf{u}_J(\mathbf{x}),$$

where the coefficients $c_0, c_1, \ldots, c_J$ define a convex linear combination[5]. The second one consists in a *sequential* update strategy, of the form

$$\mathbf{u}_{\mathrm{Seq}}(\mathbf{x}) = \big(\mathbf{u}_0 \circ \mathbf{u}_1 \circ \ldots \circ \mathbf{u}_J\big)(\mathbf{x}),$$

where $\circ$ denotes composition. Virtually any iterative update procedure can be defined using one of these two forms. Alternatively to parallel/sequential, the terms simultaneous/successive and additive/multiplicative are also widely used[6]. This distinction can be traced back to the "historical" concepts of Jacobi and Gauss-Seidel iterations.

---

[5]I.e., they are positive and such that their sum is equal to 1.

[6]Note that we have already used the word "multiplicative" in another sense for the Richardon-Lucy iteration.

While any sequential algorithm has also a parallel counterpart[7] (and vice versa), their convergence properties usually differ. In practice, sequential update methods are often found to be more efficient than parallel methods. Therefore it is not surprising that sequential methods have been derived in many different application fields of numerical analysis, leading to a prolific terminology. Important examples for the resolution of partial differential equations are the Schwarz alternating method [196, 131]—a precursor of domain-decomposition methods [41]—and multigrid algorithms [98, 32].

Sequential techniques have also played an important role for image restoration, or more broadly image reconstruction. Historically, many of them were initially developed for medical imaging applications. Let us mention just a few examples in the realm of linear reconstruction, to illustrate the above concepts (see also [38, 127]).

- A so-called coordinate-update strategy uses the one-dimensional subspaces corresponding to the individual components of the vector $\mathbf{x}$.

- The method of Ordered Subsets (OS [109]) uses subspaces corresponding to groups of components (block coordinate-update strategy).

- The Algebraic Reconstruction Technique (ART [92], also known as the Kaczmarz algorithm [115]) uses the subspaces generated by the individual columns of the "image-formation matrix" $\mathbf{H}$.

- A modification of this technique, called simultaneous algebraic reconstruction (SART [5]) uses the subspaces spanned by groups of columns.

Note that there are also stochastics variants of these methods where the subspaces are chosen randomly; this approach is still the subject of investigations (see e.g. [211] and references therein).

Finally, another well-known sequential subspace-correction method is the one of alternating projections [254], a particular case of the method of projections onto convex sets (POCS, [255]). In the context of wavelet-regularized signal restoration, an early example of a sequential optimization strategy can be found in the work of Sardy et al. on denoising [190]; drawing a parallel with projective methods, theirs could be termed an "alternating thresholding" algorithm.

---

[7]For example, the parallel version of the algebraic reconstruction technique mentioned below is called simultaneous iterative reconstruction technique (SIRT [90])—it essentially amounts to a Landweber iteration.

4.2.2 The auxiliary-problem principle

There are potentially many ways to construct the update mappings $\mathbf{u}_j$ introduced above. Perhaps the most direct approach is to project (4.3)—the characterization of the solutions to the optimization problem—onto the different subspaces E_j. Another approach is to focus on the original cost functional (4.1). We discuss the latter approach in more details because we will use it in the next two chapters. Also, it is often prefered by algorithm designers, because it is a relatively intuitive and flexible framework that leads easily to procedures with certain monotonicity properties.

From the standpoint of the original functional $\mathcal{C}$, the mapping $\mathbf{u}_j$ should provide the largest-possible decrease. In other words, we would like that $\mathbf{u}(\mathbf{x}) = \mathbf{x} + \mathbf{e}_j$, where $\mathbf{e}_j$ minimizes the functional $\mathbf{e} \mapsto \mathcal{C}(\mathbf{x}+\mathbf{e})$ over E_j. However, this minimization can rarely be performed exactly in practice. One approach for circumventing this issue is to approximate $\mathcal{C}(\mathbf{x} + \mathbf{e})$ by an auxiliary functional $\mathcal{A}_j^{\mathbf{x}}(\mathbf{e})$ that is easy to minimize. If $\mathcal{A}_j^{\mathbf{x}}(\mathbf{e})$ has a unique minimizer (typically when it is strictly convex) this uniquely defines the update mapping

$$\mathbf{u}_j(\mathbf{x}) = \mathbf{x} + \arg\min_{\mathbf{e}\in E_j} \mathcal{A}_j^{\mathbf{x}}(\mathbf{e}).$$

There are a number of techniques for deriving such functionals. In fact, generic methods from the field of optimization—such as dual formulations or barrier functions for constrained problems—can be considered part of them. The distinction between methods is often just a matter of terminology. Here we provide references for two classes of techniques that have been proposed or used in the context of image reconstruction. Again, medical imaging had an important role in the application and diffusion of these methods in image processing.

The first class of methods that we would like to mention is known under the name of "half-quadratic regularization" [87, 88]. The essential idea is to define the auxiliary functional via an additional set of variables:

$$\mathcal{A}_j^{\mathbf{x}}(\mathbf{e}) = \min_{\mathbf{e}'} \mathcal{F}(\mathbf{e}, \mathbf{e}').$$

The functional $\mathcal{F}$ is constructed in such a way that it is easily minimized with respect to $\mathbf{e}$ and $\mathbf{e}'$, taken individually. This is typically achieved thanks to a quadratic coupling term—hence the name "half-quadratic". The minimization can then be performed alternatively on both sets of variables. We refer to [245] for an overview of this method and to [163] for a detailed theoretical investigation.

The second class is refered to as "bound optimization principle", "surrogate functional optimization", or "majorize-minimize strategy" (among other denomi-

nations). The characteristic property of this class is that the auxiliary functional is an upper-bound of the original cost functional, with equality at the origin:

$$\forall \ \mathbf{e}_j \in E_j, \ \mathcal{C}(\mathbf{x} + \mathbf{e}) \leq \mathcal{A}_j^{\mathbf{x}}(\mathbf{e}) \qquad \text{and} \qquad \mathcal{C}(\mathbf{x}) = \mathcal{A}_j^{\mathbf{x}}(\mathbf{0}).$$

This framework leads to an update mapping $\mathbf{u}_j$ that has the nice property of monotonically decreasing the original cost functional. It also provides a simple interpretation for many expectation-maximization algorithms. We refer to [125, 111] for tutorial introductions and to [113] for a recent theoretical treatment in the context of image processing.

Note that the idea of splitting the cost functional into several auxiliary functionals has been formalized in numerical analysis quite some time ago—see in particular [47], which provides a very general analysis of bound-optimization techniques.

4.2.3   The scale-invariance principle

Among all multilevel schemes, those that can be formulated recursively are particurly elegant. They essentially amount to *two-level* methods, which involve only two subspaces E_1 and E_2.

Such schemes are based on the important principle of *scale-invariance*: for at least one of the subspaces, say E_2, the auxiliary functional $\mathcal{A}_2^{\mathbf{x}}$ should have the same structure as the original cost functional $\mathcal{C}$. In this case, one can apply the same type of decomposition recursively on E_2. Importantly, this allows for using the same type of update mapping—and thus a unified "tuning strategy"—at all levels.

The combination of all principles discussed so far could be summarized as a "recursive divide-and-conquer strategy".

Table 4.1 shows the subspace structure of some standard methods for the resolution of partial differential equations. Domain decomposition methods typically involve subspaces that are related to boundary conditions; in general, they are not recursive in nature. In contrast, multigrid algorithms use a hierarchy of embedded subspaces corresponding to discretizations of the equation at different resolution levels; this is an ideal playground for recursivity, allowing for a variety of standard iteration schemes (see Fig. 6.3).

The resolution of partial differential equations typically involves the inversion of high-pass operators. Thus the subspaces used in multigrid methods concentrate on low-frequency components, which are the most difficult to recover in that case. However, in deconvolution problems, the blurring operators are typically low-pass. Thanks to their spectral localization properties, wavelet subspaces (which are the

Method	Structure of the subspaces	Remarks
Domain decomposition	$E_1, \ldots, E_J$	No embedding in general (possibly "overlapping")
Multigrid	$E_0 \supset E_1 \supset \ldots \supset E_J$	Pyramidal embedding (dimension reduction)
Hierarchical-basis/Wavelet-type multilevel algorithm	$\begin{array}{ccccc} E_0 & \supset \ldots \supset & E_{J-1} & \supset & E_J \\ \cup & & \cup & & \parallel \\ E_0' & \ldots & E_J' & & E_J' \end{array}$	Comb-tree embedding in pyramidal structure
Wavelet-packet multilevel algorithm	Embedding of subspaces defined by an arbitrary tree	

Table 4.1: Comparison of subspace-correction methods.

complements of the traditional coarse-approximation subspaces) are better adapted to this situation.

This is best understood with the bandlimited Shannon wavelets, for which the different subspaces are completely decoupled with respect to a convolution operator (see Fig. 5.4). The convergence rate of the algorithm described in Chapter 5 (over $\mathbb{R}^N$) is provably better[8] than the one of the standard thresholded Landweber algorithm (4.4). Note that thanks to the decoupling property of Shannon wavelets, the algorithm of Chapter 5 is actually a parallel algorithm.

In the case of other wavelet families, the coupling between subspaces remains sufficiently weak to allow for similar convergence-rate improvements. This is shown by the numerical results of Table 6.1. The general multilevel algorithm of Chapter 6 is sequential.

The subspace structure of our multilevel algorithm is comparable to the one used in the hierarchical-basis method [13]. A possible generalization consists in using an arbitrary hierarchy of subspaces (see Table 4.1). In fact, such a structure is used implicitly in multigrid schemes with a locally-adaptive grid resolution. It is reminiscent of wavelet-packet decompositions.

---

[8]At least when $\mathbf{H}^T\mathbf{H}$ is positive definite.

4.3   Summary

4.3.1   A tentative characterization of multilevel algorithms

As argued at the beginning of the chapter, it is impossible to give a systematic definition of a multilevel algorithm. In general, however, we can say that the "multilevel philosophy" is characterized by one or several of the following principles:

- dividing a given problem into several subproblems;

- using a pyramidal (multiresolution) representation;

- relying on a recursive—and thus sequential—formulation/implementation.

But most importantly, it consists in adapting to the structure of the problem at hand.

4.3.2   A step-by-step design guide

For the algorithm designer, we provide a list of steps that could be followed for deriving a multilevel image-reconstruction method.

1. Provide a clear mathematical statement of the problem to be solved (in our case finding a minimizer of a specified cost functional). Think about possible discretizations/formulations at different resolution levels.

2. Identify a fine-level iterative procedure that solves the problem (in our case the thresholded Landweber algorithm).

3. Understand its convergence properties theoretically and/or numerically, so as to be able to fine-tune its parameters. Gain intuition from extreme situations (in our case linear deconvolution and bandlimited wavelets).

4. Choose a pyramidal subspace decomposition that is adapted to the convergence properties of the fine-level algorithm and that preserves the structure of the problem at different levels (in our case the wavelet subspaces).

5. Define a performance measure and stick to it (in our case the convergence rate). Use synthetic experiments for which solutions are known in advance in order to assess the performance of the multilevel method.

Chapter 5

The case of bandlimited wavelets

Summary

We present a fast variational deconvolution algorithm that minimizes a quadratic data term subject to a regularization on the ℓ^1-norm of the wavelet coefficients of the solution.

Previously available methods have essentially consisted in alternating between a Landweber iteration and a wavelet-domain soft-thresholding operation. While having the advantage of simplicity, they are known to converge slowly.

By expressing the cost functional in a Shannon wavelet basis, we are able to decompose the problem into a series of subband-dependent minimizations. In particular, this allows for larger (subband-dependent) step sizes and threshold levels than the previous method. This improves the convergence properties of the algorithm significantly.

We demonstrate a speed-up of one order of magnitude in practical situations. This makes wavelet-regularized deconvolution more widely accessible, even for applications with a strong limitation on computational complexity. We present promising results in 3D deconvolution microscopy, where the size of typical data sets does not permit more than a few tens of iterations.

5.1   Introduction

5.1.1   Motivation

During the past decade, biological imaging has been revolutionized by the widespread availability of novel fluorescence labeling techniques and the development of advanced 3D optical microscopy (see Chapter 2). As biological research moves towards molecular scale, the constraints in terms of resolution and light efficiency are becoming more and more stringent. In this context, scientists are increasingly relying on computational methods to take full advantage of their instrumentation [96, 161, 101, 20].

The use of deconvolution to enhance micrographs is commonly referred to as *deconvolution microscopy* [148, 242]. The results are the most striking in 3D, as demonstrated by the pioneering work of Agard and Sedat [1], which had a significant impact on the field. The main challenge in 3D resides in the size of typical data sets, which severely restricts the computational complexity of numerical restoration methods. With the present state of computer hardware, one cannot afford more than the equivalent of a few tens of FFTs on a complete 3D stack. This explains why most software packages for deconvolution microscopy are still based on relatively standard methods, at least when compared to the state-of-the-art in 2D imaging [12, 188]. For instance, the Tikhonov-Miller and Richardson-Lucy algorithms are widely used; they require 2 (respectively 4) FFTs per iteration, so that they are usually stopped after 30 to 40 iterations.

Wavelet regularization is a relatively recent technique in the area of deconvolution. Several research groups have independently proposed a restoration procedure that alternates between wavelet-coefficient thresholding and a Landweber update. Some of the ideas of the method, which we shall call the Thresholded Landweber (TL) algorithm, can be traced back to the work of Starck et al. [206, 205]. Figueiredo and Nowak presented the first formal derivation of the TL algorithm using a statistically-based Penalized Likelihood formulation [78]. The algorithm was then rederived in a deterministic variational framework and generalized to a more general class of linear operators by Daubechies et al. [58]; note that these authors also provided a general convergence proof that can be transposed to the algorithm that we present here. Bect et al. considered a similar wavelet-regularized cost function, but with an additional Total Variation term [16]. Most recently, Chaux et al. extended this type of approach to frame-based wavelet regularization [42].

The TL algorithm offers attractive features for 3D deconvolution microscopy. First, it belongs to the class of non-linear wavelet-based methods, which have al-

ready proven to be very efficient for image-restoration applications such as denoising. In a series of 2D deconvolution experiments (see [78]), the TL algorithm was actually shown to perform best against several other wavelet-based deconvolution methods. Second, the TL algorithm is simple to implement. Since it is iterative (unlike other wavelet-based methods, e.g. the ones described in [65, 116, 162]), additional constraints such as positivity can be easily incorporated. Finally, it requires only two FFTs and two wavelet transforms per iteration. This represents an intermediate computational complexity compared to the standard algorithms mentioned before.

The only weak point of the TL algorithm is its slow convergence speed. This is especially true at high SNR levels, which may require several hundred iterations. The primary purpose here is to propose a variation of the TL algorithm that minimizes the same functional but that is substantially faster computationally. Our efforts are directed towards making wavelet regularization a tractable alternative for deconvolving large data sets. In particular, we demonstrate its feasibility on a 3D stack of widefield fluorescence micrographs. To our knowledge, this is the first practical implementation of wavelet-based 3D deconvolution microscopy, and the perspectives in this area appear to be quite promising.

5.1.2   Image-formation model

In this chapter, we are concerned with the recovery of a multidimensional signal from its noisy observation through a convolutive imaging system.

Fig. 5.1 shows a model of the image-acquisiton chain that applies, for example, to fluorescence microscopy. Generally speaking, we are interested in d-dimensional signals. The observed image is the d-dimensional convolution of an original image (the characteristic function of the object of interest) with the impulse response (or point spread function, PSF) of the imaging system. In practice, we only have access to a finite number of samples of the image and these measurements are corrupted by noise.
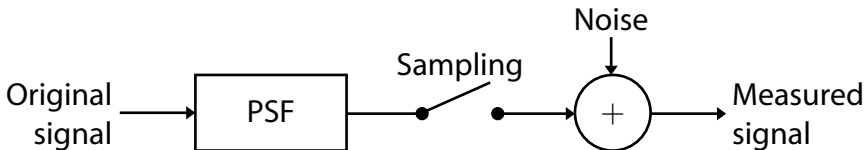
Figure 5.1: The image-formation and measurement model.

In the sequel, to simplify the presentation, we will adopt a purely discrete description of the image-formation process. We shall assume that

$$\mathbf{y} = \mathbf{H}\mathbf{x}_{\mathrm{orig}} + \mathbf{b},$$

where $\mathbf{x}_{\mathrm{orig}}$ and $\mathbf{y}$ are vectors containing uniform samples of the original and measured signals, respectively; $\mathbf{b}$ represents the measurement error. These vectors have $N = N_1 \times N_2 \times \ldots \times N_d$ components, where N_k stands for the number of samples along dimension k. $\mathbf{H}$ is a square, (block-) circulant matrix that approximates the convolution with the PSF in Fig. 5.1.

5.1.3   Deconvolution as a variational problem

The recovery of $\mathbf{x}_{\mathrm{orig}}$ from $\mathbf{y}$ is an *ill-posed problem* [17]. To obtain an estimate with reasonable accuracy and robustness to noise, one must often rely on prior information about the original signal. The estimation of $\mathbf{x}_{\mathrm{orig}}$ is then formulated as an optimization problem which incorporates this prior information. The quality of a given estimate $\mathbf{x}$ is typically measured by a *cost function* of the form

$$\mathcal{C}(\mathbf{x}) = \mathcal{D}(\mathbf{x}) + \lambda\mathcal{R}(\mathbf{x}). \tag{5.1}$$

The *data term* $\mathcal{D}(\mathbf{x})$ quantifies the "prediction error" with respect to the measurements. The *regularization term* $\mathcal{R}(\mathbf{x})$ is designed to penalize an estimate that would not exhibit the expected properties. The *regularization parameter* λ balances the contribution of both terms.   In practice, the value of λ needs to be adjusted appropriately; this is either done empirically by trial and error or by using some data-driven cross-validation method [17, 235]. The problem of deconvolution is then to find an estimate that minimizes the cost function $\mathcal{C}(\mathbf{x})$.   This functional can also interpreted as a (negative) log-likelihood in a Bayesian statistical framework, and deconvolution can then be seen as a maximum a posteriori (MAP) estimation problem [17, 78].

In the present work, we will consider cost functions that favor wavelet expansions with a small number of nonzero coefficients. This is typical of objects that are piecewise smooth [142]. The property of natural images having a sparse wavelet expansion is heavily exploited in compression standards such as JPEG2000 [45]. Here, we will concentrate on the case of an orthonormal wavelet decomposition, which will be characterized by an (orthonormal) decomposition matrix $\mathbf{W}$.   In our notation, the matrix-vector product $\mathbf{w}_{\mathbf{x}} = \mathbf{W}\mathbf{x}$ yields the coefficients of $\mathbf{x}$ in the wavelet basis, and $\mathbf{W}^T\mathbf{W}\mathbf{x}$ reconstructs the signal from these coefficients. We

implicitly include the (coarsest-scale) scaling functions when referring to the wavelet basis, unless specified otherwise. The cost function is then defined by

$$\mathcal{D}(\mathbf{x}) = \|\mathbf{y} - \mathbf{Hx}\|_2^2 \qquad \text{and} \qquad \mathcal{R}(\mathbf{x}) = \|\mathbf{Wx}\|_1. \tag{5.2}$$

Here the data term measures the residual in the image-domain using the (squared) Euclidian norm $\| \cdot \|_2$. The regularization term is the sum of the absolute values of the wavelet coefficients, which is denoted with the ℓ_1-norm $\|\mathbf{w_x}\|_1$. In contrast with traditional quadratic regularization, the ℓ_1-norm leads to a non-linear deconvolution algorithm. Compared to the ℓ_2 norm, it puts less weight on coefficients that are greater than 1, and more weight on coefficients that are smaller than 1. This tends to favor signals that are "sparse" solutions with a few large wavelet coefficients.

5.1.4 Organization of the chapter

The remainder of the chapter is organized as follows. In Section 5.2, we present the basic building blocks of the algorithm and relate them to classical image processing tasks. In Section 5.3, we recall the formalism of Daubechies et al. leading to the existing thresholded Landweber algorithm. In Section 5.4, we derive a faster algorithm that minimizes the same functional in the case of Shannon wavelets. The last section is devoted to numerical experiments that illustrate the performance of the proposed algorithm, on both simulated data and real 3D microscopy data.

5.2 Preliminaries

There are two particular cases of (5.1) and (5.2) that provide the basic building blocks of the thresholded Landweber algorithm.

5.2.1 The non-regularized case

The first one corresponds to $\lambda = 0$. The cost function then reduces to

$$\mathcal{C}(\mathbf{x}) = \|\mathbf{y} - \mathbf{Hx}\|_2^2. \tag{5.3}$$

A minimizer of this expression is the classical *least squares* estimator, which is given by $\mathbf{x} = \mathbf{H}^\dagger \mathbf{y}$, where[1]

$$\mathbf{H}^\dagger = \lim_{c \to 0, c > 0} (\mathbf{H}^T \mathbf{H} + c\mathbf{I})^{-1} \mathbf{H}^T$$

---

[1] In the case of complex vectors/matrices, ^{T} will denote Hermitian transposition.

is the Moore-Penrose pseudoinverse of $\mathbf{H}$. It is the solution with the smallest (Euclidian) norm. In practice, $\mathbf{H}^\dagger \mathbf{y}$ can be simply computed in the frequency domain (i.e. in the DFT basis), where $\mathbf{H}^\dagger$ is diagonalized.

However, this direct method is prone to noise amplification because the matrix $\mathbf{H}^T \mathbf{H}$ may have small but non-zero eigenvalues. In order to delay this amplification process, the minimization of (5.3) is often performed iteratively. Starting from an initial estimate $\mathbf{x}^{(0)}$, one constructs a sequence $(\mathbf{x}^{(n)})_{n \in \mathbb{N}}$ that converges to a minimizer of $\mathcal{C}(\mathbf{x})$. A gradient descent on (5.3) with a fixed step size τ yields the following update formula:

$$\mathbf{x}^{(n+1)} = \mathbf{x}^{(n)} + \tau \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}). \tag{5.4}$$

This algorithm is usually named after *Landweber* [124]. For an in-depth investigation of this very basic procedure—the grandfather of iterative deconvolution—we refer to [17].

5.2.2   The pure denoising case

The second particular case occurs when $\mathbf{H} = \mathbf{I}$. The cost function becomes

$$\mathcal{C}(\mathbf{x}) = \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1. \tag{5.5}$$

This corresponds to a pure denoising problem, where we want to recover the original signal from the non-blurred but noisy measurement $\mathbf{y}$ (see [40] for a complete mathematical treatment). When the wavelet basis is orthonormal—as in our case— the minimization of this expression is straightforward, since the quadratic term can be written in the wavelet domain (Parseval relation):

$$\mathcal{C}(\mathbf{x}) = \|\mathbf{W}\mathbf{y} - \mathbf{W}\mathbf{x}\|_2^2 + \lambda \|\mathbf{W}\mathbf{x}\|_1 = \sum_{n=1}^{N} |w_{\mathbf{y}\,n} - w_{\mathbf{x}\,n}|^2 + \lambda |w_{\mathbf{x}\,n}|, \tag{5.6}$$

where $w_{\mathbf{y}\,n}$ and $w_{\mathbf{x}\,n}$ are the wavelet coefficients of $\mathbf{y}$ and $\mathbf{x}$, respectively. In this form the cost function is completely decoupled, which implies that its minimization can be achieved in a coefficient-wise fashion. The solution involves the *soft-thresholding* function

$$\mathcal{T}_{\lambda/2}(w) = \mathrm{sgn}(w)\,(|w| - \lambda/2)_+\,, \tag{5.7}$$

where $(\cdot)_+$ is the positive-part function:

$$(t)_+ = \begin{cases} t & \text{if } t > 0; \\ 0 & \text{otherwise.} \end{cases}$$

The minimizer of (5.5) is simply given by

$$\mathbf{x} = \mathbf{W}^T \mathcal{T}_{\lambda/2}\{\mathbf{W}\mathbf{y}\},$$

where $\mathcal{T}_{\lambda/2}\{\cdot\}$ denotes a component-wise application of the soft-thresholding function. Note that this formalism can be adapted to complex wavelet decompositions by replacing $\text{sgn}(w)$ by $\exp(i \arg w)$ in (5.7). Interestingly, this algorithm was first proposed empirically for noise reduction in magnetic resonance imaging [246]. It has become very popular in the field, following the impulsion of Donoho and Johnstone who justified it on solid statistical grounds [67]. Several authors have also proposed a Bayesian interpretation of the method, see e.g. [159, 201, 8].

5.2.3 General principle of bound optimization algorithms

Except for simplified cases such as those just mentioned, the functional $\mathcal{C}(\mathbf{x})$ defined by (5.2) cannot be minimized directly by a coefficient-wise rule. This is primarily due to the convolution matrix $\mathbf{H}$, which makes the wavelet coefficients of $\mathbf{x}$ interdependent, hence precluding a simple coefficient-wise rule. Instead, one can use an iterative minimization technique as proposed in [78, 58, 16]. Interestingly, this involves a judicious combination of the two aforementioned methods. In the next section, we shall briefly review the derivation of Daubechies et al.; it is based on a *bound optimization* approach [111, 125], which essentially consists in replacing $\mathcal{C}(\mathbf{x})$ by a succession of auxiliary functionals that are easy to minimize.

Using the current estimate $\mathbf{x}^{(n)}$, the key-idea is to construct an auxiliary functional $\mathcal{A}_n(\mathbf{x})$ with the following properties:

- when $\mathbf{x} = \mathbf{x}^{(n)}$, $\mathcal{A}_n(\mathbf{x})$ coincides with $\mathcal{C}(\mathbf{x})$;

- when $\mathbf{x} \neq \mathbf{x}^{(n)}$, $\mathcal{A}_n(\mathbf{x})$ upper-bounds $\mathcal{C}(\mathbf{x})$.

As illustrated in Fig. 5.2 for $N = 1$, these properties guarantee that, by minimizing $\mathcal{A}_n(\mathbf{x})$, we will also decrease $\mathcal{C}(\mathbf{x})$. The general update equation is thus

$$\mathbf{x}^{(n+1)} = \arg\min_{\mathbf{x}} \mathcal{A}_n(\mathbf{x}).$$

A general discussion of convergence criteria for bound optimization methods would be outside the scope of this chapter; the reader is referred to [125] and the references therein. Moreover, as mentioned in the introduction, the paper by Daubechies et al. contains a convergence proof that can be readily extended to cover both algorithms presented in this chapter.
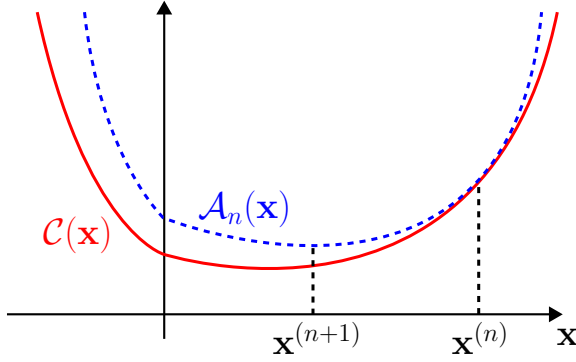
Figure 5.2: The bound-optimization principle: an estimate of the minimizer of the original cost function is constructed by minimizing an auxiliary functional. The process is repeated iteratively.

5.3 The existing thresholded Landweber algorithm

5.3.1 A bound with decoupled wavelet coefficients

Daubechies et al. [58] proposed to use functionals of the form

$$\mathcal{A}_n(\mathbf{x}) = \alpha \|\mathbf{x}^{(n)} - \mathbf{x}\|_2^2 + \mathcal{C}(\mathbf{x}) - \|\mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x}\|_2^2. \tag{5.8}$$

Here, the (real and positive) scalar α must be chosen strictly larger than the spectral radius of $\mathbf{H}^T\mathbf{H}$:

$$\alpha > \rho(\mathbf{H}^T\mathbf{H}) \qquad \text{where} \qquad \rho(\mathbf{H}^T\mathbf{H}) = \max_{\|\mathbf{v}\|_2=1} \|\mathbf{H}\mathbf{v}\|_2^2.$$

Equivalently, since $\mathbf{H}$ is a convolution matrix, $\rho(\mathbf{H}^T\mathbf{H})$ is the largest squared modulus of the DFT coefficients of $\mathbf{H}$.

Let us discuss this choice in more details. First, the constraint on α ensures that $\mathcal{A}_n(\mathbf{x})$ is a proper bound for $\mathcal{C}(\mathbf{x})$.

Property 5. *Assume that $\alpha > \rho(\mathbf{H}^T\mathbf{H})$ holds in Eq. (5.8). Then $\mathcal{A}_n(\mathbf{x}) > \mathcal{C}(\mathbf{x})$, except at $\mathbf{x} = \mathbf{x}^{(n)}$, where $\mathcal{A}_n(\mathbf{x}) = \mathcal{C}(\mathbf{x})$.*

Proof. The inequality $\alpha > \rho(\mathbf{H}^T\mathbf{H})$ ensures that $\alpha\mathbf{I} - \mathbf{H}^T\mathbf{H}$ is positive-definite. This means that, when $\mathbf{x} \neq \mathbf{x}^{(n)}$, we have

$$\alpha\|\mathbf{x}^{(n)} - \mathbf{x}\|_2^2 - \|\mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x}\|_2^2 = (\mathbf{x}^{(n)} - \mathbf{x})^T(\alpha\mathbf{I} - \mathbf{H}^T\mathbf{H})(\mathbf{x}^{(n)} - \mathbf{x}) > 0;$$

whence it follows that $\mathcal{A}_n(\mathbf{x})$ is a strict upper bound of $\mathcal{C}(\mathbf{x})$, except at $\mathbf{x} = \mathbf{x}^{(n)}$ where the equality is met. $\qquad\square$

Second, the specific form of (5.8) makes it easy to minimize $\mathcal{A}_n(\mathbf{x})$, based on the following observation.

Property 6. *Definition* (5.8) *is equivalent to*

$$\mathcal{A}_n(\mathbf{x}) = \alpha\,\|\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}) - \mathbf{x}\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1 + c,$$

where c is a constant that does not depend on $\mathbf{x}$.

Proof.

$$\begin{aligned}
\mathcal{A}_n(\mathbf{x}) &= \alpha\|\mathbf{x}^{(n)} - \mathbf{x}\|_2^2 + \|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 - \|\mathbf{H}\mathbf{x}^{(n)} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1 \\
&= \alpha\|\mathbf{x}^{(n)}\|_2^2 - 2\,\alpha\,\mathfrak{Re}\big\{\mathbf{x}^T\mathbf{x}^{(n)}\big\} + \alpha\|\mathbf{x}\|_2^2 \\
&\quad + \|\mathbf{y}\|_2^2 - 2\,\mathfrak{Re}\big\{\mathbf{x}^T\mathbf{H}^T\mathbf{y}\big\} + \|\mathbf{H}\mathbf{x}\|_2^2 \\
&\quad - \|\mathbf{H}\mathbf{x}^{(n)}\|_2^2 + 2\,\mathfrak{Re}\big\{\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}^{(n)}\big\} - \|\mathbf{H}\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}\mathbf{x}\|_1 \\
&= \alpha\|\mathbf{x}\|_2^2 - 2\,\alpha\,\mathfrak{Re}\left\{\mathbf{x}^T\left(\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})\right)\right\} + \lambda\|\mathbf{W}\mathbf{x}\|_1 + c',
\end{aligned}$$

where $c' = \alpha\|\mathbf{x}^{(n)}\|_2^2 + \|\mathbf{y}\|_2^2 - \|\mathbf{H}\mathbf{x}^{(n)}\|_2^2$ is a constant that does not depend on $\mathbf{x}$. We complete the proof by adding the constant $\alpha\|\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})\|_2^2$ (which does not depend on $\mathbf{x}$) so as to complete the quadratic term. $\qquad\square$

The above derivation reveals that the term $\|\mathbf{H}\mathbf{x}\|_2^2$, which hinders the direct minimization of $\mathcal{C}(\mathbf{x})$, is canceled by the negative term in (5.8). As a result, $\mathbf{x}$ is not premultiplied by $\mathbf{H}$ anymore in the expression of Property 6. This means that the wavelet coefficients of $\mathbf{x}$ are now completely decoupled and the minimization of $\mathcal{A}_n(\mathbf{x})$ involves a simple coefficient-wise operation in the wavelet domain.

5.3.2 The resulting algorithm

To make this more apparent, we will denote by $\mathbf{z}^{(n)}$ the expression $\mathbf{x}^{(n)} + \alpha^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$; note that this is exactly the update formula of the classical Landweber

iteration (5.4), with step size $\tau = \alpha^{-1}$. Omitting c and dividing by the (positive) constant α, the minimization of $\mathcal{A}_n(\mathbf{x})$ thus reduces to the minimization of

$$\|\mathbf{z}^{(n)} - \mathbf{x}\|_2^2 + \lambda\tau\|\mathbf{W}\mathbf{x}\|_1.$$

Now this is almost exactly the pure denoising functional of (5.5), with $\mathbf{z}^{(n)}$ as the noisy signal and $\lambda\tau$ as the regularization parameter. We have already seen that its minimizer is obtained via a simple soft-thresholding operation on the wavelet coefficients of $\mathbf{z}^{(n)}$. In the present case, the threshold level will be $\lambda\tau/2$.

To summarize, the resulting "thresholded Landweber" (TL) algorithm alternates between the following two steps, starting from an initial estimate $\mathbf{x}^{(0)}$:

- compute the Landweber iteration $\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \tau\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$, with step-size τ;

- perform the wavelet-domain denoising operation $\mathbf{x}^{(n+1)} = \mathbf{W}^T\mathcal{T}_{\lambda\tau/2}\{\mathbf{W}\mathbf{z}^{(n)}\}$, with threshold level $\lambda\tau/2$.

5.4   The fast thresholded Landweber algorithm

Using the algorithm described above, Figueiredo and Nowak have reported numerical results that are competitive with state-of-the-art 2D deconvolution methods [78]. However, the convergence of this algorithm can be rather slow, especially when one choses a small regularization parameter λ. This is typically the case for higher signal-to-noise ratios (see the numerical examples in Section 5.5), which still require regularization, but at a moderate level.

Starting from this section, we will assume that the regularization term of (5.2) is expressed in the Shannon wavelet basis (see [142], p. 223 or [210], p. 51). This wavelet family allows for larger (subband-specific) step sizes and threshold levels, resulting in a significant speed-up.

5.4.1   The Shannon wavelet basis

Besides being orthonormal, the main characteristic of Shannnon wavelets is that their spectrum is ideal and disjoint across subbands. This proves to be especially convenient when dealing with convolution operators.

Let us denote by $\left(W_j\right)_{j\in S}$ the different wavelet subspaces[2]. More precisely, our convention will be that there are $j_{\max}$ wavelet subspaces, and that $j = j_{\max} + 1$

---

[2]We use the terms subband and subspace interchangeably.

corresponds to the coarsest-scale scaling function subspace. In other words, the indexing set is $S = \{1, \ldots, j_{\max} + 1\}$. Fig. 5.3 illustrates the frequency support of these subspaces in the 1D case, where $j_{\max}$ corresponds to the number of scales of the decomposition. For higher dimensions, we use a separable extension of the Shannon wavelet basis. Note that in our finite-dimensional setting, Hermitian symmetry is violated at the mid frequency point, which implies that our basis elements are complex-valued.
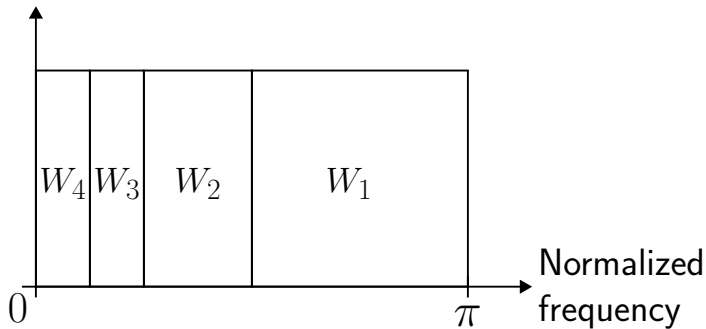
Figure 5.3: Frequency support of the Shannon-wavelet subspaces in the 1D case. The figure corresponds to $j_{\max} = 3$, which means that W_4 represents the scaling function subspace.

We use the (boldface) symbol $\mathbf{W}_j$ to denote the decomposition matrix that yields the coefficients of a signal in subspace W_j. The original signal can be reconstructed from its wavelet coefficients using the formula

$$\mathbf{x} = \sum_{j \in S} \underbrace{\mathbf{W}_j^T \mathbf{W}_j}_{\mathbf{P}_j} \mathbf{x}. \tag{5.9}$$

Note that we can use the transpose of $\mathbf{W}_j$ because the Shannon wavelet basis is orthonormal. In this equation, $\mathbf{W}_j^T \mathbf{W}_j \mathbf{x}_j$ represents the orthogonal projection of $\mathbf{x}$ on W_j.

The following result states that, for the Shannon wavelet basis, the projection matrix $\mathbf{P}_j = \mathbf{W}_j^T \mathbf{W}_j$ commutes with any convolution matrix.

Property 7. *Let* $\mathbf{C}$ *be a block-circulant matrix. Then, for the Shannon wavelet basis,*

$$\mathbf{CP}_j = \mathbf{P}_j\mathbf{C}.$$

Proof. Let $\mathbf{v}$ be an arbitrary vector. Applying (5.9) to $\mathbf{CP}_j\mathbf{v}$ yields

$$\mathbf{CP}_j\mathbf{v} = \sum_{\ell \in S} \mathbf{P}_\ell\mathbf{CP}_j\mathbf{v}.$$

Now recall that with the Shannon wavelet basis, the subspaces $(W_j)_{j \in S}$ have disjoint frequency supports. Because $\mathbf{C}$ does not modify the frequency support of a signal (since it is diagonalized in the DFT basis), it must be that $\mathbf{CP}_j\mathbf{v} \in W_j$. Therefore, $\mathbf{P}_\ell\mathbf{CP}_j\mathbf{v} = \mathbf{0}$ for $\ell \neq j$ and

$$\mathbf{CP}_j\mathbf{v} = \mathbf{P}_j\mathbf{CP}_j\mathbf{v} = \mathbf{P}_j\mathbf{C}\left(\mathbf{v} - \sum_{\ell \in S\backslash\{j\}} \mathbf{P}_\ell\mathbf{v}\right).$$

From this relation, we can reapply the same argument to obtain $\mathbf{CP}_j\mathbf{v} = \mathbf{P}_j\mathbf{Cv}$ for every $\mathbf{v}$. This is equivalent to the commutativity of $\mathbf{C}$ and $\mathbf{P}_j$. □

5.4.2   Derivation of a subband-adapted bound

Our algorithm is based on the idea of cutting the cost function into subband-specific terms, thanks to the above commutativity property.

Property 8. *When using the Shannon wavelet basis,*

$$\mathcal{C}(\mathbf{x}) = \sum_{j \in S} \|\mathbf{P}_j\mathbf{y} - \mathbf{HP}_j\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}_j\mathbf{x}\|_1,$$

where $\mathbf{P}_j$ *is the projection operator on the jth subband.*

Proof. Using (5.9), we first observe that

$$\mathbf{y} - \mathbf{Hx} = \sum_{j \in S} \mathbf{P}_j\mathbf{y} - \mathbf{HP}_j\mathbf{x}.$$

Property 7 tells us that we can rewrite the components of this sum as $\mathbf{P}_j(\mathbf{y} - \mathbf{Hx})$, which shows that they are mutually orthogonal (since the subspaces $(W_j)_{j \in S}$ are orthogonal). This implies that we can separate the data term as follows:

$$\|\mathbf{y} - \mathbf{Hx}\|_2^2 = \sum_{j \in S} \|\mathbf{P}_j\mathbf{y} - \mathbf{HP}_j\mathbf{x}\|_2^2.$$

Combining this relation with the fact that $\|\mathbf{W}\mathbf{x}\|_1 = \sum_{j \in S} \|\mathbf{W}_j \mathbf{x}\|_1$ yields the desired result. $\qquad\square$

The above result allows us to bound the cost function in a subband-specific manner. The idea is to apply the bound of Daubechies et al. to the individual terms of the sum in Property 8, leading to the following type of auxiliary functional:

$$\mathcal{A}_n(\mathbf{x}) = \mathcal{C}(\mathbf{x}) + \sum_{j \in S} \alpha_j \|\mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{P}_j \mathbf{x}\|_2^2 - \|\mathbf{H}\mathbf{P}_j \mathbf{x}^{(n)} - \mathbf{H}\mathbf{P}_j \mathbf{x}\|_2^2. \qquad (5.10)$$

The fundamental difference is that we now have subband-dependent constants $(\alpha_j)_{j \in S}$, which can be chosen significantly smaller than the constant α of the standard algorithm. More precisely, for the sum in (5.10) to be positive, it is sufficient that, for every j,

$$\alpha_j > \rho_j(\mathbf{H}^T\mathbf{H}) \qquad \text{where} \qquad \rho_j(\mathbf{H}^T\mathbf{H}) = \max_{\|\mathbf{v}\|_2=1, \mathbf{v} \in W_j} \|\mathbf{H}\mathbf{v}\|_2^2.$$

We point out that the definition of $\rho(\mathbf{H}^T\mathbf{H})$ (a squared matrix norm) is similar to that of $\rho_j(\mathbf{H}^T\mathbf{H})$. However, in the latter, the maximum is taken over subspace W_j. Therefore, we necessarily have

$$\rho_j(\mathbf{H}^T\mathbf{H}) \le \rho(\mathbf{H}^T\mathbf{H}).$$

Equivalently, since $\mathbf{H}$ is a convolution matrix, $\rho_j(\mathbf{H}^T\mathbf{H})$ is the largest squared modulus of the DFT coefficients of $\mathbf{H}$, *over the frequency support of subband W_j*. This value can be much smaller than $\rho(\mathbf{H}^T\mathbf{H})$ (the largest squared modulus over the whole spectrum). Fig. 5.4 compares choices for α and $(\alpha_j)_{j \in S}$ in a simplified 1D situation.

When the previous condition is met, $\mathcal{A}_n(\mathbf{x})$ upper-bounds $\mathcal{C}(\mathbf{x})$. Still, $\mathcal{A}_n(\mathbf{x})$ is easy to minimize (thanks to the commutativity stated in Property 7). This transposition of Properties 5 and 6 to our new auxiliary functional can be summarized as follows:

Property 9. *Assume that $\alpha_j > \rho_j(\mathbf{H}^T\mathbf{H})$ holds for every $j \in S$ in definition (5.10). Then $\mathcal{A}_n(\mathbf{x}) > \mathcal{C}(\mathbf{x})$, except at $\mathbf{x} = \mathbf{x}^{(n)}$, where $\mathcal{A}_n(\mathbf{x}) = \mathcal{C}(\mathbf{x})$. Moreover, (5.10) is equivalent to*

$$\mathcal{A}_n(\mathbf{x}) = \sum_{j \in S} \alpha_j \|\mathbf{P}_j[\mathbf{x}^{(n)} + \alpha_j^{-1}\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})] - \mathbf{P}_j\mathbf{x}\|_2^2 + \lambda\|\mathbf{W}_j\mathbf{x}\|_1 + c_j,$$

where the constants $(c_j)_{j \in S}$ do not depend on $\mathbf{x}$.
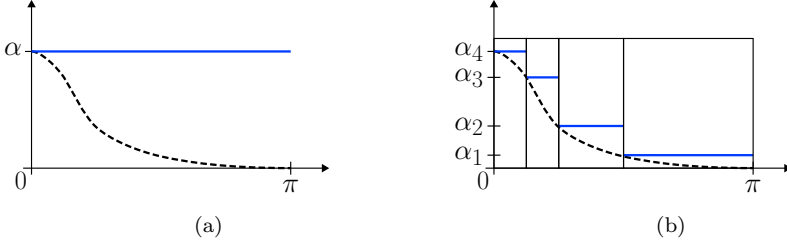
Figure 5.4: Comparison of the constants α and $(\alpha_j)_{j \in S}$ for a fictitious 1D low-pass convolution kernel. The dashed curve represents the squared modulus of its Fourier transform.

Proof. The definition of $\rho_j(\mathbf{H}^T\mathbf{H})$ implies that

$$\alpha_j\|\mathbf{P}_j\mathbf{x}^{(n)} - \mathbf{P}_j\mathbf{x}\|_2^2 - \|\mathbf{H}\mathbf{P}_j\mathbf{x}^{(n)} - \mathbf{H}\mathbf{P}_j\mathbf{x}\|_2^2 \geq 0$$

for every $j \in S$. When $\mathbf{x} \neq \mathbf{x}^{(n)}$, there is at least one j for which this inequality is strict. If $\mathbf{x} = \mathbf{x}^{(n)}$, the equality is clear. This proves the first part of the property.

Applying Property 6 to every term of the sum in (5.10) yields

$$\mathcal{A}_n(\mathbf{x}) = \sum_{j \in S} \alpha_j \left\|\mathbf{P}_j\mathbf{x}^{(n)} + \alpha_j^{-1}\mathbf{H}^T(\mathbf{P}_j\mathbf{y} - \mathbf{H}\mathbf{P}_j\mathbf{x}^{(n)}) - \mathbf{P}_j\mathbf{x}\right\|_2^2 + \lambda\|\mathbf{W}_j\mathbf{x}\|_1 + c_j.$$

Since we are using Shannon wavelets, we can use the commutativity of $\mathbf{P}_j$ with the convolution matrices $\mathbf{H}$ and $\mathbf{H}^T$ to derive the second part of the property. $\square$

5.4.3   The resulting algorithm

Let us define the constants $\tau_j = 1/\alpha_j$ and the intermediate signal

$$\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \sum_{j \in S} \tau_j \mathbf{P}_j \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)}).$$

Since the Shannon wavelet basis is orthonormal, we can replace $\mathbf{P}_j$ by $\mathbf{W}_j$ in the expression of Property 9. The minimization of $\mathcal{A}_n(\mathbf{x})$ then becomes equivalent to the minimization of

$$\sum_{j \in S} \alpha_j \left(\|\mathbf{W}_j\mathbf{z}^{(n)} - \mathbf{W}_j\mathbf{x}\|_2^2 + \lambda\tau_j\|\mathbf{W}_j\mathbf{x}\|_1\right).$$

Each term of this sum depends on the wavelet coefficients of $\mathbf{x}$ in a different subband. As a consequence, we can minimize them independently. Moreover, the individual terms have the same form as in (5.6), up to a (positive) multiplicative constant. Our new auxiliary functionals can thus be minimized using essentially the same algorithm as before—with the same cost per iteration—but with subband-dependent parameters.

The resulting "fast thresholded Landweber" (FTL) algorithm alternates between the following two steps:

- Compute the Landweber iteration

$$\mathbf{z}^{(n)} = \mathbf{x}^{(n)} + \sum_{j \in S} \tau_j \mathbf{P}_j \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$$

  with subband-dependent step sizes τ_j.

- Perform the wavelet-domain denoising operation

$$\mathbf{x}^{(n+1)} = \sum_{j \in S} \mathbf{W}_j^T \mathcal{T}_{\lambda \tau_j / 2} \{ \mathbf{W}_j \mathbf{z}^{(n)} \}$$

  with subband-dependent thresholds $\lambda \tau_j / 2$.

5.5 Numerical experiments

5.5.1 Implementation details and computational cost

The FTL algorithm is simple to implement, for it essentially consists in going back and forth between the frequency domain and the wavelet domain.

In the first step, one must start by computing the Landweber correction term $\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(n)})$. This is done in the DFT basis, where $\mathbf{H}$ is diagonalized. The actual update should a priori be performed in the wavelet domain, since the step sizes are subband-dependent. However, in the particular case of a Shannon wavelet basis, it can be carried out in the frequency domain: it is equivalent to multiplying each frequency component of the correction term by τ_j, where j refers to the wavelet subband that contains the considered frequency.

As described above, the second step just requires the application of a wavelet transform to $\mathbf{z}^{(n)}$, followed by a soft-thresholding and an inverse wavelet transform of the result. However, in practice—for both algorithms presented in this chapter— we use the random-shift method described in [78]. The main motivation is to reduce

unpleasant artifacts that are common side-effects of thresholding operations in non-redundant wavelet bases. Still, we would like to do so without having to resort to a fully shift-invariant (undecimated) transform. The compromise consists—for every iteration—in applying a random (circular) shift to the intermediate estimate $\mathbf{z}^{(n)}$, before performing its wavelet decomposition. After the thresholding and the inverse transform, the new estimate is shifted back to the original position. By doing this, the effect of thresholding is "averaged out" over multiple shifts during the iteration process. Note that a shift in the image domain corresponds to a simple modulation in the frequency domain. This means that the random-shift method can be used at no significant additional cost.

The computational cost is therefore essentially the same as for the classical thresholded Landweber algorithm, which also requires two (one direct and one inverse) wavelet transforms per iteration. We use a frequency-domain implementation of the wavelet transform [25]. Thus, every iteration essentially amounts to computing two FFTs, which is no more than the standard algorithms mentioned in the introduction.

One last remark relating to the scaling-function subspace: to simplify the presentation, we have not distinguished it from the wavelet subspaces so far. However, the scaling-function coefficients are usually not included in the regularization term, which means that they are not thresholded in practice.

5.5.2   Evaluation protocol

To compare the performance of the thresholded Landweber (TL) and the fast thresholded Landweber (FTL) algorithms, we performed several experiments on synthetic and real data. The literal way of assessing the convergence speedup of our algorithm would be to compare the intermediate estimates to the true minimizer of the cost function. Generally, however, this true minimizer is not available and its determination would involve an overwhelming amount of computations. Therefore, following the common practice in the field, we will use the quadratic error between the intermediate estimates and the original image (gold-standard). This is justified from at least two standpoints.

1. The user standpoint: ultimately, what counts in practice is the difference between the deconvolution result and the original signal.

2. The theoretical standpoint: the true minimizer can be shown to be "close" to the original signal in the following sense. When the noise level and the regularization parameter tend to zero, the minimizer of the cost function

> tends to the original signal (see the paper by Daubechies et al. [58] for a rigorous formulation and proof).

For each experiment, we used the same regularization parameter for the TL and FTL algorithms. For the phantom experiments, this value was optimized to yield the result closest to the original signal after a large number of FTL iterations (so as to approach the true minimizer reasonably well). In the real data cases, it was adjusted empirically for best visual appearance.

5.5.3   Visual comparison in 2D: results for a limited number of iterations

In the first experiment, we took a 512×512 image of a neuron cell acquired on a confocal microscope as our (ground-truth) original image. We then simulated an optical defocusing blur produced by a widefield microscope. We used a standard diffraction-limited PSF model [94, 236], which is represented in Fig. 5.5.3 (b). Gaussian white noise was added to the result; the variance σ^2 was set such that the *blurred signal-to-noise ratio* (BSNR) was equal to 40 dB. The BSNR is used to quantify the noise level with respect to the original signal strength, while taking into account the attenuation introduced by the simulated imaging device. It is defined as

$$\mathrm{BSNR} = 10 \log_{10} \left(\frac{\|\mathbf{H}\mathbf{x}_{\mathrm{orig}}\|_2^2 - N \, \mathrm{mean}(\mathbf{H}\mathbf{x}_{\mathrm{orig}})^2}{N\sigma^2} \right).$$

Here, N is the total number of pixels per image and $\mathrm{mean}(\mathbf{H}\mathbf{x}_{\mathrm{orig}})$ stands for the average gray level of the blurred original image.

Fig. 5.5.3 shows the original image (a) and the simulated measurement (c). Estimates at different stages of the TL and FTL iterations are displayed in Fig. 5.6. For both algorithms, we used the same initial estimate (the measured image $\mathbf{y}$), the same number of decomposition levels ($j_{\mathrm{max}} = 5$) and the same regularization parameter. It is seen that the images obtained after 10 or 30 iterations of the TL algorithm are less sharp than the one obtained with 10 iterations of the FTL algorithm. This is especially true for the details inside the object which exhibit better contrast than in the latter. It is only after a larger number of iterations (100) that the output of the TL algorithm gets closer to the FTL result with 10 iterations, at least from a visual standpoint. We did not include the results of the FTL algorithm with more than 10 iterations, since the differences are not perceptible visually because the algorithm has already essentially reached convergence.

(a) Original

(b) PSF

(c) Measured
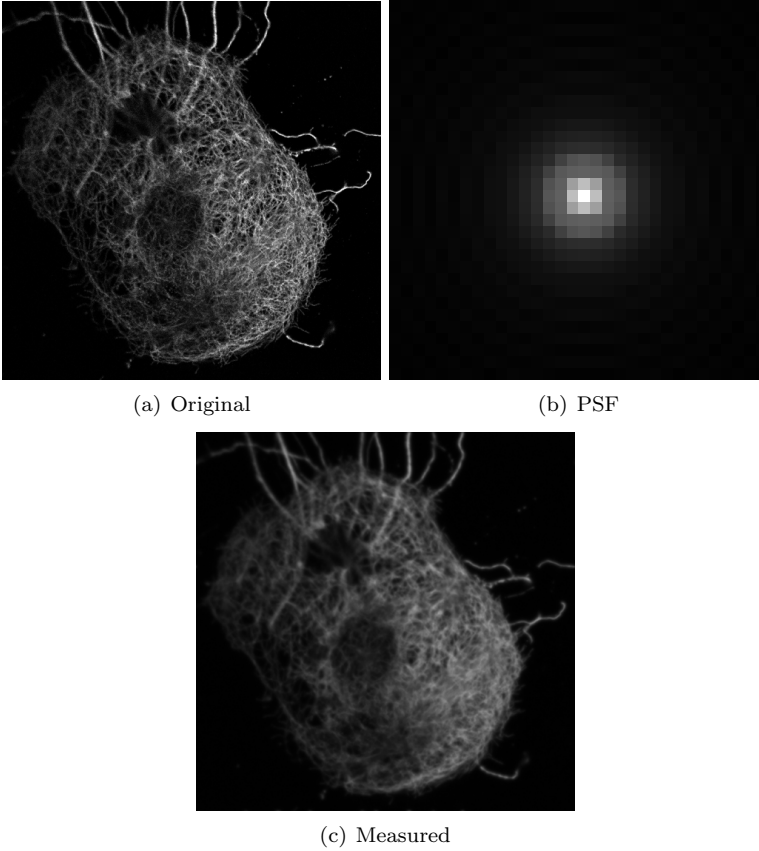
Figure 5.5: (a) The original confocal image (courtesy of Bertrand Vileno, LNNME, EPFL) and (c) the simulated widefield image of Section 5.5.3. The square root of the PSF is shown in (b), with a 16× zoom factor compared to (a) and (c).

(a) Original   (b) Measured   (c) FTL, 10 iterations

(d) TL, 10 iterations   (e) TL, 30 iterations   (f) TL, 100 iterations
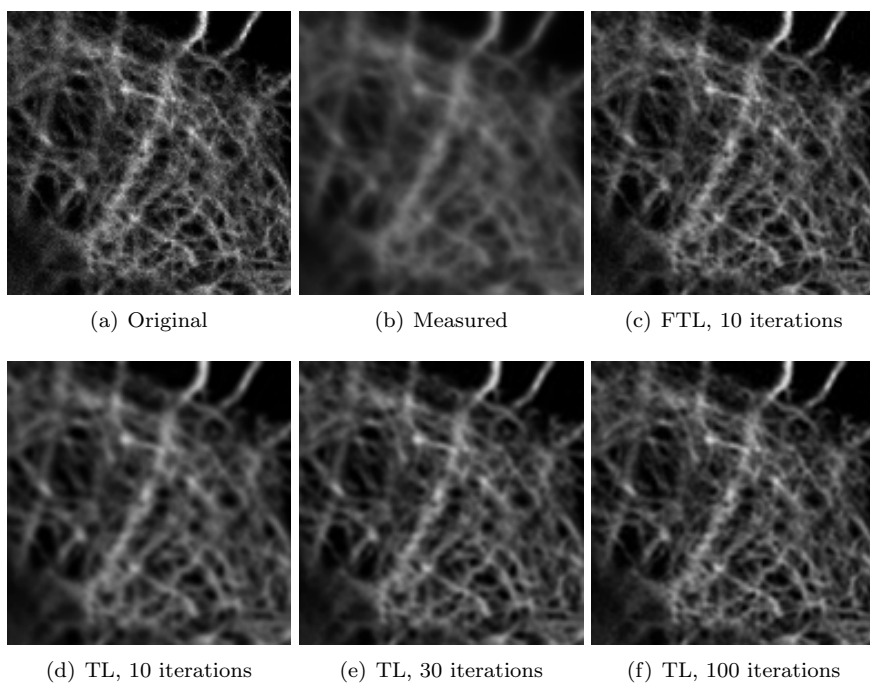
Figure 5.6: Deconvolution results for the setup of Fig. 5: comparison on a 128×128 region of interest.
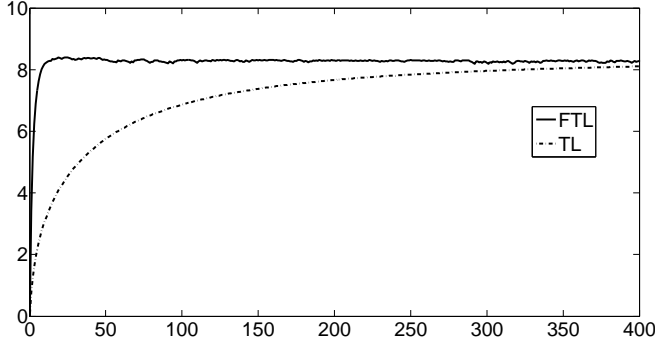
Figure 5.7: SER gain as a function of the iteration number for the experiment of Section 5.5.3.

5.5.4   Quantitative comparison in 2D: number of iterations to reach a given level of SER gain

The above results strongly suggest that the FTL algorithm requires fewer iterations than the TL algorithm to reach a given level of restoration quality. Fig. 5.7 gives a quantitative insight: it shows the evolution of the *signal-to-error ratio* (SER) as a function of the iteration number. For an estimate $\mathbf{x}$, the SER is defined as

$$\mathrm{SER}(\mathbf{x}) = 10 \log_{10} \left(\frac{\|\mathbf{x}_{\mathrm{orig}}\|_2^2}{\|\mathbf{x} - \mathbf{x}_{\mathrm{orig}}\|_2^2} \right).$$

For each estimate $\mathbf{x}^{(n)}$, we compare this figure to the SER of the measured signal, leading to the SER *gain*:

$$\mathrm{SERG}(\mathbf{x}^{(n)}) = \mathrm{SER}(\mathbf{x}^{(n)}) - \mathrm{SER}(\mathbf{y}).$$

One can observe that the FTL algorithm requires less than 10 iterations to reach an improvement of 8 dB. This is roughly thirty times less than the TL algorithm, which takes about 300 iterations to reach the same level.

We performed a series of more extensive experiments on 4 standard test-images (Cameraman, MRI, House, Bird). We convolved these images with a 9×9 uniform blur kernel and added white Gaussian noise to the results in order to replicate the experimental setup of Figueiredo and Nowak. We considered different noise

levels (BSNR $= 10, 20, 30, 40, 50$ dB) and we averaged the SER gains over 30 noise realizations. For each test case, we used the same value of λ and the same initial estimate for both algorithms. The value of λ was optimized to yield the best possible SER gain after 300 iterations of the FTL algorithm. The initial estimate was obtained using the same Wiener-type filter as Figueiredo and Nowak [78]:

$$\mathbf{x}^{(0)} = (\mathbf{H}^T\mathbf{H} + 10^{-3}\sigma^2\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}.$$

The results are summarized in Table 5.1. For each image, we indicate the SER gain of the FTL algorithm after 10 and 30 iterations. We also give the number of iterations of the TL algorithm required to reach the same SER improvement. The corresponding acceleration factors (number of TL iterations, divided by number of FTL iterations) are listed in the 5th and 8th column. The acceleration factors vary between 1.5 to 4.7 for low BSNR levels (10-20 dB) and 32.3 to 191.5 for high BSNR levels (40-50 dB).

Note that the accelerated algorithm that has been derived here is specific to Shannon wavelets and is not directly transposable to other types of basis functions. Nevertheless, the results obtained for the Cameraman image are comparable to those reported by Figueiredo and Nowak in terms of restoration quality[3], even though these authors used different wavelets than ours. In fact, they observed that the performance of their algorithm was only very mildly dependent on the choice of a particular type of wavelet. Still, the present results suggest that there should be ways of accelerating the convergence with other types of wavelets as well, which calls for further investigation.

5.5.5 Visual comparison in 3D: experiment on fluorescence microscopy data

For the last experiment, we used 3D image-stacks of a biological sample. The sample contained fibroblast cells stained with a DiO dye from Invitrogen Corporation (Carlsbad, USA). This dye is predominantly retained in the cell membrane. In addition, fluorescent microbeads were introduced into the sample medium. The dye and the microbeads were chosen to have approximately the same peak excitation and emission wavelengths; a 505-530 nm bandpass filter was used to delimit the detection spectrum.

The sample was observed on a Zeiss LSM 510 confocal microscope with a $63\times$, 1.4 NA oil-immersion objective. We first acquired a stack of images with the pinhole

---

[3]Table I of [78] indicates a gain of 6.33 dB at 40 dB BSNR, while our experiments yielded 6.03 and 6.61 dB after 10 and 30 iterations of the FTL algorithm respectively.

	BSNR	SERG FTL 10 iter.	Number of TL iter.	Accel. factor	SERG FTL 30 iter.	Number of TL iter.	Accel. factor
Cameraman	10	2.90	36	3.6	2.94	61	2.0
	20	2.62	39	3.9	2.74	87	2.9
	30	4.31	121	12.1	4.43	172	5.7
	40	6.03	541	54.1	6.61	972	32.4
	50	7.80	1876	187.6	8.38	> 2000	> 66.7
MRI	10	3.68	40	4.0	3.87	99	3.3
	20	4.53	47	4.7	4.81	91	3.0
	30	6.79	167	16.7	6.98	211	7.0
	40	7.94	724	72.4	9.55	1344	44.8
	50	9.53	1915	191.5	10.70	> 2000	> 66.7
House	10	3.73	22	2.2	3.82	58	1.9
	20	3.81	27	2.7	3.99	49	1.6
	30	6.25	170	17.0	6.39	199	6.6
	40	7.30	559	55.9	9.07	969	32.3
	50	7.49	1616	161.6	8.90	> 2000	> 66.7
Bird	10	4.65	20	2.0	4.70	45	1.5
	20	3.67	34	3.4	3.81	51	1.7
	30	5.76	180	18.0	5.93	216	7.2
	40	5.75	555	55.5	8.24	1004	33.5
	50	5.40	1562	156.2	7.12	> 2000	> 66.7

Table 5.1: Number of iterations required to reach a given level of SER gain (SERG).

completely open. In this configuration, the confocal effect is not used and the system becomes essentially equivalent to a widefield microscope. This results in images with out-of-focus blur due to the poor localization of the widefield PSF along the z-dimension. The second set of images was acquired with the pinhole radius set to 1 Airy unit. In this configuration, much of the out-of-focus light is rejected by the system, resulting in significantly sharper images. We then used the confocal data set as a reference for comparing the performance of the TL and FTL algorithms on the 3D widefield data set.

Maximum intensity projections (along the z axis) of both data sets are shown

(a) Widefield image   (b) Confocal reference

(c) TL, 10 iterations   (d) FTL, 10 iterations
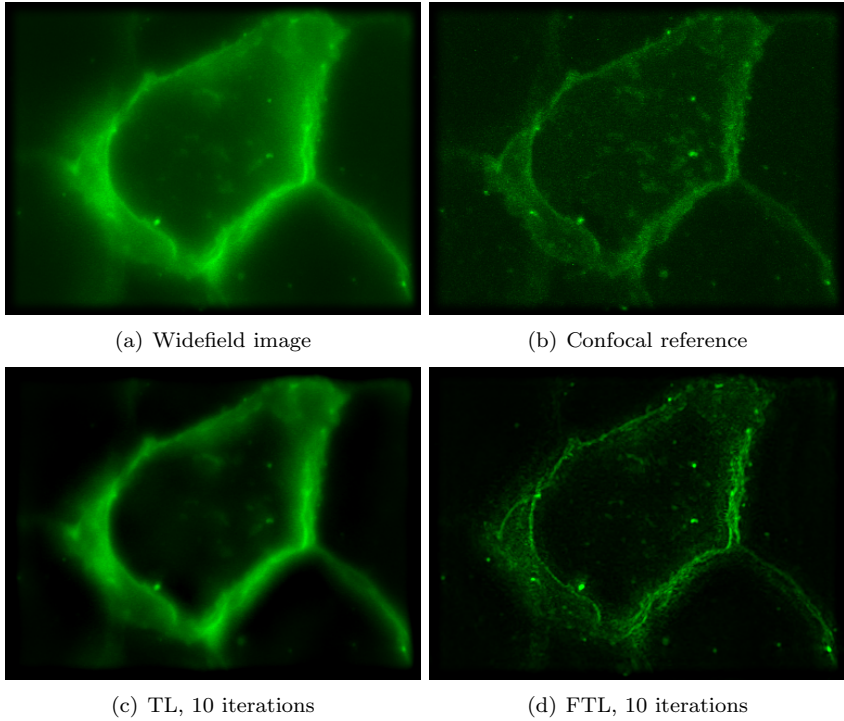
Figure 5.8: Maximum intensity projections of $384 \times 288 \times 16$ image stacks. (a, b) Widefield and confocal image of the same object. (c, d) Results of the deconvolution algorithms applied to the widefield image.

in Fig. 5.8. Below are the results of 10 iterations of the TL and FTL algorithms on the widefield stack; this corresponds to roughly 1 minute and 30 seconds of computation time on a 2.66 GHz Intel Xeon workstation. We used a 3D version of the diffraction-limited model mentioned above (with parameters corresponding to manufacturer specification: $NA = 1.4$, $n_i = 1.518$) to generate the PSF and we took the measured widefield image as the initial estimate.

The TL algorithm cannot produce a visible deconvolution effect within 10 iterations; its output is very similar to the original widefield data. On the other hand, the FTL algorithm rapidly produces an estimate that is significantly sharper. In particular, the thickness of the cell membranes is comparable to the confocal image.

Also, the fluorescent microbeads are brighter than in the widefield image. The results were shown to biologists who were positively impressed.

5.6   Conclusion

We have presented a fast algorithm for wavelet-regularized variational deconvolution. It is based on the use of a Shannon wavelet basis, which allows for a subband-adapted minimization strategy. As a result, the update introduced by each step of the FTL algorithm is much more efficient than with the standard TL algorithm. For BSNR levels above 30 dB, we have obtained acceleration factors of one order of magnitude or better. This corresponds to relatively standard acquisition conditions for widefield deconvolution microscopy, at least when considering fixed specimens[4]. Thanks to this substantial speed increase, we have demonstrated the feasibility of 3D wavelet-based deconvolution microscopy on a widefield fluorescence image stack.

A direct extension of the proposed algorithm is to consider subband-dependent regularization parameters [58] and/or other sparsity-enforcing penalizations (that is, other thresholding functions, see [78]). The principle of the method could also be applied to other types of operators that are block-diagonal in the basis that is used for regularization. The possibilities of this type of procedure are numerous and our current experimentations suggest that there is room for improvement, which calls for further investigations and more extensive comparative studies.

---

[4]Current cooled CCD cameras are capable of much wider dynamic ranges of 12 or even 16 bits per pixel (1 bit $\approx$ 6 dB).

Chapter 6

A general multilevel algorithm

Summary

We present a multi-level extension of the popular "thresholded Landweber" algorithm for wavelet-regularized image restoration that yields an order of magnitude speed improvement over the standard fixed-scale implementation. The method is generic and targeted towards large-scale linear inverse problems, such as 3D deconvolution microscopy.

The algorithm is derived within the framework of bound optimization. The key idea is to successively update the coefficients in the various wavelet channels using fixed, subband-adapted iteration parameters (step sizes and threshold levels). The optimization problem is solved efficiently via a proper chaining of basic iteration modules. The higher-level description of the algorithm is similar to that of a multigrid solver for PDEs, but there is one fundamental difference: the latter iterates though a sequence of multiresolution versions of the original problem, while, in our case, we cycle through the wavelet subspaces corresponding to the difference between successive approximations. This strategy is motivated by the special structure of the problem and the preconditioning properties of the wavelet representation. We establish that the solution of the restoration problem corresponds to a fixed point of our multilevel optimizer. We also provide experimental evidence

that the improvement in convergence rate is essentially determined by the (unconstrained) linear part of the algorithm, irrespective of the type of wavelet. Finally, we illustrate the technique with some image deconvolution examples, including some real 3D fluorescence microscopy data.

6.1 Introduction

6.1.1 Motivation and originality of the present work

Inverse problems arise in various imaging applications such as biomicroscopy [236, 188], medical imaging [136, 21] or astronomy [158, 207]. An increasingly important issue for reconstruction and restoration tasks is the mass of data that is now routinely produced in these fields. The instrumentation typically allows for high-dimensional and multimodal imaging, fostering the evolution of experimental practices towards more quantitative and systematic investigations. This trend will arguably persist over the forthcoming years, and, as a result, computation time will remain a serious bottleneck for restoration methods, despite the progress of computer hardware. In this context, advanced (non-linear) restoration methods that were developed for traditional 2D imaging cannot be applied directly; larger-scale problems require more efficient algorithmic implementations.

The concept of "sparsity" has drawn considerable interest recently, leading to a new and successful paradigm for the regularization of inverse problems. The main idea is to constrain the restored image to have only a few non-zero coefficients in a suitable transform domain. Based on this principle, a simple and elegant iterative algorithm—which we shall call the "thresholded Landweber" (TL) algorithm—was independently derived by several research groups [78, 58, 16]. The method has the advantage of being very general. However, it is known to converge slowly when applied to ill-conditioned inverse problems (see [135, 23] and Chapter 5), which restrains its suitability for large data sets.

In the present work, we construct a multilevel version of the TL algorithm that is significantly faster; this allows us to apply the method for the restoration of real 3D multichannel fluorescence micrographs. To do so, we specifically consider the case where the sparsity constraint is enforced in the wavelet domain, which was shown to yield state-of-the-art results for 2D image restoration (see [78]). From a numerical standpoint, the advantage of using wavelet representations is twofold. First, their *tree structure* naturally leads to efficient computational schemes in the spirit of Mallat's Fast Wavelet Transform [141]. Second, the *spectral localization* properties of wavelets make them suitable for preconditioning, that is, for partly

compensating the poor conditioning of the inverse problem.

The structure and the convergence speed of our multilevel algorithm make it comparable to multigrid schemes [97, 32]. These schemes belong to the most efficient known methods for the numerical resolution of partial differential equations; they are typically one order of magnitude faster than standard iterative methods. In fact, the connection between wavelet and multigrid theory was recognized early on [126, 35, 56, 31]. Surprisingly, though, the potential of wavelet-based multilevel methods for image restoration has hardly been exploited so far. An exception is the paper by Wang et al. [244], which is however restricted to linear restoration and based on a relatively empirical reformulation of the image-formation model in the wavelet domain.

Our approach is based on a non-quadratic variational formulation (leading to a non-linear restoration method) and on the principle of bound optimization [125, 111]. This principle also underlies the derivation of [58] and is known under several alternative denominations, such as optimization transfer, surrogate functional optimization or majorize-minimize (MM) strategy.

Our method can be related to the family of "block-alternating MM algorithms" [113]. In the context of statistical signal processing, one of the earliest representatives of this family is the "Space-Alternating Generalized EM" (SAGE) algorithm of Fessler and Hero [76]. More recently, a bound-optimization approach was also used by Oh et al. to derive a multigrid inversion method for non-linear problems [167]. While the works [113, 76, 167] do not involve wavelets, the latter can be related to the so-called lazy wavelet transform [142], which itself corresponds to the anterior concept of hierarchical basis in the finite element and multigrid literature [256, 13]. Similarly, our work can be related to generalizations of the hierarchical-basis method [35, 232].

To achieve our goal, we construct a family of bounds that allow us to divide the original variational problem into a collection of smaller problems, corresponding to the different scales of the wavelet decomposition. The bounds can be made particularly tight for specific subbands. This leads to subband-dependent iteration parameters (step sizes and threshold levels), which are the key to faster convergence. The bound optimization framework provides a rationale for choosing these parameters in a consistent manner. At the same time, this framework is simple to deploy and guarantees that the underlying cost functional is monotonically decreased.

6.1.2 Image-formation model

We will be concerned with the recovery of signals that are distorted by a linear measurement device and noise. Throughout the chapter, we will use a discrete

description where the measured signal is given by the algebraic relation

$$\mathbf{y} = \mathbf{H}\mathbf{x}_{\text{orig}} + \mathbf{b}.$$

Here, the vector $\mathbf{x}_{\text{orig}} \in \mathbb{R}^N$ holds lexicographically ordered samples of the original D-dimensional signal ($N = N_1 \times N_2 \times \ldots \times N_D$ is the product of the number of samples along each dimension). $\mathbf{H}$ is a transform matrix modeling the image-formation device and $\mathbf{b}$ represents the noise component.

The estimation of the original signal $\mathbf{x}_{\text{orig}}$ from the measurement $\mathbf{y}$ is an ill-posed inverse problem [17]. Most approaches for overcoming this ill-posedness can be described in a variational framework, where one looks for an estimate that minimizes a predefined cost functional. This functional is typically the sum of a *data term* and a *regularization term*. Without going into the details of a Bayesian interpretation [158, 78], the former term enforces a certain level of consistency between the estimate and the measured signal (with respect to the image-formation model). The latter term prevents overfitting—and thus instability—by favoring estimates that are close to some desirable class of solutions (according to some regularity measure).

6.1.3 Regularized inversion using a wavelet-domain sparsity constraint

The discovery that natural images can be well approximated using only a few large wavelet coefficients can be traced back to the seminal work of Mallat [141] and is, for example, exploited in the JPEG2000 compression format [45]. Following several recent works (see below), we will use a regularization term that promotes estimates with a sparse wavelet expansion; the data term will be a standard quadratic criterion.

In the sequel, we assume that the reader is familiar with the filter-bank implementation of the wavelet transform [142]. We will denote by $\mathbf{w}$ the vector that contains the coefficients of an estimate $\mathbf{x}$ in a preassigned wavelet basis; we shall refer to this basis as the *synthesis* wavelet basis. Introducing the *synthesis matrix* $\mathbf{W}$, whose columns are the elements of this basis, we can write that

$$\mathbf{x} = \mathbf{W}\mathbf{w}.$$

Later in this chapter, we will also use the *analysis matrix* $\tilde{\mathbf{W}}$, whose columns are the elements of the dual wavelet basis. The perfect-reconstruction condition can be expressed as $\mathbf{W}\tilde{\mathbf{W}}^T = \mathbf{I}$, where ^{T} denotes transposition (or Hermitian transposition in the case of a complex wavelet transform). Note that the present formulation also

includes the case of overcomplete wavelet representations ($\mathbf{W}$ and $\tilde{\mathbf{W}}$ are then non-square matrices).

With these notations, we consider that a solution to the inverse problem is given by $\mathbf{x} = \mathbf{W}\mathbf{w}$, where $\mathbf{w}$ minimizes the functional

$$\mathcal{C}(\mathbf{w}) = \|\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 + \lambda\|\mathbf{w}\|_1. \tag{6.1}$$

Here, $\|\mathbf{w}\|_1$ represents the ℓ_1-norm of the wavelet coefficients, that is, the sum of their absolute values. Compared to the standard Euclidian norm (denoted by $\|\cdot\|_2$), the ℓ_1-norm puts more weight on small coefficients, and less weight on large coefficients. Thus, depending on the magnitude of the regularization parameter λ, it favors estimates whose energy is mostly concentrated in a few large wavelet coefficients. Note that in general the coarsest-scale scaling-function coefficients are not included in the ℓ_1 regularization term (see Section 6.3.3 for more details).

An algorithm for the minimization of (6.1) has been derived in [78, 58, 16], as well as in the earlier works [157, 166]. A similar procedure is also described in [205, 206]. The beauty of the method resides in its simplicity: it essentially consists in alternating between a Landweber iteration [124] and a wavelet-domain thresholding operation [246]—hence the name "thresholded Landweber" (TL) algorithm. When $\mathbf{H}$ is adequately normalized and $\mathbf{W}$ is orthonormal (implying that $\tilde{\mathbf{W}}^T = \mathbf{W}^T$), the TL algorithm can be described by the recursive update rule[1]

$$\mathbf{x} \leftarrow \mathbf{W}\mathcal{T}_{\lambda/2}\Big\{\mathbf{W}^T\Big(\mathbf{x} + \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x})\Big)\Big\}, \tag{6.2}$$

starting from some arbitrary initial estimate $\mathbf{x}$. Here, $\mathcal{T}_\theta\{\cdot\}$ stands for a pointwise application of the well-known soft-thresholding function [68], which can be defined for $w \in \mathbb{C}$ as

$$\mathcal{T}_\theta(w) = \text{sgn}(w)\max(|w| - \theta, 0) \qquad \text{where} \qquad \text{sgn}(w) = \begin{cases} 0 & \text{if } w = 0; \\ \frac{w}{|w|} & \text{otherwise.} \end{cases}$$

The presence of $\mathcal{T}_{\lambda/2}$ in (6.2) guarantees that a certain fraction of wavelet coefficients will be set to zero, depending on the magnitude of the regularization parameter λ.

---

[1]To keep the notations simple, we do not introduce a specific index to distinguish between the individual estimates. Instead we use the assignment operator "$\leftarrow$" whenever a quantity (such as the estimate) is updated. The algorithmic signification of this operator is that the expression on the right-hand side is evaluated and the result is stored in the left-hand side variable.

6.1.4 Recent relevant work and objectives

The present work represents a substantial extension of the algorithm presented in Chapter 5, which was specific to convolutive image-formation operators and to a sparsity constraint in the (bandlimited) Shannon wavelet basis [142]. Here, the goal is to derive a comparably fast algorithm for an arbitrary wavelet basis, without making the assumption that the image-formation operator leaves the different subbands uncoupled. The approach described in the present chapter differs from the previous one in that it is based on a sequential update of the wavelet subbands, instead of a parallel update. This requires a more sophisticated multilevel algorithm.

Similarly to what is done in some presentations of the multigrid methodology—where a "model problem" is often used to convey the intuition [32]—we will motivate and illustrate our approach in the context of deconvolution. In this case, $\mathbf{H}$ can be thought of as a (block-)circulant matrix corresponding to a given convolution kernel; our multilevel method is then particularly efficient, thanks to the shift-invariant structure of the wavelet subspaces. However, its principle can be applied to more general inverse problems. The most direct extension concerns inverse problems for which $\mathbf{H}^T\mathbf{H}$ can be approximated by a convolution matrix—specifically tomographic image-reconstruction, where $\mathbf{H}$ corresponds to a discretized Radon transform. The subclass of inverse problems involving a unitary image-formation operator (such that $\mathbf{H}^T\mathbf{H} = \mathbf{I}$)—e.g., denoising, reconstruction from K-space (frequency-domain) samples or digital holography microscopy [130]—may also benefit from the method. In the present work, we have tried to provide a general and modular pseudo-code description of the multilevel TL algorithm that is readily transposable to machine implementation.

Several works have already extended the standard TL algorithm (which was originally formulated only for orthonormal bases) to more general decompositions, including overcomplete wavelet representations [135, 42]. Nevertheless, the principle and the convergence properties of the algorithm were not fundamentally changed in these settings (although [42] is based on a quite different proximal thresholding interpretation).

Faster methods for the minimization of (6.1) have only been proposed very recently. We are aware of two-step methods [22, 23, 77], line-search methods [72, 73, 79], coordinate-descent methods ([84] and also [72, 73]) and a domain-decomposition method [80]. The latter is based on a well-established concept from the finite-element literature, so that it is arguably the closest to our approach. However it is not specific to wavelets and relies entirely on dimension-reduction effects for decreasing the computational complexity.

The above methods differ with respect to the number and the determination of their step sizes. Among the fixed-step-size strategies, the domain-decomposition approach [80] uses the same step size for all subspaces, whereas the coordinate-descent methods described in [72, 73, 84] use step sizes that are adapted to each atom individually. The methods of Bioucas-Dias, Figueiredo and Nowak [22, 23, 77] have the advantage of simplicity, because they use only two iteration parameters that are also determined a priori (however these parameters may require some hand tuning based on the outcome of a small number of preliminary iterations). The principle of the line-search methods [72, 73, 79] is that the step sizes are adjusted depending on the context, which involves additional computations at every iteration. Our algorithm is somewhere in-between all these approaches: the step sizes are adjusted at the level of individual wavelet subbands, they can be precomputed for a given image-formation operator and wavelet family, and they remain fixed during the entire algorithm.

In summary, to the best of our knowledge, a wavelet-based multilevel method comparable to ours—which combines cyclic updates of the different resolution levels with the preconditioning effect of subband-specific iteration parameters—has not been proposed so far. Therefore, we have chosen to focus on the derivation and the experimental validation of our algorithm. A theoretical study of its convergence properties and a comparison with the aforementioned techniques is a research subject in its own right that will certainly be investigated in the future.

The remainder of the chapter is organized as follows. In Section 6.2, we revisit the derivation of the TL algorithm (6.2) which was presented in [58], introducing additional degrees of freedom into the bound optimization framework. This leads to our multilevel algorithm, described in Section 6.3. Section 6.4 is dedicated to numerical experiments.

6.2 Divide—The thresholded Landweber algorithm, revisited

6.2.1 Notations

In this section, we will primarily be interested in the subspace structure of the wavelet representation. The tree-structure of the wavelet transform—that is, the embedding of the underlying scaling-function subspaces—will become important for the algorithmic considerations of the next section. To account for both aspects, we introduce the following notations, which are illustrated in Fig. 6.1. Throughout this chapter, we shall use the terms "scale", "resolution level", "decomposition level"

and "level" interchangeably.

- J: number of resolution levels of the wavelet representation (j: scale index).

- M_j: number of *wavelet* subbands at scale j, excluding the scaling-function subband (m: subband index).

- $s = (j, m)$: general subband index. Our convention will be that $(j, 0)$ corresponds to the scaling-function subband at scale j; however, for the sake of conciseness, we will often simply write j instead of $(j, 0)$. The context will indicate whether we are referring to the scaling-function subband or to the decomposition level.

- $S_j = \{(j, m), m = 1, \dots, M_j\}$: indexing set for all wavelet subbands at a scale $j \leq J - 1$. At the coarsest level, we include the scaling-function subband: $S_J = \{(j, m), m = 0, \dots, M_J\}$.

- S: indexing set for all subbands produced by a J-scale decomposition (including the coarsest-scale scaling-function subband):

$$S = \bigcup_{j=1}^{J} S_j.$$

- $\mathbf{w}_s$: wavelet or scaling-function coefficients of the current estimate corresponding to subband s. $\mathbf{w}$ is the concatenation of $\mathbf{w}_s$ for every $s \in S$. Note that $\mathbf{w}_0$ is an alias for $\mathbf{x}$.

- $\mathbf{G}_{j,m}$: matrix corresponding to the reconstruction part (upsampling and filtering using the synthesis filters) of the mth channel of the filter bank at scale j (maps the wavelet subband (j, m) into the scaling-function subband at scale $j - 1$).

- $\mathbf{W}_s$: "restriction" of the synthesis matrix $\mathbf{W}$ to subband s, such that

$$\mathbf{W}\mathbf{w} = \sum_{s \in S} \mathbf{W}_s \mathbf{w}_s. \tag{6.3}$$

More precisely, this is a cascade of upsampling and filtering operations defined recursively by

$$\begin{cases} \mathbf{W}_{1,m} = \mathbf{G}_{1,m}; \\ \mathbf{W}_{j,m} = \mathbf{W}_{j-1} \times \mathbf{G}_{j,m} \text{ for } j \geq 2. \end{cases} \tag{6.4}$$
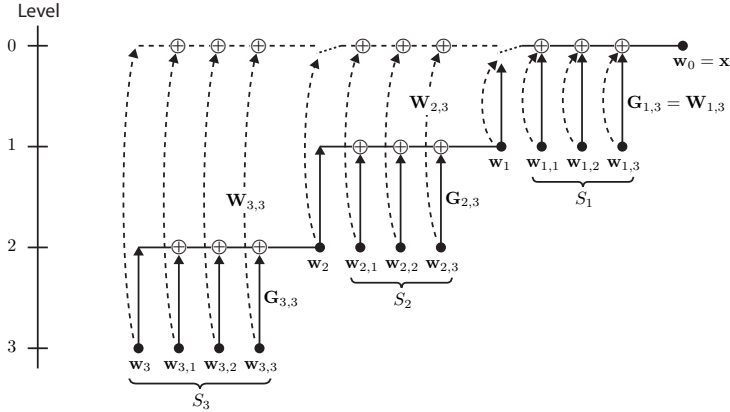
Figure 6.1: Complementary notations reflecting the tree structure (notations associated with continuous lines) and the subspace structure (notations associated with dashed lines) of the wavelet representation. Here the number of decomposition levels is $J = 3$. The number of subbands is $M_j = 3$ at every scale j, which is typical for a 2D separable wavelet representation.

6.2.2 Estimation of the cost functional using subband-dependent bounds

Our algorithm is based on the availability of a wavelet-domain estimation of $\mathbf{H}$ that takes the following form: we assume that there are constants $(\alpha_s)_{s \in S}$ such that

$$\|\mathbf{HWw}\|_2^2 \leq \sum_{s \in S} \alpha_s \|\mathbf{w}_s\|_2^2. \tag{6.5}$$

We shall assume for now that this inequality holds for an arbitrary vector of wavelet coefficients $\mathbf{w}$, and we shall revisit the derivation of the bound-optimization algorithm of Daubechies et al. [58]. Rather than directly considering the original cost functional $\mathcal{C}(\mathbf{w})$, the idea is to iteratively construct a series of auxiliary functionals that are easy to minimize.

Given an estimate of the minimizer of $\mathcal{C}(\mathbf{w})$, say $\mathbf{w}^{\text{old}}$, we define

$$\mathcal{A}(\mathbf{w}) = \sum_{s \in S} \left(\alpha_s \| \mathbf{w}_s^{\text{old}} - \mathbf{w}_s \|_2^2 \right) + \mathcal{C}(\mathbf{w}) - \| \mathbf{H}\mathbf{W}(\mathbf{w}^{\text{old}} - \mathbf{w}) \|_2^2. \qquad (6.6)$$

This functional has three important characteristics:

1. When $\mathbf{w} = \mathbf{w}^{\text{old}}$, $\mathcal{A}(\mathbf{w})$ takes the same value as $\mathcal{C}(\mathbf{w})$.

2. For all other values of $\mathbf{w}$, $\mathcal{A}(\mathbf{w})$ is an upper-bound of $\mathcal{C}(\mathbf{w})$, by virtue of (6.5).

3. $\mathcal{A}(\mathbf{w})$ admits a minimizer with a closed-form expression.

The first two properties imply that, if we find a new estimate $\mathbf{w}^{\text{new}}$ that minimizes (or at least decreases) $\mathcal{A}(\mathbf{w})$, we also decrease $\mathcal{C}(\mathbf{w})$. We simply have to observe that

$$\mathcal{C}(\mathbf{w}^{\text{new}}) \leq \mathcal{A}(\mathbf{w}^{\text{new}}) \leq \mathcal{A}(\mathbf{w}^{\text{old}}) = \mathcal{C}(\mathbf{w}^{\text{old}}).$$

The third property allows us to actually construct such a new estimate. It originates from the negative (rightmost) term in (6.6), which cancels out the coupling of the wavelet coefficients in $\mathcal{C}(\mathbf{w})$. As a result, the auxiliary functional can be rewritten as

$$\mathcal{A}(\mathbf{w}) = c + \sum_{s \in S} \alpha_s \Big(\| \mathbf{w}_s^{\text{old}} + \alpha_s^{-1} \underbrace{\mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{\text{old}})}_{\mathbf{r}_s} - \mathbf{w}_s \|_2^2 + \lambda \alpha_s^{-1} \| \mathbf{w}_s \|_1 \Big),$$

$$(6.7)$$

where the constant c does not depend on $\mathbf{w}$. This expression reveals that the auxiliary functional is essentially a weighted sum of "subfunctionals" that depend on distinct subbands. Furthermore, the wavelet coefficients appear to be completely decoupled in each subfunctional. This means that once we have computed $\mathbf{r}_s = \mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{\text{old}})$ for every subband s, we can minimize each subfunctional using solely pointwise operations.

This minimization procedure can be related to two standard image-restoration methods. First, the computation of $\mathbf{w}_s^{\text{old}} + \alpha_s^{-1} \mathbf{r}_s$ may be seen as a wavelet-domain Landweber iteration [124, 17]: the wavelet decomposition of the "reblurred residual" $\mathbf{H}^T (\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^{\text{old}})$ serves as a correction-term, which is applied with a (subband-dependent) step size α_s^{-1}. Let us point out, however, that the decomposition of the residual must be performed using the *synthesis* basis. Second, each subfunctional can be interpreted as a denoising functional where $\mathbf{w}_s^{\text{old}} + \alpha_s^{-1} \mathbf{r}_s$ represents the wavelet coefficients of a signal to be denoised and $\lambda \alpha_s^{-1}$ is a regularization parameter (again subband-dependent). The minimizer of such a functional is *unique* and is obtained by soft-thresholding the coefficients of the noisy signal, with a threshold level equal to half the regularization parameter [40].

6.2.3 Relation with the standard thresholded Landweber algorithm

Iterating the previous minimization scheme produces a sequence of estimates that are guaranteed to monotonically decrease the cost functional. The procedure can be summarized by the following two-step update rule:

$$\begin{cases} \text{(a) For every } s \in S, \mathbf{r}_s \leftarrow \mathbf{W}_s^T \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}); \\ \text{(b) For every } s \in S, \mathbf{w}_s \leftarrow \mathcal{T}_{\lambda \alpha_s^{-1}/2}\{\mathbf{w}_s + \alpha_s^{-1}\mathbf{r}_s\}. \end{cases} \quad (6.8)$$

Note that the threshold levels must be adjusted proportionally to the inverse of the bound constants.

In particular, when the bounds are the same for all subbands ($\alpha_s = \alpha$ for every s), one obtains the standard "thresholded Landweber" (TL) algorithm. This algorithm uses the same step size (α^{-1}) and the same threshold level ($\lambda \alpha^{-1}/2$) for all subbands. It is relatively easy to obtain an admissible value for α when $\mathbf{W}$ is an orthonormal matrix. We can then write that, for an arbitrary vector of wavelet coefficients $\mathbf{w}$,

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 \leq \rho(\mathbf{H}^T\mathbf{H})\|\mathbf{W}\mathbf{w}\|_2^2 = \rho(\mathbf{H}^T\mathbf{H})\|\mathbf{w}\|_2^2.$$

Here, $\rho(\mathbf{H}^T\mathbf{H})$ denotes the spectral radius of $\mathbf{H}^T\mathbf{H}$; when $\mathbf{H}$ is a convolution matrix, this is simply the maximum over the squared modulus of its frequency response. Thus, for (6.5) to hold, it is sufficient to choose $\alpha_s = \rho(\mathbf{H}^T\mathbf{H})$ for every s. Note that (6.2), which corresponds to $\rho(\mathbf{H}^T\mathbf{H}) = 1$, is a space-domain reformulation of the TL algorithm that is made possible by using an orthonormal basis. This description is quite natural, since eventually we are interested in $\mathbf{x} = \mathbf{W}\mathbf{w}$.

However, we have already mentioned that the TL algorithm converges slowly, especially when the image-formation matrix $\mathbf{H}$ is ill-conditioned. This can be explained intuitively by the fact that using the same bound α for all subbands can only give a very limited account of the spectral characteristics of $\mathbf{H}$. The corresponding auxiliary functionals will thus be relatively poor approximations of the original cost functional, and many intermediate minimization steps will be required before getting a reasonable estimate of the minimizer.

6.2.4 The single-level thresholded Landweber algorithm

Our motivation for introducing subband-dependent bounds is to design auxiliary functionals that better reflect the behavior of the underlying cost functional by exploiting the spectral localization properties of the wavelet basis. Specifically,

we would like to use an estimate (6.5) that is tighter—i.e., that involves smaller constants α_s—than the aforementioned bound for the standard TL algorithm.

In the sequel, $\rho_{s_2\,s_1}$ will denote the largest singular value of the matrix $\mathbf{W}_{s_2}^T\mathbf{H}^T\mathbf{H}\mathbf{W}_{s_1}$. In particular, $\rho_{s\,s} = \rho(\mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_s)$ is the spectral radius of $\mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_s$; when $\mathbf{H}$ is a convolution matrix, this is the upper Riesz bound of the *filtered version* of the wavelet that spans subspace s. Note that $\rho_{s\,s}$ can be significantly smaller than $\rho(\mathbf{H}^T\mathbf{H})$. As an intuitive example, one could imagine the case where $\mathbf{H}$ corresponds to a low-pass filter and s is a high-frequency wavelet subband.

The quantity $\rho_{s\,s}$ is important because it represents a lower limit for α_s. Indeed, for a vector $\mathbf{w}$ with a single non-zero wavelet subband, say $\mathbf{w}_s$, (6.5) reduces to $\|\mathbf{H}\mathbf{W}_s\mathbf{w}_s\|_2^2 \leq \alpha_s\,\|\mathbf{w}_s\|_2^2$. For this inequality to hold for every $\mathbf{w}_s$, we must choose $\alpha_s \geq \rho(\mathbf{W}_s^T\mathbf{H}^T\mathbf{H}\mathbf{W}_s)$.

A particular case arises when the subspaces spanned by the matrices $\mathbf{H}\mathbf{W}_s$ are mutually orthogonal. We can then use exactly the value $\alpha_s = \rho_{s\,s}$, since

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 = \sum_{s\in S}\|\mathbf{H}\mathbf{W}_s\mathbf{w}_s\|_2^2 \leq \sum_{s\in S}\rho_{s\,s}\,\|\mathbf{w}_s\|_2^2.$$

Chapter 5 was based on the fact that the bandlimited Shannon wavelet basis exhibits this decorrelation property with respect to convolution operators. In such a situation we can directly apply algorithm (6.8).

When considering arbitrary wavelet families and image-formation operators, we must a priori bound numerous cross-subband correlation terms, since in general

$$\|\mathbf{H}\mathbf{W}\mathbf{w}\|_2^2 = \sum_{s_1\in S}\sum_{s_2\in S}\langle\mathbf{H}\mathbf{W}_{s_2}\mathbf{w}_{s_2}, \mathbf{H}\mathbf{W}_{s_1}\mathbf{w}_{s_1}\rangle. \tag{6.9}$$

This would require constants α_s that are significantly larger than $\rho_{s\,s}$. However, we can make the following observation: if we impose that $\alpha_s = \rho_{s\,s}$, inequality (6.5) remains valid for all vectors $\mathbf{w}$ that have at most one non-zero subband. This means that (6.6) would define a valid upper-bound of the cost functional under the constraint that $\mathbf{w}$ and $\mathbf{w}^{\text{old}}$ differ by only one subband, that is, under the constraint that we update only one subband at a time.

In practice, owing to the structure of the wavelet representation, it is algorithmically more efficient to be able to update all subbands at a given scale simultaneously. We thus propose to replace (6.8) by

$$\begin{cases} \text{(a) For every } s\in S_j, \mathbf{r}_s \leftarrow \mathbf{W}_s^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}); \\ \text{(b) For every } s\in S_j, \mathbf{w}_s \leftarrow \mathcal{T}_{\lambda\alpha_s^{-1}/2}\{\mathbf{w}_s + \alpha_s^{-1}\mathbf{r}_s\}. \end{cases} \tag{6.10}$$

This choice only requires taking into account correlations between a small number of subbands (those located at the same scale), so that the resulting constants α_s are still close to $\rho_{s\,s}$. More precisely, the following property provides a valid upper-bound under the constraint that we update only a single scale.

Property 10. *If we set*

$$\alpha_s = \sum_{s_0 \in S_j} \rho_{s_0\,s} \tag{6.11}$$

for every s, then inequality (6.5) holds for an arbitrary vector of wavelet coefficients $\mathbf{w}$ *satisfying the following constraint: there is a scale j such that for all subbands* $s \notin S_j$, $\mathbf{w}_s = \mathbf{0}$.

Proof. Equality (6.9) reduces to

$$\|\mathbf{HWw}\|_2^2 = \sum_{s_1 \in S_j} \sum_{s_2 \in S_j} \langle \mathbf{HW}_{s_2}\mathbf{w}_{s_2}, \mathbf{HW}_{s_1}\mathbf{w}_{s_1} \rangle.$$

Combining this with the fact that, for every $\mathbf{w}_{s_1}$, $\mathbf{w}_{s_2}$,

$$|\langle \mathbf{HW}_{s_2}\mathbf{w}_{s_2}, \mathbf{HW}_{s_1}\mathbf{w}_{s_1} \rangle| \le \rho_{s_2\,s_1} \|\mathbf{w}_{s_2}\|_2 \|\mathbf{w}_{s_1}\|_2, \tag{6.12}$$

we obtain

$$\|\mathbf{HWw}\|_2^2 \le \sum_{s_1 \in S_j} \sum_{s_2 \in S_j} \rho_{s_2\,s_1} (\|\mathbf{w}_{s_2}\|_2^2 + \|\mathbf{w}_{s_1}\|_2^2)/2 = \sum_{s \in S} \alpha_s \|\mathbf{w}_s\|_2^2.$$

$\square$

Appendix 6.6.1 describes an algorithm for computing the constants $\rho_{s_0\,s}$ in the convolutive case; for more general operators one may use the power method [91]. Let us emphasize that, under the condition of Property 10, every application of (6.10) is guaranteed to decrease the auxiliary functional (and thus the original cost functional), despite the fact that only a subset of subbands is updated. This follows from (6.7), which shows that the minimization of the auxiliary functional can always be divided into a collection of subband-specific—hence independent—minimization problems. Of course, by letting j vary at every iteration, we can successively update the subbands at all scales. The next section describes an efficient method for doing this.
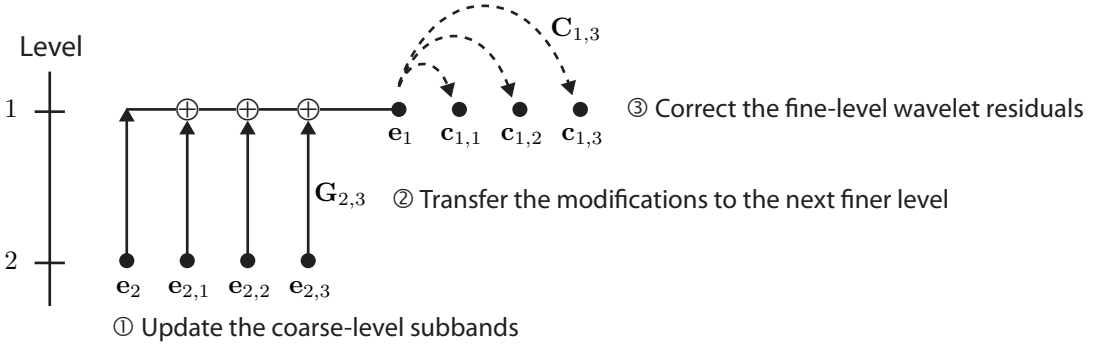
Figure 6.2: Principle of the coarse-to-fine thresholded Landweber (CFTL) algorithm for a two-level decomposition.

6.3 Conquer—The multilevel thresholded Landweber algorithm

The general idea behind our multilevel scheme is to interlace the computation of the residual and the minimization procedure. To give the reader the intuition of this principle, we focus on the description of a simplified strategy that consists in applying (6.10) successively from the coarsest-scale to the finest-scale subbands.

6.3.1 A coarse-to-fine update strategy

Let $\mathbf{r} = \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w})$ be the residual corresponding to the current estimate. Assume that $\mathbf{w}$ is modified at scale j by applying procedure (6.10). In general, this will imply a modification of the residual in all subbands (due to the matrix $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$, which couples the subbands). If the next iteration is performed at scale $j-1$, it is however not necessary to recompute the entire residual; instead, one can simply update the subbands $s \in S_{j-1}$. Denoting by $(\mathbf{e}_s)_{s \in S_j}$ the modifications that have been applied to the estimate, the corresponding correction that must be applied to the residual in a subband $s_0 \in S_{j-1}$ is

$$\mathbf{c}_{s_0} = -\mathbf{W}_{s_0}^T\mathbf{H}^T\mathbf{H}\sum_{s \in S_j}\mathbf{W}_s\mathbf{e}_s = -\mathbf{W}_{s_0}^T\mathbf{H}^T\mathbf{H}\mathbf{W}_{j-1}\sum_{s \in S_j}\mathbf{G}_s\mathbf{e}_s.$$

The above equality stems from the cascade implementation of the wavelet transform—see (6.4). Thus, an updated version of the residual at scale $j - 1$ is obtained as follows:

1. transfer all modifications to the scaling-function subband at the next finer scale $j - 1$;

2. apply the "correction matrices" $\mathbf{C}_s = \mathbf{W}_s^T \mathbf{H}^T \mathbf{H} \mathbf{W}_{j-1}$ for every $s \in S_{j-1}$;

3. subtract the results from the respective subbands.

This principle is illustrated in Fig. 6.2; its recursive application leads to the "coarse-to-fine thresholded Landweber" (CFTL) algorithm. A pseudo-code description is given below. Note that all modifications—including those from subbands located at coarser scales than the current scale j—are progressively transfered to finer scales. The CFTL algorithm depends on the "single-level thresholded Landweber" (SLTL) procedure, which essentially corresponds to the updating rule (6.10). The only difference is that the modifications that are applied to the estimate are stored in intermediate variables so as to be able to update the residual. For simplicity, the variables $\mathbf{w}_s$, $\mathbf{e}_s$ and $\mathbf{r}_s$ are considered to be global (for every subband s) in all pseudo-code descriptions given in this chapter.

We emphasize that the correction steps 2) and 3) above are the only additional operations compared to the standard TL algorithm. These steps should require little computational effort at coarse levels, thanks to the pyramidal structure of (decimated) wavelet representations. In other words, they can be implemented efficiently if the computational complexity of evaluating $\mathbf{H}^T \mathbf{H}$ (the forward image-formation model followed by the corresponding "back-projection") scales well with this data-size reduction.

A particular case arises when $\mathbf{H}^T \mathbf{H}$ is a convolution matrix. The shift-invariant structure of wavelet subspaces then implies that the correction steps essentially reduce to filtering operations. Here we refer to Appendix 6.6.1, which also gives a recursive method for precomputing the correction filters: the procedure is akin to a wavelet decomposition of the convolution kernel corresponding to $\mathbf{H}^T \mathbf{H}$ and is easily implementable in the frequency domain (see also [25]). The correction steps can be implemented with a linear cost provided that we store the DFTs of the individual wavelet subbands; the actual wavelet coefficients are only needed for the thresholding operations and can be computed efficiently using the FFT algorithm. In terms of computational work, one iteration of our algorithm is therefore equivalent to two FFTs per subband, which amounts to two FFTs at the signal level (level 0). The overall complexity of a full coarse-to-fine run is thus on the same

---

Algorithm 1 SLTL(j)

For every $s \in S_j$:

- $\mathbf{e}_s \leftarrow \mathbf{w}_s$

- $\mathbf{w}_s \leftarrow \mathcal{T}_{\lambda \alpha_s^{-1}/2} \left\{ \mathbf{w}_s + \alpha_s^{-1} \mathbf{r}_s \right\}$

- $\mathbf{e}_s \leftarrow \mathbf{w}_s - \mathbf{e}_s$

---

order as one run of the standard TL algorithm, which also requires two FFTs per iteration in the convolutive case.

With a slight anticipation of the next subsection, we conclude this part by noting that multigrid methodologies sometimes advocate the approximate resolution of coarse-level problems [32]. In the particular situations where the wavelet subbands are weakly coupled by the image-formation operator, we have indeed observed that the CFTL algorithm converges even if the correction steps are not applied; that is, if the residual is only updated at the beginning of the iteration loop. This amounts to applying (6.8) using fairly optimistic bound constants—without guarantee that the cost functional is monotonically decreased—and calls for further investigation. This approach may turn out to be useful when dealing with complex image-formation models that can not be evaluated easily at coarse levels.

6.3.2   A general multilevel scheme

With the previous algorithm in mind, one can conceive of more general multilevel strategies for updating the different scales in a more flexible manner. In Appendix 6.6.2, we provide a pseudo-code description of a method that is strongly inspired by the multigrid paradigm. However, there is one fundamental difference: traditional multigrid schemes typically cycle through nested subspaces corresponding to increasingly coarse discretizations of the original inverse problem [253]. In the present context, we successively update the wavelet subbands at every scale; that is, we reinterpret the different scales of the wavelet transform as a multilevel representation of the inverse problem. The corresponding subspaces are not nested—they contain the oscillating components corresponding to the difference between successive coarse-level approximations. Incidentally, early attempts to apply the multigrid paradigm to image-restoration problems remained relatively unsuccessful because they were concentrating on slowly oscillating components [259, 170].

We have tried to specify the "multilevel thresholded Landweber" (MLTL) al-

---

Algorithm 2 CFTL

- Initialization:
  - Choose some initial estimate $\mathbf{x}$
  - Compute its wavelet decomposition: $\mathbf{w} \leftarrow \tilde{\mathbf{W}}^T \mathbf{x}$

- Repeat K times:
  - Compute the residual: for every $s \in S$, $\mathbf{r}_s \leftarrow \mathbf{W}_s^T \mathbf{H}^T (\mathbf{y} - \mathbf{Hx})$
  - Update the subbands from coarse to fine levels, i.e., for $j = J, J-1, J-2, \ldots, 1$:
    - $*$ Update the subbands at the current level: $\mathrm{SLTL}(j)$
    - $*$ Transfer the modifications to the next finer level: $\mathbf{e}_{j-1} \leftarrow \sum_{s \in S_j \cup \{j\}} \mathbf{G}_s \mathbf{e}_s$
    - $*$ If $j \geq 2$, correct the residual for the wavelet subbands at the next finer level:
      for every $s \in S_{j-1}$, $\mathbf{r}_s \leftarrow \mathbf{r}_s - \mathbf{C}_s \mathbf{e}_{j-1}$
  - Set $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{e}_0$

- Return $\mathbf{x}$

---

gorithm in a modular way that is relatively close to machine implementation. Its main building block, UpdateLevel(j), depends on three parameters so as to be able to mimic typical multigrid schemes (see Fig. 6.3). The parameters are thus named η_1, η_2 and μ, following the conventions of the multigrid literature [98, 32]. In the particular case $\eta_1 = 0$, $\eta_2 = 1$ and $\mu = 1$, one retrieves the coarse-to-fine update described in the preceding subsection (Fig. 6.3a). However, we should note that the MLTL algorithm is numerically (slightly) more stable, because the current estimate is explicitly reconstructed from its wavelet coefficients at every iteration.

The different modules of Appendix 6.6.2 can be summarized as follows:

- UpdateResidual(j): updates the residual for the subbands at scale j (if needed). Uses the correction principle of Fig. 6.2 if the wavelet subbands have not been modified so far. Otherwise the update is performed by temporarily moving up to the next finer scaling-function subband.

- UpdateLevel(j): recursive procedure which

(a) Coarse-to-fine ($\mu = 1$; $\eta_1 = 0$; $\eta_2 = 1$)   (b) V-cycles ($\mu = 1$; $\eta_1 = \eta_2 = 1$)

(c) W-cycles ($\mu = 2$; $\eta_1 = \eta_2 = 1$)

Figure 6.3: Examples of multigrid-like updating schemes made possible by the general MLTL algorithm. Each dot corresponds to an application of (6.10) at the corresponding level.

– updates the subbands at coarser scales by calling itself μ times;

– updates the subbands at scale j by calling SLTL(j) (the number of updates before and after a recursive call are fixed by η_1 and η_2 respectively).

Note that the procedure must compute the current residual for the scaling-function subband at scale j before calling itself. This is either done by applying the correction principle of Fig. 6.2 in the opposite direction (from the wavelet subbands to the scaling-function subband), or by going to the next finer scale using UpdateResidual(j).

• MLTL: main routine that performs initialization tasks followed by several iterations of the update procedure. One may devise even more general—e.g. "full multigrid" [97]—schemes by adapting this routine.

6.3.3   A fixed-point property

A comprehensive study of the convergence properties of the MLTL algorithm is well beyond the scope of the present work. In particular, obtaining *tight* theoretical convergence-rate estimates is a difficult problem even for linear subspace-correction

methods [253]. In the next section, we thus propose a numerical study of the convergence rate of the MLTL algorithm, based on the following concise characterization of the minimizer(s) of the cost functional (6.1). We provide a proof in Appendix 6.6.3 for completeness (see also [58] and the general results in [49, 42]).

Property 11. $\mathbf{w}^*$ *is a minimizer of* $\mathcal{C}(\mathbf{w})$ *if and only if it is a fixed point of the standard TL algorithm, that is, if and only if there is an arbitrary step size* $\tau > 0$ *such that* $\mathbf{w}^* = \mathcal{T}_{\lambda\tau/2}\{\mathbf{w}^* + \tau\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)\}$. *Furthermore the minimizer is unique if* $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$ *is positive definite.*

A similar property can be obtained for the MLTL algorithm. This ensures that we obtain a minimizer of the cost functional whenever the MLTL algorithm converges, which was always the case in our numerical experiments (see Section 6.4). In the sequel, $[\mathbf{v}]_n$ stands for the n-th component of a vector $\mathbf{v}$.

Property 12. $\mathbf{w}^*$ *is a minimizer of* $\mathcal{C}(\mathbf{w})$ *if and only if it is a fixed point of the MLTL algorithm; that is, if and only if it is not modified by a sequence of successive applications of* (6.10) *at different scales, such that every subband is updated at least once.*

Proof. • *Necessary part:* $\mathbf{w}^*$ is assumed to be a minimizer of $\mathcal{C}(\mathbf{w})$.
The characterization of Property 11 is equivalent to the following statement: there is a $\tau > 0$ such that, for every component n of $\mathbf{w}^*$,

$$\begin{cases} \text{either } [\mathbf{w}^*]_n = 0 \text{ and } |[\tau\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| \leq \frac{\lambda\tau}{2}; \\ \text{or } [\mathbf{w}^*]_n \neq 0 \text{ and } [\tau\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n = \frac{\lambda\tau}{2}\text{sgn}([\mathbf{w}^*]_n). \end{cases} \quad (6.13)$$

Multiplying by a suitable constant shows that the step size $\tau > 0$ can actually be chosen arbitrarily. In particular, if n corresponds to a wavelet coefficient of subband s, (6.13) also holds for $\tau = \tau_s$. Using this argument for all wavelet coefficients shows that $\mathbf{w}^*$ is a fixed point of (6.10) at any scale j.

• *Sufficient part:* $\mathbf{w}^*$ is assumed to be invariant under a sequence of successive applications of (6.10), such that every scale is visited at least once.
We first observe that, for a given scale j, (6.10) computes the minimizer of the auxiliary functional considered as a function of $(\mathbf{w}_s)_{s \in S_j}$ only. Because this minimizer is unique, the result of applying (6.10) is either to leave the estimate unchanged, or to strictly decrease the auxiliary functional—and thus the original cost functional by construction.

The fixed-point assumption excludes the latter case. Therefore it must be that $\mathbf{w}^*$ is invariant under each individual application of (6.10). Let n be an

arbitrary component of $\mathbf{w}^*$, e.g. corresponding to a subband $s \in S_j$. Since we assume that (6.10) was applied at scale j at least once, (6.13) holds with $\tau = \tau_s$; in fact, it holds for an arbitrary $\tau > 0$. Using this argument for all wavelet coefficients, one retrieves the characterization of Property 11.

<div align="right">□</div>

Before proceeding to the experimental part of this work, we mention two straightforward extensions of the MLTL algorithm. First, it is clear that the algorithm (and the above results) can be extended to a cost functional with *subband-specific* regularization parameters λ_s. In particular, the coarsest-scale saling-function subband $\mathbf{w}_J$ is not thresholded in practice, i.e., $\lambda_J = 0$. Second, one can also replace the ℓ_1-regularization in (6.1) by another coefficient-wise penalization in the wavelet-domain. This will essentially amount to changing the thresholding function; as long as the regularization term is convex, the same bound-optimization framework can be deployed.

6.4   Numerical experiments

In the experiments presented below, we use an ℓ_1 norm for the regularization term. Unless specified otherwise, we use the same regularization parameter for all wavelet subbands; the scaling-function subband is never penalized.

6.4.1   Asymptotic convergence (1D experiments)

To evaluate the convergence behavior of the MLTL algorithm, we designed an experiment where the true minimizer of the cost functional is used as a gold standard. Each test case was constructed as follows. The standard "bumps" signal (Fig. 6.4a) is convolved with an N-periodic low-pass kernel defined by $h[n] \propto \exp\left(-|n|/2\right)$ for $n = -N/2, \ldots, N/2 - 1$; the corresponding convolution matrix $\mathbf{H}$ is normalized such that $\rho(\mathbf{H}^T\mathbf{H}) = 1$. White Gaussian noise is added to the result, so as to simulate a measurement $\mathbf{y}$. The standard TL algorithm (with a step size $\tau = 1$) is then initialized with this measurement and run for 50000 iterations in order to obtain our reference solution $\mathbf{w}^*$. Since $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$ is positive-definite for the convolution kernel defined above (the smallest DFT coefficient of $\mathbf{H}^T\mathbf{H}$ being 0.06), Property 11 implies that this minimizer is unique. Figures 6.4b and 6.4c show an example of the measurement $\mathbf{y}$ and of the corresponding minimizer $\mathbf{x}^*$. Fig. 6.4d shows the locations of the non-zero wavelet coefficients of the solution (we use a 3-level wavelet decomposition and the upper plot represents the finest-resolution subband).

(a) Original signal $\mathbf{x}$

(b) Measurement $\mathbf{y}$

(c) Minimizer $\mathbf{x}^* = \mathbf{W}\mathbf{w}^*$

(d) Active wavelet coefficients of $\mathbf{x}^*$

(e) Cost function: $10\log_{10}(\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*))$
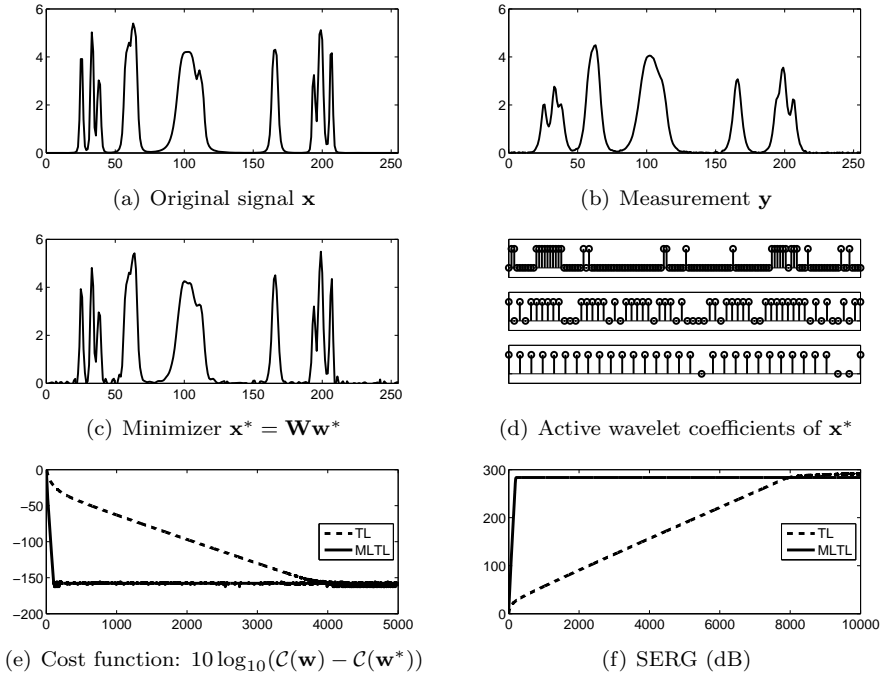
(f) SERG (dB)

Figure 6.4: Experiment with a known minimizer (Symlet8, $\lambda = 0.005$).

We then used this reference to compare the asymptotic behavior of the TL and MLTL algorithms. To this end, we performed a series of experiments where the algorithms are applied to the minimization of (6.1) and initialized with the measurement $\mathbf{y}$. Although the asymptotic convergence rates that are presented here may not be directly relevant to practical situations, they give a quantitative indication of the acceleration potential of the MLTL algorithm. Our asymptotic study required several thousand iterations of the TL and MLTL algorithms in various configurations, which is why we resorted to a small-scale problem ($D = 1$ and $N = 256$). The MLTL algorithm was used with the parameters $\mu = 1$, $\eta_1 = 0$ and $\eta_2 = 1$ (coarse-to-fine strategy). The computational cost of one complete MLTL iteration is then essentially the same as the cost of one TL iteration (each subband is updated once per iteration). This allows for a direct comparison of both algorithms in terms of number of iterations.

The decay of the cost functional towards its minimal value $\mathcal{C}(\mathbf{w}^*)$ is represented in Fig. 6.4e. This decay is only limited by the numerical precision of the computer environment (we used Matlab on a 64-bit Intel Xeon workstation). To reach this limit with the MLTL algorithm, the number of iterations is divided by more than 10 compared to the TL algorithm. We also display the distance between the estimate and the minimizer (Fig. 6.4f); here the "signal-to-error-ratio gain" is defined as SERG $= 20 \log_{10}(\|\tilde{\mathbf{W}}^T \mathbf{y} - \mathbf{w}^*\|_2 / \|\mathbf{w} - \mathbf{w}^*\|_2)$. As expected, both algorithms converge to the minimizer, but the MLTL algorithm is again faster by more than one order of magnitude for reaching the level of numerical precision.

To obtain a more quantitative insight, we repeated the experiment in several test cases and computed the slope of the SERG curves between 100 and 250 dB. This measurement gives an estimate of the asymptotic convergence rate, in dB per iteration. The results are summarized in Table 6.1, for various orthonormal wavelet bases and different values of the regularization parameter, corresponding to different noise levels (from top to bottom, the values of λ correspond to BSNR noise levels of 60, 50, 40, 30, 20, 10 dB respectively—see Chapter 5 for the definition of BSNR).

For validation purposes, we computed a theoretical convergence-rate estimate in the case where $\lambda = 0$. With this particular choice, both algorithms reduce to linear (least squares) restoration procedures: since the thresholding step disappears, the TL algorithm reduces to the standard Landweber iteration and the MLTL algorithm corresponds to a wavelet-based multilevel implementation of the Landweber iteration. For this type of linear iterations, the asymptotic convergence rate can be estimated using the spectral radius of the so-called iteration matrix (see e.g. [98]). This spectral radius can be obtained directly for the TL algorithm because we consider orthonormal wavelet bases (the computation is essentially the same as

λ	Cv. rates	Algo.	Haar	Daub2	CubicSpline	Sym8	Shannon
0	Theoretical	TL	0.031	0.031	0.031	0.031	0.031
		MLTL	0.376	0.761	1.302	1.301	1.301
0	Measured	TL	0.032	0.032	0.032	0.032	0.032
		MLTL	0.383	0.769	1.321	1.321	1.321
0.00025	Measured	TL	0.033	0.032	0.032	0.032	0.032
		MLTL	0.393	0.803	1.324	1.331	1.335
0.002	Measured	TL	0.032	0.033	0.032	0.032	0.033
		MLTL	0.408	0.818	1.336	1.326	1.379
0.01	Measured	TL	0.039	0.038	0.035	0.033	0.038
		MLTL	0.570	1.178	1.347	1.368	1.486
0.05	Measured	TL	0.035	0.058	0.051	0.055	0.051
		MLTL	0.433	1.317	1.559	1.512	1.595
0.25	Measured	TL	0.132	0.173	0.154	0.184	0.165
		MLTL	1.054	1.570	1.837	2.013	1.978

Table 6.1: Convergence rates (in dB per iteration) for different values of the regularization parameter and various orthonormal wavelet bases.

in Section 6.2.3). For the MLTL algorithm, the small dimension of the problem allows us to explicitly construct the iteration matrix in order to evaluate its spectral radius. The resulting theoretical convergence-rate estimates (expressed in dB per iteration for comparison purposes) are reported in the first row of Table 6.1.

The theoretical and the measured values are in good agreement for $\lambda = 0$, suggesting that our experimental method for measuring the asymptotic convergence rate is reliable. The results for $\lambda \neq 0$ corroborate the former observation (see [78] and Chapter 5) that the TL algorithm tends to converge faster for higher values of λ. This can be explained by the fact that the variational problem is more constrained, thus compensating for the unfavorable conditioning of the convolution kernel. Nevertheless, the convergence rates of the MLTL algorithm are consistently one order of magnitude larger than those of the TL algorithm. The figures suggest that the strongest acceleration is generally obtained for higher-order wavelets, which can be related to their improved frequency selectivity. The Shannon wavelet basis provides perfect frequency selectivity, a property that was exploited in Chapter 5.

(a) Original   (b) Measured   (c) Restored (MLTL after 1
                                                        sec.)

(d) SERG vs. computation time   (e) SERG vs. computation time (zoom)
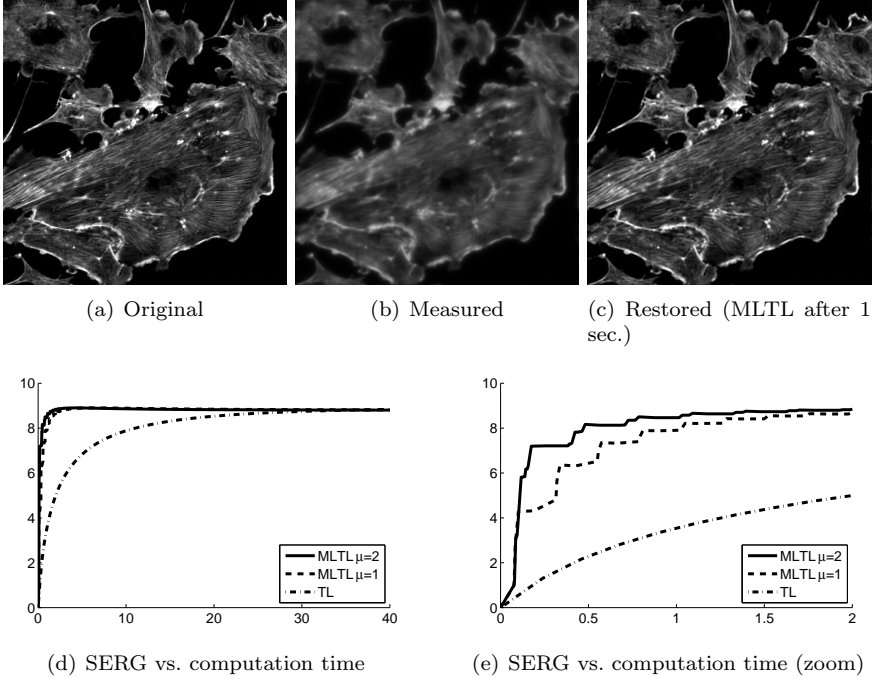
Figure 6.5: Computation time comparison (in seconds).

6.4.2   Computation time (2D experiments)

In our second series of experiments, we evaluated the performance of the MLTL
algorithm in terms of computation time. This type of assessment is most relevant in
practical situations, but it depends on computer hardware parameters. Therefore
we always provide a comparison with the standard TL algorithm.

We first simulated the effect of a defocusing blur on a 512×512 test image (Fig.
6.5a). We used a standard diffraction-limited point spread function (PSF) model for
widefield fluorescence microscopy [236]. The result was then corrupted by additive
white Gaussian noise with a BSNR of 40 dB (Fig. 6.5b). We restored this simulated
measurement using the TL and MLTL algorithms. Both were initialized with the
measurement. We used a separable orthonormalized cubic spline wavelet basis
with four decomposition levels. The regularization parameter $\lambda = 0.2$ was the
same for both algorithms; it was adjusted using multiple trials, so as to give the

Wavelet basis	MLTL SERG after 1 sec.	TL computation time to reach same SERG	MLTL SERG after 4 sec.	TL computation time to reach same SERG
Haar	7.13 dB	31.7 sec.	7.53 dB	41.6 sec.
9/7	6.96 dB	31.6 sec.	7.30 dB	46.0 sec.
Shannon	6.50 dB	29.5 sec.	6.74 dB	41.3 sec.

Table 6.2: Computation time required to reach a given level of restoration quality, for the second experiment of Section 6.4.2.

best restoration quality after the MLTL algorithm had converged. Figures 6.5d and 6.5e show the evolution of the restoration quality measure SERG $= 20 \log_{10}(\|\mathbf{y} - \mathbf{x}_{\mathrm{orig}}\|_2 / \|\mathbf{x} - \mathbf{x}_{\mathrm{orig}}\|_2)$, where $\mathbf{x}_{\mathrm{orig}}$ stands for the original signal and $\mathbf{x} = \mathbf{W}\mathbf{w}$ is the estimate. One can observe that the coarse-to-fine MLTL algorithm requires 1 second of computations to reach an improvement of 8 dB (result shown in Fig. 6.5c). The TL algorithm needs approximately 10 seconds to reach the same figure. We found that the performance of the MLTL algorithm can be further improved by using $\mu = 2$, i.e., with a modified W-cycle iteration. This makes sense since natural images tend to have mostly low-frequency content; thus, iterating on coarse-scale subbands brings the largest improvement in the beginning, unless the algorithm is initialized with a very accurate estimate.

We used a similar protocol for the second part of our 2D experiments, where we replicated the standard test case used by Figueiredo and Nowak in [78] (Cameraman image convolved with a 9×9 uniform blur; additive white Gaussian noise with a BSNR of 40 dB; initialization with a Wiener-type estimate). In particular, we introduced a random shift of the estimate at the beginning of every TL and MLTL iteration; the authors found that this method gave optimal results with non-translation-invariant wavelet transforms. We present results for three wavelet bases, including biorthogonal 9/7 wavelets. We always used 3 decomposition levels and the same regularization parameter $\lambda = 0.04$. The second and fourth columns of Table 6.2 show the restoration quality of the MLTL algorithm after 1 and 4 seconds, respectively. The third and fifth columns give the minimum TL computation time that is required to reach the same restoration quality. Again, the MLTL algorithm provides an acceleration of roughly one order of magnitude. Our results confirm the superiority of the Haar basis among separable wavelet bases for 2D image restoration; this fact is already known from denoising applications [24]. In

(a) Original widefield stack (input data)

(b) Deconvolution result after 15 TL iterations

(c) Deconvolution result after 15 MLTL iterations

(d) Confocal reference stack
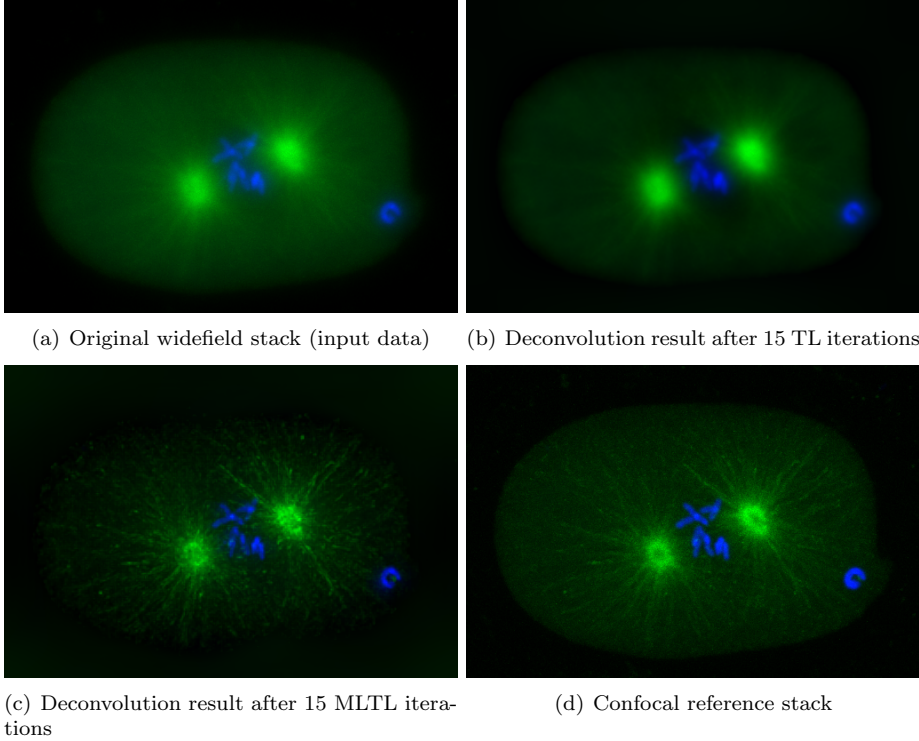
Figure 6.6: 3D deconvolution results (maximum-intensity projections of 512×352×96 image stacks).

general, the ability to use other wavelet bases than the Shannon wavelet basis leads to substantial improvements over our previous work (Chapter 5). In summary, the MLTL algorithm can yield state-of-the-art results (similar to those obtained in [78]) in a substantially shorter time than the TL algorithm.

6.4.3 Application to real fluorescence-microscopy data (3D experiments)

To conclude this experimental part, we applied the MLTL algorithm to real 3D fluorescence microscopy data. Similarly to our experiment in Chapter 5, we ac-

quired two image-stacks of the same sample (a C. Elegans embryo), one of them serving as a visual reference to assess the restoration quality. For the present work we acquired a much larger, two-channel data set.

Both data sets were acquired on a confocal microscope, which has the ability to reject out-of-focus light using a small aperture in front of the detector. This creates a relatively sharp but noisy image (Fig. 6.6d). When the aperture is opened the signal intensity is improved, but the measurement gets blurred by the contributions of defocused objects. This results in hazy images that are characteristic for widefield microscopes (Fig. 6.6a).

We applied the MLTL algorithm to the widefield-type stack, with three decomposition levels. To account for the anisotropic sampling scheme of the microscope, we used an orthonormalized linear spline wavelet for the X-Y dimensions, and a a Haar wavelet for the Z dimension. We kept the random-shift method of [78] and we used scale-dependent regularization parameters that were adjusted with the confocal stack as a visual reference. The channels were processed independently, using computer-generated PSFs based on a three-dimensional version of the diffraction-limited model used in the previous subsection. The parameters of this model were adjusted according to manufacturer-provided specifications of the objective, the immersion oil and the fluorescent dyes (NA, refractive index, emission wavelength).

The result is shown in Fig. 6.6c: the restored image-stack provides significantly better contrast than the original widefield image, especially for the filaments (green channel). The chromosomes and their centromeres (blue channel) appear almost as sharp as in the confocal image.

For completeness, we have included the result of the TL algorithm after the same number of iterations (Fig. 6.6b). It is seen that the resulting image-stack is still very hazy; the TL algorithm fails to produce a visible deconvolution effect within the assigned budget of iterations. The computation time was on the order of 5 minutes for both algorithms.

6.5   Discussion and conclusion

We have presented a wavelet-based multilevel image-restoration algorithm inspired from multigrid techniques. The method is one order of magnitude faster than the standard algorithm for sparsity-constrained restoration, whose results belong to the state-of-the-art in the field of image processing.

The MLTL algorithm allows for typical multigrid iteration schemes such as V-cycles and W-cycles. However, it differs from textbook multigrid schemes, which iterate on nested subspaces corresponding to different "resolution levels" of the in-

verse problem. Our algorithm, by contrast, iterates on the wavelet subbands, which are the complements of the standard multigrid spaces. Yet our algorithm takes advantage of the underlying multiresolution structure, which greatly contributes to the efficiency of the method. The other key point is the preconditioning effect of wavelets.

We have provided theoretical convergence rates for the linear parts of the TL and MLTL algorithms, giving a quantitative insight into the convergence acceleration of the latter. Our experimental results show that one can achieve the same kind of acceleration in the general non-linear case (with thresholding).

Our method is directly applicable to separable wavelet bases in arbitrary dimensions. We obtained promising results in the context of 3D fluorescence microscopy using such bases, extending the work of Chapter 5. Nevertheless we have tried to provide a sufficiently general description that should require little adaptation for more "exotic" wavelet representations, e.g. with quincunx subsampling schemes [223], or non-stationary refinement filters [118, 237]. The algorithm is readily implementable using our modular specification and standard wavelet-decomposition/reconstruction building blocks.

We are currently investigating the benefits of the MLTL algorithm for redundant wavelet representations. We have not specifically explored this possibility here because we were primarily interested in high-dimensional inverse problems that do typically not allow for redundant decompositions. We have already obtained promising results for medical applications (specifically fMRI signal restoration and tomographic image reconstruction) which will be the subject of forthcoming reports.

6.6   Appendix

6.6.1   A method for precomputing the bound constants and the correction filters

To keep the presentation simple, we will consider the one-dimensional ($D = 1$) situation where $\mathbf{H}^T\mathbf{H}$ is a (positive) circulant matrix. Its eigenvalues (DFT coefficients) are real and positive and can thus be denoted $|\hat{h}[\nu]|^2$, $\nu = 0, \ldots, N - 1$. Furthermore, we will consider a wavelet decomposition with a dyadic subsampling scheme. The method presented below can easily be extended to higher dimensions and more general wavelet representations.

To compute the bound constants defined in (6.11), one must essentially estimate

the inner products of (6.12). These can be rewritten in the frequency domain as

$$\langle \mathbf{HW}_{s_2} \mathbf{w}_{s_2}, \mathbf{HW}_{s_1} \mathbf{w}_{s_1} \rangle = \frac{1}{N} \sum_{\nu=0}^{N-1} |\hat{h}[\nu]|^2 \, \hat{\psi}^*_{s_2}[\nu] \, \hat{w}^*_{s_2}[\nu] \, \hat{\psi}_{s_1}[\nu] \, \hat{w}_{s_1}[\nu],$$

where we use the following conventions.

- $\hat{w}_s[\nu]$ denotes the DFT of the wavelet coefficients corresponding to a subband s at a given level j. Since we assume a dyadic subsampling scheme, $\hat{w}_s[\nu]$ can be seen as an N_j-periodic sequence, with $N_j = N/2^j$.

- $\hat{\psi}_s[\nu]$ denotes the DFT of the wavelet or scaling function that spans the subspace associated with subband s. Note that if we define $\hat{\psi}_0[\nu] = 1$, the discrete version of the standard scaling relation [142] can be stated as

$$\hat{\psi}_s[\nu] = \hat{g}_s[\nu]\hat{\psi}_{j-1}[\nu] \text{ for } s \in S_j \cup \{j\}, \tag{6.14}$$

where $\hat{g}_s[\nu]$ is the N_j-periodic filter corresponding to $\mathbf{G}_s$ (see below).

Our multilevel method only requires the explicit value of the constants $\rho_{s_2 s_1}$ when s_2 and s_1 are subbands located at the same level, that is, when $s_2, s_1 \in S_j \cup \{j\}$ for some j. In this case $\hat{w}_{s_2}[\nu]$ and $\hat{w}_{s_1}[\nu]$ have the same period N_j. We can thus write that

$$\langle \mathbf{HW}_{s_2} \mathbf{w}_{s_2}, \mathbf{HW}_{s_1} \mathbf{w}_{s_1} \rangle = \frac{1}{N_j} \sum_{\nu=0}^{N_j-1} \hat{w}^*_{s_2}[\nu] \, \hat{w}_{s_1}[\nu] \hat{c}_{s_2 s_1}[\nu],$$

where

$$\hat{c}_{s_2 s_1}[\nu] = \frac{1}{2^j} \sum_{k=0}^{2^j-1} |\hat{h}[\nu + kN_j]|^2 \, \hat{\psi}^*_{s_2}[\nu + kN_j] \, \hat{\psi}_{s_1}[\nu + kN_j]. \tag{6.15}$$

Inequality (6.12) is then obtained by defining $\rho_{s_2 s_1} = \max_\nu |\hat{c}_{s_2 s_1}[\nu]|$.

When $\mathbf{H}^T\mathbf{H}$ is a circulant matrix and $s \in S_j \cup \{j\}$, the matrix $\mathbf{C}_s = \mathbf{W}_s^T \mathbf{H}^T \mathbf{H} \mathbf{W}_j$ is also circulant and its DFT coefficients are precisely given by $\hat{c}_{s\,j}[\nu]$, for $\nu = 0, \ldots, N_j$. This property allows for an efficient frequency-domain implementation of the residual correction steps in the CFTL and MLTL algorithms.

To prove the property, we introduce the general notation $\mathbf{C}_{s_2 s_1} = \mathbf{W}_{s_2}^T \mathbf{H}^T \mathbf{H} \mathbf{W}_{s_1}$ for arbitrary subbands s_2, s_1. We also define $\mathbf{C}_{0\,0} = \mathbf{H}^T\mathbf{H}$. We can then proceed by recurrence. For $j = 0$, $\mathbf{C}_{0\,0}$ is circulant and its DFT coefficients are given by

(6.15) with $s_2 = s_1 = 0$. For $j \geq 1$, we assume that $\mathbf{C}_{j-1\,j-1}$ is a circulant matrix whose DFT coefficients are $\hat{c}_{j-1\,j-1}[\nu]$. The cascade structure of the wavelet representation (6.4) implies that, for $s_2, s_1 \in S_j \cup \{j\}$,

$$\mathbf{C}_{s_2\,s_1} = \mathbf{G}_{s_2}^T \mathbf{C}_{j-1\,j-1} \mathbf{G}_{s_1}.$$

For a given subband s, the algorithmic interpretation of $\mathbf{G}_s$ is 1) dyadic upsampling, followed by 2) filtering with $\hat{g}_s[\nu]$. Its transpose $\mathbf{G}_s^T$ stands for 1) filtering with $\hat{g}_s[\nu]^*$, followed by 2) dyadic downsampling. Therefore, $\mathbf{C}_{s_2\,s_1}$ is also a circulant matrix with DFT coefficients

$$\frac{1}{2} \sum_{k=0}^{1} \hat{g}_{s_2}[\nu + kN_j]^* \, \hat{c}_{j-1\,j-1}[\nu + kN_j] \, \hat{g}_{s_1}[\nu + kN_j] = \hat{c}_{s_2\,s_1}[\nu]. \qquad (6.16)$$

The equality stems from definition (6.15) for $\hat{c}_{j-1\,j-1}[\nu]$, and from the scaling relation (6.14); this completes the proof by recurrence.

Note that relation (6.16) provides a way to recursively compute the filters $\hat{c}_{s_2\,s_1}[\nu]$ and the corresponding constants $\rho_{s_2\,s_1}$ (with $s_2, s_1 \in S_j \cup \{j\}$).

6.6.2   Pseudo-code description of the general MLTL algorithm

As in the previous subsection, we use the notation $\mathbf{C}_{s_2\,s_1} = \mathbf{W}_{s_2}^T \mathbf{H}^T \mathbf{H} \mathbf{W}_{s_1}$. The general MLTL algorithm uses both the matrices $\mathbf{C}_{j\,s}$ and $\mathbf{C}_{s\,j}$, for $s \in S_j \cup \{j\}$. It is useful to observe that $\mathbf{C}_{j\,s} = \mathbf{C}_{s\,j}^T$: in the convolutive case, this implies that $\hat{c}_{j\,s}[\nu] = \hat{c}_{s\,j}[\nu]^*$.

Algorithm 3 UpdateResidual(j)

- If $\mathbf{e}_s \neq \mathbf{0}$ for some $s \in S_j$:
    - $\mathbf{r}_{j-1} \leftarrow \mathbf{r}_{j-1} - \mathbf{C}_{j-1} \sum_{s \in S_j} \mathbf{G}_s \mathbf{e}_s$
    - For every $s \in S_j \cup \{j\}$, $\mathbf{r}_s \leftarrow \mathbf{G}_s^T \mathbf{r}_{j-1}$
- Otherwise, if $\mathbf{e}_j \neq \mathbf{0}$: for every $s \in S_j$, $\mathbf{r}_s \leftarrow \mathbf{r}_s - \mathbf{C}_{s\,j} \mathbf{e}_j$

Algorithm 4 UpdateLevel(j)

- Initialization:
  - For every $s \in S_j \cup \{j\}$, $\mathbf{e}_s \leftarrow \mathbf{0}$
  - For every $s \in S_j \cup \{j\}$, $\mathbf{r}_s \leftarrow \mathbf{G}_s^T \mathbf{r}_{j-1}$
- Repeat μ times:
  - Repeat η_1 times:
    * UpdateResidual(j)
    * SLTL(j)
  - If $j < J$:
    * If $\mathbf{e}_s \neq \mathbf{0}$ for some $s \in S_j$:
      · If $\mathbf{e}_j \neq \mathbf{0}$, UpdateResidual($j$)
      · Otherwise, $\mathbf{r}_j \leftarrow \mathbf{r}_j - \sum_{s \in S_j} \mathbf{C}_{j\,s} \mathbf{e}_s$
    * $\mathbf{e}_j \leftarrow \mathbf{w}_j$
    * UpdateLevel($j+1$)
    * $\mathbf{e}_j \leftarrow \mathbf{w}_j - \mathbf{e}_j$
  - Repeat η_2 times:
    * UpdateResidual(j)
    * SLTL(j)
- $\mathbf{w}_{j-1} \leftarrow \sum_{s \in S_j \cup \{j\}} \mathbf{G}_s \mathbf{w}_s$

Algorithm 5 MLTL

- Initialization:
  - Choose some initial estimate $\mathbf{x}$ and set $\mathbf{w}_0 \leftarrow \mathbf{x}$
  - Compute its wavelet decomposition (keeping the coarse approximations): for $j = 1, \ldots, J$, for every $s \in S_j \cup \{j\}$, $\mathbf{w}_s = \tilde{\mathbf{G}}_s^T \mathbf{w}_{j-1}$
- Repeat K times:
  - $\mathbf{r}_0 \leftarrow \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{w}_0)$
  - UpdateLevel(1)
- Set $\mathbf{x} \leftarrow \mathbf{w}_0$ and return $\mathbf{x}$

6.6.3　Proof of Property 11

A short computation reveals that

$$\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*) = \|\mathbf{HW}(\mathbf{w} - \mathbf{w}^*)\|_2^2 - 2\Re\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*), \mathbf{w} - \mathbf{w}^*\rangle + \lambda\|\mathbf{w}\|_1 - \lambda\|\mathbf{w}^*\|_1, \tag{6.17}$$

where $\mathbf{w}$ and $\mathbf{w}^*$ are arbitrary vectors of wavelet coefficients.

- *Necessary part:* we assume that $\mathcal{C}(\mathbf{w}) \geq \mathcal{C}(\mathbf{w}^*)$ for every $\mathbf{w}$.

  - Suppose that $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*)]_n| > \lambda/2$ for some n. Given a real and strictly positive constant ϵ, we define the vector $\mathbf{w}$ by

    $$[\mathbf{w}]_{n'} = \begin{cases} [\mathbf{w}^*]_{n'} + \epsilon\,\mathrm{sgn}\,[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*)]_n & \text{if } n' = n; \\ [\mathbf{w}^*]_{n'} & \text{otherwise.} \end{cases}$$

    In view of (6.17), ϵ can always be chosen such that $\mathcal{C}(\mathbf{w}) < \mathcal{C}(\mathbf{w}^*)$, a contradiction. Therefore it must be that $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*)]_n| \leq \lambda/2$ for every n.

  - Choosing $\epsilon \in\,]0, 1]$ and inserting $\mathbf{w} = (1 - \epsilon)\mathbf{w}^*$ into (6.17) gives the necessary condition

    $$\epsilon^2\|\mathbf{HWw}^*\|_2^2 + 2\epsilon\Re\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*), \mathbf{w}^*\rangle - \epsilon\lambda\|\mathbf{w}^*\|_1 \geq 0.$$

    If it were true that $2\Re\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*), \mathbf{w}^*\rangle - \lambda\|\mathbf{w}^*\|_1 < 0$, we could find a sufficiently small ϵ such that this necessary condition is violated. Thus it must be that $2\Re\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*), \mathbf{w}^*\rangle - \lambda\|\mathbf{w}^*\|_1 \geq 0$. Since $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*)]_n| \leq \lambda/2$ for every n, it follows that $[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*)]_n = \lambda/2\,\mathrm{sgn}\,[\mathbf{w}^*]_n$ whenever $[\mathbf{w}^*]_n \neq 0$.

  The combination of both results is equivalent to the fixed-point property.

- *Sufficient part:* $\mathbf{w}^*$ is assumed to be a fixed point of the TL algorithm. We use the same equivalence:

  - Since $[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*)]_n = \lambda/2\,\mathrm{sgn}\,[\mathbf{w}^*]_n$ whenever $[\mathbf{w}^*]_n \neq 0$, we know that $2\Re\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*), \mathbf{w}^*\rangle - \lambda\|\mathbf{w}^*\|_1 = 0$ and (6.17) reduces to

    $$\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*) = \|\mathbf{HW}(\mathbf{w} - \mathbf{w}^*)\|_2^2 - 2\Re\langle \mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{HWw}^*), \mathbf{w}\rangle + \lambda\|\mathbf{w}\|_1.$$

– Since $|[\mathbf{W}^T\mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{w}^*)]_n| \leq \lambda/2$ for every n, it follows that

$$\mathcal{C}(\mathbf{w}) - \mathcal{C}(\mathbf{w}^*) \geq \|\mathbf{H}\mathbf{W}(\mathbf{w} - \mathbf{w}^*)\|_2^2 \geq 0.$$

This also shows the unicity of the minimizer when $\mathbf{W}^T\mathbf{H}^T\mathbf{H}\mathbf{W}$ is positive definite.

Part III

Application of risk-estimation methods to wavelet-based image restoration

Estimations are approximations.
Stéphane Mallat

Chapter 7

An introduction to risk-estimation techniques under quadratic loss

In the previous chapters, our starting point was always the variational definition of a restoration method. We will now change perspective and focus on the functional specification of a given restoration algorithm. In general, the restoration result depends on the measured image $\mathbf{y}$ and on a parameter θ (or several parameters, in which case θ is vector-valued). For example, θ could represent the regularization parameter, the number of iterations and/or the step size. Thus, we shall think of the restoration algorithm as a function $\mathbf{f}_\theta(\mathbf{y})$.

In this part, we will be concerned with the problem of selecting a "good" parameter θ. This is a recurring and important issue in image restoration. Our assumption is that the original image $\mathbf{x}$ is *deterministic*; only the measurement $\mathbf{y}$ is viewed as a random variable, which models the uncertainty introduced by the measurement process. In this sense, we take a step back from standard Bayesian approaches that assume a probabilistic model on the original image.

Nevertheless, in the deterministic framework of approximation theory, the methods presented in Part I—in particular wavelet-domain regularization (see Chapter (II))—can still be justified. In this sense, we use approximation theory and estimation theory as complementary tools. Interestingly, both theories can be formalized using the same "calculus" (functional analysis and in particular Fourier theory), which lets them appear similar in many ways.

7.1 Probabilistic framework and statistical concepts

7.1.1 Remarks on measure and distribution theory

The classical foundation of probability is measure theory in its abstract set-theoretic formulation [213]. A probability density function (PDF) is then viewed as a measure, that is, a countably additive function over some σ-algebra.

Alternatively (under additional topological assumptions), measure theory can also be developed in the framework of functional analysis [26, 219]. This is made possible by the Riesz representation theorem, which establishes a correspondence between finite measures[1] and continuous linear functionals over the space $\mathcal{C}_0^0$ of compactly-supported continuous functions.

In the sequel we will use the latter functional-analytic approach, since our discussion will be based on an operator formalism. In particular we will think of a PDF as a *distribution*, in the sense of generalized function. Note that the restriction of a measure to the space $\mathcal{C}_0^\infty$ of compactly-supported infinitely-differentiable functions defines a distribution [195]. Since $\mathcal{C}_0^\infty$ is dense in $\mathcal{C}_0^0$, two measures are equal if and only if they coincide on $\mathcal{C}_0^\infty$.

Similarly, a tempered measure[2] defines a distribution over the Schwartz space $\mathcal{S}$ of infinitely-differentiable functions with rapid decay [95]. Since $\mathcal{S}$ contains $\mathcal{C}_0^\infty$, a tempered measure is entirely characterized by its values on $\mathcal{S}$.

Based on these considerations, we introduce the following definition.

Definition 4. *A distribution $p \in \mathcal{S}'$ is a* probability distribution *if it is a positive measure such that* 1 *is p-integrable and $\int 1 \mathbf{d}p = 1$.*

From now on, we will systematically identify the random vector $\mathbf{y} \in \mathbb{R}^N$ with its probability distribution $p \in \mathcal{S}'(\mathbb{R}^N)$. The expectation operator is then defined as an integral with respect to the measure p. For a p-integrable function f (not necessarily in $\mathcal{S}$), we shall extend the bracket notation as follows:

$$\langle p, f \rangle = \int f \mathbf{d}p = \mathsf{E}[f(\mathbf{y})].$$

The probability distribution of $\mathbf{y}$ depends on the original image $\mathbf{x}$. We will use $p_{\mathbf{x}}$ instead of p to make this dependence apparent.

---

[1]A measure μ is finite if $|\mu(K)| < +\infty$ whenever K is compact. In the sequel, we use the word measure to refer to a finite measure.

[2]A measure μ is tempered if the function $\mathbf{y} \mapsto (1 + \|\mathbf{y}\|_2^2)^{-p}$ is μ-integrable for some $p > 0$. With respect to Definition 4, note that if 1 is μ-integrable then μ is tempered.

7.1.2 Parametric estimation under quadratic loss

The problem of parameter adjustment depends on the choice of a performance measure. In this work, we adopt the mean squared error (MSE) which is arguably the most widespread measure of restoration quality in image processing. The popularity of the MSE can be explained by its good mathematical properties, which greatly facilitate its manipulation and its optimization in comparison with other quality measures.

In statistics, the measure of restoration quality is called *loss function*. For the MSE, the definition is

$$\ell_\theta(\mathbf{y}) = \frac{1}{N}\|\mathbf{x} - \mathbf{f}_\theta(\mathbf{y})\|_2^2.$$

Ideally, we would like to adjust the parameter θ such that the loss function is minimized. With the exception of synthetic experiments, however, the exact value of the loss function is not available, since it requires the knowledge of the original image $\mathbf{x}$. For this reason, the loss function is termed an *oracle* quantity. Similarly, an adjustment of θ based on the knowledge of $\mathbf{x}$ can be qualified as an oracle adjustment.

A common approach in estimation theory is thus to replace the loss function by its statistical expectation, which is also known as the *risk*:

$$r_\theta = \mathsf{E}[\ell_\theta(\mathbf{y})].$$

This can be justified (at least in simple cases) by the law of large numbers: for large-sample data, the loss function is close to its statistical average. While the risk is not directly accessible either, one can usually construct a statistical estimate of this quantity using a realization of $\mathbf{y}$ (again based on the law of large numbers). Therefore, minimizing this *risk estimate* should provide a good approximation to an oracle adjustment.

The risk can be decomposed as follows:

$$r_\theta = \frac{1}{N}\Big(\mathsf{E}[\|\mathbf{x}\|_2^2] - 2\mathsf{E}[\mathbf{x}^T\mathbf{f}_\theta(\mathbf{y})] + \mathsf{E}[\|\mathbf{f}_\theta(\mathbf{y})\|_2^2]\Big).$$

Note that the first term $\mathsf{E}[\|\mathbf{x}\|_2^2]$ does not depend on θ and is thus irrelevant for optimizing this parameter. The last term $\mathsf{E}[\|\mathbf{f}_\theta(\mathbf{y})\|_2^2]$ can be estimated empirically from the data by computing $\|\mathbf{f}_\theta(\mathbf{y})\|_2^2$. Only the cross-term $\mathsf{E}[\mathbf{x}^T\mathbf{f}_\theta(\mathbf{y})]$ is problematic because it depends on the unknown $\mathbf{x}$. In the remainder, we will thus focus on the estimation of this cross-term. To simplify the presentation, we omit the index θ.

7.2   Operator formulation of risk estimation

7.2.1   The scalar case

We will first consider the case where $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{f}(\mathbf{y})$ are scalar-valued. We will distinguish this case from the vector-valued case by the use of non-boldface notations.

We define the linear operator L_x by $L_x\{f\}(y) = xf(y)$. We can then write that

$$\mathsf{E}[xf(y)] = \mathsf{E}[L_x\{f\}(y)].\tag{7.1}$$

In many situations it turns out that it is possible to replace L_x in (7.1) by a linear operator L that does *not* depend on x. In this case we have the equality

$$\mathsf{E}[xf(y)] = \mathsf{E}[L\{f\}(y)].\tag{7.2}$$

This means that $L\{f\}(y)$ can serve as an empirical estimate of the cross-term that is only based on the measured data y.

Let us re-express property (7.2) using the probability distribution p_x, in the case where f is a test function: we would like to find a continuous linear operator $L : \mathcal{S} \to \mathcal{S}$ such that, for every $\varphi \in \mathcal{S}$,

$$\langle p_x, x\,\varphi \rangle = \langle p_x, L\{\varphi\} \rangle.\tag{7.3}$$

Introducing the notation L^* for the dual operator of L, (7.3) is equivalent to

$$\langle L^*\{p_x\}, \varphi \rangle = \langle x\, p_x, \varphi \rangle$$

for every test function φ. As first pointed out by Raphan and Simoncelli [182], this amounts to the eigenfunction equation

$$L^*\{p_x\} = x\, p_x.$$

In other words, p_x must be an "eigendistribution" of L^* corresponding to the eigenvalue x.

The above approach raises interesting mathematical questions—in particular:

- when does such an operator L exist?

- if it exists, can it be extended—along with equality (7.3)—to functions $\varphi \notin \mathcal{S}$?

The answer to these questions is specific to each family of probability distributions p_x. A comprehensive study would go beyond the scope of this introduction, but we can make the following general observation (which is a standard result from linear algebra).

Property 13. *Assume that $L^*\{p_x\} = x\, p_x$ for every x; then the probability distributions p_x are linearly independent.*

Property 13 provides a necessary condition for the existence of an operator L. In particular, it requires the mapping $x \mapsto p_x$ to be injective. This result can be interpreted as follows: *if the measurement process involves a "loss of information" about x, an estimation of the form* (7.3) *is not possible.*

7.2.2   Some scalar examples

We will now illustrate the above approach with a few examples. A more comprehensive list including pointers to the statistics literature can be found in [128, 182].

Gaussian distribution: $y \sim \mathcal{N}(x, \sigma^2)$

$$p_x(y) = g(y - x) \qquad \text{where} \qquad g(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{y^2}{2\sigma^2}}.$$

We will use the following property of a Gaussian:

$$y\, g(y) = -\sigma^2 D\{g\}(y) \tag{7.4}$$

where D denotes the derivative operator. It follows that

$$x\, p_x(y) = y\, p_x(y) - (y - x)\, p_x(y) = y\, p_x(y) + \sigma^2 D\{p_x\}(y) = L^*\{p_x\}(y).$$

Therefore:

$$L\{\varphi\}(y) = y\, \varphi(y) - \sigma^2 D\{\varphi\}(y).$$

This result is at the core of Stein's Unbiased Risk Estimate (SURE) [209].

Poisson distribution: $y \sim \mathcal{P}(x)$

$$p_x = e^{-x} \sum_{k \in \mathbb{N}} \frac{x^k}{k!} \delta(\cdot - k).$$

We can observe that

$$e^{-x} \sum_{k \in \mathbb{N}} \frac{x^{k+1}}{k!} \delta(\cdot - k) = e^{-x} \sum_{k \in \mathbb{N}^*} \frac{x^k}{(k-1)!} \delta(\cdot + 1 - k)$$

$$= e^{-x} \sum_{k \in \mathbb{N}} k \frac{x^k}{k!} \delta(\cdot + 1 - k). \tag{7.5}$$

Therefore $x \, p_x = (\cdot + 1) \, p_x(\cdot + 1) = L^*\{p_x\}$ and

$$L\{\varphi\}(y) = y \, \varphi(y - 1).$$

This result can for example be found in [108].

Gaussian distribution with signal-dependent variance: $y \sim \mathcal{N}(x, x)$

$$p_x(y) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{(y-x)^2}{2x}}, \text{ where } x > 0.$$

Adapting (7.4) to this case yields $(y - x) \, p_x(y) = -D\{x \, p_x\}(y)$ and thus

$$y \, p_x(y) = (I - D)\{x \, p_x\}(y) = (I - D) \circ L^*\{p_x\}(y).$$

Note that $(I - D)$ is a shift-invariant operator that is invertible in $\mathcal{S}'$; the convolution kernel corresponding to its inverse is specified in the Fourier domain by $1/(1 - j\omega)$. Taking into account the duality, we can conclude that

$$L\{\varphi\}(y) = y \, (h * \varphi)(y),$$

where $h(y) = u(-y)e^y$.

While we will not use it later, this result is interesting from several standpoints. First it could reveal useful in applications where the previous Poissonian model is approximated by a Gaussian. Second, it illustrates the role of Fourier analysis for deriving the operator L; obviously, Fourier analysis is relevant whenever L has a shift-invariant "part". This has already been observed in the case of additive (signal-independent) noise [182]. Here the noise statistics are signal-dependent but Fourier analysis is still useful. Third, it shows a practical limitation of risk-estimation techniques: while it seems that risk estimates can be derived for a wide range of conceivable noise models, they are not always computationally tractable. For example, the above result implies that one must be able to compute the convolution of the restoration function with the kernel h. This is not always straightforward, unless the restoration function is designed to facilitate this computation.

Poisson+Gaussian mixture: $y \sim \gamma \mathcal{P}(x) + \mathcal{N}(\mu, \sigma^2)$

This example is of particular interest to us, as it is closely related to the general image-acquisiton model (3.1) introduced in Chapter 3. The probability distribution

of the sum of two independent random variables is the convolution of their probability distributions. Here we have the convolution of a (rescaled) Poisson distribution with a Gaussian:

$$p_x = e^{-x} \sum_{k \in \mathbb{N}} \frac{x^k}{k!} g(\cdot - \mu - \gamma k).$$

A computation similar to (7.5) leads to

$$(\cdot - \mu + \gamma) \, p_x(\cdot + \gamma) - \gamma \, x \, p_x = e^{-x} \sum_{k \in \mathbb{N}} (\cdot - \mu + \gamma - \gamma k) \frac{x^k}{k!} g(\cdot - \mu + \gamma - \gamma k)$$

$$= -e^{-x} \sum_{k \in \mathbb{N}} \frac{x^k}{k!} \sigma^2 D\{g\}(\cdot - \mu + \gamma - \gamma k).$$

Thus we have the equality

$$\gamma \, x \, p_x = (\cdot - \mu + \gamma) \, p_x(\cdot + \gamma) + \sigma^2 D\{p_x\}(\cdot + \gamma) = \gamma \, L^*\{p_x\}$$

and we can conclude that

$$L\{\varphi\}(y) = \frac{1}{\gamma} \Big((y - \mu) \, \varphi(y - \gamma) - \sigma^2 D\{\varphi\}(y - \gamma) \Big).$$

7.2.3 Extension to the vectorial case

In general, the parameter to be estimated is vector-valued (hence we return to the initial boldface notations). The cross-term can be written as

$$\mathsf{E}[\mathbf{x}^T \mathbf{f}(\mathbf{y})] = \sum_{n=1}^{N} \mathsf{E}[x_n f_n(\mathbf{y})],$$

where n indexes the components of the vectors. Thus, the multivariate case can be treated by determining N linear operators L_n that satisfy eigenfunction equations of the form (7.3); the only difference is that the probability distributions $p_\mathbf{x}$ are now indexed by the *vector* $\mathbf{x}$. Due to these structural similarities, the necessary conditon of Property 13 extends to the vectorial case as well (i.e., the probability distributions $p_\mathbf{x}$ are linearly independent if such operators exist).

The previous examples can be generalized as follows.

White Gaussian noise: $\mathbf{y} \sim \mathcal{N}(\mathbf{x}, \sigma^2 \mathbf{I})$

$$\mathsf{E}[\mathbf{x}^T \mathbf{f}(\mathbf{y})] = \mathsf{E}[\mathbf{y}^T \mathbf{f}(\mathbf{y}) - \sigma^2 \, \mathrm{div}\{\mathbf{f}\}(\mathbf{y})]. \tag{7.6}$$

Independent Poisson counts: $\mathbf{y} \sim \mathcal{P}(\mathbf{x})$

$$\mathsf{E}[\mathbf{x}^T \mathbf{f}(\mathbf{y})] = \mathsf{E}[\mathbf{y}^T \mathbf{g}(\mathbf{y})] \quad \text{where} \quad g_n(\mathbf{y}) = f_n(\mathbf{y} - \mathbf{e}_n). \tag{7.7}$$

Poisson+Gaussian mixture: $\mathbf{y} \sim \gamma\mathcal{P}(\mathbf{x}) + \mathcal{N}(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$

$$\mathsf{E}[\mathbf{x}^T \mathbf{f}(\mathbf{y})] = \frac{1}{\gamma}\mathsf{E}[(\mathbf{y}-\boldsymbol{\mu})^T \mathbf{g}(\mathbf{y}) - \sigma^2 \operatorname{div}\{\mathbf{g}\}(\mathbf{y})] \quad \text{where} \quad g_n(\mathbf{y}) = f_n(\mathbf{y}-\gamma\mathbf{e}_n). \tag{7.8}$$

7.3   Discussion

The estimate (7.8) for the mixture model includes the estimates (7.6) and (7.7) for the Gaussian and Poissonian models as particular cases[3]. Nevertheless, these examples will be useful to discuss some practical issues.

7.3.1   Algorithmic aspects

As mentioned previously, the application of an estimate such as (7.8) can pose computational problems. The primary difficulty lies in the definition of $\mathbf{g}$, which follows from the Poissonian part of the model. This definition implies that the N components of the restoration function $\mathbf{f}$ must be evaluated with different arguments (the vectors $\mathbf{y} - \gamma\mathbf{e}_n$). Depending on the definition of $\mathbf{f}$, this task may not be computationally tractable. Another difficulty lies in the evaluation of the divergence term—which is related to the Gaussian part of the model.

A simplification occurs when $\mathbf{f}$ is a linear function, i.e., when there is a matrix $\mathbf{F}$ such that $\mathbf{f}(\mathbf{y}) = \mathbf{F}\mathbf{y}$. In this case, $\mathbf{g}(\mathbf{y}) = \mathbf{F}\mathbf{y} - \gamma\mathbf{F}\mathbf{1}$ and the divergence of $\mathbf{g}$ is the trace of $\mathbf{F}$. Therefore, (7.8) reduces to

$$\mathsf{E}[\mathbf{x}^T \mathbf{F}\mathbf{y}] = \frac{1}{\gamma}\mathsf{E}[(\mathbf{y} - \boldsymbol{\mu})^T (\mathbf{F}\mathbf{y} - \gamma\mathbf{F}\mathbf{1}) - \sigma^2 \operatorname{Tr}\{\mathbf{F}\}]. \tag{7.9}$$

In particular, the inner product $(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{F}\mathbf{1}$ can be precomputed.

We have successfully applied (7.9) to linear deconvolution in [181] (not reproduced here). To our knowledge, [181] represents the first treatment (at least in the context of fluorescence microscopy) of a Poisson+Gaussian mixture model in image restoration that is not based on an approximation of the noise distribution. Previous approaches have typically relied on Gaussian approximations—see e.g. [160, 146, 59].

---

[3]See our remark at the end of Section 3.2.

Another simple case arises when $\mathbf{f}$ is a pointwise non-linearity, that is, when $f_n(\mathbf{y})$ only depends on y_n. The function $\mathbf{g}$ then reduces to $\mathbf{g}(\mathbf{y}) = \mathbf{f}(\mathbf{y} - \gamma\mathbf{1})$ and the divergence term only involves the derivatives of univariate functions.

When $\mathbf{f}$ is a general non-linear function, even the estimate (7.6) corresponding to a purely Gaussian model is not straightforward to evaluate. So far, the most generic approaches rely on *randomized* estimates. This includes Ramani et al.'s Monte-Carlo SURE [180] as well as the recursive risk-estimation method that we describe in Chapter 9.

Another promising approach is to make deterministic approximations on the restoration function [139].

7.3.2 Reliability

Besides the aforementioned computational aspects, the reliability of the proposed risk estimates is an important issue. In this subsection, we focus on the estimate (7.6) for the white Gaussian noise model. We study its accuracy in terms of the expected quadratic estimation error. In the formulas below, we use the identity $\text{div}\{\mathbf{f}\} = \text{Tr}\{\mathbf{J_f}\}$, where $\mathbf{J_f}$ stands for the Jacobian matrix[4] of $\mathbf{f}$.

Property 14. *Assume that* $\mathbf{y} \sim \mathcal{N}(\mathbf{x}, \sigma^2\mathbf{I})$ *and that the components of* $\mathbf{f}$ *are twice continuously differentiable. Then,*

$$\mathsf{E}\left[\left((\mathbf{y}-\mathbf{x})^T\mathbf{f}(\mathbf{y}) - \sigma^2\text{Tr}\{\mathbf{J_f}(\mathbf{y})\}\right)^2\right] = \mathsf{E}\left[\sigma^2\|\mathbf{f}(\mathbf{y})\|^2 + \sigma^4\text{Tr}\{\mathbf{J_f}(\mathbf{y})^2\}\right]. \quad (7.10)$$

Proof. Introducing the notation $\mathbf{b} = \mathbf{y} - \mathbf{x}$, we can decompose the average squared error into

$$\mathsf{E}\left[(\mathbf{b}^T\mathbf{f}(\mathbf{y}))^2\right] - 2\sigma^2\mathsf{E}\left[\mathbf{b}^T\mathbf{f}(\mathbf{y})\ \text{Tr}\ \mathbf{J_f}(\mathbf{y})\right] + \sigma^4\mathsf{E}\left[\left(\text{Tr}\ \mathbf{J_f}(\mathbf{y})\right)^2\right].$$

---

[4]The entry of $\mathbf{J_f}(\mathbf{y})$ at row k and column n is the derivative of f_n with respect to its k-th argument, evaluated at $\mathbf{y}$.

Let us expand the first term:

$$\mathsf{E}\Big[\sum_{k,n} b_k b_n f_k(\mathbf{y}) f_n(\mathbf{y})\Big] = \sigma^2 \mathsf{E}\Big[\sum_k f_k(\mathbf{y})^2\Big]$$

$$+ \sigma^2 \mathsf{E}\Big[\sum_{k,n} b_n \partial_k f_k(\mathbf{y}) f_n(\mathbf{y}) + b_n f_k(\mathbf{y}) \partial_k f_n(\mathbf{y})\Big]$$

$$= \sigma^2 \mathsf{E}\Big[\|\mathbf{f}(\mathbf{y})\|^2\Big]$$

$$+ \sigma^4 \mathsf{E}\Big[\sum_{k,n} \partial_n \partial_k f_k(\mathbf{y}) f_n(\mathbf{y}) + \partial_k f_k(\mathbf{y}) \partial_n f_n(\mathbf{y})$$

$$+ \partial_n f_k(\mathbf{y}) \partial_k f_n(\mathbf{y}) + f_k(\mathbf{y}) \partial_n \partial_k f_n(\mathbf{y})\Big]$$

$$(*) = \sigma^2 \mathsf{E}\Big[\|\mathbf{f}(\mathbf{y})\|^2\Big] + 2\sigma^4 \mathsf{E}\Big[\sum_{k,n} \partial_n \partial_k f_k(\mathbf{y}) f_n(\mathbf{y})\Big]$$

$$+ \sigma^4 \mathsf{E}\Big[\big(\mathrm{Tr}\{\mathbf{J_f}(\mathbf{y})\}\big)^2\Big] + \sigma^4 \mathsf{E}\Big[\sum_{k,n} \partial_n f_k(\mathbf{y}) \partial_k f_n(\mathbf{y})\Big]$$

$$= \sigma^2 \mathsf{E}\Big[\|\mathbf{f}(\mathbf{y})\|^2\Big] + 2\sigma^2 \mathsf{E}\Big[\mathbf{b}^T \mathbf{f}(\mathbf{y}) \,\mathrm{Tr}\{\mathbf{J_f}(\mathbf{y})\}\Big]$$

$$- \sigma^4 \mathsf{E}\Big[\big(\mathrm{Tr}\{\mathbf{J_f}(\mathbf{y})\}\big)^2\Big] + \sigma^4 \mathsf{E}\Big[\mathrm{Tr}\{\mathbf{J_f}(\mathbf{y})^2\}\Big].$$

Note that equality $(*)$ holds because the functions f_n are assumed to be C^2 (which allows for changing the order of the partial derivatives). After inserting this last expression in the first one, four terms cancel out and one obtains the desired result. □

The above result potentially allows for computing an *estimate of the estimation error* from the measured data (since the expression inside the right-hand expectation of (7.10) only depends on $\mathbf{y}$).

7.3.3   Extension to ill-conditioned linear inverse problems

In this subsection, we assume that $\mathbf{y}$ depends on $\mathbf{Hx}$ rather than $\mathbf{x}$, where $\mathbf{H}$ is an invertible but potentially ill-conditioned matrix. To make the presentation more concrete, we focus on the standard model

$$\mathbf{y} \sim \mathcal{N}(\mathbf{Hx}, \sigma^2 \mathbf{I}).$$

The estimate (7.6) can then be extended as follows:

$$\mathsf{E}[\mathbf{x}^T\mathbf{f}(\mathbf{y})] = \mathsf{E}[(\mathbf{Hx})^T\mathbf{H}^{-T}\mathbf{f}(\mathbf{y})] = \mathsf{E}[\mathbf{y}^T\mathbf{H}^{-T}\mathbf{f}(\mathbf{y}) - \sigma^2 \operatorname{div}\{\mathbf{H}^{-T}\mathbf{f}\}(\mathbf{y})].$$

In the favorable situation where the restoration function is an algorithm of the form $\mathbf{f}(\mathbf{y}) = \mathbf{H}^T\mathbf{g}(\mathbf{y})$, this reduces to

$$\mathsf{E}[\mathbf{x}^T\mathbf{f}(\mathbf{y})] = \mathsf{E}[\mathbf{y}^T\mathbf{g}(\mathbf{y}) - \sigma^2 \operatorname{div}\{\mathbf{g}\}(\mathbf{y})].$$

This is for example the case for linear deconvolution solutions such as (3.11), if one exploits the commutativity property of convolution matrices.

Similarly, the equivalent of the expected quadratic estimation error given in (7.10) is now

$$\sigma^2\mathsf{E}\left[\|\mathbf{H}^{-T}\mathbf{f}(\mathbf{y})\|^2\right] + \sigma^4\mathsf{E}\left[\|\operatorname{Tr}\{\mathbf{H}^{-T}\mathbf{J}_{\mathbf{f}}(\mathbf{y})^2\}\|\right]. \tag{7.11}$$

In practice the restoration algorithm very often depends on $\mathbf{H}^T\mathbf{y}$ rather than directly on $\mathbf{y}$ (this applies for example to Wiener filtering or to the thresholded Landweber algorithm with a Wiener-type initialization). This means that $\mathbf{f}$ takes the functional form $\mathbf{f}(\mathbf{y}) = \mathbf{g}(\mathbf{H}^T\mathbf{y})$, which implies that

$$\mathbf{H}^{-T}\mathbf{J}_{\mathbf{f}}(\mathbf{y}) = \mathbf{H}^{-T}\mathbf{H}^T\mathbf{J}_{\mathbf{g}}(\mathbf{H}^T\mathbf{y}) = \mathbf{J}_{\mathbf{g}}(\mathbf{H}^T\mathbf{y}).$$

Therefore, the second term in (7.11) can be considered as "stable". However, the form of the first term shows that the statistical error can be large if $\mathbf{H}$ is ill-conditioned.

The extreme case is when $\mathbf{H}$ is not invertible, which means that the mapping $\mathbf{x} \mapsto p_{\mathbf{x}}$ is not injective. As observed in Section 7.2, this does not allow for an estimation of the form (7.2). This is why we propose to introduce a small bias when considering ill-posed inverse problems. Preliminary results with this approach are presented in Chapter 9.

7.4   Summary

The present chapter provides a concise and yet self-contained introduction to risk-estimation techniques. It also led to the following extensions and insights, which have motivated the work presented in the next two chapters.

- The generalization of the Gaussian and Poissonian risk estimates to the mixture model is novel and particularly relevant to image restoration in fluorescence microscopy. The algorithm described in Chapter 8 was initially developped for the Poisson-only model, but it can be extended to the mixture model at virtually the same computational cost.

- While other works on risk estimation have been recently published [182, 75], there have been few comments on the reliability of SURE-like techniques. We have proposed preliminary insights. In particular, we have found that it is illusory to aim at an *unbiased* risk estimate for ill-posed linear inverse problems (which involve a loss of information about the original image). Another way to state this is the following: the more the image-formation model is ill-conditioned, the more the variance of the estimate increases. Therefore, one must aim at a good bias-variance tradeoff. Chapter 9 presents preliminary results in this direction.

- From the algorithmic standpoint there are also tradeoffs, because the "risk-estimation operators" are not always computationally tractable. Of particular interest to us is the Poisson+Gaussian mixture, which is tractable for linear algorithms. For non-linear algorithms, adaptations are required. A relatively standard approach is to approximate the mixture by a Gaussian model with known, space-varying variance; this gave promising results with the method of Chapter 9 applied to the thresholded Landweber algorithm (which will be the subject of a future report). Another option is to design the algorithm such that the operator is easily applicable; this is the approach we have chosen in Chapter 8.

Chapter 8

A Haar-based algorithm for denoising shot-noise-limited images

Summary

We extend the SURE-LET denoising methodology to image restoration in the presence of Poisson noise. Our approach is based on 1) the minimization of a rigorous estimate of the denoising MSE (or "risk") for Poisson noise (PURE) comparable to Stein's estimate (SURE), 2) a linear parametrization of the denoising process (LET) and 3) a fundamental property of the maximally decimated unnormalized Haar wavelet transform with respect to Poisson statistics.

The minimization of the MSE estimate is performed independently in each Haar-wavelet subband for computational efficiency, but this is equivalent to performing a global image-domain minimization. This is an important difference with standard Poisson noise-removal methods and in particular those that rely on a non-linear preprocessing of the data to stabilize the variance. A further advantage of our method is the concept of linear expansion of thresholds (LET), which allows for a fast and direct computation of the optimal parameters.

The chapter includes a series of numerical experiments performed on standard grayscale images. We show that our non-redundant wavelet method compares favor-

---

This chapter is based on [238].

ably with standard variance-stabilizing schemes, even when the latter are applied in a translation-invariant setting (cycle-spinning). Our algorithm also achieves a quality similar to a state-of-the-art multiscale method that was specially developed for Poisson data. Considering that the computational complexity of our method is orders of magnitude lower, it is actually a very competitive alternative.

The proposed approach is particularly promising in the context of low signal intensities and/or large data sets. This is illustrated experimentally with the denoising of low-count fluorescence micrographs of a biological sample.

8.1   Introduction

8.1.1   Motivation

"Additive white Gaussian noise" is a ubiquitous model in the context of statistical image restoration. In many applications, however, the current trend towards quantitative imaging calls for less generic models that better account for the physical acquisition process. The need for such models is particularly stringent in biomicroscopy, where live samples are often observed at very low light levels, due to acquisition-time and phototoxicity constraints (see Chapter 2). In this regime, the performance of the imaging device is typically shot-noise limited, which is strongly signal-dependent. For this reason, opting for a non-additive, non-Gaussian model can yield significant improvements in restoration quality in such applications.

Yet, Gaussian priors are particularly attractive in a Bayesian framework, where they easily lead to a closed-form solution. In comparison, alternative noise models appear to be less tractable and typically require expectation-maximization schemes for computing the solution (see e.g. [226] for the Poisson case); a significant body of research is concerned with the acceleration of such methods (see [151] and references therein). When using non-Bayesian approaches (typically, when we do not want to make statistical hypotheses on the images to denoise), however, the advantage of a Gaussian model is not as clearcut.

8.1.2   The PURE-LET approach

Blu and Luisier have recently proposed a novel non-Bayesian methodology to process noisy images based on the knowledge of the noise corruption *only* [24]: the SURE-LET strategy. The present work extends its validity to an instance of non-additive, non-Gaussian noise processes: the Poisson noise. We first derive a statistical estimate of the Mean Square Error (MSE), or "risk", between the (unknown)

noiseless image and the processed noisy image, under the Poisson noise hypothesis: the Poisson Unbiased Risk Estimate (PURE). We then *minimize* this MSE estimate over a collection of "acceptable" denoising processes to find the best one, in the sense of the Signal-to-Noise Ratio (SNR), which is a widespread measure of restoration quality [19]. To our knowledge, this is actually the first reported use of an (unbiased) MSE estimate in the Poisson-noise case for image processing.

The efficiency of our method stems from the concept of Linear Expansion of Thresholds (LET, see [24]): the "acceptable" denoising processes are expressed as a linear combination of elementary denoising processes, from which only the weights are unknown. It is these weights that are then computed by minimizing the PURE, through the resolution of a simple linear system of equations. This approach is further powered by the use of a simple (unnormalized) Haar wavelet transform.

For each subband, our restoration functions involve several parameters, which provides more flexibility than standard single-parameter thresholding functions. Importantly, the thresholds are adapted to local estimates of the (signal-dependent) noise variance. These estimates are derived from the corresponding low-pass coefficients at the same scale; the latter are also used to incorporate interscale relationships into the denoising functions. The resulting procedure can be easily integrated into the wavelet decomposition, which is *non-redundant.* The MSE estimate is optimized independently for each subband by exploiting the orthogonality of the Haar wavelet basis.

As a result, our algorithm has low computational complexity and modest memory requirements. These are valuable features for denoising large data sets, such as those typically produced in astronomy or in fluorescence microscopy. Importantly, this computational efficiency is not traded for quality. On the contrary, the algorithm yields improved results compared to traditional Gaussian-inspired approaches, and it performs competitively with a state-of-the-art multiscale method that was specially developed for Poisson data.

8.1.3   Related work

The first wavelet-based techniques that were specifically designed for Poisson intensity estimation appeared in the fields of Astrophysics [119] and Seismology [33].

Several subsequent works were based on the fact that Poisson statistics are preserved across scales in the low-pass channels of an unnormalized Haar wavelet transform. This fundamental property was for example used by Timmermann and Nowak [218] to construct a multiscale Bayesian model of the signal; an extension for estimating all parameters of the model using the expectation-maximization algorithm was derived in [137]. A similar model was proposed independently by

Kolaczyk [120] using the concept of recursive dyadic partition. A generalization to three families of distributions (Gaussian, Poisson and multinomial) was described in [123], along with an asymptotic minimax analysis.

The aforementioned property was also used within a user-calibrated hypothesis-testing approach in [122]. The idea of using hypothesis testing to handle Poisson statistics was initially adapted from the Gaussian case by Kolaczyk [121], who proposed to use (pairs of) level-dependent thresholds. A complementary study of the theoretical asymptotic properties of wavelet estimators was presented in [172].

Notice that the Bayesian framework has also been used in conjunction with more involved multiscale transformations than the Haar transform. For example, Sardy et al. [189] considered a general ℓ_1-penalized-likelihood framework for arbitrary wavelet bases and noise models, including the Poisson case. More recently, Willett and Nowak have proposed a platelet-based penalized-likelihood estimator that was shown to be very efficient for denoising piecewise-smooth images [249].

A widespread alternative to the direct handling of Poisson statistics is to apply variance-stabilizing transforms (VSTs)—with the underlying idea of exploiting the broad class of denoising methods that are based on a Gaussian noise model [64]. Since the seminal work of Anscombe [7], more involved VSTs have been proposed, such as the Haar-Fisz transform [85]. Such approaches belong to the state-of-the-art for 1D wavelet-based Poisson noise removal [19, 85]. They have been combined with various other methodologies, e.g. Bayesian multiscale likelihood models that can be applied to arbitrary wavelet transforms [114]. Very recently, a hybrid approach that combines VSTs, hypothesis testing, ℓ_1-penalized reconstruction and advanced redundant multiscale representations has been proposed by Zhang et al. [257].

With the exception of cross-validation methods [165, 9, 189], however, the potential of purely data-driven techniques seems to have remained under-exploited for the wavelet-based restoration of images corrupted by Poisson noise.

8.1.4   Organization of the chapter

This chapter is organized as follows. In Section 8.2, we derive an unbiased risk estimate for a broad class of Poisson denoising algorithms formulated in the Haar-wavelet domain. In Section 8.3, we specify several such algorithms based on our previous work and experimental insights.   In Section 8.4, we compare our approach to typical variance-stabilizing methods, as well as a recent algorithm specifically designed for Poisson statistics. Finally we present results obtained with real fluorescence-microscopy data in Section 8.5.

8.2   Theory

Recall that m is a Poisson random variable of intensity $\mu \in \mathbb{R}^+$ if and only if

$$P(m = k) = e^{-\mu} \frac{\mu^k}{k!} \tag{8.1}$$

for every $k \in \mathbb{N}$; we use the standard notation $m \sim \mathcal{P}(\mu)$. Throughout this chapter, for a given vector $\mathbf{v} \in \mathbb{R}^N$, we use the notation v_n to refer to its n-th component, where $n \in \{1, \ldots, N\}$. $\mathbf{m}$ will denote a vector of N independent Poisson random variables m_n of underlying intensities μ_n, i.e., $m_n \sim \mathcal{P}(\mu_n)$. A realization of $\mathbf{m}$ can be thought of as a noisy measurement of the intensity signal $\boldsymbol{\mu}$. Note that in contrast with Bayesian approaches, $\boldsymbol{\mu}$ is considered to be *deterministic* in the present work. Based on the measurement $\mathbf{m}$, our goal is to find an estimate $\hat{\boldsymbol{\mu}}$ that is the closest possible to the original signal in the minimum *mean squared error* (MSE) sense; that is, we want to minimize

$$\text{MSE} = \frac{1}{N} \|\hat{\boldsymbol{\mu}} - \boldsymbol{\mu}\|^2 = \frac{1}{N} \sum_{n=1}^{N} (\hat{\mu}_n - \mu_n)^2.$$

8.2.1   Some useful properties of the Poisson distribution

The Poisson distribution enjoys the following useful properties.

Property 15. *The sum of independent Poisson random variables is also a Poisson random variable, whose intensity is the sum of the original intensities.*

For example, $m_1 + m_2 \sim \mathcal{P}(\mu_1 + \mu_2)$.

Property 16. *If $m \sim \mathcal{P}(\mu)$ and θ: $\mathbb{R} \to \mathbb{R}$ is a real function such that $\mathsf{E}\left[|\theta(m)|\right] < \infty$, then*

$$\mathsf{E}\left[\mu\theta(m)\right] = \mathsf{E}\left[m\theta(m-1)\right],$$

where $\mathsf{E}\left[\cdot\right]$ stands for the mathematical expectation operator.

Proof:

$$
\begin{aligned}
\mathsf{E}\left[\mu\theta(m)\right] &= \sum_{k\in\mathbb{N}} \mu\theta(k)\frac{\mu^k}{k!}e^{-\mu} \\
&= \sum_{k\in\mathbb{N}\setminus\{0\}} \theta(k-1)\frac{\mu^k}{(k-1)!}e^{-\mu} \\
&= \sum_{k\in\mathbb{N}} k\theta(k-1)\frac{\mu^k}{k!}e^{-\mu} \\
&= \mathsf{E}\left[m\theta(m-1)\right].
\end{aligned}
$$

Note that Property 16 is the Poisson equivalent of Stein's lemma for Gaussian statistics [209]; similar results can be found e.g. in [108, 221]. As a straightforward application of this property, we note that we have

$$
\mu^2 = \mathsf{E}\left[m(m-1)\right].
$$

Finally we state a result dating back to [212] that will be useful for the experimental part of this work (Section 8.5).

Property 17 (binomial selection). *Let $m \sim \mathcal{B}(\ell, \eta)$ be a binomial random variable, where $\eta \in [0,1]$ represents the probability of success. If the number of trials $\ell \in \mathbb{N}$ is random and follows a Poisson distribution with mean λ, then m is itself Poisson distributed with mean $\mu = \eta\lambda$.*

Proof. We show that m follows the Poisson distribution defined in (8.1):

$$
\begin{aligned}
P(m=k) &= \sum_{n\in\mathbb{N}} P(m=k|\ell=n)\, P(\ell=n) \\
&= \sum_{n\geq k} \frac{n!}{(n-k)!\,k!}\eta^k(1-\eta)^{n-k}\, e^{-\lambda}\frac{\lambda^n}{n!}.
\end{aligned}
$$

Performing the change of index $n' = n - k$ yields

$$
\begin{aligned}
P(m=k) &= \sum_{n'\in\mathbb{N}} \frac{1}{n'!\,k!}\eta^k(1-\eta)^{n'}\, e^{-\lambda}\lambda^{n'+k} \\
&= e^{-\eta\lambda}\frac{(\eta\lambda)^k}{k!}\underbrace{\sum_{n'\in\mathbb{N}} \frac{1}{n'!}(1-\eta)^{n'}\, e^{-(1-\eta)\lambda}\lambda^{n'}}_{=1}.
\end{aligned}
$$

$\square$
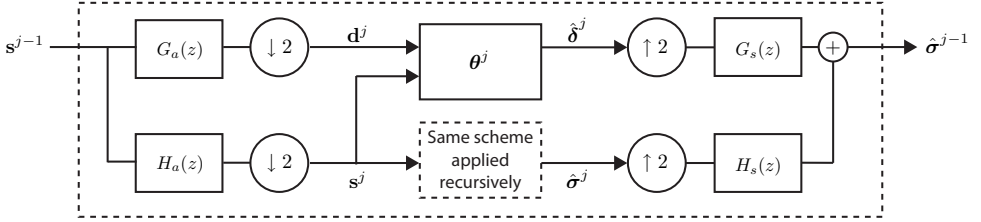
Figure 8.1: Filter bank implementation of the unnormalized discrete Haar wavelet transform and principle of the class of denoising algorithms described by (8.2). The scheme is applied recursively on the low-pass channel output.

8.2.2 The unnormalized Haar discrete wavelet transform

The unnormalized Haar discrete wavelet transform (DWT) can be seen as a standard two-channel filterbank (see Fig. 8.1). The analysis pair of lowpass/highpass filters is given in the z-transform domain by

$$\begin{cases} H_a(z) = 1 + z^{-1}; \\ G_a(z) = 1 - z^{-1}. \end{cases}$$

The corresponding synthesis pair is

$$\begin{cases} H_s(z) = \frac{1}{2}H_a(z^{-1}); \\ G_s(z) = \frac{1}{2}G_a(z^{-1}). \end{cases}$$

In this work, the unnormalized Haar scaling coefficients of the measurement $\mathbf{m}$ at scales $j = 1, \ldots, J$ are denoted by $\mathbf{s}^j \in \mathbb{R}^{N_j}$, where $N_j = N/2^j$, and $\mathbf{d}^j \in \mathbb{R}^{N_j}$ stands for the associated wavelet coefficients (we assume that the signal dimension is divisible by 2^J). Setting $\mathbf{s}^0 = \mathbf{m}$, these coefficients are obtained from the following sums and differences:

$$\begin{cases} s_n^j = s_{2n}^{j-1} + s_{2n-1}^{j-1}, \\ d_n^j = s_{2n}^{j-1} - s_{2n-1}^{j-1}, \end{cases} \quad \text{for } j = 1, \ldots, J.$$

The original sequence $\mathbf{m} = \mathbf{s}^0$ is simply recovered by computing

$$\begin{cases} s_{2n}^{j-1} = (s_n^j + d_n^j)/2, \\ s_{2n-1}^{j-1} = (s_n^j - d_n^j)/2, \end{cases} \quad \text{for } j = J, \ldots, 1.$$

Similarly, we denote by $\boldsymbol{\sigma}^j$ and $\boldsymbol{\delta}^j$ the scaling and wavelet coefficients of the original signal $\boldsymbol{\mu}$ at a given scale j. Note that, by linearity of the wavelet transform, we have $\mathsf{E}\left[d_n^j\right] = \delta_n^j$ and $\mathsf{E}\left[s_n^j\right] = \sigma_n^j$.

The key properties of the unnormalized Haar DWT are the following.

1. It is an *orthogonal* transform. In particular, we can split the MSE into subband-specific error terms:

$$\mathrm{MSE} = \frac{2^{-J}}{N}\left\|\hat{\boldsymbol{\sigma}}^J - \boldsymbol{\sigma}^J\right\|^2 + \sum_{j=1}^{J}\frac{2^{-j}}{N}\left\|\hat{\boldsymbol{\delta}}^j - \boldsymbol{\delta}^j\right\|^2.$$

   This implies that we can minimize the MSE for each subband independently, while ensuring a global signal-domain MSE minimization.

2. At a given scale j, the scaling coefficients of an input vector of independent Poisson random variables are also independent Poisson random variables, thanks to Property 15 of the Poisson distribution.

8.2.3   PURE: a Haar-wavelet-domain Poisson unbiased risk estimate

In principle, the estimate $\hat{\boldsymbol{\delta}}^j$ may be constructed using all available subbands of the measurement $\boldsymbol{\mu}$; in practice, however, standard thresholding techniques only use the corresponding wavelet subband of the measurement, $\mathbf{d}^j$. In the sequel, we will consider denoising algorithms where $\hat{\boldsymbol{\delta}}^j$ also depends on the scaling coefficients at the same scale $\mathbf{s}^j$. This means that we have the following functional relationship:

$$\hat{\boldsymbol{\delta}}^j = \boldsymbol{\theta}^j(\mathbf{d}^j, \mathbf{s}^j). \tag{8.2}$$

As usual, the lowpass residual is not processed, i.e., $\hat{\boldsymbol{\sigma}}^J = \mathbf{s}^J$. Our algorithmic framework is illustrated in Fig. 8.1.

The above choice is advantageous from a computational standpoint because such a restoration procedure can be implemented in parallel with the wavelet decomposition, which yields the scaling coefficients $\mathbf{s}^j$ as a by-product. Furthermore, this framework comprises advanced denoising schemes that exploit interscale dependencies via the scaling coefficients $\mathbf{s}^j$ (see Section 8.3.2). Finally, it allows us to minimize $\mathrm{MSE}_j = \|\boldsymbol{\theta}^j(\mathbf{d}^j, \mathbf{s}^j) - \boldsymbol{\delta}^j\|^2/N_j$ independently for each wavelet subband. We will thus focus on a fixed scale and drop the superscript ^{j} to simplify the notations.

The MSE is a very popular measure of restoration quality in phantom experiments, where the ground-truth signal is known. However, it is not accessible in real situations where the original noise-free signal is unknown. Our approach is thus to construct a statistical estimate of this quantity that solely depends on the measurement. This type of estimator is very accurate in the context of image processing, because it can rely on a large number of samples. In the sequel, we borrow the term risk from the statistics literature, to refer to the expectation of the MSE.

In the Gaussian case, the scaling coefficients $\mathbf{s}^j$ are statistically independent of the wavelet coefficients $\mathbf{d}^j$ (in an orthogonal wavelet domain). This is in contrast with the Poisson case, for which $\mathbf{d}^j$ and $\mathbf{s}^j$ are statistically dependent and even correlated. This makes the derivation of a bivariate (i.e. involving both $\mathbf{d}^j$ and $\mathbf{s}^j$) SURE-like MSE estimate less obvious. In the following theorem, we give an expression of an unbiased estimate of the MSE in a given subband j of the unnormalized Haar DWT defined in Section 8.2.2. This result serves as a data-dependent quantitative measure to be minimized for Poisson intensity estimation. The family of vectors $(\mathbf{e}_n)_{n=1,\ldots,N_j}$ denotes the canonical basis of $\mathbb{R}^{N_j}$, i.e. all components of $\mathbf{e}_n$ are zero, except for the n-th component, which is equal to one.

Theorem 1. *Let* $\boldsymbol{\theta}(\mathbf{d},\mathbf{s}) = \boldsymbol{\theta}^j(\mathbf{d}^j,\mathbf{s}^j)$ *be an estimate of the noise-free wavelet coefficients* $\boldsymbol{\delta} = \boldsymbol{\delta}^j$. *Define* $\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})$ *and* $\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})$ *by*

$$\begin{cases} \theta_n^+(\mathbf{d},\mathbf{s}) = \theta_n(\mathbf{d}+\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n), \\ \theta_n^-(\mathbf{d},\mathbf{s}) = \theta_n(\mathbf{d}-\mathbf{e}_n,\mathbf{s}-\mathbf{e}_n). \end{cases} \tag{8.3}$$

Then the random variable

$$\epsilon_j = \frac{1}{N_j}\Big(\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\|^2+\|\mathbf{d}\|^2-\mathbf{1}^T\mathbf{s}-\mathbf{d}^T\big(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})+\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\big)-\mathbf{s}^T\big(\boldsymbol{\theta}^-(\mathbf{d},\mathbf{s})-\boldsymbol{\theta}^+(\mathbf{d},\mathbf{s})\big)\Big)$$
$$\tag{8.4}$$

is an unbiased *estimate of the MSE for the subband under consideration, i.e.,* $\mathsf{E}\left[\epsilon_j\right] = \mathsf{E}\left[MSE_j\right]$.

Proof. The proof relies centrally on the fact that, within a given scale, the scaling coefficients are independent Poisson random variables. We consider the case where $j=1$, so that we can use $\mathbf{m} = \mathbf{s}^{j-1}$ and $\boldsymbol{\mu} = \boldsymbol{\sigma}^{j-1}$ to avoid superscripts.

We first develop the squared error between $\boldsymbol{\delta}$ and its estimate $\boldsymbol{\theta}(\mathbf{d},\mathbf{s})$, using the fact that $\boldsymbol{\delta}$ is a deterministic quantity:

$$\mathsf{E}\left[\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})-\boldsymbol{\delta}\|^2\right] = \mathsf{E}\left[\|\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\|^2\right] + \underbrace{\|\boldsymbol{\delta}\|^2}_{(\mathrm{I})} - 2\underbrace{\mathsf{E}\left[\boldsymbol{\delta}^T\boldsymbol{\theta}(\mathbf{d},\mathbf{s})\right]}_{(\mathrm{II})}. \tag{8.5}$$

Now, we can evaluate the two expressions (I,II) that involve the unknown data $\boldsymbol{\delta}$.

(I) $\quad \|\boldsymbol{\delta}\|^2 = \displaystyle\sum_{n=1}^{N_j} (\delta_n)^2$:

We notice that

$$(\delta_n)^2 = \mathsf{E}\left[\mu_{2n}(m_{2n} - m_{2n-1})\right] + \mathsf{E}\left[\mu_{2n-1}(m_{2n-1} - m_{2n})\right].$$

By applying Property 2 for $\theta(m) = m - m_{2n-1}$ and for $\theta(m) = m - m_{2n}$, by using the fact that m_{2n} and m_{2n-1} are statistically independent, we get

$$(\delta_n)^2 = \mathsf{E}\left[(m_{2n} - m_{2n-1})^2 - (m_{2n} + m_{2n-1})\right] = \mathsf{E}\left[(d_n)^2 - s_n\right].$$

Therefore $\|\boldsymbol{\delta}\|^2 = \mathsf{E}\left[\|\mathbf{d}\|^2 - \mathbf{1}^T\mathbf{s}\right]$.

(II) $\quad \mathsf{E}\left[\boldsymbol{\delta}^T\boldsymbol{\theta}(\mathbf{d}, \mathbf{s})\right] = \displaystyle\sum_{n=1}^{N_j} \mathsf{E}\left[\delta_n\theta_n(\mathbf{d}, \mathbf{s})\right]$:

We can successively write

$$
\begin{aligned}
\mathsf{E}\left[\delta_n\theta_n(\mathbf{d}, \mathbf{s})\right] \quad &= \quad \mathsf{E}\left[\mu_{2n}\theta_n(\mathbf{d}, \mathbf{s})\right] - \mathsf{E}\left[\mu_{2n-1}\theta_n(\mathbf{d}, \mathbf{s})\right] \\
&\overset{\text{Prop. 16}}{=} \quad \mathsf{E}\left[m_{2n}\theta_n(\mathbf{d} - \mathbf{e}_n, \mathbf{s} - \mathbf{e}_n)\right] - \mathsf{E}\left[m_{2n-1}\theta_n(\mathbf{d} + \mathbf{e}_n, \mathbf{s} - \mathbf{e}_n)\right] \\
&= \quad \mathsf{E}\left[\frac{m_{2n} - m_{2n-1}}{2}\left(\theta_n^-(\mathbf{d}, \mathbf{s}) + \theta_n^+(\mathbf{d}, \mathbf{s})\right)\right] + \\
&\qquad \mathsf{E}\left[\frac{m_{2n} + m_{2n-1}}{2}\left(\theta_n^-(\mathbf{d}, \mathbf{s}) - \theta_n^+(\mathbf{d}, \mathbf{s})\right)\right] \\
&= \quad \frac{1}{2}\mathsf{E}\left[d_n\left(\theta_n^-(\mathbf{d}, \mathbf{s}) + \theta_n^+(\mathbf{d}, \mathbf{s})\right)\right] + \frac{1}{2}\mathsf{E}\left[s_n\left(\theta_n^-(\mathbf{d}, \mathbf{s}) - \theta_n^+(\mathbf{d}, \mathbf{s})\right)\right].
\end{aligned}
$$

Thus $2\mathsf{E}\left[\boldsymbol{\delta}^T\boldsymbol{\theta}(\mathbf{d}, \mathbf{s})\right] = \mathsf{E}\left[\mathbf{d}^T\left(\boldsymbol{\theta}^-(\mathbf{d}, \mathbf{s}) + \boldsymbol{\theta}^+(\mathbf{d}, \mathbf{s})\right)\right] + \mathsf{E}\left[\mathbf{s}^T\left(\boldsymbol{\theta}^-(\mathbf{d}, \mathbf{s}) - \boldsymbol{\theta}^+(\mathbf{d}, \mathbf{s})\right)\right].$

Putting these results back into (8.5) gives the desired equality. For $j \geq 2$, the proof is based on the same idea. □

We will refer to (8.4) as the *Poisson unbiased risk estimate* (PURE). This estimate involves finite differences of the restoration function (instead of derivatives in the Gaussian case). It can be used to evaluate the restoration quality of any algorithm of the form (8.2) in terms of MSE. In the next section, we will consider algorithms that depend on a set of parameters, and we will minimize the PURE in order to obtain their optimal values.

8.2.4   Extension to multidimensional signal denoising

While the above result (8.4) is stated in a one-dimensional setting, it can easily be extended to arbitrary dimensions when using the separable Haar DWT. Indeed, Theorem 1 essentially relies on the fact that the wavelet coefficients are differences of Poisson random variables. In higher dimensions, it suffices to observe that the wavelet coefficients are differences of *sums of* scaling coefficients; these sums are still Poisson distributed, according to Property 15.

8.3   PURE-based wavelet thresholding

In this section we will define a series of thresholding functions for Poisson denoising. In general, thresholding operations can be expected to work well on signals that are piecewise smooth. Such signals have good energy compaction in the wavelet domain, i.e., they have only a few large wavelet coefficients. These significant coefficients can be efficiently distinguished from noise-only coefficients using an appropriate threshold level.

8.3.1   PUREshrink

As a benchmark for illustrating our approach, we propose a wavelet-domain estimator which consists in applying the popular *soft-threshold* with a PURE-optimized threshold. Our *PUREshrink* estimator can be viewed as the transposition of Donoho and Johnstone's *Gaussian SUREshrink* [66] to Poisson noise removal. An important difference is that the method described in [66] forces the threshold T to be smaller than the universal threshold (otherwise it is set to the value of the universal threshold); this is known to be suboptimal for image-denoising applications [140]. Our threshold optimization totally relies on the minimization of the PURE (without restrictions).

Contrary to the Gaussian case, where the noise is stationary and completely described by its variance[1], for Poisson data, the amount of noise directly depends on the intensity we want to estimate. Thus, for deciding on the amount of shrinkage, we use a threshold T that is proportional to the square root of the scaling coefficient at the same location and scale. This quantity is an estimate of the local noise standard deviation, so that it is a good reference for assessing the significance of a wavelet coefficient. Indeed, each wavelet coefficient of the unnormalized Haar transform follows a Skellam distribution[202], whose variance is equal to the sum of

---

[1]The noise is usually assumed to be zero-mean.

the two underlying Poisson intensities, i.e. approximately the corresponding scaling coefficient. Note that in the *Haar-Fisz* transform [85], the scaling coefficients are also considered as an estimate of the local noise variance and thus used to stabilize it. The *PUREshrink* estimator is therefore defined as

$$\theta_n^{\text{PUREshrink}}(\mathbf{d}, \mathbf{s}; a) = \text{sign}(d_n) \max(|d_n| - a\sqrt{|s_n|}, 0), \tag{8.6}$$

where, for each wavelet subband, the parameter a is set to the value that minimizes the PURE (8.4) with $\boldsymbol{\theta}(\mathbf{d}, \mathbf{s}) = \boldsymbol{\theta}^{\text{PUREshrink}}(\mathbf{d}, \mathbf{s}; a)$.

8.3.2   PURE-LET

Following our recently devised *SURE-LET* strategy[24], we propose to consider a wavelet estimator that is formulated as a *linear expansion of thresholds* (LET), i.e.

$$\boldsymbol{\theta}^{\text{LET}}(\mathbf{d}, \mathbf{s}; \mathbf{a}) = \sum_{k=1}^{K} a_k \boldsymbol{\theta}_k(\mathbf{d}, \mathbf{s}).$$

Thanks to this linear parameterization, the unbiased estimate of the MSE in (8.4) is *quadratic* with respect to the parameters $\mathbf{a} \in \mathbb{R}^K$. Therefore, its minimization boils down to the resolution of a *linear* system of equations with *small* dimension K:

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{c}, \tag{8.7}$$

where for $1 \leq k, \ell \leq K$,

$$\begin{cases} c_k &= [\mathbf{d}^T(\boldsymbol{\theta}_k^-(\mathbf{d}, \mathbf{s}) + \boldsymbol{\theta}_k^+(\mathbf{d}, \mathbf{s})) + \mathbf{s}^T(\boldsymbol{\theta}_k^-(\mathbf{d}, \mathbf{s}) - \boldsymbol{\theta}_k^+(\mathbf{d}, \mathbf{s}))]/2 \\ M_{k,\ell} &= \boldsymbol{\theta}_k(\mathbf{d}, \mathbf{s})^T \boldsymbol{\theta}_\ell(\mathbf{d}, \mathbf{s}) \end{cases} \tag{8.8}$$

The definition of $\boldsymbol{\theta}_k^+(\mathbf{d}, \mathbf{s})$ and $\boldsymbol{\theta}_k^-(\mathbf{d}, \mathbf{s})$ is similar to (8.3).

Basic thresholding function

Similarly to [140], we propose a linearly-parameterized thresholding function with $K = 2$ parameters (a_1 and a_2), whose n-th component is defined by

$$\theta_n^{\text{LET0}}(\mathbf{d}, \mathbf{s}; [a_1\, a_2]^T) = \left(a_1 + a_2 \exp\left(-\frac{d_n^2}{2T^2}\right)\right) d_n. \tag{8.9}$$

As in the *PUREshrink* estimator, the threshold T is directly linked to the local noise variance, estimated from the magnitude of the corresponding scaling coefficient $|s_n|$.

However, thanks to the degrees of freedom provided by the two *linear* parameters a_1 and a_2, the value of this threshold does not need to be optimized. By running several simulations, we found that $T^2 = 6|s_n|$ constituted a good choice, inducing no significant loss compared to a subband-optimized threshold. Our experiments (see Fig. 8.3) show that the above thresholding function is already more efficient (approx. +0.25 dB) than the previously presented *PUREshrink* (8.6).

Interscale sign dependencies

The integration of interscale dependencies has already been shown to bring a substantial improvement in the context of additive Gaussian white noise removal [140]. Therefore, we propose here an analogous interscale wavelet thresholding, but for Poisson intensity estimation. The idea is to exploit the scaling coefficients **s** to "predict" and reinforce the significant wavelet coefficients of **d** at the same scale. Indeed, the scaling coefficients offer improved SNR because they arise from Poisson random variables with summed intensities. Thanks to the use of *Haar* filters, there is no group delay between the lowpass H_a and highpass G_a analysis filters. An interscale predictor $\tilde{\mathbf{d}}$ of **d** can thus be obtained simply by applying the *anti-symmetric* filter $G(z) = z^{-1} - z$ to the scaling coefficients **s**. This filter G is the shortest filter that ensures a perfect feature alignement, while transforming the lowpass subband into an approximate highpass subband, as illustrated in the 2D example of Fig. 8.2.

Lowpass: **s** Interscale predictor: $\tilde{\mathbf{d}}$ Highpass subband: **d**
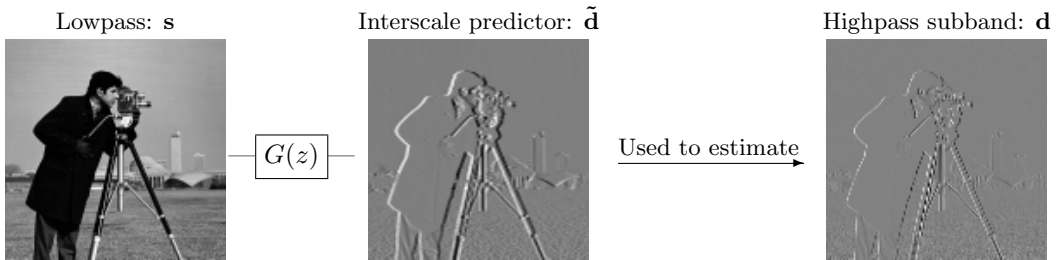
$G(z)$ Used to estimate

Figure 8.2: The interscale predictor of subband HL_j (resp. LH_j, resp. HH_j) is obtained by horizontally (resp. vertically, resp. horizontally and vertically) filtering the same-scale lowpass subband LL_j with the filter $G(z) = z^{-1} - z$.

By taking a closer look at Fig. 8.2, it can be observed that the signs of the inter-

scale predictor coefficients are consistent with those of the corresponding highpass subband. This suggests adding a term proportional to this interscale predictor into the simple thresholding function (8.9), leading to

$$\theta_n^{\text{LET1}}(\mathbf{d}, \mathbf{s}; [a_1\, a_2\, a_3]^T) = \theta_n^{\text{LET0}}(\mathbf{d}, \mathbf{s}; [a_1\, a_2]^T) + a_3 \underbrace{(s_{n-1} - s_{n+1})}_{\tilde{d}_n}. \qquad (8.10)$$

This simple strategy brings significant improvements (approx. +0.5 dB). This was confirmed by multiple experiments on standard grayscale images; some of these simulations are reported in Fig. 8.3.

Interscale amplitude-sign dependencies

Further improvements can be obtained by grouping together wavelet coefficients of similar magnitudes. This grouping is based on the magnitude of a smoothed version $\mathbf{p}$ of the previously defined interscale predictor $\tilde{\mathbf{d}}$ to increase the robustness toward noise. The smoothed version of the interscale predictor is simply obtained by applying a normalized Gaussian kernel on the absolute value of $\tilde{\mathbf{d}}$, i.e., $p_n = \sum_k |\tilde{d}_k| f_{n-k}$, where $f_k = e^{-k^2/2}/\sqrt{2\pi}$. The proposed interscale wavelet thresholding is thus finally defined as

$$
\begin{aligned}
\theta_n^{\text{LET2}}(\mathbf{d}, \mathbf{s}; \mathbf{a}, \mathbf{b}) &= \exp\left(-\frac{p_n^2}{12|s_n|}\right) \theta_n^{\text{LET1}}(\mathbf{d}, \mathbf{s}; \mathbf{a}) \\
&\quad + \left(1 - \exp\left(-\frac{p_n^2}{12|s_n|}\right)\right) \theta_n^{\text{LET1}}(\mathbf{d}, \mathbf{s}; \mathbf{b}). \qquad (8.11)
\end{aligned}
$$

In Fig. 8.3 and 8.4, it is seen that this latter interscale wavelet estimator clearly gives the best results, both quantitatively and visually, among all estimators presented here. Note that the PURE-based adjustment of the parameters $\mathbf{a}$ and $\mathbf{b}$ gives a SNR gain that is very close to the optimum (which is obtained from an *oracle* adjustment of these parameters using the knowledge of the original image).

8.4   Results on simulated data

We now propose to compare our PURE-based approach with the following multiscale methods in simulated experiments:
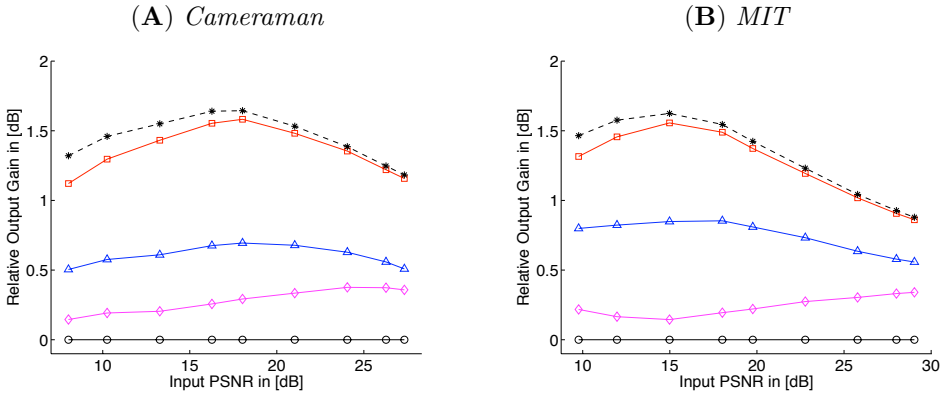
- A standard procedure, which consists of three steps:

(A) *Cameraman*   **(B)** *MIT*

Figure 8.3: PSNR improvements brought by the *PURE-LET* strategy over *PUREshrink* (Formula (8.6): "∘" markers), for two standard grayscale images. Formula (8.9): "⋄" markers. Formula (8.10): "△" markers. Formula (8.11): "□" markers). Oracle results for Formula (8.11) are shown with "∗" markers.

1. Apply a variance-stabilizing transformation (VST) on the Poisson-distributed data. In the experiments, we have tried the Anscombe[7] and the Haar-Fisz[85] transforms.

2. Use any available denoiser designed for additive Gaussian white noise removal: we have chosen Donoho and Johnstone's (subband-dependent) *SUREshrink* [66].

3. Apply the corresponding inverse variance-stabilizing transformation to the denoised data to finally get an estimation of the underlying Poisson intensities.

This type of approach has the advantage of giving relatively good results at low computational cost; moreover it is easily reproducible .

• A more involved algorithm: we have retained Willett and Nowak's *Platelet* approach[2][249], which stands among the state-of-the-art algorithms for Poisson intensity estimation [257].

---

[2]Matlab code downloadable at: `http://www.ee.duke.edu/∼willett/Research/platelets.html`

(**A**)   (**B**)

(**C**)   (**D**)

(**E**)   (**F**)

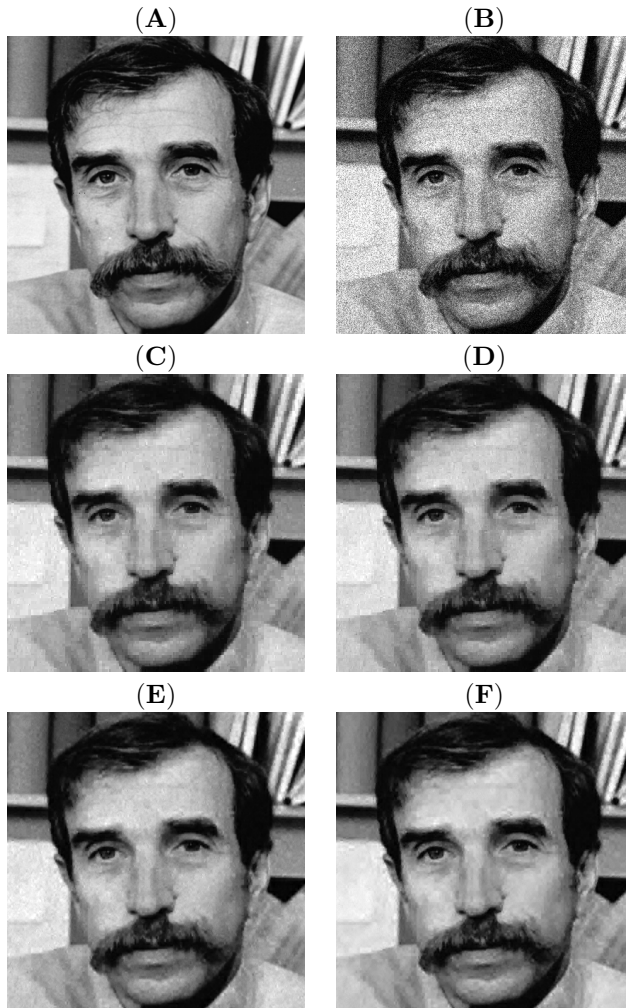**Figure 8.4:** Visual quality of the various estimators presented in Section 8.3. (**A**) Part of the original *Al* image ($I_{\max} = 40$). (**B**) Noisy realization of it: PSNR = 18.83 dB. (**C**) Denoised by *PUREshrink* (8.6): PSNR = 27.61 dB. (**D**) Denoised by *PURE-LET0* (8.9): PSNR = 27.81 dB. (**E**) Denoised by *PURE-LET1* (8.10): PSNR = 29.12 dB. (**F**) Denoised by *PURE-LET2* (8.11): PSNR = 29.69 dB.

For all wavelet-based methods presented in this section, we use the same number of decomposition levels, i.e. 4 (resp. 5) for 256×256 (resp. 512×512) images. The input peak signal-to-noise ratios (PSNR[3]) are adjusted by rescaling the original test images[4], from a maximum intensity of 120 to a minimum of 1.

Table 8.1 summarizes the PSNRs obtained by the various algorithms, both in a non-redundant and in a redundant framework (using cycle spinning [48]). It can be observed that the PURE-based approach clearly outperforms (around $+1.5$ dB, on average) the standard VST-based wavelet denoiser applied in an *orthonormal* wavelet basis. Note that the retained basis (Daubechies' *symlets* with eight vanishing moments[57]) is smoother—and thus more suitable for image-restoration tasks—than the basic *Haar* wavelets that we use. Our solution also gives significantly better PSNRs ($\sim 0.5 - 1$ dB) than the non-redundant version of the *Platelet* approach. In most cases, the proposed algorithm gives even better results (between $0.5 - 1.5$ dB, on average) than the VST-based thresholding applied in a *redundant* cycle-spinning framework, and results similar to two cyclic shifts of *Platelets*.

Table 8.2 gives more insights concerning the tradeoff between the degree of redundancy and the computation time of the various algorithms. The *Platelet* procedure achieves the best PSNRs when considering a high number (25) of cyclic shifts. However, these results are obtained at a prohibitive computational cost. Cyclic shifts of our PURE-based approach also brings some gains (around 1 dB), despite the fact that an independent "shift-wise" PURE minimization is sub-optimal (as shown in [24, 183] for the Gaussian case). There is probably room for improvement by deriving a rigorous unbiased estimate of the MSE for redundant processing of Poisson data; but this is outside the scope of the present work.

As shown in Fig. 8.5, our interscale *PURE-LET* algorithm removes most of the Poisson noise, without over-smoothing the underlying intensities. Moreover, from a computational point of view, it takes only ~ 0.5s to denoise a 512×512 image with the current Matlab implementation of our algorithm; this corresponds to the optimization of 90 parameters. Under the same conditions, the execution of the cycle-spinning *SUREshrink* (25 cyclic shifts) combined with the *Haar-Fisz* variance-stabilizing transform requires ~ 5.5s, while a single shift of the *Platelets* lasts 150s, on average.

We also compared the proposed *PURE-LET* algorithm with our previously described interscale *SURE-LET* strategy, specifically devised for Gaussian noise statistics[140]. For a fair comparison, we used an adapted implementation of the

---

[3]Defined as: PSNR $= 10 \log_{10} \dfrac{I_{\max}^2}{\text{MSE}}$, where $I_{\max}$ is the maximum intensity of the noise-free image.

[4]Available at: `http://bigwww.epfl.ch/luisier/Test-Images.zip`

(**A**)

(**B**)

(**C**)

(**D**)

(**E**)
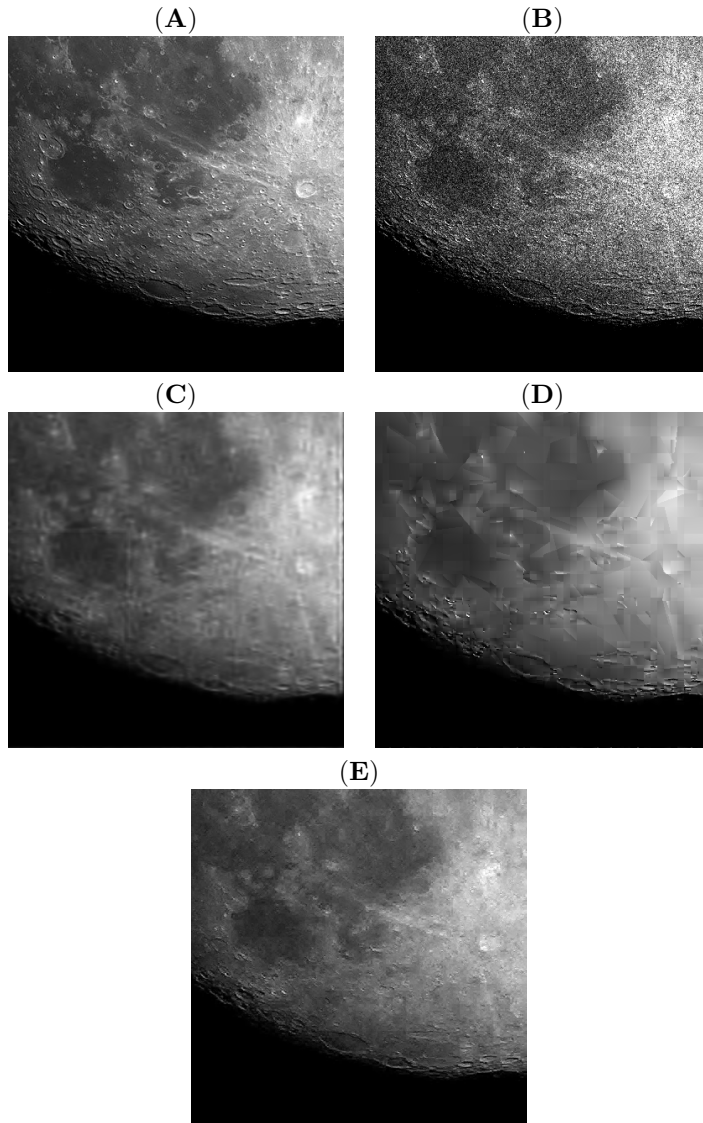
Figure 8.5: (**A**) Part of the original *Moon* image ($I_{\max} = 15$), by courtesy of the following website: http://www.galaxyimages.com. (**B**) Degraded *Moon*: PSNR $=$ 17.24 dB. (**C**) Denoised with *Haar-Fisz + SUREshrink* (25 cyclic shifts of OWT *sym8*): PSNR $=$ 24.15 dB in 5.7 s. (**D**) Denoised with *Platelet* (2 cyclic shifts): PSNR $=$ 24.70 dB in 218 s. (**E**) Denoised with our interscale *PURE-LET* (OWT *Haar*): PSNR $=$ 25.44 dB in 0.65 s.

Table 8.1: Comparison of Multiscale Poisson Noise Removal Algorithms

Cameraman 256 × 256

	Images	120	60	30	20	10	5	1
	I_{max}							
	Input PSNR	24.05	21.03	18.03	16.27	13.27	10.25	3.28
Non-redundant	Anscombe + *SUREshrink* (OWT *sym8*)	28.53	26.37	24.26	23.29	21.81	19.61	15.36
	Haar-Fisz + *SUREshrink* (OWT *sym8*)	28.49	26.40	24.36	23.69	22.39	20.93	18.57
	Platelet	28.29	26.79	25.44	24.60	23.24	21.49	18.70
	Intescale *PURE-LET* (OWT *Haar*)	**30.07**	**28.28**	**26.54**	**25.55**	**23.94**	**22.42**	**19.18**
Redundant	Haar-Fisz + *SUREshrink* (25 cyclic shifts of OWT *sym8*)	29.35	27.30	25.09	24.46	23.02	21.46	18.71
	Platelet (2 cyclic shifts)	29.34	27.84	26.21	25.32	24.05	22.35	19.34
	Intescale *PURE-LET* (2 cyclic shifts of OWT *Haar*)	**30.36**	**28.56**	**26.87**	**25.89**	**24.32**	**22.76**	**19.67**

Moon 512 × 512

	Images	120	60	30	20	10	5	1
	I_{max}							
	Input PSNR	26.27	23.25	20.23	18.48	15.47	12.46	5.47
Non-redundant	Anscombe + *SUREshrink* (OWT *sym8*)	29.00	26.92	24.83	24.09	22.99	21.38	18.37
	Haar-Fisz + *SUREshrink* (OWT *sym8*)	29.03	26.89	25.01	24.33	23.55	22.78	21.98
	Platelet	27.16	26.01	25.05	24.60	23.96	23.63	22.97
	Intescale *PURE-LET* (OWT *Haar*)	**29.62**	**27.97**	**26.56**	**25.87**	**24.92**	**24.23**	**23.16**
Redundant	Haar-Fisz + *SUREshrink* (25 cyclic shifts of OWT *sym8*)	29.61	27.39	25.40	24.68	23.83	23.06	22.09
	Platelet (2 cyclic shifts)	27.99	26.76	25.60	25.09	24.28	23.74	22.69
	Intescale *PURE-LET* (2 cyclic shifts of OWT *Haar*)	**29.77**	**28.09**	**26.70**	**25.97**	**24.99**	**24.28**	**23.19**

Peppers 256 × 256

	Images	120	60	30	20	10	5	1
	I_{max}							
	Input PSNR	23.92	20.92	17.91	16.17	13.14	10.12	3.14
Non-redundant	Anscombe + *SUREshrink* (OWT *sym8*)	29.29	27.27	24.85	23.56	21.47	19.77	15.32
	Haar-Fisz + *SUREshrink* (OWT *sym8*)	29.27	27.30	24.54	23.82	21.98	21.03	18.69
	Platelet	29.07	27.44	25.73	24.92	23.23	21.57	18.17
	Intescale *PURE-LET* (OWT *Haar*)	**30.28**	**28.51**	**26.72**	**25.70**	**23.81**	**21.99**	**18.92**
Redundant	Haar-Fisz + *SUREshrink* (25 cyclic shifts of OWT *sym8*)	30.29	28.37	25.45	24.49	22.43	21.53	18.90
	Platelet (2 cyclic shifts)	30.61	28.52	27.08	26.11	24.34	22.13	18.88
	Intescale *PURE-LET* (2 cyclic shifts of OWT *Haar*)	**30.79**	**29.07**	**27.27**	**26.18**	24.29	**22.52**	**19.33**

MIT 256 × 256

	Images	120	60	30	20	10	5	1
	I_{max}							
	Input PSNR	25.78	22.78	19.76	18.02	14.98	11.97	4.99
Non-redundant	Anscombe + *SUREshrink* (OWT *sym8*)	28.80	26.70	24.10	22.58	21.21	18.14	13.93
	Haar-Fisz + *SUREshrink* (OWT *sym8*)	28.78	26.74	24.25	22.88	21.78	20.00	16.03
	Platelet	28.25	26.44	24.70	23.47	21.65	20.18	16.24
	Intescale *PURE-LET* (OWT *Haar*)	**30.03**	**27.92**	**25.96**	**24.85**	**23.01**	**21.18**	**17.43**
Redundant	Haar-Fisz + *SUREshrink* (25 cyclic shifts of OWT *sym8*)	29.67	27.67	25.21	23.77	22.64	20.89	16.59
	Platelet (2 cyclic shifts)	29.23	27.66	25.64	24.63	22.99	21.27	17.27
	Intescale *PURE-LET* (2 cyclic shifts of OWT *Haar*)	**30.47**	**28.41**	**26.55**	**25.34**	**23.49**	**21.63**	**17.82**

<u>Note:</u> Output PSNRs have been averaged over ten noise realizations, except for the *Platelet* approach.

Table 8.2: Cycle-spinning: PSNR improvement and computation time

Cameraman at 17.25 dB						
Algorithms	**Non-redundant**		**2 cyclic shifts**		**25 cyclic shifts**	
	PSNR [dB]	**Time [s]**	**PSNR [dB]**	**Time [s]**	**PSNR [dB]**	**Time [s]**
Haar-Fisz + _SUREshrink_	24.01	0.06	24.49	0.11	24.83	1.3
Interscale _PURE-LET_	26.09	0.19	26.45	0.37	26.75	4.6
Platelet	25.01	44.5	25.83	89	27.17	1112
MIT at 18.98 dB						
Algorithms	**Non-redundant**		**2 cyclic shifts**		**25 cyclic shifts**	
	PSNR [dB]	**Time [s]**	**PSNR [dB]**	**Time [s]**	**PSNR [dB]**	**Time [s]**
Haar-Fisz + _SUREshrink_	23.47	0.06	24.10	0.11	24.64	1.3
Interscale _PURE-LET_	25.48	0.19	26.02	0.37	26.43	4.6
Platelet	24.10	36	25.33	72	26.80	891

<u>Note:</u> Output PSNRs and computation times have been averaged over ten noise realizations, except for the Platelet approach.

SURE-LET algorithm that involved the same number of parameters as the present _PURE-LET_ method. We also considered the same wavelet transformation, i.e. OWT Haar, for both techniques. As can be seen in Fig. 8.6, applying the _SURE-LET_ strategy in the VST-domain is less efficient for small intensities (over 0.5 dB loss for intensities lower than 10). This can be attributed to the rigorous minimization of an estimate of the actual MSE that is performed by the proposed _PURE-LET_ algorithm.

8.5   Application to real biological data

In this last section, we describe the application of our denoising algorithm to real fluorescence-microscopy images of biological samples. The images were acquired on a Leica TCS SP5 confocal microscope at the Imaging Center of the IGBMC (Institut de Génétique et de Biologie Moléculaire et Cellulaire, Illkirch, France). This microscope is equipped with a water-cooled scan-head and low-noise PMTs[5] (Hamamatsu R 9624).

---

[5]PMT stands for photomultiplier tube.

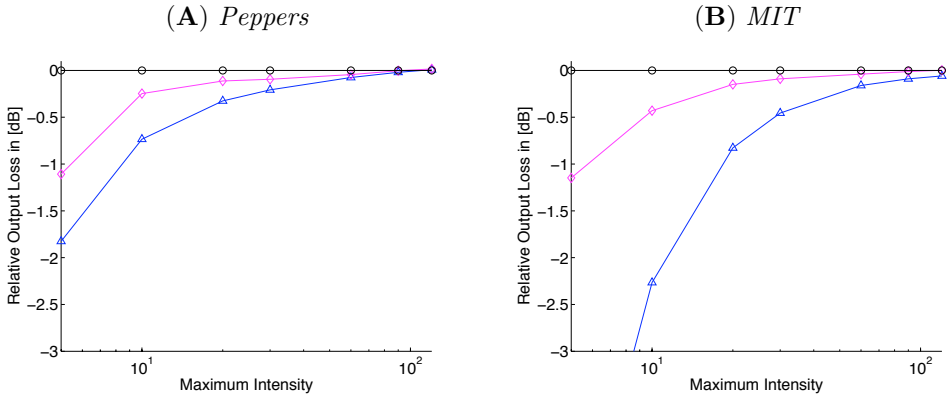**(A)** *Peppers*   **(B)** *MIT*

Figure 8.6: PSNR improvements brought by the direct handling of Poisson statistics (*PURE-LET* (OWT Haar): "○" markers) over VST + *SURE-LET* schemes (*Anscombe + SURE-LET* (OWT Haar): "△" markers; *Haar-Fisz + SURE-LET* (OWT Haar): "◇" markers).

8.5.1   Brief description of the acquisition process

The measurement process is the same for every scan position (pixel); it is illustrated in Fig. 8.7 and can be summarized as follows (see e.g. [174] for a more detailed description). The number of photons arriving at the PMT during the integration time follows a Poisson distribution of mean λ determined by the source intensity. Each photon may traverse the protection window, penetrate the photocathode and be converted to an electron with a certain probability η; this probability is known as the quantum efficiency of the PMT and is on the order of 30% for the best models. The conversion process can be seen as a binomial selection [212, 14] and according to Prop. 17, the number of electrons at the output of the photocathode (= photoelectrons) follows a Poisson distribution of mean $\mu = \eta\lambda$. The number of photoelectrons represents a shot-noise-corrupted measurement m of the intensity signal μ in our framework.

The electrons are then multiplied (via several amplification stages) and converted to an electric current that is integrated and quantized. The recorded signal is essentially proportional to the number of photoelectrons; although the amplification factor may fluctuate in practice, recent work [251] suggests that the newest
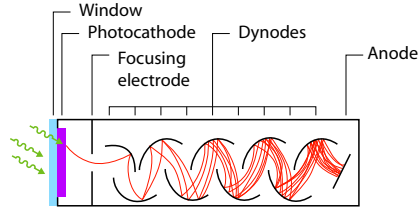
Figure 8.7: Schematic representation of a photomultiplier tube (PMT).

PMTs with high first-stage gain have the ability to discriminate between multi-photoelectron events, at least for low count numbers. Under this assumption the measurement process is purely shot-noise limited and one can estimate the number of photoelectrons by adequately renormalizing the data.

8.5.2 Data preprocessing and verification of the statistical model

Our normalization procedure is based on the characteristic property of a Poisson random variable that its mean is equal to its variance. Thus, by dividing the data by the amplification factor (gain), we should approximately retrieve this behavior.

To validate our approach, we acquired a set of 100 images of the same object (a *C. elegans* embryo) under low illumination intensity. We could thus compute estimates of the mean and variance for every pixel. The amplification factor (gain) was determined by fitting a linear function to these mean-variance measurements. After dividing by the gain (and subtracting a constant corresponding to the offset of the detector), the frequency distribution of the pixel values was found to be in good agreement with Poisson statistics. Fig. 8.8 shows the histogram of the normalized pixel values for those pixels whose mean was equal to a given value μ ($\pm 5\%$).

In practice one cannot use multiple realizations of the same image to obtain the amplification gain. We found that estimates of the local mean and variance based on spatial averaging can yield a good estimate of the gain. This approach has been used for the results presented in the next subsection.
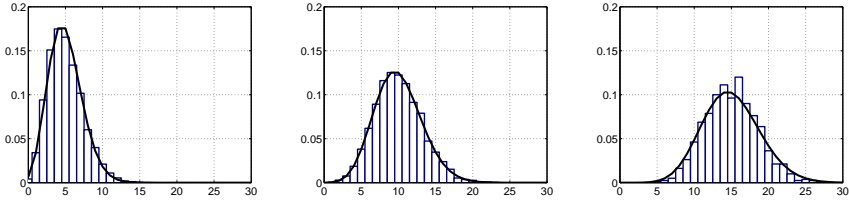
Figure 8.8: Comparison of the Poisson distribution with the histogram of the normalized data for $\mu = 5, 10, 15$.

8.5.3 Denoising of color fluorescence micrographs

Fig. 8.9 (**A**) shows the 2D confocal image of a *C. elegans* embryo labeled with three fluorescent dyes. We reproduced very noisy experimental conditions by reducing the laser power and using short exposure times. The three color channels were processed separately.

The result of the SUREShrink algorithm with a Haar-Fisz transform and 25 cyclic shifts is shown in Fig. 8.9 (**C**)). The result of our algorithm is sharper and shows less artifacts (Fig. 8.9 (**D**)). Besides, it is less noisy than the image shown in Fig. 8.9 (**B**), which corresponds (virtually) to a four times longer exposure time.

8.6 Conclusion

The above results suggest that our PURE-based approach is a promising alternative for denoising Poisson-corrupted images. Although our PURE-LET method is based on a maximally decimated Haar wavelet transform, it yields results that are comparable or superior to standard translation-invariant approaches. At the same time, our algorithm has substantially lower computational complexity and smaller memory requirements than the latter.

These are appealing features for applications in dynamic fluorescence microscopy, where biologists must often acquire large time-lapse image series under very low light conditions. Our method potentially allows for a reduction in exposure time and/or laser intensity, which are critical parameters when imaging fast-moving live samples.

We are currently working on an extension of the algorithm so as to allow for arbitrary (redundant) wavelet decompositions. It would also be interesting to test
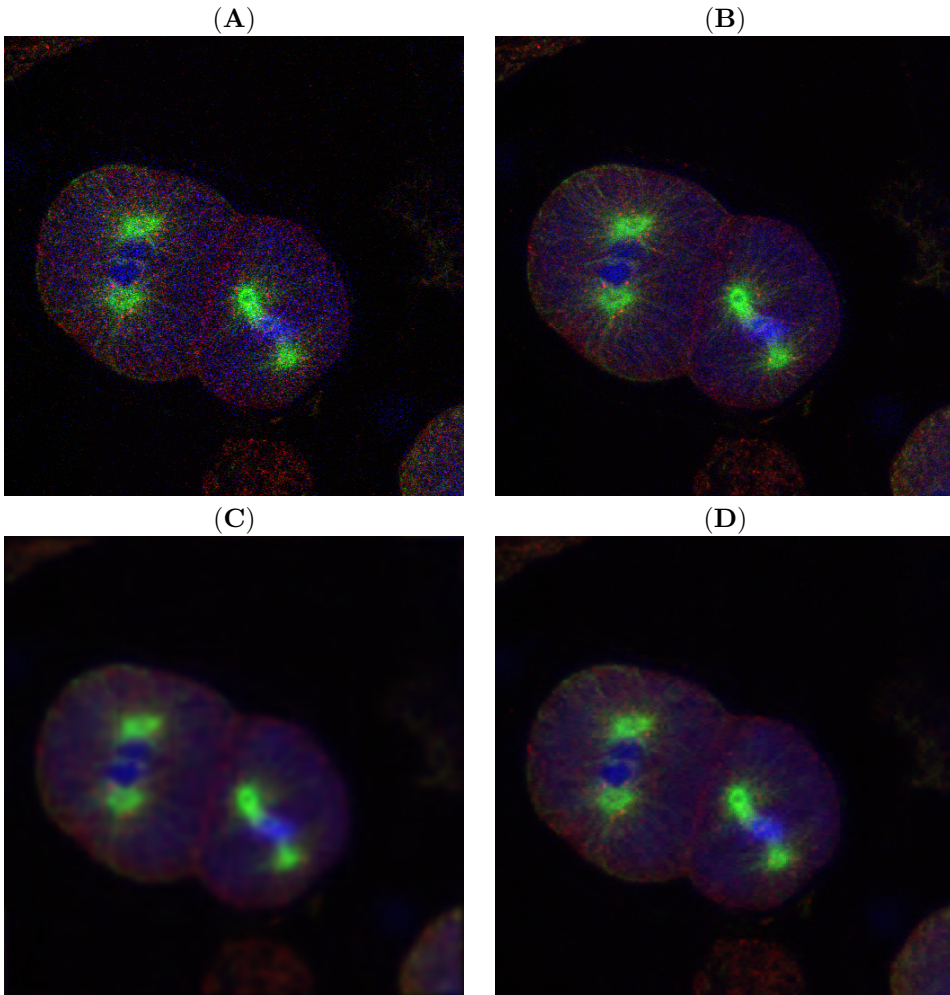
Figure 8.9: (**A**) Input image for the denoising algorithms. (**B**) Reference image: average of four images of the same sample. (**C**) Denoising result with *Haar-Fisz + SUREshrink* (25 cyclic shifts of OWT *sym8*). (**D**) Denoising result with our interscale *PURE-LET* (OWT *Haar*).

our method on a confocal microscope equipped with a photon-counting module, instead of a standard device measuring the output current.

Chapter 9

Recursive risk estimation and its application to wavelet-regularized deconvolution

Summary

We propose a recursive data-driven risk-estimation method for non-linear iterative deconvolution. Our two main contributions are 1) a solution-domain risk-estimation approach that is applicable to non-linear restoration algorithms for ill-conditioned inverse problems; and 2) a risk estimate for a state-of-the-art iterative procedure, the thresholded Landweber iteration, which enforces a wavelet-domain sparsity constraint. Our method can be used to estimate the SNR improvement at every step of the algorithm; e.g., for stopping the iteration after the highest value is reached. It can also be applied to estimate the optimal threshold level for a given number of iterations.

9.1   Introduction

Deconvolution software plays an increasingly important role in various imaging applications, such as optical microscopy, medical imaging, astronomy, or satellite imaging. While there are many different deconvolution algorithms available [17, 235], virtually all of them depend on a set of parameters that must be properly adjusted to obtain satisfying results. Depending on the type of algorithm, one may have to tune various parameters such as regularization factor, step size, number of iterations, etc. Many end-users do not feel comfortable with this task, unless they have a good practical experience of the algorithm. As a result, they often resort to a subjective choice based on purely qualitative visual criteria. This reduces the reproducibility of the results and hinders comparison between data sets.

Consequently, the problem of parameter adjustment is of considerable importance for deconvolution-software manufacturers. There are various methods for automating this task based on quantitative criteria, such as the discrepancy principle, the L-curve method, the C_L method or generalized cross-validation (see for example Chapter 7 of [235]). However, the underlying performance criteria are expressed in the measurement domain, which gives only a partial account of the actual restoration quality. Signal-domain approaches are less-common; to our knowledge, they have only been applied to linear deconvolution (see [86] and our recent work [181]).

Blu and Luisier [24] recently revitalized the principle of data-driven image restoration in the context of denoising, based on a risk estimate introduced by Stein [209]. In this chapter, we use a similar estimate for ill-conditioned deconvolution problems. Specifically, we construct a recursive risk estimate—which depends only on the measurement—for a non-linear iterative deconvolution algorithm: the so-called "thresholded-Landweber" (TL) algorithm (see [78, 58, 16] and Chapter 5. This wavelet-based method belongs to the state-of-the-art in image deconvolution (we refer the reader to the experimental results presented in [78]). Nevertheless, the optimal result is typically obtained before the algorithm has fully converged, that is, for a particular number of iterations (see the example in Fig. 9.1). In such a situation, our method can be applied to monitor the (estimated) SNR improvement during the execution, so as to automatically stop the algorithm after the highest value has been reached. Our experiments show that it can also be used to adjust the threshold level.

In this chapter, we consider the algebraic image-formation equation

$$\mathbf{y} = \mathbf{Hx} + \mathbf{b},$$

where $\mathbf{H}$ is a block-circulant matrix that represents the effect of a convolutive image-

formation device. $\mathbf{b} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is a vector that models the errors introduced by the measurement device (e.g. a CCD camera); we assume that its entries are realizations of IID Gaussian random variables of mean 0 and variance σ^2.

A deconvolution method can be seen as a function $\mathbf{f} : \mathbb{R}^N \to \mathbb{R}^N$ that returns an estimate of the original image $\mathbf{x}$ from the measurement $\mathbf{y}$ (N stands for the dimension of these vectors, that is, the number of pixels of the corresponding images). The design of a good deconvolution method is a difficult task, because the inversion of $\mathbf{H}$ is typically an ill-posed problem [17]. Most approaches are based on variational principles or on statistical assumptions whose discussion would go beyond the scope of this chapter. The performance of a deconvolution method $\mathbf{f}$ is usually measured in terms of the quadratic error (or "risk") $r = \|\mathbf{x} - \mathbf{f}(\mathbf{y})\|^2$, where $\|\cdot\|$ is the ℓ_2 norm. Of course, evaluating the risk is only possible in phantom experiments where the original image $\mathbf{x}$ is known. Such a quantity (which requires the knowledge of the ground-truth image) is often refered to as an "oracle" measure.

9.2   A general risk-estimation approach

In the sequel, we consider the case where the deconvolution algorithm depends on a parameter p; we thus use the notation $\mathbf{f}_p(\mathbf{y})$ instead of $\mathbf{f}(\mathbf{y})$. Ideally, the parameter p should be selected so as to minimize the corresponding risk

$$r_p = \|\mathbf{x} - \mathbf{f}_p(\mathbf{y})\|^2.$$

Since the ground-truth is obviously not available in real situations, our first goal is to derive an estimate of this quantity that depends only on the available information about $\mathbf{x}$, that is, the measurement $\mathbf{y}$. We have already constructed and successfully employed such a risk estimate in the case where $\mathbf{f}_p$ corresponds to a (linear) Wiener-type inverse filter with regularization parameter p [181]. Here, we consider a general (possibly non-linear) deconvolution algorithm. To obtain an estimate in this case, we rewrite the risk as

$$\begin{aligned} r_p &= \|\mathbf{x}\|^2 - 2\mathbf{x}^T \mathbf{f}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2 \\ &= \|\mathbf{x}\|^2 - 2(\mathbf{H}_{\text{inv}}\mathbf{H}\mathbf{x})^T \mathbf{f}_p(\mathbf{y}) \\ &\quad - 2([\mathbf{I} - \mathbf{H}_{\text{inv}}\mathbf{H}]\mathbf{x})^T \mathbf{f}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2 \\ &= \|\mathbf{x}\|^2 - 2\mathbf{y}^T \mathbf{g}_p(\mathbf{y}) + 2\mathbf{b}^T \mathbf{g}_p(\mathbf{y}) \\ &\quad - 2([\mathbf{I} - \mathbf{H}_{\text{inv}}\mathbf{H}]\mathbf{x})^T \mathbf{f}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2, \end{aligned}$$

where $\mathbf{H}_{\text{inv}}$ is a stabilized approximation of the inverse of $\mathbf{H}$ (see e.g. [17]) and $\mathbf{g}_p(\mathbf{y}) = \mathbf{H}_{\text{inv}}^T \mathbf{f}_p(\mathbf{y})$. There are three terms that are related to unknown quantities

($\mathbf{x}$ and $\mathbf{b}$) in the last expression. The first one, $\|\mathbf{x}\|^2$, is a constant[1] with respect to p and is thus irrelevant for our purpose (we are only interested in *minimizing* the risk). The second one, $\mathbf{b}^T \mathbf{g}_p(\mathbf{y})$, can be estimated using a technique that is quite common in the field of statistics, which follows from the equality

$$\mathsf{E}[\mathbf{b}^T \mathbf{g}_p(\mathbf{y})] = \sigma^2 \mathsf{E}[\text{div } \mathbf{g}_p(\mathbf{y})],$$

where div is the divergence operator (see [209] or [24], which also state adequate conditions on $\mathbf{g}$). If $\mathbf{H}_{\text{inv}}\mathbf{H}$ is reasonably close to $\mathbf{I}$, then the third term ($[\mathbf{I} - \mathbf{H}_{\text{inv}}\mathbf{H}]\mathbf{x})^T \mathbf{f}_p(\mathbf{y})$ is negligible.

Based on these considerations, we propose to estimate the risk using the quantity

$$e_p = c - 2\mathbf{y}^T \mathbf{g}_p(\mathbf{y}) + 2\sigma^2 \text{div } \mathbf{g}_p(\mathbf{y}) + \|\mathbf{f}_p(\mathbf{y})\|^2.$$

This quantity only depends on the measurement $\mathbf{y}$ (here, c represents an arbitrary constant, which does not influence the minimum of e_p with respect to p).

9.3   Recursive risk estimation

The general approach described in the previous section must be further specified in the case of an iterative method such as the TL algorithm. For simplicity, we consider the case of soft-thresholding in an orthonormal wavelet basis; however, the following ideas can readily be extended to a more general setting (biorthogonal or redundant decompositions; other thresholding functions).

When it is run until convergence, the classical Landweber iteration [17] induces considerable noise amplification for ill-conditioned inverse problems such as deconvolution. The TL algorithm can be understood as a way to compensate for this effect using wavelet-coefficient thresholding, a technique known for its effectiveness in image-denoising applications. Starting from some initialization $\mathbf{x}^{(0)}$, it produces a sequence of images $\mathbf{x}^{(k)}$ according to the update rule[2]

$$\mathbf{x}^{(k+1)} = \mathbf{W}\mathcal{T}_\theta\Big\{\mathbf{W}^T\big[\mathbf{x}^{(k)} + \mathbf{H}^T(\mathbf{y} - \mathbf{H}\mathbf{x}^{(k)})\big]\Big\}.$$

Here, $\mathbf{W}$ is an orthonormal wavelet reconstruction matrix (thus $\mathbf{W}^T$ is the corresponding decomposition matrix) and $\mathcal{T}_\theta\{\cdot\}$ denotes a component-wise application of the so-called "soft-thresholding" function (with threshold level θ): $\mathcal{T}_\theta(w) = \text{sign}(w) \max(|w| - \theta, 0)$. This operation guarantees that the deconvolved images

---

[1]Recall that $\mathbf{x}$ is assumed to be a deterministic signal.
[2]Assuming that $\mathbf{H}$ is adequately normalized—see Chapter 5.

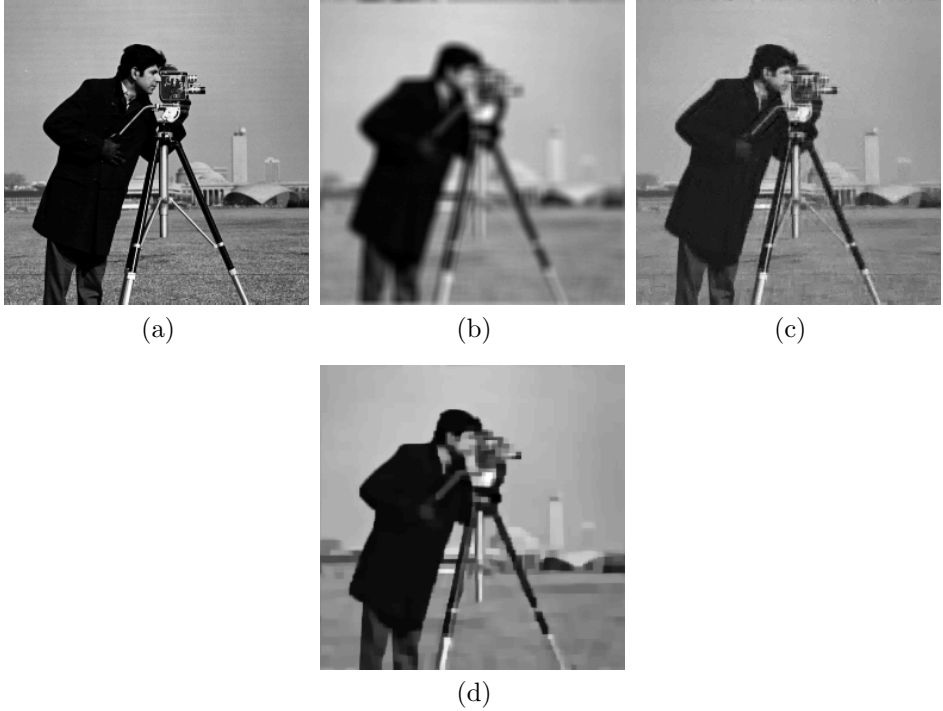(a)   (b)   (c)

(d)

Figure 9.1: (a) Original image. (b) Blurred and noisy image (9×9 uniform blur, Gaussian white noise with BSNR = 40 dB). (c) Restored image for the optimal number of iterations (16) and (d) after 300 iterations.

have a certain degree of sparsity in the wavelet domain (depending on the value of θ). This property makes the TL algorithm very attractive, because many natural images can be well approximated using only a few non-zero coefficients in a given wavelet basis.

Our goal is now to obtain a risk estimate for every iteration k of the TL algorithm. In this case, we can interpret the number of iterations as our parameter ($p = k$). The only difficulty for applying the estimate e_p lies in the computation of the divergence term. We propose to use the identity div $\mathbf{g}_p(\mathbf{y}) = \mathrm{Tr}\, \mathbf{J}_{\mathbf{g}_p}(\mathbf{y})$, where Tr is the trace operator and $\mathbf{J}_{\mathbf{g}_p}(\mathbf{y})$ denotes the Jacobian matrix of $\mathbf{g}_p$, evaluated

at $\mathbf{y}$. In our case, $\mathbf{J}_{\mathbf{g}_p}(\mathbf{y}) = \mathbf{H}_{\text{inv}}^T \mathbf{J}_{\mathbf{f}_p}(\mathbf{y})$; using a functional description of the TL algorithm, we obtain an explicit expression of this matrix. Let us define

$$\mathbf{t}(\mathbf{u}, \mathbf{v}) = \mathbf{W} \mathcal{T}_\theta \Big\{ \mathbf{W}^T (\mathbf{I} - \mathbf{H}^T \mathbf{H}) \mathbf{u} + \mathbf{W}^T \mathbf{H}^T \mathbf{v} \Big\}.$$

Then the result at iteration $p = k$ is given by $\mathbf{x}^{(k)} = \mathbf{f}_k(\mathbf{y})$, where the function $\mathbf{f}_k$ is defined recursively by

$$\begin{cases} \mathbf{f}_0 & \text{(initialization method);} \\ \mathbf{f}_{k+1} = \mathbf{t}(\mathbf{f}_k, \cdot) & \text{(update rule).} \end{cases}$$

This leads to a recursive relation between the Jacobian matrices of the functions $\mathbf{f}_k$:

$$\begin{aligned} \mathbf{J}_{\mathbf{f}_{k+1}}(\mathbf{y}) &= \mathbf{J}_{\mathbf{t}(\mathbf{f}_k, \cdot)}(\mathbf{y}) \\ &= \mathbf{J}_{\mathbf{t}(\mathbf{f}_k, \mathbf{y})}(\mathbf{y}) + \mathbf{J}_{\mathbf{t}(\mathbf{f}_k(\mathbf{y}), \cdot)}(\mathbf{y}) \\ &= \mathbf{J}_{\mathbf{t}(\cdot, \mathbf{y})}(\mathbf{f}_k(\mathbf{y})) \, \mathbf{J}_{\mathbf{f}_k}(\mathbf{y}) + \mathbf{J}_{\mathbf{t}(\mathbf{f}_k(\mathbf{y}), \cdot)}(\mathbf{y}) \\ &= \mathbf{W} \mathbf{D}_k \mathbf{W}^T ([\mathbf{I} - \mathbf{H}^T \mathbf{H}] \mathbf{J}_{\mathbf{f}_k}(\mathbf{y}) + \mathbf{H}^T). \end{aligned}$$

Here $\mathbf{D}_k = \mathbf{J}_{\mathcal{T}_\theta}(\mathbf{W}^T[\mathbf{I} - \mathbf{H}^T \mathbf{H}]\mathbf{f}_k(\mathbf{y}) + \mathbf{W}^T \mathbf{H}^T \mathbf{y})$ is a diagonal matrix, since thresholding is a pointwise operation.

This result can not be applied directly, however. The Jacobian matrices $\mathbf{J}_{\mathbf{f}_k}(\mathbf{y})$ are indeed extremely large for typical image sizes, so that it is unrealistic to compute and store them explicitly. Instead, we propose to exploit the following property: if $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ is independent of $\mathbf{J}$, then $\mathsf{E}[\mathbf{n}^T \mathbf{J} \mathbf{n}] = \text{Tr } \mathbf{J}$ (where the expectation is taken over $\mathbf{n}$). In other words, given a normally distributed noise realization $\mathbf{n}$, we propose to compute the *vector* $\mathbf{J}_{\mathbf{f}_k}(\mathbf{y})\mathbf{n}$ rather than the *matrix* $\mathbf{J}_{\mathbf{f}_k}(\mathbf{y})$. This vector is updated recursively using the above recursion relation, which is computationally equivalent to performing one iteration of the TL algorithm. To estimate the divergence term at a given iteration k, we take the inner product with the filtered noise realization $\mathbf{H}_{\text{inv}}\mathbf{n}$. Thus, the risk estimate that we use in practice is

$$e_p' = c - 2(\mathbf{H}_{\text{inv}}\mathbf{y})^T \mathbf{f}_p(\mathbf{y}) + 2\sigma^2 (\mathbf{H}_{\text{inv}}\mathbf{n})^T \mathbf{J}_{\mathbf{f}_p}(\mathbf{y})\mathbf{n} + \|\mathbf{f}_p(\mathbf{y})\|^2.$$

Typically, one realization of $\mathbf{n}$ is enough to yield an accurate estimate, since the number of samples N is relatively large for images.

9.4   Practical applications of the method

We performed two phantom experiments to illustrate the potential applications of the method. These experiments were conducted under the following general

conditions. The value of σ^2 was assumed to be known; in practical situations it can be estimated using a method similar to the one described in [181]. The TL algorithm was applied with four decomposition levels of the orthonormal Haar wavelet transformation. In addition to the description given in Section 9.3, we applied the random-shift strategy proposed in [78] which was found to yield the best deconvolution results (for non-redundant decompositions). Our risk-estimation method can be easily adapted to account for this additional processing step, without notably changing its computational complexity (which is of the same order as the TL algorithm itself). The initialization $\mathbf{x}^{(0)}$ was generated with the same method as in [78].

The estimate was computed using the simplest form of regularized inverse filter, $\mathbf{H}_{\mathrm{inv}} = (\mathbf{H}^T\mathbf{H} + \epsilon\mathbf{I})^{-1}\mathbf{H}^T$; the constant ϵ was set to $\sigma^2/m_{\mathbf{y}}^2$, where $m_{\mathbf{y}}$ stands for the average (spatial) intensity of the measurement. For graphical-representation purposes, the constant c was chosen such that the risk and our estimate would coincide for the initialization $\mathbf{x}^{(0)}$.

9.4.1   Optimal number of iterations

In our first experiment, we convolved the standard Cameraman image with a 9×9 uniform-blur kernel and added noise to the result; this is illustrated in Fig. 9.1 for a BSNR of 40 dB. We considered the case of a threshold level that is suboptimal for a large number of iterations, which is representative of practical situations. We looked at the evolution of the true risk r_p and of our estimate e'_p as a function of the parameter $p = k$ (number of iterations). The results (signal-to-noise improvement in dB) are presented in Fig. 9.2; $\mathrm{SNRI} = 10\log_{10}(\|\mathbf{y} - \mathbf{x}\|^2/q_p)$, where q_p stands for one of the two previous quantities. The optimal value of the parameter according to the oracle and the estimate are indicated with a star and a circle, respectively.

It is seen that the curves of the true risk and of the estimate are in good agreement. In particular, the maximum of the estimate is sufficiently close to the maximum of the risk to allow for a prediction of the optimal number of iterations; such a prediction would yield the optimal number with an accuracy of about 0.1 dB in terms of SNRI. For BSNR levels of 40 and 30 dB, one would even obtain the optimal result (shown in Fig. 9.1 for a BSNR of 40 dB). In the 20 dB case, using the estimate would result in stopping the algorithm a few iterations before the optimum, which is favorable from a computational standpoint.
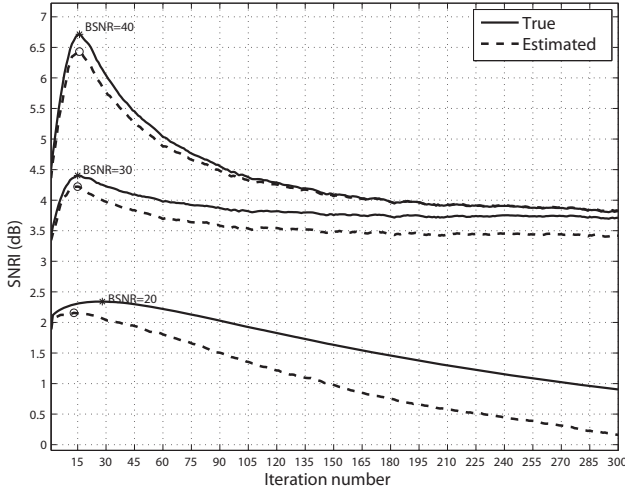
Figure 9.2: Estimation of the SNR improvement as a function of the number of iterations, for different noise levels. In each case, the continuous line corresponds to the true value of the risk (oracle measure) and the dashed one to the estimated risk.

9.4.2   Optimal threshold level

We applied a similar protocol for the parameter $p = \theta$ (threshold level) and a fixed number of iterations ($k = 50$). This time we used the MRI image and a Gaussian convolution kernel of width $\sigma_{\mathrm{kernel}} = 2$ pixels.

The results are shown in Fig. 9.3. The curves of the true risk and of the estimate have the same shape. In particular, the maxima of the curves are again situated within a range of about 0.1 dB; this would allow for a prediction of the optimal threshold level based on the proposed risk estimate.

9.5   Conclusion

We have presented a promising data-driven method for estimating the risk during the execution of a recent non-linear iterative deconvolution algorithm: the thresholded Landweber iteration. Our results suggest that the method is suitable for
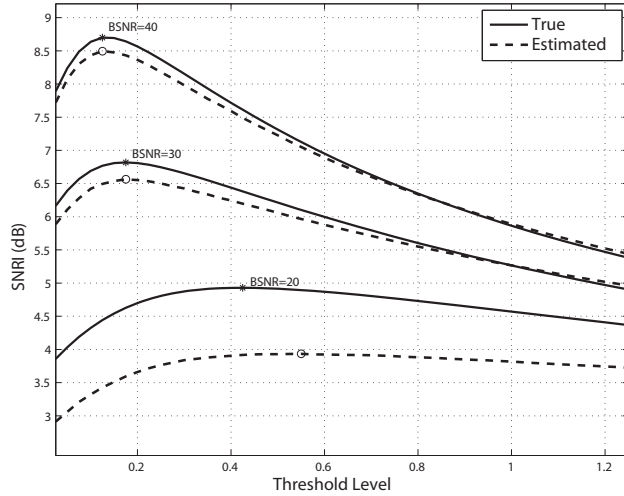
Figure 9.3: Estimation of the SNR improvement as a function of the threshold level, for a fixed number of iterations.

estimating the optimal number of iterations, or for estimating the optimal threshold level. Future work will be concerned with a more extensive study of the accuracy of the method, and with the online estimation of the optimal threshold level (during the execution of the algorithm).

Chapter 10

Conclusion

10.1   Summary of contributions

In this thesis, we have presented several novel algorithmic solutions for image restoration. These procedures correspond to different tradeoffs in terms of computational complexity, restoration quality and automation. Their benefits for fluorescence microscopy can be summarized as follows.

For applications involving weak light intensities (Poisson noise model), we proposed a fully self-adjusting denoising method with very low computational complexity. The method is suitable for automatically processing large quantities of data, such as time-lapse sequences.

For applications with sufficient signal-to-noise ratio, we designed a new multi-level deconvolution algorithm that is based on wavelet-domain ℓ_1 regularization. This regularization technique belongs to the current state of the art in variational image restoration, and its application to fluorescence microscopy is novel. Using our algorithm, its computational cost can typically be brought down to a few dozen FFTs, which is competitive with existing solutions for deconvolution microscopy.

Finally we described a numerical scheme for estimating the signal-to-noise ratio during the execution of an arbitrary iterative image-restoration algorithm. This allows for stopping the algorithm at the optimal number of iterations, and for adjusting other parameters—such as regularization—automatically. Thus our method reduces the need for user interaction, while increasing the reproducibility of the results.

From a methodological standpoint, our developments rely extensively on wavelet

representations and multilevel concepts. Our guiding principles have been to exploit the structure of the problem at hand, and to preserve this structure accross scales. While the wavelet-based multilevel scheme of Part II was primarily developed for image restoration, it can be applied to more general inverse problems. In particular, we have been involved in preliminary investigations concerning fMRI signal restoration, tomographic inversion and MR image reconstruction.

We also made a pedagogical effort in summarizing the principles of fluorescence microscopy and risk-estimation techniques for a signal-processing audience. We highlighted certain aspects of risk estimation that are likely to be relevant for future developments, in particular with respect to ill-posed inverse problems.

10.2   Open algorithmic and conceptual questions

The present work poses a series of questions that open the way for future investigations.

An alternative to wavelet-coefficient penalization that is also considered state-of-the-art in variational image restoration is total-variation regularization. This formulation is computationally more challenging (it amounts to a constrained optimization problem). Although there have been various attempts to apply multilevel methods to total-variation minimization—among the most recent ones, see [44, 81] and references therein—it might be worth trying to adapt our wavelet-based approach.

If restoration quality is a priority, one should assess the benefits of using more advanced non-separable wavelet representations. Especially appealing are recently-proposed wavelet frames that are motivated by invariance principles and promise to be more adapted to the geometry of natural images [224]. However, in our context of high-dimensional image restoration, one should carefully consider the tradeoffs in terms of computational complexity and memory consumption (filter computation and storage; redundancy of the representation).

Another way to improve restoration quality is to introduce more degrees of freedom into the restoration procedure. With this approach, even the simple Haar wavelet transform can yield competitive denoising results, as suggested by the numerical experiments of Chapter 8 and previous results [24]. For iterative methods, however, this comes at the expense of automation, since there are more parameters to adjust. In practice, generic adjustment methods such as the one presented in Chapter 9 are computationally infeasible for many parameters.

These considerations lead to another algorithmic challenge for deconvolution: is it possible to design a non-iterative technique that would reach the restoration

quality of recent variational methods while allowing for a complete self-adjustment of its parameters? A further advantage with respect to iterative techniques would be a significant reduction of the computational cost. We are planning to investigate this issue in the near future.

From a conceptual standpoint, there is a computational "trick" that remains to be explained: in our numerical experiments, the best results were systematically obtained using random shifts, a cheap version of cycle spinning [48] for iterative methods. There is currently no variational interpretation for this operation, although cycle spinning has been justified on theoretical grounds [183]. This poses a more fundamental question with respect to the paradigm of sparsity; in fact, a very recent investigation [74] suggests that sometimes "a weighted average of sparse representations is better than the sparsest one alone"...

Finally, the recent switch of mass-market computer manufacturers to multicore architectures and the possibility of using video-card hardware via specific libraries (e.g. CUDA) will inevitably set new standards for algorithm designers. It may eventually be worth investigating parallel versions of the sequential methods discussed here.

Bibliography

[1] D. A. Agard and J. W. Sedat. Three-dimensional architecture of a polytene nucleus. *Nature*, 302(5910):676–681, April 1983.

[2] F. Aguet, D. Van De Ville, and M. Unser. A maximum-likelihood formalism for sub-resolution axial localization of fluorescent nanoparticles. *Optics Express*, 13:10503–10522, 2005.

[3] F. Aguet, C. Vonesch, J.-L. Vonesch, and M. Unser. *Microscopic image analysis for life science applications*, chapter An introduction to fluorescence microscopy: basic principles, challenges and opportunities, pages 85–114. Artech House, July 2008.

[4] G. Allaire. *Analyse numérique et optimisation*. Éditions de l'École Polytechnique, Paris, France, October 2005.

[5] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm. *Ultrasonic Imaging*, 6(1):81–94, January 1984.

[6] R. Ando, H. Hama, M. Yamamoto-Hino, H. Mizuno, and A. Miyawaki. An optical marker based on the UV-induced green-to-red photoconversion of a fluorescent protein. *Proceedings of the National Academy of Sciences*, 99(20):12657–12662, October 2002.

[7] F. J. Anscombe. The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4):246–254, December 1948.

[8] A. Antoniadis and J. Fan. Regularization of wavelet approximations. *Journal of the American Statistical Association*, 96(455):939–967, September 2001.

[9] A. Antoniadis and T. Sapatinas. Wavelet shrinkage for natural exponential families with quadratic variance functions. *Biometrika*, 88(3):805–820, 2001.

[10] M. Arigovindan, M. Sühling, P. Hunziker, and M. Unser. Variational image reconstruction from arbitrarily spaced samples: A fast multiresolution spline solution. *IEEE Transactions on Medical Imaging*, 14(4):450–460, April 2005.

[11] G. B. Avinash. Data-driven simultaneous blur and image restoration in 3-D fluorescence microscopy. *Journal of Microscopy*, 183(2):145–157, August 1996.

[12] M. R. Banham and A. K. Katsaggelos. Digital image restoration. *IEEE Signal Processing Magazine*, 14(2):24–41, March 1997.

[13] R. E. Bank, T. F. Dupont, and H. Yserentant. The hierarchical basis multigrid method. *Numerische Mathematik*, 52(4):427–458, July 1988.

[14] H. H. Barrett and K. J. Myers. *Foundations of image science*. Wiley, 2003.

[15] M. Bates, Bo. Huang, G. T. Dempsey, and X. Zhuang. Multicolor super-resolution imaging with photo-switchable probes. *Science*, 317:1749–1753, 2007.

[16] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle. A ℓ^1-unified variational framework for image restoration. *Lecture Notes in Computer Science - Proceedings of ECCV 2004 Part IV*, 3024:1–13, 2004.

[17] M. Bertero and P. Boccacci. *Introduction to inverse problems in imaging*. Institute of Physics Publishing, 1998.

[18] D. P. Bertsekas, A. Nedic, and A. E. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, April 2003.

[19] P. Besbeas, I. De Feis, and T. Sapatinas. A comparative simulation study of wavelet shrinkage estimators for Poisson counts. *International Statistical Review*, 72(2):209–237, August 2004.

[20] E. Betzig, G. H. Patterson, R. Sougrat, O. Wolf Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313(5793):1642–1645, September 2006.

[21] J. Beutel, H. L. Kundel, and R. L. Van Metter, editors. *Handbook of Medical Imaging: Physics and Psychophysics*, volume 1. SPIE Press, February 2000.

[22] J. M. Bioucas-Dias. Bayesian wavelet-based image deconvolution: a GEM algorithm exploiting a class of heavy-tailed priors. *IEEE Transactions on Image Processing*, 15(4):937–951, April 2006.

[23] J. M. Bioucas-Dias and M. A. T. Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Transactions on Image Processing*, 16(12):2992–3004, December 2007.

[24] T. Blu and F. Luisier. The SURE-LET approach to image denoising. *IEEE Transactions on Image Processing*, 16(11):2778–2786, November 2007.

[25] T. Blu and M. Unser. The fractional spline wavelet transform: Definition and implementation. In *Proceedings of the Twenty-Fifth IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume I, pages 512–515, June 2000.

[26] N. Bourbaki. *Intégration*. Hermann, second edition, 1965.

[27] J. Boutet 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. *Biophysical Journal*, 80(5):2455–2470, May 2001.

[28] J. Boutet de Monvel, E. Scarfone, S. Le Calvez, and M. Ulfendahl. Image-adaptive deconvolution for three-dimensional deep biological imaging. *Biophysical Journal*, 85:3991–4001, December 2003.

[29] W. S. Boyle and G. E. Smith. Charge coupled semiconductor devices. *Bell System Technical Journal*, 49(4):587–593, April 1970.

[30] K. Bredies and D. A. Lorenz. Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications*, In press.

[31] W. L. Briggs and V. E. Henson. Wavelets and multigrid. *SIAM Journal on Scientific Computing*, 14(2):506–510, March 1993.

[32] W. L. Briggs, V. E. Henson, and S. F. McCormick. *A Multigrid Tutorial*. Society for Industrial and Applied Mathematics, second edition, 2000.

[33] D.R. Brillinger. Some wavelet analyses of point process data. *Signals, Systems & Computers, 1997. Conference Record of the Thirty-First Asilomar Conference on*, 2:1087–1091 vol.2, Nov 1997.

[34] L. G. Brown. A survey of image registration techniques. *ACM Comput. Surv.*, 24:325–376, 1992.

[35] Z. Cai and W. E. Hierarchical method for elliptic problems using wavelet. *Communications in Applied Numerical Methods*, 8(11):819–825, November 1992.

[36] G. Carrero, D. McDonald, E. Crawford, G. de Vries, and M. J. Hendzel. Using FRAP and mathematical modeling to determine the in vivo kinetics of nuclear proteins. *Methods*, 29(1):14–28, January 2003.

[37] W. A. Carrington. Image restoration in 3-D microscopy with limited data. In *Proceedings of the SPIE*, volume 1205, pages 72–83, 1990.

[38] Y. Censor. Row-action methods for huge and sparse systems and their applications. *SIAM Review*, 23(4):444–466, October 1981.

[39] M. Chalfie, Y. Tu, G. Euskirchen, W. Ward, and D. C. Prasher. Green fluorescent protein as a marker for gene expression. *Science*, 263(5148):802–805, February 1994.

[40] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335, March 1998.

[41] T. F. Chan and T. P. Mathew. Domain decomposition algorithms. *Acta Numerica*, 3:61–143, January 1994.

[42] C. Chaux, P. L. Combettes, J.-C. Pesquet, and V. R. Wajs. A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4):1495–1518, August 2007.

[43] G. H.-G. Chen and R. T. Rockafellar. Convergence rates in forward-backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, May 1997.

[44] K. Chen and X.-C. Tai. A nonlinear multigrid method for total variation minimization from image restoration. *Journal of Scientific Computing*, 33(2):115–138, November 2007.

[45] C. Christopoulos, A. Skodras, and T. Ebrahimi. The JPEG2000 still image coding system: an overview. *IEEE Transactions on Consumer Electronics*, 46(4):1103–1127, November 2000.

[46] D. M. Chudakov, V. V. Verkhusha, D. B. Staroverov, E. A. Souslova, S. Lukyanov, and K. A. Lukyanov. Photoswitchable cyan fluorescent protein for protein tracking. *Nature Biotechnology*, 22(11):1435–1439, November 2004.

[47] G. Cohen. Auxiliary problem principle and decomposition of optimization problems. *Journal of Optimization Theory and Applications*, 32(3):277–305, November 1980.

[48] R. R. Coifman and D. L. Donoho. Translation invariant de-noising. In *Lecture Notes in Statistics: Wavelets and Statistics*, volume 103, pages 125–150, Springer Verlag, NewYork, 1995.

[49] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling and Simulation*, 4(4):1168–1200, 2005.

[50] J.-A. Conchello. Superresolution and convergence properties of the expectation-maximization algorithm for maximum-likelihood deconvolution of incoherent images. *Journal of the Optical Society of America A*, 15(10):2609–2619, October 1998.

[51] J.-A. Conchello and M. E. Dresser. Extended depth-of-focus microscopy via constrained deconvolution. *Journal of Biomedical Optics*, 12(6):064026, November/December 2007.

[52] J.-A. Conchello and E. W. Hansen. Enhanced 3D reconstruction from confocal scanning microscope images. 1: Deterministic and ML reconstructions. *Applied Optics*, 29(26):3795–3804, September 1990.

[53] J.-A. Conchello, J. J. Kim, and E. W. Hansen. Enhanced 3D reconstruction from confocal scanning microscope imagess. 2: Depth discrimination versus signal-to-noise ratio in partially confocal images. *Applied Optics*, 33(17):3740–3750, June 1994.

[54] J.-A. Conchello and J. G. McNally. Fast regularization technique for expectation maximization algorithm for optical sectioning microscopy. In *Proceedings of the SPIE*, volume 2655, pages 199–208, 1996.

[55] A. H. Coons and M. H. Kaplan. Localization of antigen in tissue cells. II. Improvements in a method for the detection of antigen by means of fluorescent antibody. *Journal of Experimental Medicine*, 91(1):1–13, 1950.

[56] W. Dahmen and A. Kunoth. Multilevel preconditioning. *Numerische Mathematik*, 63(1):315–344, December 1992.

[57] I. Daubechies. *Ten lectures on wavelets*, volume 61 of *CBMS-NSF Regional Conference series in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, March 1992.

[58] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457, August 2004.

[59] S. Delpretti, F. Luisier, S. Ramani, T. Blu, and M. Unser. Multiframe SURE-LET denoising of timelapse fluorescence microscopy images. In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging (ISBI): From Nano to Macro*, Paris, France, May 14-17, 2008.

[60] W. Denk, J. H. Strickler, and W. W. Webb. Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76, April 1990.

[61] R. A. DeVore. Nonlinear approximation. *Acta Numerica*, pages 51–150, 1998.

[62] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. 3D microscopy deconvolution using Richardson-Lucy algorithm with total variation regularization. Technical report, Institut National de Recherche en Informatique et en Automatique, 2004.

[63] N. Dey, L. Blanc-Féraud, C. Zimmer, P. Roux, Z. Kam, J.-C. Olivo-Marin, and J. Zerubia. Richardson-Lucy algorithm with total variation regularization for 3D confocal microscope deconvolution. *Microscopy Research and Technique*, 69(4):227–301, April 2006.

[64] D. L. Donoho. Nonlinear wavelet methods for recovery of signals, densities, and spectra from indirect and noisy data. In I. Daubechies, editor, *Different Perspectives on Wavelets*, volume 47 of *Proceedings of Symposia in Applied Mathematics*, pages 173–205, Providence, RI, USA, 1993. American Mathematical Society.

[65] D. L. Donoho. Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition. *Applied and Computational Harmonic Analysis*, 2(2):101–126, April 1995.

[66] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[67] D. L. Donoho and J. M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, August 1994.

[68] D.L. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[69] B. Dubertret, P. Skourides, D. J. Norris, V. Noireaux, A. H. Brivanlou, and A. Libchaber. In vivo imaging of quantum dots encapsulated in phospholipid micelles. *Science*, 298:1759–1762, 2002.

[70] J. Eckstein. *Splitting methods for monotone operators with applications to parallel optimization.* PhD thesis, Massachusetts Institute of Technology, 1989.

[71] B. Eicke. Iteration methods for convexly constrained ill-posed problems in Hilbert spaces. *Numerical Functional Analysis and Optimization*, 13(5-6):413–429, 1992.

[72] M. Elad. Why simple shrinkage is still relevant for redundant representations? *IEEE Transactions on Information Theory*, 52(12):5559–5569, December 2006.

[73] M. Elad, B. Matalon, and M. Zibulevsky. Coordinate and subspace optimization methods for linear least squares with non-quadratic regularization. *Applied and Computational Harmonic Analysis*, 23(3):346–367, November 2007.

[74] M. Elad and I. Yavneh. A weighted average of sparse representations is better than the sparsest one alone. Preprint.

[75] Y. C. Eldar. Generalized SURE for exponential families: Applications to regularization. *IEEE Transactions on Signal Processing*, in press.

[76] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Transactions on Signal Processing*, 42(10):2664–2677, October 1994.

[77] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak. Majorization-minimization algorithms for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 16(12):2980–2991, December 2007.

[78] M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, August 2003.

[79] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of Selected Topics in Signal Processing*, 2007(4):586–597, December 2007.

[80] M. Fornasier. Domain decomposition methods for linear inverse problems with sparsity constraints. *Inverse Problems*, 23(6):2505–2526, December 2007.

[81] M. Fornasier and C.-B. Schönlieb. Subspace correction methods for total variation and ℓ_1-minimization. Preprint.

[82] B. Forster, D. Van De Ville, J. Berent, D. Sage, and M. Unser. Complex wavelets for extended depth-of-field: A new method for the fusion of multi-channel microscopy images. *Microscopy Research and Technique*, 65(1-2):33–42, September 2004.

[83] R. Frey. Fitting of 3D-PSF models to fluorescent microbead stacks. Master's thesis, Ecole Polytechnique Fédérale de Lausanne, February 2007.

[84] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani. Pathwise coordinate optimization. *Annals of Applied Statistics*, 1(2):302–332, December 2007.

[85] P. Fryzlewicz and G. P. Nason. A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of Computational and Graphical Statistics*, 13(3):621–638, September 2004.

[86] N. P. Galatsanos and A. K. Katsaggelos. Methods for choosing the regularization parameter and estimating the noise variance in image restoration and their relation. *IEEE Transactions on Image Processing*, 1(3):322–336, July 1992.

[87] D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383, March 1992.

[88] D. Geman and C. Yang. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*, 4(7):932–946, July 1995.

[89] S. F. Gibson and F. Lanni. Experimental test of an analytical model of aberration in an oil-immersion objective lens used in three-dimensional light microscopy. *Journal of the Optical Society of America A*, 8(10):1601–1613, October 1991.

[90] P. Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1):105–117, July 1972.

[91] G. H. Golub and C. F. van Loan. *Matrix Computations*. Johns Hopkins University Press, 1996.

[92] R. Gordon, R. Bender, and G. T. Herman. Algebraic reconstruction techniques (art) for three-dimensional electron microscopy and x-ray photography. *Journal of Theoretical Biology*, 29(3):471–481, December 1970.

[93] B. A. Griffin, S. R. Adams, and R. Y. Tsien. Specific covalent labeling of recombinant protein molecules inside live cells. *Science*, 281(5374):269–272, July 1998.

[94] M. Gu. *Principles of three-dimensional imaging in confocal microscopes*. World Scientific, December 1996.

[95] I. M. Guelfand and N. Y. Vilenkin. *Les distributions. Tome 4. Applications de l'analyse harmonique*. Collection universitaire de mathématiques. Dunod, 1967.

[96] M. G. L. Gustafsson, D. A. Agard, and J. W. Sedat. I^{5}M: 3D widefield light microscopy with better than 100 nm axial resolution. *Journal of Microscopy*, 195(1):10–16, July 1999.

[97] W. Hackbusch. *Multi-grid methods and applications*, volume 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, 1985.

[98] W. Hackbusch. *Iterative Solution of Large Sparse Systems of Equations*, volume 95 of *Applied Mathematical Sciences*. Springer-Verlag, 1994.

[99] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for ℓ_1-minimization: Methodology and convergence. *SIAM Journal on Optimization*, 19(3):1107–1130, 2008.

[100] R. Heintzmann. Estimating missing information by maximum likelihood deconvolution. *Micron*, 38(2):136–144, February 2007.

[101] R. Heintzmann, T. M. Jovin, and C. Cremer. Saturated patterned excitation microscopy—a concept for optical resolution improvement. *Journal of the Optical Society of America A*, 19(8):1599–1609, August 2002.

[102] R. Heintzmann and C. J. R. Sheppard. The sampling limit in fluorescence microscopy. *Micron*, 38(2):145–149, February 2007.

[103] S. W. Hell and E. H. K. Stelzer. Fundamental improvement of resolution with a 4pi-confocal fluorescence microscope using two-photon excitation. *Optics Communications*, 93(5-6):277–282, October 1992.

[104] S. W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters*, 19(11):780–782, June 1994.

[105] A. E. Hoerl. Application of ridge analysis to regression problems. *Chemical Engineering Progress*, 58:54–59, 1962.

[106] T. J. Holmes and Y.-H. Liu. Acceleration of maximum-likelihood image restoration for fluorescence microscopy and other noncoherent imagery. *Journal of the Optical Society of America A*, 8(6):893–907, June 1991.

[107] M. R. P. Homem, N. D. A. Mascarenhas, L. F. Costa, and C. Preza. Biological image restoration in optical-sectioning microscopy using prototype image constraints. *Real-Time Imaging*, 8:475–490, 2002.

[108] H. M. Hudson. A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics*, 6(3):473–484, May 1978.

[109] H. M. Hudson and R. S. Larkin. Accelerated image reconstruction using ordered subsets of projection data. *IEEE Transactions on Medical Imaging*, 13(4):601–609, December 1994.

[110] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. Stelzer. Optical sectioning deep inside live embryos by selective plane illumination microscopy. *Science*, 305(5686):1007–9, August 2004.

[111] D. R. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58(1):30–37, February 2004.

[112] M. Jacob and M. Unser. Design of steerable filters for feature detection using Canny-like criteria. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8):1007–1019, August 2004.

[113] M. W. Jacobson and J. A. Fessler. An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Transactions on Image Processing*, 16(10):2411–2422, October 2007.

[114] M. Jansen. Multiscale Poisson data smoothing. *Journal of the Royal Statistical Society B*, 68(1):27–48, 2006.

[115] S. Kaczmarz. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin de l'Académie Polonaise des Sciences et des Lettres A*, 6:355–357, 1937.

[116] J. Kalifa and S. Mallat. Thresholding estimators for linear inverse problems and deconvolutions. *Annals of Statistics*, 31(1):58–109, 2003.

[117] H. Kano, H. T. M. van der Voort, M. Schrader, G. M. P. van Kempen, and S. W. Hell. Avalanche photodiode detection with object scanning and image restoration provides 2-4 fold resolution increase in two-photon fluorescence microscopy. *Bioimaging*, 4(3):187–197, September 1996.

[118] I. Khalidov and M. Unser. From differential equations to the construction of new wavelet-like bases. *IEEE Transactions on Signal Processing*, 54(4):1256–1267, April 2006.

[119] E. D. Kolaczyk. Non-parametric estimation of gamma-ray burst intensities using Haar wavelets. *The Astrophysical Journal*, 483:340–349, 1997.

[120] E. D. Kolaczyk. Bayesian multi-scale models for Poisson processes. *Journal of the American Statistical Association*, 94(447):920–933, September 1999.

[121] E. D. Kolaczyk. Wavelet shrinkage estimation of certain Poisson intensity signals using corrected thresholds. *Statistica Sinica*, 9(1):119–135, January 1999.

[122] E. D. Kolaczyk. Nonparametric estimation of intensity maps using Haar wavelets and Poisson noise characteristics. *The Astrophysical Journal*, 534:490–505, 2000.

[123] E. D. Kolaczyk and R. D. Nowak. Multiscale likelihood analysis and complexity penalized estimation. *Annals of statistics*, 32(2):500–527, 2004.

[124] L. Landweber. An iterative formula for Fredholm integral equations of the first kind. *American Journal of Mathematics*, 73(3):615–624, July 1951.

[125] K. Lange, D. R. Hunter, and I. Yang. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*, 9(1):1–20, March 2000.

[126] W. M. Lawton. Multiresolution properties of the wavelet Galerkin operator. *Journal of Mathematical Physics*, 32(6):1440–1443, June 1991.

[127] R. Leahy and C. Byrne. Recent developments in iterative image reconstruction for PET and SPECT. *IEEE Transactions on Medical Imaging*, 19(4):257–260, April 2000.

[128] E. L. Lehmann and G. Casella. *Theory of Point Estimation*. Springer, second edition, 1998.

[129] M. Liebling, A. S. Forouhar, M. Gharib, S. E. Fraser, and M. E. Dickinson. 4-dimensional cardiac imaging in living embryos via post-acquisition synchronization of nongated slice-sequences. *Journal of Biomedical Optics*, 10(5), 2005.

[130] M. Liebling and M. Unser. Autofocus for digital Fresnel holograms by use of a fresnelet-sparsity criterion. *Journal of the Optical Society of America A*, 21(12):2424–2430, December 2004.

[131] P. L. Lions. On the Schwarz alternating method. i. In R. Glowinski, G. Golub, G. Meurant, and J. Periaux, editors, *Proceedings of the First International Symposium of Domain Decomposition Methods for Partial Differential Equations*, pages 1–42, Philadelphia, PA, 1988. Society for Industrial and Applied Mathematics.

[132] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear algorithms. *SIAM Journal on Numerical Analysis*, December(6):964–979, December 1979.

[133] J. Lippincott-Schwartz, A.-B. Nihal, and G. H. Patterson. Photobleaching and photoactivation: following protein dynamics in living cells. *Nature Cell Biology*, 5(9):Suppl:S7–S14, September 2003.

[134] J. Lippincott-Schwartz and G. H. Patterson. Development and use of fluorescent protein markers in living cells. *Science*, 300(5616):87–91, April 2003.

[135] I. Loris, G. Nolet, I. Daubechies, and F. A. Dahlen. Tomographic inversion using ℓ_1-regularization of wavelet coefficients. *Geophysical Journal International*, 170(1):359–370, July 2007.

[136] A. K. Louis. Medical imaging: state of the art and future development. *Inverse Problems*, 8(5):709–738, October 1992.

[137] H. Lu, Y. Kim, and John M. M. Anderson. Improved Poisson intensity estimation: denoising application using Poisson data. *IEEE Transactions on Image Processing*, 13(8):1128–1135, August 2004.

[138] L. B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal*, 79(6):745–754, June 1974.

[139] F. Luisier and T. Blu. Poisson intensity estimation in additive Gaussian white noise. Preprint.

[140] F. Luisier, T. Blu, and M. Unser. A new SURE approach to image denoising: Interscale orthonormal wavelet thresholding. *IEEE Transactions on Image Processing*, 16(3):593–606, March 2007.

[141] S. Mallat. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 11(7):674–693, July 1989.

[142] S. Mallat. *A wavelet tour of signal processing*. Academic Press, 1998.

[143] J. S. Marchant, G. E. Stutzmann, M. A. Leissring, F. M. LaFerla, and I. Parker. Multiphoton-evoked color change of DsRed as an optical highlighter for cellular and subcellular labeling. *Nature Biotechnology*, 19(7):645 – 649, July 2001.

[144] J. Markham and J.-A. Conchello. Parametric blind deconvolution: a robust method for the simultaneous estimation of image and blur. *Journal of the Optical Society of America A*, 16(10):2377–2391, October 1999.

[145] J. Markham and J.-A. Conchello. Artefacts in restored images due to intensity loss in three-dimensional fluorescence microscopy. *Journal of Microscopy*, 204(2):93–98, November 2001.

[146] J. Markham and J.-A. Conchello. Fast maximum-likelihood image-restoration algorithms for three-dimensional fluorescence microscopy. *Journal of the Optical Society of America*, 18(5):1062–1071, May 2001.

[147] L. McMahon, R. Legouis, J.-L. Vonesch, and M. Labouesse. Assembly of C. elegans apical junctions involves positioning and compaction by LET-413 and protein aggregation by the MAGUK protein DLG-1. *Journal of Cell Science*, 114(12):2265–2277, June 2001.

[148] J. G. McNally, T. Karpova, J. Cooper, and J. A. Conchello. Three-dimensional imaging by deconvolution microscopy. *Methods*, 19(3):373–385, November 1999.

[149] E. Meijering, M. Jacob, J.-C.F. Sarria, P. Steiner, H. Hirling, and M. Unser. Design and validation of a tool for neurite tracing and analysis in fluorescence microscopy images. *Cytometry Part A*, 58A(2):167–176, April 2004.

[150] E. Meijering, I. Smal, and G. Danuser. Tracking in molecular bioimaging. *IEEE Signal Processing Magazine*, 23(3):46–53, 2006.

[151] X.-L. Meng and D. A. van Dyk. The EM algorithm—an old folk-song sung to a fast new tune. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):511–567, 1997.

[152] Y. Meyer. *Ondelettes et opérateurs*. Hermann, 1990.

[153] X. Michalet, F. F. Pinaud, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, A. M. Wu, S. S. Gambhir, and S. Weiss. Quantum dots for live cells, in vivo imaging, and diagnostics. *Science*, 307(5709):538–544, January 2005.

[154] M. Minsky. Memoir on inventing the confocal scanning microscope. *Scanning*, 10(4):128–138, August 1988.

[155] J. Mitić, T. Anhut, M. Meier, M. Ducros, A. Serov, and T. Lasser. Optical sectioning in wide-field microscopy obtained by dynamic structured light illumination and detection based on a smart pixel detector array. *Optics Letters*, 28(9):698–700, May 2003.

[156] A. Miyawaki, A. Sawano, and T. Kogure. Lighting up cells: labelling proteins with fluorophores. *Nature Cell Biology*, 5 (supplement):S1–S7, September 2003.

[157] C. De Mol and M. Defrise. A note on wavelet-based inversion algorithms. In M. Zuhair Nashed and O. Scherzer, editors, *Inverse problems, image analysis and medical imaging*, volume 313 of *Contemporary Mathematics*, pages

85–96, Providence, RI, November 2002. American Mathematical Society. Proceedings of the Special Session on Interaction of Inverse Problems and Image Analysis. AMS Joint Mathematics Meetings. New Orleans, LA. January 10-13, 2001.

[158] R. Molina, J. Nunez, F. J. Cortijo, and J. Mateos. Image restoration in astronomy: a Bayesian perspective. *IEEE Signal Processing Magazine*, 18(2):11–29, March 2001.

[159] P. Moulin and J. Liu. Analysis of multiresolution image denoising schemes using generalized Gaussian and complexity priors. *IEEE Transactions on Information Theory*, 45(3):909–919, April 1999.

[160] F. Murtagh, J.-L. Starck, and A. Bijaoui. Image restoration with noise suppression using a multiresolution support. *Astronomy and Astrophysics Supplement Series*, 112:179–189, July 1995.

[161] M. Nagorni and S. W. Hell. Coherent use of opposing lenses for axial resolution increase. II. Power and limitation of nonlinear image restoration. *Journal of the Optical Society of America A*, 18(1):49–54, January 2001.

[162] R. Neelamani, H. Choi, and R. Baraniuk. ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems. *IEEE Transactions on Signal Processing*, 52(2):418– 433, February 2004.

[163] M. Nikolova and M. K. Ng. Analysis of half-quadratic minimization methods for signal and image recovery. *SIAM Journal on Scientific Computing*, 27(3):937–966, 2005.

[164] P. R. Norton. *Handbook of optics*, volume 1, chapter Photodetectors, pages 15.3–15.100. Mc Graw-Hill, 1995.

[165] R. D. Nowak and R. G. Baraniuk. Wavelet-domain filtering for photon imaging systems. *IEEE Transactions on Image Processing*, 8(5):666–678, May 1999.

[166] R. D. Nowak and M. A. T. Figueiredo. Fast wavelet-based image deconvolution using the EM algorithm. In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers. Pacific Grove, CA. November 4-11, 2001.*, volume 1, 2001.

[167] S. Oh, A. B. Milstein, C. A. Bouman, and K. J. Webb. A general framework for nonlinear multigrid inversion. *IEEE Transactions on Image Processing*, 14(1):125–140, January 2005.

[168] J.-C. Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989–1996, September 2002.

[169] J. M. Ortega and W. C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables*. Academic Press, New York, 1970.

[170] T.-S. Pan and A. E. Yagle. Numerical study of multigrid implementations of some iterative image reconstruction algorithms. *IEEE Transactions on Medical Imaging*, 10(4):572–288, December 1991.

[171] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72:383–390, 1979.

[172] P. N. Patil and A. T.A. Wood. Counting process intensity estimation by orthogonal wavelet methods. *Bernoulli*, 10(1):1–24, 2004.

[173] G. H. Patterson and J. Lippincott-Schwartz. A photoactivatable GFP for selective photolabeling of proteins and cells. *Science*, 297(5588):1873–1877, September 2002.

[174] J. Pawley. *Three-dimensional confocal microscopy: volume investigation of biological specimens*, chapter Sources of noise in three-dimensional microscopical data sets, pages 47–94. Academic Press, 1994.

[175] M. Petran, M. Hadravsky, M.D. Egger, and R. Galambos. Tandem-scanning reflected-light microscope. *Journal of the Optical Society of America*, 58(5):661–664, 1968.

[176] D. L. Phillips. A technique for the numerical solution of certain integral equations of the first kind. *Journal of the ACM*, 9(1):84–97, January 1962.

[177] D. C. Prasher, V. K. Eckenrode, W. W. Ward, F. G. Prendergast, and M. J. Cormier. Primary structure of the Aequorea victoria green-fluorescent protein. *Gene*, 111(2):229–233, February 1992.

[178] C. Preza and J. A. Conchello. Depth-variant maximum-likelihood restoration for three-dimensional fluorescence microscopy. *Journal of the Optical Society of America A*, 21(9):1593–1601, September 2004.

[179] R. C. Puetter, T. R. Gosnell, and A. Yahil. Digital image reconstruction: deblurring and denoising. *Annual Review of Astronomy and Astrophysics*, 43:139–194, September 2005.

[180] S. Ramani, T. Blu, and M. Unser. Monte-Carlo SURE: A black-box optimization of regularization parameters for general denoising algorithm. *IEEE Transactions on Image Processing*, 17(9):1540–1554, September 2008.

[181] S. Ramani, C. Vonesch, and M. Unser. Deconvolution of 3D fluorescence micrographs with automatic risk minimization. In *Proceedings of the 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, Paris, France*, pages 732–735, May 14-17, 2008.

[182] M. Raphan and E. P. Simoncelli. Learning to be Bayesian without supervision. In B. Schölkopf, J. Platt, and T. Hofmann, editors, *Advances in Neural Information Processing Systems*, volume 19, pages 1145–1152, Cambridge, MA, 2007. MIT Press.

[183] M. Raphan and E. P. Simoncelli. Optimal denoising in redundant representations. *IEEE Transactions on Image Processing*, 17(8):1342–1352, Aug. 2008.

[184] W. H. Richardson. Bayesian-based iterative method of image restoration. *Journal of the Optical Society of America*, 62(1):55–59, January 1972.

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

[186] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, November 1992.

[187] D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser, and M. Unser. Automatic tracking of individual fluorescence particles: Application to the study of chromosome dynamics. *IEEE Transactions on Image Processing*, page (in press), 2005.

[188] P. Sarder and A. Nehorai. Deconvolution methods for 3-D fluorescence microscopy images. *IEEE Signal Processing Magazine*, 23(3):32–45, May 2006.

[189] S. Sardy, A. Antoniadis, and P. Tseng. Automatic smoothing with wavelets for a wide class of distributions. *Journal of Computational & Graphical Statistics*, 13(2):399–423, June 2004.

[190] S. Sardy, A. G. Bruce, and P. Tseng. Block coordinate relaxation methods for nonparametric wavelet denoising. *Journal of Computational and Graphical Statistics*, 9(2):361–379, June 2000.

[191] L. H. Schaefer, D. Schuster, and H. Herz. Generalized approach for accelerated maximum likelihood based image restoration applied to three-dimensional fluorescence microscopy. *Journal of Microscopy*, 204(2):99–107, November 2001.

[192] G. Schmit. Deconvolution in 3D: an ImageJ plugin. Master's thesis, Ecole Polytechnique Fédérale de Lausanne, February 2007.

[193] G. Schmit, C. Vonesch, S. Ramani, and M. Unser. DeconvolutionLab: an ImageJ plugin for fast and automated 3D image restoration. Preprint.

[194] M. Schrader, S. W. Hell, and H. T. M. van der Voort. Three-dimensional super-resolution with a 4Pi-confocal microscope using image restoration. *Journal of Applied Physics*, 84(8):4033–4042, October 1998.

[195] L. Schwartz. *Théorie des distributions*. Hermann, 1966.

[196] H. A. Schwarz. Ueber einen Grenzübergang durch alternirendes Verfahren. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zürich*, 15:272–286, May 1870.

[197] P. Selvin. The renaissance of fluorescence resonance energy transfer. *Nature Structural Biology*, 7(9):730–734, 2000.

[198] J. W. Shaevitz and D. A. Fletcher. Enhanced three-dimensional deconvolution microscopy using a measured depth-varying point-spread function. *Journal of the Optical Society of America A*, 24(9):2622–2627, September 2007.

[199] O. Shimomura. The discovery of aequorin and green fluorescent protein. *Journal of Microscopy*, 217(1):3–15, January 2005.

[200] J.-B. Sibarita. *Microscopy Techniques*, volume 95 of *Advances in Biochemical Engineering/Biotechnology*, chapter Deconvolution Microscopy, pages 201–243. Springer, 2005.

[201] E. P. Simoncelli. *Bayesian Inference in Wavelet Based Models*, volume 141 of *Lecture Notes in Statistics*, chapter Bayesian denoising of visual images in the wavelet domain, pages 292–308. Springer-Verlag, New York, June 1999.

[202] J. G. Skellam. The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society*, 109(3):296, 1946.

[203] C. Ó. Sánchez Sorzano, P. Thévenaz, and M. Unser. Elastic registration of biological images using vector-spline regularization. *IEEE Transactions on Biomedical Engineering*, 52(4):652–663, April 2005.

[204] A. Squire and P. I. H. Bastiaens. Three dimensional image restoration in fluorescence lifetime imaging microscopy. *Journal of Microscopy*, 193(1):36–49, January 1999.

[205] J. L. Starck, D. L. Donoho, and E. J. Candès. Astronomical image representation by the curvelet transform. *Astronomy and Astrophysics*, 398(2):785–800, February 2003.

[206] J.-L. Starck, M. K. Nguyen, and F. Murtagh. Wavelets and curvelets for image deconvolution: a combined approach. *Signal Processing*, 83(10):2279–2283, October 2003.

[207] J. L. Starck, E. Pantin, and F. Murtagh. Deconvolution in astronomy: A review. *Publications of the Astronomical Society of the Pacific*, 114(800):1051–1069, October 2002.

[208] T. Stearns. Green fluorescent protein. the green revolution. *Current Biology*, 5(3):262–264, March 1995.

[209] C. M. Stein. Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, 9(6):1135–1151, November 1981.

[210] G. Strang and T. Nguyen. *Wavelets and filter banks*. Wellesley-Cambridge Press, 1996.

[211] T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, Preprint.

[212] L. Takács. Über die Wahrscheinlichkeitstheoretische Behandlung der Anodenstromschwankungen von Elektronenröhren. *Acta physica Academiae scientiarum hungaricae*, 7:25–50, 1957.

[213] J. C. Taylor. *An Introduction to Measure and Probability*. Springer, 1998.

[214] P. Thévenaz and M. Unser. User-friendly semiautomated assembly of accurate image mosaics in microscopy. *Microscopy Research and Technique*, 70(2):135–146, 2007.

[215] D. Thomann, J. Dorn, P. K. Sorger, and G. Danuser. Automatic fluorescent tag localization II: Improvement in super-resolution by relative tracking. *Journal of Microscopy*, 211(3):230–248, September 2003.

[216] 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):1–63, January 2006.

[217] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics - Doklady*, 4:1035–1038, 1963.

[218] K. E. Timmermann and R. D. Nowak. Multiscale modeling and estimation of Poisson processes with application to photon-limited imaging. *IEEE Transactions on Information Theory*, 45(3):846–862, April 1999.

[219] T. Tjur. *Probability based on Radon measures*. Probability and mathematical statistics. Wiley, 1980.

[220] V. Tom, T. Quatieri, M. Hayes, and J. McClellan. Convergence of iterative nonexpansive signal reconstruction algorithms. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(5):1052–1058, October 1981.

[221] K.-W. Tsui and S. J. Press. Simultaneous estimation of several Poisson parameters under K-normalized squared error loss. *The Annals of Statistics*, 10(1):93–100, March 1982.

[222] H. Tsurui, H. Nishimura, S. Hattori, S. Hirose, K. Okumura, and T. Shirai. Seven-color fluorescence imaging of tissue samples based on Fourier spectroscopy and singular value decomposition. *Journal of Histochemistry and Cytochemistry*, 48(5):653–662, May 2000.

[223] D. Van De Ville, T. Blu, and M. Unser. Isotropic polyharmonic B-splines: Scaling functions and wavelets. *IEEE Transactions on Image Processing*, 14(11):1798–1813, November 2005.

[224] D. Van De Ville and M. Unser. Complex wavelet bases, steerability, and the Marr-like pyramid. *IEEE Transactions on Image Processing*, 17(11):2063–2080, November 2008.

[225] H. T. M. van der Voort and K. C. Strasters. Restoration of confocal images for quantitative image analysis. *Journal of Microscopy*, 178(2):165–181, May 1995.

[226] D. A. van Dyk, A. Connors, V. L. Kashyap, and A. Siemiginowska. Analysis of energy spectra with low photon counts via Bayesian posterior simulation. *The Astrophysical Journal*, 548(1):224–243, 2001.

[227] G. M. P. van Kempen, H. T. M. van der Voort, J. G. J. Bauman, and K. C. Strasters. Comparing maximum likelihood estimation and constrained Tikhonov-Miller restoration. *IEEE Engineering in Medicine and Biology Magazine*, 15(1):76–83, January-February 1996.

[228] G. M. P. van Kempen and L. J. van Vliet. Background estimation in nonlinear image restoration. *Journal of the Optical Society of America A*, 17(3):425–433, March 2000.

[229] 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. *Journal of Microscopy*, 198(1):63–75, April 2000.

[230] G. M. P. van Kempen, L. J. van Vliet, P. J. Verveer, and H. T. N. van der Voort. A quantitative comparison of image restoration methods for confocal microscopy. *Journal of Microscopy*, 185:354–365, 1997.

[231] V. V. Vasin and A. L. Ageev. *Ill-Posed Problems With a Priori Information*. Inverse and Ill-Posed Problems. VSP, Utrecht, The Netherlands, December 1995.

[232] P. S. Vassilevski and J. Wang. Stabilizing the hierarchical basis by approximate wavelets I: Theory. *Numerical Linear Algebra with Applications*, 4(2):103–126, March/April 1998.

[233] 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. *Journal of Microscopy*, 193(1):50–61, January 1999.

[234] P. J. Verveer and T. M. Jovin. Acceleration of the ICTM image restoration algorithm. *Journal of Microscopy*, 188(3):191–195, December 1997.

[235] C. Vogel. *Computational methods for inverse problems*. Society for Industrial and Applied Mathematics, 2002.

[236] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser. The colored revolution of bioimaging. *IEEE Signal Processing Magazine*, 23(3):20–31, May 2006.

[237] C. Vonesch, T. Blu, and M. Unser. Generalized Daubechies wavelet families. *IEEE Transactions on Signal Processing*, 55(9):4415–4429, September 2007.

[238] C. Vonesch, F. Luisier, T. Blu, and M. Unser. Wavelet-domain PURE-LET for Poisson intensity estimation. Submitted.

[239] C. Vonesch, S. Ramani, and M. Unser. Recursive risk estimation for non-linear image deconvolution with a wavelet-domain sparsity constraint. In *Accepted for the 15th IEEE International Conference on Image Processing (ICIP)*, San Diego CA, USA, 2008.

[240] C. Vonesch and M. Unser. A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Transactions on Image Processing*, 17(4):539–549, April 2008.

[241] C. Vonesch and M. Unser. A fast multilevel algorithm for wavelet-regularized image restoration. *IEEE Transactions on Image Processing*, 18(3):509–523, March 2009.

[242] W. Wallace, L. H. Schaefer, and J. R. Swedlow. A workingperson's guide to deconvolution in light microscopy. *Biotechniques*, 31(5):1076–1097, November 2001.

[243] J. E. Wampler and K. Kutz. Quantitative fluorescence microscopy using photomultiplier tubes and imaging detectors. *Methods Cell Biol.*, 29:239–67, 1989.

[244] G. Wang, J. Zhang, and G.-W. Pan. Solution of inverse problems in image processing by wavelet expansion. *IEEE Transactions on Image Processing*, 4(5):579–593, May 1995.

[245] Y. Wang, J. Yang, W. Yin, and Y. Zhang. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272, 2008.

[246] J. B. Weaver, Y. S. Xu, D. M. Healy, and L. D. Cromwell. Filtering noise from images with wavelet transforms. *Magnetic Resonance in Medicine*, 21(2):288–295, October 1991.

[247] M. Weinstein and K. R. Castleman. Reconstructing 3-D specimens from 2-D section images. In *Proceedings of the SPIE*, volume 26, pages 131–137, 1971.

[248] N. Wiener. *Extrapolation, interpolation, and smoothing of stationary time series : with engineering applications.* Technology Press of the Massachusetts Institute of Technology, 1949.

[249] R. M. Willett and R. D. Nowak. Multiscale Poisson intensity and density estimation. *IEEE Transactions on Information Theory*, 53(9):3171–3187, September 2007.

[250] K. I. Willig, J. Keller, M. Bossi, and S. W. Hell. STED microscopy resolves nanoparticle assemblies. *New Journal of Physics*, 8(106):24131–24134, June 2006.

[251] A. G. Wright. The statistics of multi-photoelectron pulse-height distributions. *Nuclear Instruments and Methods in Physics Research A*, 579(3):967–972, September 2007.

[252] J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, December 1992.

[253] J. Xu. The method of subspace corrections. *Journal of Computational and Applied Mathematics*, 128(1-2):335–362, March 2001.

[254] D. C. Youla. Generalized image restoration by the method of alternating orthogonal projections. *IEEE Transactions on Circuits and Systems*, 25(9):694–702, September 1978.

[255] D. C. Youla and H. Webb. Image restoration by the method of convex projections: Part 1—theory. *IEEE Transactions on Medical Imaging*, 1(2):81–94, October 1982.

[256] H. Yserentant. On the multi-level splitting of finite element spaces. *Numerische Mathematik*, 49:379–412, 1986.

[257] B. Zhang, J. M. Fadili, and J.-L. Starck. Wavelets, ridgelets and curvelets for Poisson noise removal. *IEEE Transactions on Image Processing*, 17(7):1093–1108, July 2008.

[258] J. Zhang, R. E. Campbell, A. Y. Ting, and R. Y. Tsien. Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology*, 3(12):906–918, December 2002.

[259] K. Zhou and C. K. Rushforth. Image restoration using multigrid methods. *Applied Optics*, 30(20):2906–2912, July 1991.

[260] X. Zhou and S. T. C. Wong. Informatics challenges of high-throughput microscopy. *IEEE Signal Processing Magazine*, 23(3):63–72, 2006.

[261] C. Zimmer, E. Labruyère, V. Meas-Yedid, N. Guillén, and J.-C. Olivo-Marin. Segmentation and tracking of migrating cells in videomicroscopy with parametric active contours: a tool for cell-based drug testing. *IEEE Transactions on Medical Imaging*, 21(10):1212–1221, October 2002.

[262] C. Zimmer, B. Zhang, A. Dufour, A. Thébaud, S. Berlemont, V. Meas-Yedid, and J.-C. Olivo Marin. On the digital trail of mobile cells. *IEEE Signal Processing Magazine*, 23(3):54–62, 2006.

[263] T. Zimmermann, J. Rietdorf, and R. Pepperkok. Spectral imaging and its applications in live cell microscopy. *FEBS Letters*, 546(1):87–92, July 2003.

Biography

Cédric Vonesch was born on April 14, 1981 in Obernai (France). He graduated from EPFL, the Swiss Federal Institute of Technology in Lausanne (Switzerland), in 2004. He is currently with the Biomedical Imaging Group at EPFL, where he is working towards a PhD. His research interests include multiresolution and wavelet analysis, inverse problems and applications to bioimaging.