
Sparsity-Inducing Reconstruction Methods for Fluorescence Diffuse Optical Tomography

Jean-Charles Baritoux

Thèse N° 5272 (March 2012)

*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. Pierre Vandergheynst, *président*
Prof. Michael Unser, *directeur de thèse*
Dr. Franoise Peyrin, *rapporteur*
Prof. Jorge Ripoll, *rapporteur*
Prof. Dimitri Van De Ville, *rapporteur*

École polytechnique fédérale de Lausanne—2012

Cover design by Annette Unser
Printing and binding by Repro-EPFL
Typeset with L^AT_EX
Copyright © 2012 by Jean-Charles Baritoux
Available at <http://bigwww.epfl.ch/>

Abstract

The recent advances in fluorescent molecular probes, photon detection instrumentation, and photon propagation models in tissue, have facilitated the emergence of innovative molecular imaging technologies such as Fluorescence Diffuse Optical Tomography (FDOT). FDOT has already successfully been applied on its own to resolve specific molecular targets and pathways *in vivo*. Recent attempts have tried to integrate FDOT with structural imaging technology into a multi-modal system. This thesis is amongst the first academic-industrial partnerships to develop a hybrid X-Ray Computed Tomography/FDOT platform for small animal imaging. In this study, we focus on image reconstruction algorithms for FDOT, and consider three main research directions.

First, we investigate innovative discretization techniques for FDOT, in the context of variational reconstruction methods. Our goal is to treat in a consistent manner the propagation equation, and the regularization, needed in the reconstruction algorithm. This study uses the diffusion approximation for modeling the propagation of photons in tissue. We solve the diffusion equation with a novel Finite Volumes Method that is specifically designed to be compatible with the sparsity-inducing regularization developed in the second part of this work. Our approach results in a computationally efficient implementation of the forward model.

Next, we study a novel regularization approach for FDOT which is based on sparsity-promoting regularization. This approach is expected to enhance the reconstruction when the fluorescence distribution presents a sparsity pattern in a suitable representation, or after linear transformation. We propose a variational framework based on ℓ_p -norm penalization in order to incorporate the sparsity *a priori* in the inversion algorithm. We demonstrate this approach on experimental data, using computationally efficient algorithms relying on convex optimization principles.

Finally, we focus on hybrid imaging systems that combine FDOT with another imaging technique. The latter is employed to acquire a high-resolution image of the anatomy of the specimen under investigation, which provides valuable information for the FDOT reconstruction. We propose a new regularization method that allows the incorporation of the anatomical information in the FDOT reconstruction. Our method is based on the concept of group sparsity, and implemented in practice using $\ell_{2,1}$ -norm penalization. We extend our sparsity-inducing algorithms to this case, and study the performance of the resulting implementation on experimental data.

Keywords: molecular imaging, optical imaging, multi-modal imaging, optical tomography, fluorescence, inverse problems, reconstruction algorithms, variational methods, convex optimization

Résumé

Les découvertes récentes en matière de marqueurs moléculaires fluorescents, de capteurs optiques, et de modèles de propagation de la lumière dans les tissus, ont conduit à l'émergence de technologies d'imagerie moléculaires novatrices, telles que la Tomographie Optique de Fluorescence dans les milieux Diffusants (FDOT). Utilisée seule, la FDOT permet déjà de visualiser un grand nombre de molécules cibles *in vivo*. À présent, la tendance est à l'intégration de plusieurs technologies d'imagerie, fonctionnelles et structurelles, au sein de systèmes multi-modaux. Cette thèse est l'un des premiers partenariats académie-industrie visant à développer une plateforme d'imagerie hybride Tomographie par Rayons X/FDOT pour petits animaux. Dans ce travail, nous nous concentrons sur les algorithmes de reconstruction, et considérons trois axes de recherche.

Le premier est le développement de méthodes de discrétisation pour la FDOT, destinées à être employées au sein d'un principe de reconstruction variationnel. Notre but est de traiter de manière unifiée à la fois le modèle de propagation, et le terme de régularisation utilisés pendant la reconstruction. Ce travail se base sur l'approximation de diffusion pour modéliser la propagation des photons dans les tissus. Nous résolvons l'équation de diffusion avec une nouvelle Méthode de Volumes Finis conçue spécialement pour prendre en compte la parcimonie des signaux, que nous utilisons comme critère de régularisation dans le problème inverse. Sur ces bases, nous élaborons une implémentation du modèle direct à faible coût calcul.

Dans un deuxième temps, nous étudions une nouvelle technique de régularisation basée sur la sélection des signaux parcimonieux. Cette approche a pour but une meilleure reconstruction des signaux dont la parcimonie peut être mise en évidence par transformation linéaire. Nous proposons un cadre général basé sur la pénalisation de normes ℓ_p , pour intégrer *a priori* de parcimonie dans l'algorithme de reconstruc-

tion. Cette formulation variationnelle du problème de reconstruction nous permet d'employer des principes d'optimisation convexe et les algorithmes à faible cout calcul associés. Nous implémentons la méthode proposée, et présentons des résultats de reconstructions obtenus a partir de données expérimentales.

Finalement, nous nous intéressons aux systèmes d'imagerie hybrides combinant la FDOT avec une autre modalité. Cette dernière a pour but d'acquérir une image haute résolution de l'anatomie du spécimen étudié, fournissant ainsi une information supplémentaire utilisée dans la reconstruction FDOT. Nous proposons une nouvelle méthode de régularisation qui permet d'incorporer l'information anatomique dans la reconstruction. Notre méthode est basée sur la notion de parcimonie de groupe, que nous implémentons en pratique en pénalisant une norme mixte $\ell_{2,1}$. Nous étendons les algorithmes développés dans la première partie de cette thèse au cas de la parcimonie de groupe. Nous étudions la performance de la méthode proposée sur des données expérimentales.

Mots clés : imagerie moléculaire, imagerie optique, imagerie multi-modale, tomographie optique, fluorescence, problème inverse, algorithmes de reconstruction, méthodes variationnelles, optimisation convexe.

To my family

Acknowledgement

This work was made possible by a number of actors that I wish to thank here. First of all I thank the two persons who initiated this project, namely my advisor Prof. Michael Unser from EPFL, and Peter Niederer from SCANCO Medial AG (and former professor at ETHZ). In particular, I am grateful to Michael for his generosity and his guidance at scientific and communication levels (which I very much need to improve). My main collaborator during this project was Dr. Kai Hassler from SCANCO Medical, to whom I owe most of the applied part of this work. He is the one who engineered the measurement system, and provided me with FDOT data. I also thank him for the various insights he gave about the physics involved in fluorescence imaging. I acknowledge as well Sebanti Sanyal, who contributed to the forward model code, and shared her technical expertise in software engineering. I am grateful to my fellow members in the lab, who helped me indirectly by exchanging interesting ideas, and more generally made the working atmosphere enjoyable. Among them, a special thanks to Matthieu and Pouya, I appreciated their patience and consideration for discussing scientific and practical matters. I also have a thought for my successive office mates Florian, Masih, Ayush and Satish who kept a cheerful ambiance at our working place. Finally I wish to thank the permanent staff of the lab, Manuelle, Daniel and Philippe, for their flawless management of administrative and computer related issues, not to mention the sound expertise accumulated over years of engineering image processing algorithms.

The funding of this research comes from several entities. We received support from the Swiss Innovation Promotion Agency (CTI 9601.PFLS-LS), the National Competence Center in Biomedical Imaging, and the STI Faculty of EPFL.

Last but not least, I express my gratitude to my family and my girlfriend Sandrine for their unconditional support. I thank them for everything. This work is

dedicated to them.

Contents

Abstract	i
Résumé	iii
Acknowledgement	vii
Introduction	1
0.1 Contributions	2
0.1.1 An efficient forward model	3
0.1.2 General ℓ_p regularization	3
0.1.3 Anatomical <i>a priori</i> using $\ell_{2,1}$ regularization	4
0.2 Organization of the thesis	4
1 Tissue Optics and Diffuse Optical Imaging	5
1.1 Scattering and absorption	5
1.2 Radiative transfer	6
1.2.1 Governing equation	6
1.2.2 Boundary conditions	7
1.2.3 Measured quantity	7
1.3 Diffusion approximation	8
1.3.1 The diffusion model	8
1.3.2 Boundary conditions	9
1.3.3 Measured quantity in the diffusion model	9
1.3.4 Validity of the diffusion approximation	10

1.4	Tomographic Diffuse Optical Imaging	10
1.4.1	General considerations	10
1.4.2	Diffuse Optical Tomography	12
1.4.3	Fluorescence Diffuse Optical Tomography	12
2	Theoretical basis of FDOT reconstruction	15
2.1	Inverse-problem framework	15
2.2	Diffusion based forward model in fluorescence tomography	16
2.2.1	General diffusion model	16
2.2.2	Helmholtz model	18
2.2.3	Source terms and sampling operator	19
2.2.4	Mathematical model for the FDOT inverse-problem	20
2.2.5	Multiple sources, tomographic setup	22
2.3	Well-posedness, Ill-posedness	23
2.4	Properties of the FDOT forward model	23
2.4.1	Some properties of the Helmholtz operator	23
2.4.2	Some properties of the forward model	26
2.5	FDOT inverse problem, regularization and variational approach	28
2.5.1	Ill-posedness	28
2.5.2	Regularization, Variational reconstruction approach	31
2.5.3	Regularization and statistical inference	32
2.6	Remarks	33
2.6.1	Ill-posedness and uniqueness	33
2.6.2	Partial differential equations and coefficient identification	33
2.6.3	Optimize-discretize vs. discretize-optimize	33
2.6.4	Other regularization methods	34
3	Numerical model for FDOT	35
3.1	Notation	36
3.2	The discrete variational reconstruction method	36
3.3	Discretization grid	38
3.4	Discrete Helmholtz operator	39
3.4.1	A Finite-Volumes Method for the Helmholtz equation	39
3.4.2	Numerical implementation	43
3.4.3	Multiple right-hand-sides	44
3.4.4	Accuracy of the Finite-Volumes Method	44

3.4.5	Sensitivity relations	44
3.5	Discrete forward model	45
3.5.1	Sampling	45
3.5.2	Structure of the discrete forward model	45
3.5.3	Multiple sources and implementation	46
3.5.4	Performance	51
3.5.5	Misfit rescaling	52
3.6	Regularization of the discrete problem	52
3.6.1	General considerations	52
3.6.2	Discrete gradient and application to regularization	54
3.6.3	Application to Tikhonov regularization	55
3.6.4	Application to total-variation	56
4	General ℓ_p-regularization for Fluorescence Diffuse Optical Tomography	57
4.1	Notation	59
4.2	Sparsity-inducing regularization potentials	59
4.3	Iterative optimization methods for sparsity-inducing regularization	62
4.3.1	Limitations of conventional methods	63
4.3.2	Optimization transfer methods	63
4.3.3	Dual and primal-dual methods	71
4.3.4	Practical performance on FDOT reconstruction	76
4.4	Reconstruction experiments : methods	79
4.4.1	Phantom experiments	79
4.4.2	Simulations	80
4.4.3	Contrast-to-noise ratio	80
4.4.4	Reconstruction algorithm	81
4.5	Reconstruction experiments : results	82
4.5.1	Experiment 1 : synthetic data, a single inclusion.	82
4.5.2	Experiment 2 : phantom data, a single inclusion.	83
4.5.3	Experiment 3 : phantom data, two inclusions.	86
4.5.4	Experiment 4 : CNR study, phantom data, a single inclusion.	88
4.5.5	Experiment 5 : CNR study, synthetic data, a single inclusion of varying size.	89
4.5.6	Experiment 6 : Quantification, synthetic data	89

5	Sparsity-driven reconstruction for FDOT with anatomical priors	93
5.1	Introduction	94
5.2	Group sparsity, anatomical <i>a priori</i> , and $\ell_{2,1}$ penalty	95
5.3	Reconstruction algorithm	96
5.4	Results	97
5.4.1	Experiment 1 : two dimensions, synthetic data, multiple inclusions	98
5.4.2	Experiment 2 : three dimensions, experimental data, two inclusions, accurate structural information	99
5.4.3	Experiment 3 : three dimensions, experimental data, two inclusions, inaccurate structural information	101
5.4.4	Experiment 4 : three dimensions, experimental data, acquired <i>a priori</i> information	101
5.4.5	Experiment 5 : three dimensions, experimental data, quantification	103
5.5	Conclusion	103
6	Conclusion	113
6.1	Summary	113
6.2	Outlook for future research	114
A	Derivation of the mathematical models for photon migration in tissue	117
A.1	Some moments on the unit sphere	117
A.2	Derivation of the transport equation	118
A.3	Derivation of the diffusion equation	120
B	Constrained proximal mappings for the $\ell_{2,1}$-norm	125
	Bibliography	128
	Curriculum Vitæ	139

Introduction

MOLECULAR imaging has come to mean the use of engineered biomarkers to visualize non-invasively specific molecular targets and pathways *in vivo*. This complements the traditional imaging techniques that reveal the anatomy of an organism, and shifts the way imaging is employed in practice, from imaging structure to imaging function. Initially limited to nuclear imaging modalities, such as positron emission tomography, the capability of visualizing individual molecules *in vivo* was implemented in optical imaging modalities as a result of revolutionary advances in the development of optical molecular probes. It is now possible to create 'smart' fluorescent markers that are targeted to specific molecules, and activated using an optical signal. Combined with recent progress in photon detection instrumentation, mathematical modeling of photon migration, and inversion algorithms, the use of these targetable optical probes developed into *fluorescence diffuse optical tomography* (FDOT), an innovative modality for resolving molecular activity deep inside living tissue.

The use of light to investigate properties of living tissue is reported in the scientific literature as early as the 1920s [1]. However, it is only in the early 1990s that tomographic reconstruction of optical properties deep inside tissue was considered and formulated as an inverse problem, as in X-Ray computed tomography [2]. These early studies were limited to intrinsic contrast of tissue, such as absorption and scattering coefficients. Yet, the development of this technique, termed *diffuse optical tomography* (DOT), initiated a new imaging modality and paved the way to implementations using exogenous fluorescent contrast agents. In the late 1990s, the first experiments involving reconstruction of a fluorescent molecule distribution in scattering media were reported [3], and the first *in vivo* results appeared in the early 2000s [4]. This new imaging technology started to develop under the names of

fluorescence mediated tomography (FMT) and FDOT. In the last ten years, FDOT saw a number of breakthroughs which led to the launching of the first commercial systems in the early 2010s. The prevailing application is small animal imaging, but medical imaging implementations are anticipated.

This thesis originates from a collaboration between the Biomedical Imaging Group of EPFL, and the micro computed-tomography company SCANCO Medical AG, to develop an interoperable X-Ray/FDOT system from scratch. In this context, our goal was to design integrated software providing an innovative solution to image reconstruction in FDOT. The image reconstruction problem for FDOT in the continuous-wave mode is briefly stated as follows. We are given an object composed of a highly scattering material and containing a fluorescent inclusion. A number of measurements are performed, in a similar manner as in X-Ray tomography, by illuminating the object on one side and collecting the out-coming fluorescence light on the other side. In the continuous-wave mode, the illumination has a steady amplitude. The task is to estimate the fluorophore distribution inside the object, starting from these projections. Independently of the practical limitations of the acquisition system, the physical factors making this problem difficult are the strong scattering and attenuation experienced by photons as they propagate in the medium. The corresponding imaging kernel contains an important smoothing component which makes the inverse problem severely ill-posed. Our goal is to investigate novel inversion algorithms with improved reconstruction accuracy and execution time.

0.1 Contributions

We mainly focus on three topics. First, we present the design and implementation of a computationally efficient forward model. Then, we propose a novel regularization technique which enables one to incorporate general sparsity *a priori* in the inversion. And finally, in the context of a hybrid acquisition system, we provide an efficient strategy to incorporate anatomical information obtained from the dual modality, in the reconstruction. The end product of this study is a practical reconstruction software entirely developed in our lab, with state-of-the-art performance in terms of computational cost and localization accuracy.

0.1.1 An efficient forward model

All FDOT reconstruction algorithms implement a *forward model* that is used to compute the propagation of photons in the investigated object. Most algorithms consist in iteratively fitting the forward model to the observations, which makes it a performance bottleneck. For the majority of FDOT applications, the most accurate mathematical model of photon migration which remains computationally tractable, is the diffusion equation. In this work, we introduce a novel approach based on the *finite volumes method* (FVM), as an alternative to the finite elements method, in order to solve numerically the diffusion equation. We develop this method into a fast implementation of the forward model, that handles arbitrary source and detector patterns; not necessarily points. One important point is that our design choices were guided by the type of regularization policies that we want to employ. Therefore, we obtain a unified framework that allows for a consistent treatment of the propagation model and of the regularization. It is also noteworthy that our forward model accommodates arbitrary tetrahedral grids, which is not the case of many FVM.

0.1.2 General ℓ_p regularization

Due to the ill-posedness of the FDOT inverse problem, it is necessary to integrate some kind of regularization technique in the reconstruction algorithms. This is commonly implemented by early stopping of iterative algorithms [5], or by incorporating a Tikhonov-like quadratic penalty in variational methods [6]. These approaches are known to achieve regularization by selecting candidate solutions with low-frequency content, which corresponds to a smoothing constraint. In this work, building upon the recent breakthroughs in the sparse representation of signals, we propose to apply sparsity-promoting constraints to the reconstruction problem. We see sparsity as a form of *a priori* which is well suited to a class of signals often encountered in practice. Apart from a particular case [7], sparsity has not been considered as a general regularization paradigm for FDOT. In this study, we enforce general sparsity constraints by using a variational method with a ℓ_p regularization term, where $p \in [1, 2[$. This formulation leads to fast reconstruction algorithms relying on convex optimization principles.

0.1.3 Anatomical *a priori* using $\ell_{2,1}$ regularization

As hybrid acquisition systems are being developed, a natural idea is to employ the information obtained from another modality, in order to enhance the FDOT reconstruction. We propose a regularization method that allows to incorporate anatomical *a priori* information in the inversion. To that end, we extend the sparsity-promoting framework developed in the first part of this work, to the more general concept of group sparsity. While sparsity is usually understood as sparsity of coefficients, this study considers sparsity of groups of coefficients. Based on the anatomical information, we partition the reconstruction grid into groups of voxels. The effect of promoting group sparsity in that case, is to automatically select the segments containing the fluorescent probe. As a result, the reconstruction is focussed on the active part of the grid. This approach is implemented in practice by using a variational method with a $\ell_{2,1}$ regularization term. Efficient algorithms and experimental results are discussed.

0.2 Organization of the thesis

This thesis is organized as follows. In Chapter 1, we present the most commonly employed models of photon propagation in tissue. We review the advantages and limitations of the diffusion approximation. In Chapter 2, we describe abstractly the mathematical model underlying this work. We attempt to approach the general theory of linear inverse problems from the angle of FDOT. Then, in Chapter 3, we present the discretized model that we have developed in this work. After that, in Chapter 4, we discuss an inversion method that is based on ℓ_p regularization. We introduce sparsity-inducing regularization, and derive efficient algorithms. We present experimental results obtained with tissue phantoms, and inquire the proposed ℓ_1 regularization approach in contrast with more conventional ℓ_2 regularization. Finally, in Chapter 5, we describe an approach to incorporate anatomical information in FDOT reconstruction. We discuss $\ell_{2,1}$ penalization, and demonstrate the resulting algorithm on phantom data.

Chapter 1

Tissue Optics and Diffuse Optical Imaging

WE start this chapter by reviewing the physical models employed for the propagation of light in tissue. Biological tissue belongs to the class of turbid media which are characterized by their strong photon absorption and photon scattering properties. The most adequate model developed in scattering theory for the propagation of light in turbid media is the equation of radiative transfer [8], or (photon) transport equation. In the case of a weakly absorbing, but strongly scattering material, the transport equation is readily approximated by a diffusion equation. The photon transport models presented in this chapter bear close similarities with those encountered in neutron transport [9]. In the second part of this chapter, we explain how diffuse photon waves are used in practice in tomographic systems, in order to visualize endogenous contrast of optical properties, or exogenous fluorescent contrast agents.

1.1 Scattering and absorption

As photons travel in tissue they are subjected to interactions with matter. In tissue optics, one usually considers two types of interactions: *absorption* and elastic *scattering*. Elastic scattering refers to a change in photon direction with conservation

of energy. This phenomenon arises from inhomogeneities in refractive index due to the presence of cell membranes and sub-cellular structures. Scatterer density in tissue is high, and therefore, the scattering process is appropriately described by the multiple scattering theory which assumes that light undergoes many scattering events before reaching the observer. Because of the high number of randomly distributed scatterers, with very different sizes and shapes, one is only able to make a description of tissue in terms of an equivalent scattering object. The latter is characterized by its *scattering cross-section* μ_s [m^{-1}] representing the probability per unit length for a photon to be scattered, and its *scattering phase-function* $p(\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}')$ [sr^{-1}] which gives the probability per unit of solid angle for a scattered photon coming from direction $\hat{\mathbf{s}}$ to be redirected in direction $\hat{\mathbf{s}}'$. Note that $p(\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}')$ depends solely on the cosine of the scattering angle. It is also noteworthy that the inverse of the scattering cross-section μ_s^{-1} gives the scattering *mean free-path*; that is, the mean distance traveled by a photon before being scattered. In a similar way, absorption in tissue is described by the *absorption cross-section* μ_a [m^{-1}].

1.2 Radiative transfer

1.2.1 Governing equation

It is common to employ the *radiance* $I(\mathbf{r}, \hat{\mathbf{s}}, t)$ [$\text{W m}^{-2} \text{sr}^{-1}$] for describing a light field in tissue. It is the density of light energy at point \mathbf{r} , flowing in direction $\hat{\mathbf{s}}$, per unit of surface, per unit of solid angle. The *radiative transfer equation* (or *photon transport equation*) governs the radiance in tissue

$$\begin{aligned} \frac{1}{c} \partial_t I(\mathbf{r}, \hat{\mathbf{s}}, t) + \hat{\mathbf{s}} \cdot \nabla_{\mathbf{r}} I(\mathbf{r}, \hat{\mathbf{s}}, t) + \mu_t I(\mathbf{r}, \hat{\mathbf{s}}, t) \\ = \mu_s \int_{4\pi} p(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) I(\mathbf{r}, \hat{\mathbf{s}}', t) d\hat{\mathbf{s}}' + q(\mathbf{r}, \hat{\mathbf{s}}, t) \end{aligned} \quad (1.1)$$

This equation is obtained by balancing incoming and out-coming energy in an infinitesimal volume at every point. It does not account for coherence and polarization of light, which is intuitively justified by the fact that there is a high density of randomly-located, randomly-oriented scatterers. In (1.1), $\mu_t = \mu_s + \mu_a$ [m^{-1}] is the *transport cross-section*, c [m s^{-1}] is the speed of light in the medium, and $q(\mathbf{r}, \hat{\mathbf{s}}, t)$ [$\text{W m}^{-3} \text{sr}^{-1}$] is a source term. The reader is referred to Appendix A for a derivation of the radiative transfer equation.

1.2.2 Boundary conditions

Boundary conditions account for light injected in the medium, as well as internal reflexion at interfaces where the refraction indices of the medium and the outside world do not match. We denote Ω the domain, $\partial\Omega$ its boundary, and \mathbf{n} the unit normal vector of the boundary pointing outward. In this absence of external illumination, we obtain

$$I(\mathbf{r}, \hat{\mathbf{s}}, t) = R(\hat{\mathbf{s}}')I(\mathbf{r}, \hat{\mathbf{s}}', t), \quad \mathbf{r} \in \partial\Omega, \hat{\mathbf{s}} \cdot \mathbf{n} < 0 \quad (1.2)$$

$$I(\mathbf{r}, \hat{\mathbf{s}}, t) = T(\hat{\mathbf{s}}'')I(\mathbf{r}, \hat{\mathbf{s}}'', t), \quad \mathbf{r} \in \partial\Omega, \hat{\mathbf{s}} \cdot \mathbf{n} > 0 \quad (1.3)$$

where $R(\hat{\mathbf{s}}')$ and $S(\hat{\mathbf{s}}'')$ are the Fresnel reflection and transmission coefficients, respectively, $\hat{\mathbf{s}}'$ gives the inner reflection direction with respect to $(\hat{\mathbf{s}}, \mathbf{n})$, and $\hat{\mathbf{s}}''$ gives the refraction direction (Snell's law).

1.2.3 Measured quantity

We are concerned here with non-contact measurements. For a detector with a narrow field of view $\delta\omega$ [sr], pointing at position $\mathbf{r} \in \partial\Omega$ of the object, and facing direction $\hat{\mathbf{s}}$, it can be established that the measured quantity is proportional to the radiance at that point, in that direction : $MQ(\mathbf{r}, \hat{\mathbf{s}}, t) \propto I_o(\mathbf{r}, \hat{\mathbf{s}}, t)\delta\omega$. This applies in particular to setups employing a camera; the observed image intensity is proportional to the radiance, up to distortions introduced by the optical system. One also has to account for the mismatch in optical index at the interface, which leads to

$$MQ(\mathbf{r}, \hat{\mathbf{s}}, t) \propto T(\hat{\mathbf{s}}')I_o(\mathbf{r}, \hat{\mathbf{s}}', t) \quad (1.4)$$

where $T(\hat{\mathbf{s}}')$ is the Fresnel transmission coefficient, and $\hat{\mathbf{s}}'$ is related to $\hat{\mathbf{s}}$ by Snell's law.

Equation (1.1) is a linear integro-differential equation in space, time and angle. Although some authors considered this model for photon propagation in optical tomography [10], angular coupling makes it difficult to work with in practice. The main hurdle is the high computational cost incurred for evaluating numerical solutions. We will now turn our attention to the diffusion approximation of this equation, which leads to lighter calculations.

1.3 Diffusion approximation

Under conditions specified in the sequel, one can employ an accurate approximation of the radiative transfer equation that does not include an angular variable. It is called the *diffusion approximation*. This way of eliminating the angle variable is a standard assumption in practice, which leads to a formulation of the inverse problem that is computationally tractable.

1.3.1 The diffusion model

The diffusion model involves two variables $u(\mathbf{r}, t)$ [W m^{-2}] and $\mathbf{J}(\mathbf{r}, t)$ [W m^{-2}] related to the radiance in the following way

$$u(\mathbf{r}, t) = \int_{4\pi} I(\mathbf{r}, \hat{\mathbf{s}}, t) d\hat{\mathbf{s}} \quad (1.5)$$

$$\mathbf{J}(\mathbf{r}, t) = \int_{4\pi} I(\mathbf{r}, \hat{\mathbf{s}}, t) \hat{\mathbf{s}} d\hat{\mathbf{s}} \quad (1.6)$$

Variable $u(\mathbf{r}, t)$ is known as the *fluence rate*. It corresponds to the density of optical energy flowing through an infinitesimal sphere at every point. The other quantity $\mathbf{J}(\mathbf{r}, t)$ is the *energy flux density*. It is defined as follows. For a surface element $d\vec{A}$ located at point \mathbf{r} , $\mathbf{J}(\mathbf{r}, t) \cdot d\vec{A}$ is the flux of light energy through $d\vec{A}$.

The diffusion model of light propagation is stated as the first-order differential system

$$\partial_t u(\mathbf{r}, t) + \nabla \cdot \mathbf{J}(\mathbf{r}, t) + \mu_a u(\mathbf{r}, t) = Q(\mathbf{r}, t) \quad (1.7)$$

$$\mathbf{J}(\mathbf{r}, t) + D \nabla u(\mathbf{r}, t) = 0 \quad (1.8)$$

where

$$Q(\mathbf{r}, t) = \int_{4\pi} q(\mathbf{r}, \hat{\mathbf{s}}, t) d\hat{\mathbf{s}}$$

We refer to Appendix A for the derivation of this equation. In (1.8) we have introduced the spatially-varying *diffusion coefficient* D [m] :

$$D = \frac{1}{3(\mu_a + (1-g)\mu_s)} \quad (1.9)$$

It is defined using g , the mean cosine of the scattering angle

$$g = \int_{4\pi} (\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) p(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) d\hat{\mathbf{s}}. \quad (1.10)$$

It becomes apparent that the natural quantity that characterizes scattering in the diffusion approximation is the *reduced scattering coefficient* $\mu'_s = (1 - g) \mu_s$. This coefficient incorporates part of the angular information in the diffusion model. We see that the reduced scattering mean-free-path $1/\mu'_s$ is modified by the mean cosine of the scattering angle. Intuitively, when photons tend to be scattered forward, the situation is closer to an absence of scattering, and the mean free path is larger.

The first equation of the diffusion model is a continuity equation for the optical energy. Until there, no approximation is made. The approximation lies within the second equation, namely Fick's law, which states that the energy flux density is proportional to the gradient of the energy density, and directed in the opposite direction. Eliminating the flux variable, one can restate the diffusion model using a single diffusion equation

$$\frac{1}{c} \partial_t u - \nabla \cdot (D \nabla u) + \mu_a u = q \quad (1.11)$$

1.3.2 Boundary conditions

The question of boundary conditions has been extensively studied [11]. In this work we use the Robin boundary condition

$$u + \beta \mathbf{n} \cdot \nabla u = 0, \quad \text{on } \partial\Omega, \quad (1.12)$$

where Ω is the bounded domain enclosing the object under investigation, and

$$\beta = 2 \frac{1 + R_{\text{eff}}}{1 - R_{\text{eff}}} D, \quad (1.13)$$

with R_{eff} the reflexion coefficient.

1.3.3 Measured quantity in the diffusion model

Under the assumption that the surface of the sample is lambertian (the radiance does not depend on the direction of observation), it is established that $MQ(\mathbf{r}, \hat{\mathbf{s}}, t) \propto u(\mathbf{r}, t)$. In this expression, the use of the proportionality symbol rather than an equality accounts for the detector characteristics and the refractive index mismatch.

1.3.4 Validity of the diffusion approximation

The diffusion model is valid provided the following conditions are met :

- scattering is high enough to ensure that the multiple scattering model is valid.
- scattering dominates absorption ($\mu_a \ll \mu_s$). This condition ensures that the radiance is weakly dependent on the angle (see Appendix A for a precise meaning of that statement).
- the relative variation of the photon current in time is negligible compared to the interaction frequency

$$\frac{1}{\mathbf{J}(\mathbf{r}, t)} \frac{\partial}{\partial t} \mathbf{J}(\mathbf{r}, t) \ll c(\mu_a + \mu'_s) \quad (1.14)$$

Note that this condition is void in steady state, which is the case that we consider in the sequel.

- away from interfaces between regions having mismatching refractive indices. The characteristic distance involved here is $\mu'_s{}^{-1}$.
- the sources are isotropic.

1.4 Tomographic Diffuse Optical Imaging

Similar to X-ray computed tomography in spirit, diffuse optical tomography infers inner properties of tissue from observations of scattered waves. Line integrals of X-ray tomography are replaced by samples of diffuse waves. In this section we present *diffuse optical tomography* (DOT) and *fluorescence diffuse optical tomography* (FDOT) setups and applications. As this is the focus of our work, we limit ourselves to the *continuous-wave* mode of operation.

1.4.1 General considerations

Provided that the diffusion approximation is valid, optical wave propagation in living tissue is aptly described by a diffusion equation. The diffuse wave is entirely specified by the pair (D, μ_a) , where D is the diffusion coefficient of the medium, and

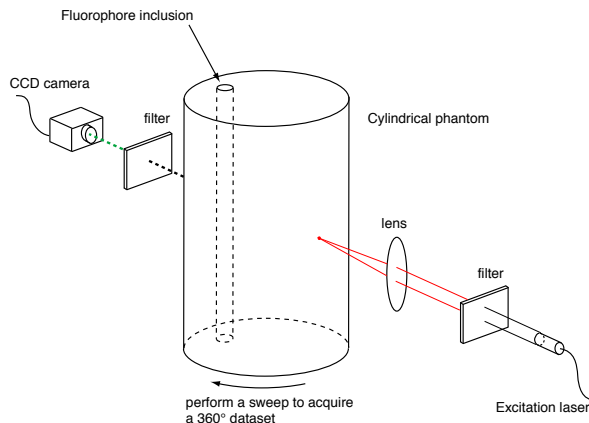


Figure 1.1: Data acquisition setup.

μ_a is the absorption coefficient, which are both assumed to be spatially-varying. It is thus natural to use such waves in order to gain insight about the inner diffusion and absorption properties of an object. As long as one can relate scattering and absorption contrast to biological features, one obtains valuable information about tissue. If secondary sources, fluorescent inclusions for instance, are present in the medium, diffuse waves can also be used to sense these sources.

A typical non-contact data acquisition setup employed in diffuse optical imaging is depicted Fig 1.1. A light source, usually a laser beam, is pointed at the specimen under investigation. Opposite to the light source, a detector consisting of an optical filter and a CCD camera records the light transmitted through the specimen. Data obtained this way is the light intensity, *on the boundary* of the specimen. Intuitively, such a system has a higher sensitivity close to the locations of the sources and detectors. In order to make the sensitivity more homogeneous, a tomographic acquisition is performed by rotating the specimen in between the source-detector setup, and scanning in the z direction. The use of optical filters allows for multi-spectral data acquisition. In this study, we consider the continuous-wave mode of operation in which the laser source emits a steady-state wave [12, 13]. Other setups employ pulsed sources [14, 15], or amplitude modulated sources [16, 17].

The performance of such an imaging system is limited by the strong attenuation experienced by photons as they propagate through the specimen. This results in a weak transmitted signal, potentially improper for imaging applications. Fortunately, tissue absorption attains a minimum in the *near-infrared* (NIR) range of the electromagnetic spectrum (600 – 950 nm). In this wavelength range, tissue absorption is low enough to allow for propagation of photons over several centimeters; see Fig 1.2. A feasibility study for fluorescence imaging was made in [18].

1.4.2 Diffuse Optical Tomography

Diffuse optical tomography relies on diffuse photon waves in turbid medium to image the diffusion and absorption coefficients. The three primary absorbers in tissue are water, oxy-, and desoxy-hemoglobin, with distinct absorption signatures shown in Fig 1.2. Absorption contrast is therefore mainly related to blood concentration and its oxygen content. Imaging of absorption contrast has several interesting medical applications. One of the most investigated is probably breast cancer detection and follow up [19, 20]. The contrast mechanism is the alteration of blood oxygenation and angiogenesis caused by aggressive growth of tumors. Another promising application is cerebral activity monitoring [21, 22]. Here, the principle is the same as in functional magnetic resonance imaging which detects variations in blood oxygenation incidental to brain function. It was also shown that scattering properties of cerebral tissue are correlated with brain function [23], which opens another potential avenue for DOT. From an abstract point of view, these applications pose the problem of estimating absorption and diffusion maps from peripheral light measurements.

1.4.3 Fluorescence Diffuse Optical Tomography

Imaging diffusion and absorption coefficients with DOT reveals intrinsic properties of tissue. By contrast, molecular imaging employs exogenous biomarkers in order to visualize specific molecular targets and pathways. The fluorescent probes used in FDOT are similar in spirit to the radio-tracers of PET. They are injected in the investigated specimen prior to data acquisition, where they bind to predefined molecular targets. FDOT consists in sensing them optically with diffuse photon waves. Genetically modified organisms naturally expressing fluorescent proteins can also be studied using the same principle.

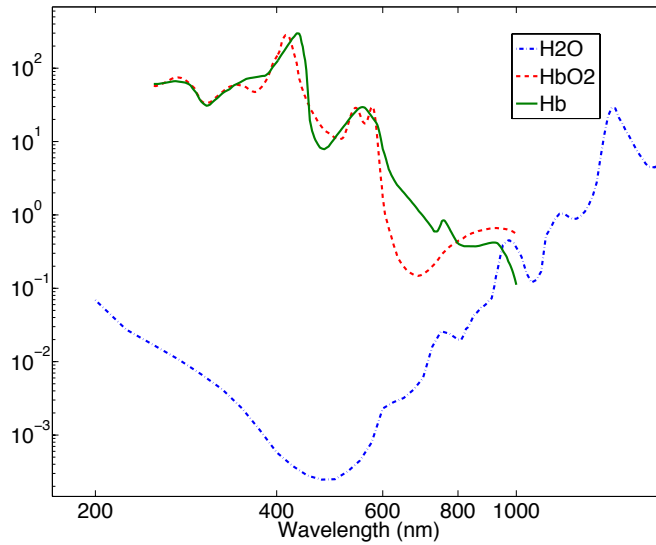


Figure 1.2: Absorption spectra of H₂O (in cm⁻¹), oxy-hemoglobin Hb (in mm⁻¹), and deoxy-hemoglobin HbO₂ (in mm⁻¹). From data compiled by the Oregon Medical Laser Center.

In fluorescence imaging, one uses a laser source in order to excite the fluorescent dye present in the specimen. This is achieved if the wavelength of the source belongs to the excitation spectrum of the fluorophore. In this case, the dye partly absorbs the excitation wave and the fluorescent molecule reaches an excited energy state. It instantly decays to ground state and becomes a secondary source that radiates light in its emission spectrum. By employing optical filters, one separates two sets of measurements :

- transmission measurements of light at the excitation wavelength, similar to those obtained in DOT,
- fluorescence light measurements.

In FDOT, the *distribution* of the fluorescent molecule has to be estimated from this boundary data. Ideally a quantitative imaging system outputs the concentration of fluorophore at each point.

Research employing small animal models is the main field of application of FDOT. So far, the technology has been successfully applied in oncology [24], studies on inflammation [25], cardiovascular diseases, pharmacokinetics, and bone metabolism [26]. Medical imaging applications are anticipated, in particular for breast cancer detection [27].

Chapter 2

Theoretical basis of FDOT reconstruction

THE goal of this second chapter is twofold. First we intend to give a description of the mathematical concepts related to reconstruction. We try to approach the general theory of inverse problems from the angle of FDOT reconstruction. To that end, we recall a selected number of results that are relevant to our problem. We emphasize the theoretical limitations of the model, and justify the regularization approach employed in the rest of this work. Our second goal is to bridge the abstract inverse problem framework with a signal processing formulation in terms of a sampling operator. The main outcome of this re-interpretation is the result in Proposition 2.4.7, which identifies linear functionals that specify the FDOT forward model. This original formulation extends the commonly employed reciprocity property of Green's functions, to arbitrary source and detector functions. The practical implication of this result is to provide an alternative to the direct implementation of the forward model with cascaded diffusion equations.

2.1 Inverse-problem framework

FDOT image reconstruction is performed in the *inverse-problem* framework, as is depicted in diagram Fig 2.1 (inspired from [28]). The underlying spaces and opera-

tors of interest are as follows. P is the *parameter space*, containing the parameters we want to reconstruct. M is the *measurements space*. It contains the observations $\tilde{\mathbf{y}}$ made with the imaging system. In signal processing terminology, P and M are sometimes named *image domain* and *measurement domain*, respectively. Physical modeling of the imaging system enables to define the *forward model* \mathcal{F} which is an operator mapping P to M . Given a parameter $\mathbf{p} \in P$, the forward model predicts measurement values $\mathbf{y} = \mathcal{F}\mathbf{p}$. Due to measurement non-idealities (modeling error, measurement noise, ...), the observed quantities $\tilde{\mathbf{y}}$ do not belong necessarily to the range of \mathcal{F} . The problem of reconstructing parameter values from measurements is called the *inverse problem*. It amounts in some sense to solving the equation $\mathcal{F}\mathbf{p} = \tilde{\mathbf{y}}$, or equivalently, to finding a suitable inverse \mathcal{F}^{-1} of \mathcal{F} .

The forward model is further described using a propagation operator \mathcal{K} , and a sampling operator \mathcal{S} . Most imaging modalities are based on probing the sample under investigation using a number of *sources*. Q denotes the space of sources. Given a source term and a parameter value, the operator \mathcal{K} propagates the effect of the source inside the sample, which results in a field belonging to U , the *solution space* of \mathcal{K} . The resulting field is then sampled using operator \mathcal{S} and sampling functions from space S . Typical data acquisition involves several source terms and sampling functions.

In order to reconstruct \mathbf{p} , one monitors the *misfit* between forward model predictions and observations. If mean-square error is employed, the misfit reads

$$\frac{1}{2} \|\tilde{\mathbf{y}} - \mathcal{F}\mathbf{p}\|_2^2 \quad (2.1)$$

Most reconstruction procedures compute a parameter $\hat{\mathbf{p}}$ that minimizes the misfit. In the following we specify the different elements of this general framework in the case of FDOT.

2.2 Diffusion based forward model in fluorescence tomography

2.2.1 General diffusion model

As we saw in the previous chapter, fluorescence diffuse optical tomography measurements are obtained by illuminating the sample with a laser and collecting the

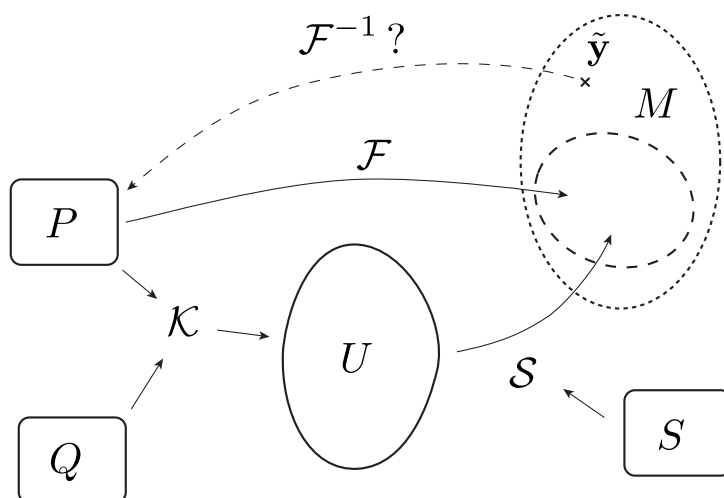


Figure 2.1: Inverse problem framework

outgoing light using a camera. Since light propagation in tissue is aptly modeled by a diffusion equation, it is natural to use the diffusion equation as propagation operator. Specifically, let us assume a tissue sample containing a fluorophore inclusion with excitation wavelength λ_x and emission wavelength λ_m . Let D_x (resp. D_m) be the diffusion coefficient at excitation (resp. emission) wavelength, and μ_{ax} (resp. μ_{am}) the absorption coefficient at excitation (resp. emission) wavelength. Coefficients $D_x, \mu_{ax}, D_m, \mu_{am}$ are combined quantities accounting for endogenous properties of the medium, as well as fluorophore contribution. As an example, for absorption at excitation wavelength, we have $\mu_{ax} = \mu_{ax}^i + \mu_{ax}^f$, where μ_{ax}^i is the endogenous absorption and μ_{ax}^f the fluorophore absorption. Modeling the excitation laser by a source term q , fluorescence excitation-emission is described by two cascaded diffusion equations

$$-\nabla \cdot (D_x \nabla u_x) + \mu_{ax} u_x = q \quad (2.2)$$

$$-\nabla \cdot (D_m \nabla u_m) + \mu_{am} u_m = \mu_{ax}^f \eta u_x \quad (2.3)$$

where u_x is the excitation field and u_m the emission field. In the second equation, η is the quantum yield of the fluorophore. Partial-current boundary conditions are applied to both equations. Fluorophore concentration is directly proportional to the fluorophore absorption μ_{ax}^f . In this setting, the propagation operator is given by diffusion equations (2.2), (2.3). The parameter \mathbf{p} consists of the vector of space-varying coefficients $(D_x, \mu_{ax}, D_m, \mu_{am}, \mu_{ax}^f)$.

2.2.2 Helmholtz model

Because we are operating in a relatively narrow wavelength range (650-700nm), we assume that $D_x \approx D_m$ and $\mu_{ax} \approx \mu_{am}$. This leads to the following simplified propagation operator

$$-\nabla \cdot (D \nabla u_x) + \mu_a u_x = q \quad (2.4)$$

$$-\nabla \cdot (D \nabla u_m) + \mu_a u_m = \mu_a^f \eta u_x \quad (2.5)$$

where the parameter vector is $\mathbf{p} = (D, \mu_a, \mu_a^f)$, instead of $\mathbf{p} = (D_x, D_m, \mu_{ax}, \mu_{am}, \mu_{ax}^f)$ for the general model. D and μ_a encode the medium heterogeneity, whereas μ_a^f encodes the fluorophore distribution.

One can express the diffusion equations (2.4), (2.5) in Helmholtz form by posing $\gamma = \sqrt{D}$ and making the change of variable $\phi_x = \gamma u_x$, $\phi_m = \gamma u_m$ [29, 30]. Since $\nabla \cdot (\gamma^2 \nabla u) = \gamma (\nabla^2(\gamma u) - (\nabla^2 \gamma)u)$, system (2.4), (2.5) is equivalent to

$$-\nabla^2 \phi_x + \kappa^2 \phi_x = \hat{q} \quad (2.6)$$

$$-\nabla^2 \phi_m + \kappa^2 \phi_m = c \phi_x \quad (2.7)$$

where $\kappa^2 = \frac{\nabla^2 \gamma}{\gamma} + \frac{\mu_a}{\gamma}$, $\hat{q} = \frac{q}{\gamma}$ and $c = \frac{\mu_a^f \eta}{\gamma^2}$.

We recall that u_x (resp. u_m) satisfy partial-current boundary conditions; that is, $u_x + \beta \mathbf{n} \cdot \nabla u_x = 0$, on $\partial\Omega$. By assuming that D is constant (or slowly varying) close to the boundary, the same boundary conditions remain valid for ϕ_x (resp. ϕ_m).

In the sequel we use the propagation operator defined by (2.6), (2.7), and denote it \mathcal{K} . In this case, the parameter is $\mathbf{p} = (\kappa^2, \hat{c})$. The main advantage of this formulation is that we have a single heterogeneity parameter κ^2 which combines μ_a and D . On the downside, the fluorescence parameter μ_a^f is not accessible directly. Instead we have the ratio μ_a^f/D . If the distribution of μ_a^f is localized, this handicap can be compensated using the mean diffusion coefficient.

Forward models based on the Helmholtz equation are commonly employed in DOT or FDOT [31, 5]. Note that some authors assume that D is constant, or that ∇D is negligible [32, 33]. This is in fact a special case of the Helmholtz model, and the same restrictions apply. We also mention the Born normalization approach [12, 34] which also uses a Helmholtz forward model, although in this case, tissue heterogeneity is not estimated but compensated by the normalization.

2.2.3 Source terms and sampling operator

In most FDOT setups, the source is a laser beam pointed at the sample, which is accurately described by a point source. There are mainly two ways of modeling point sources proposed in the literature :

- a Dirac distribution inside the domain Ω , located at a distance $1/\mu'_s$ from the intersection point between the laser and the boundary. This leads to a source term $q(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_s)$ in the Helmholtz equation, with $\mathbf{r}_s \in \Omega$.

- a Dirac distribution on the boundary $\partial\Omega$ of the domain, located at the intersection point of laser and boundary. In this case the source term of the Helmholtz equation is 0, but the boundary conditions are modified to become $u_x + \beta \mathbf{n} \cdot \nabla u_x = \delta(\mathbf{r} - \mathbf{r}_s)$ with $\mathbf{r}_s \in \partial\Omega$.

Here we consider the first modeling strategy with a slight generalization. We describe an arbitrary illumination pattern by a corresponding source function $q(\mathbf{r})$ defined in Ω and whose support lies in the vicinity of the boundary.

In this work we consider linear measurements, which means that the sampling operator \mathcal{S} is described using a number of linear forms. In practice, sampling occurs on the boundary $\partial\Omega$ of the domain. We approximate this by considering linear forms defined in Ω , whose support lie in the vicinity of the boundary.

Note that we choose to model source terms and sampling operator by functions defined on Ω , rather than $\partial\Omega$, to simplify the mathematical model presented thereafter. One could consider boundary source functions and sampling functions at the expense of additional mathematical technicalities, without changing the spirit of the argument. In practice both approaches lead to similar implementations and results.

2.2.4 Mathematical model for the FDOT inverse-problem

We now summarize the mathematical model employed in this work. For the parameter space we consider $P = P_1 \times P_2$ where

$$P_1 = \{ \kappa^2 : \Omega \rightarrow \mathbb{R}, \text{ measurable ; } \exists \kappa_1^2 \geq \kappa_0^2 > 0 \text{ with } \kappa_0^2 \leq \kappa^2(x) \leq \kappa_1^2 \text{ a.e.} \} \quad (2.8)$$

$$P_2 = \{ c : \Omega \rightarrow \mathbb{R}, \text{ measurable ; } \exists c_0 > 0 \text{ with } c_0 \leq c(x) \text{ a.e.} \} \quad (2.9)$$

The source space is chosen to be $Q = L^2(\Omega)$, the space of square-integrable functions. We assume that source functions have support in the neighborhood of $\partial\Omega$.

We denote \mathcal{H} the Helmholtz operator defined by

$$\mathcal{H} : P_1 \times L^2(\Omega) \rightarrow L^2(\Omega) \quad (2.10)$$

$$(\kappa^2, q) \mapsto u \quad (2.11)$$

where u satisfies

$$-\nabla^2 u + \kappa^2 u = q, \quad \text{in } \Omega, \quad (2.12)$$

$$u + \beta \mathbf{n} \cdot \nabla u = 0, \quad \text{on } \partial\Omega \quad (2.13)$$

Here we used the Sobolev embedding $H^1(\Omega) \subset L^2(\Omega)$, where $H^1(\Omega)$ is the first-order Sobolev space of square-integrable functions. With this definition, the natural choice for the solution space is $U = L^2(\Omega) \times L^2(\Omega)$, and the propagation operator is given by

$$\mathcal{K} : P \times L^2(\Omega) \rightarrow U \quad (2.14)$$

$$(\kappa^2, c, q) \mapsto (u_x, u_m) = (\mathcal{H}(\kappa^2; q), \mathcal{H}(\kappa^2; c u_x)) \quad (2.15)$$

The first component of \mathcal{K} corresponds to the excitation field u_x , while the second corresponds to the emission field u_m .

To each source term q , we associate a sampling operator on $U = L^2(\Omega) \times L^2(\Omega)$. Here, we use the same sampling for u_x and u_m , which means that the sampling operator is of the form $\mathcal{S}_{q, xm} = \mathcal{S}_q \otimes \mathcal{S}_q$, where \mathcal{S}_q is a sampling operator on $L^2(\Omega)$ for source q . \mathcal{S}_q is specified by N_q linear forms $(f_1, \dots, f_{N_q}) \in (L^2(\Omega))^{N_q}$ that are identified to elements of $L^2(\Omega)$ since it is a Hilbert space. As discussed previously, the sampling functions are assumed to be localized in the neighborhood of the boundary. Thus, the sampling operator for source q and excitation (resp. emission) is defined by

$$\mathcal{S}_q : L^2(\Omega) \rightarrow \mathbb{R}^{N_q} \quad (2.16)$$

$$u \mapsto \begin{pmatrix} \langle f_1, u \rangle_{L^2(\Omega)} \\ \vdots \\ \langle f_{N_q}, u \rangle_{L^2(\Omega)} \end{pmatrix} \quad (2.17)$$

and the overall sampling operator (for both excitation and emission) is

$$\mathcal{S}_{q, mx} : U = L^2(\Omega) \times L^2(\Omega) \rightarrow \mathbb{R}^{2N_q} \quad (2.18)$$

$$(u_x, u_m) \mapsto \begin{pmatrix} \mathcal{S}_q u_x \\ \mathcal{S}_q u_m \end{pmatrix} \quad (2.19)$$

In this setting, the forward model for a single source term q is

$$\mathcal{F} : P \rightarrow \mathbb{R}^{2N_q} \quad (2.20)$$

$$(\kappa^2, c) \mapsto \begin{pmatrix} \mathcal{S}_q u_x \\ \mathcal{S}_q u_m \end{pmatrix} = \begin{pmatrix} \mathcal{S}_q \mathcal{H}(\kappa^2; q) \\ \mathcal{S}_q \mathcal{H}(\kappa^2; c u_x) \end{pmatrix} \quad (2.21)$$

The partial forward model related to excitation (resp. emission) is denoted \mathcal{F}_x (resp. \mathcal{F}_m), which enables to write $\mathcal{F} = \begin{pmatrix} \mathcal{F}_x \\ \mathcal{F}_m \end{pmatrix}$.

2.2.5 Multiple sources, tomographic setup

The forward model for a single source presented in the previous paragraph is straightforwardly extended to multiple sources employed in a tomographic setup. This is described by the following notation. Let us call N_s the number of source positions in the tomographic scan. We denote q_i , $i = 1..N_s$ the corresponding source terms, and \mathcal{S}_i , $i = 1..N_s$ the sampling operators. For source position i , the forward model is given by

$$(\kappa^2, c) \mapsto \begin{pmatrix} \mathcal{F}_x^i \\ \mathcal{F}_m^i \end{pmatrix} = \begin{pmatrix} \mathcal{S}_i \mathcal{H}(\kappa^2; q_i) \\ \mathcal{S}_i \mathcal{H}(\kappa^2; c u_x^i) \end{pmatrix} \quad (2.22)$$

where u_x^i is the i -th excitation wave. Now, let us regroup all the terms modeling excitation under the notation \mathcal{F}_x , which leads to

$$\mathcal{F}_x : \kappa^2 \mapsto \begin{pmatrix} \mathcal{F}_x^1(\kappa^2) \\ \vdots \\ \mathcal{F}_x^{N_s}(\kappa^2) \end{pmatrix} \quad (2.23)$$

Similarly, we denote \mathcal{F}_m the terms modeling emission. We have

$$\mathcal{F}_m : (\kappa^2, c) \mapsto \begin{pmatrix} \mathcal{F}_m^1(\kappa^2, c) \\ \vdots \\ \mathcal{F}_m^{N_s}(\kappa^2, c) \end{pmatrix} \quad (2.24)$$

Note that \mathcal{F}_x is functionally dependent on κ^2 only, while \mathcal{F}_m is functionally dependent both on κ^2 and c . Overall, our forward model \mathcal{F} is obtained by stacking \mathcal{F}_m and \mathcal{F}_x

$$\mathcal{F} : (\kappa^2, c) \mapsto \begin{pmatrix} \mathcal{F}_x(\kappa^2) \\ \mathcal{F}_m(\kappa^2, \beta) \end{pmatrix} \quad (2.25)$$

From a practical point of view, \mathcal{F} consists of $2 N_s$ Helmholtz equations followed by sampling.

2.3 Well-posedness, Ill-posedness

In both the forward and the inverse problems of FDOT, we need to solve functional equations. The forward problem involves solving the Helmholtz equation, while the inverse problem is to solve the equation $\mathcal{F}\mathbf{p} = \tilde{\mathbf{y}}$. The concept of well-posedness, and its converse ill-posedness, characterize existence and stability of solutions.

Definition 2.3.1. Consider a mapping $\mathcal{K} : H_1 \rightarrow H_2$. Equation $\mathcal{K}u = v$ is *well-posed* in the sense of Hadamard if the three following conditions hold :

1. there exists a solution for all $v \in H_2$,
2. the solution is unique,
3. the solution depends continuously on v .

If one or more conditions fail, the problem is said *ill-posed*.

A well-posed problem can be expected to be solved exactly and stably by numerical techniques. For an ill-posed problem, one has to consider generalized solutions since exact solutions might not exist. Practical numerical techniques also need to include schemes that select one candidate among multiple possibilities, and ensure continuity of the inversion scheme. This is termed *regularization*.

2.4 Properties of the FDOT forward model

In this section we summarize some general mathematical results about the Helmholtz equation and their implications on the forward model. We emphasize in particular that the forward problem is well posed. The mathematical results presented here are classical in functional analysis. Proposition 2.4.7 is an original formulation that expresses the forward model explicitly as a sampling operator.

2.4.1 Some properties of the Helmholtz operator

Theorem 2.4.1. Let $\kappa^2 \in P_1$. The Helmholtz operator

$$\mathcal{H}(\kappa^2; \cdot) : L^2(\Omega) \rightarrow L^2(\Omega) \tag{2.26}$$

$$q \mapsto \mathcal{H}(\kappa^2; q) \tag{2.27}$$

is well-defined. It is linear, injective, bounded, and compact.

We remind that for a linear operator injectivity is equivalent to the null-space being reduced to 0, and boundedness is equivalent to continuity. In addition \mathcal{H} is compact. This essentially means that \mathcal{H} is very similar to a finite-dimensional operator in the sense that it admits an eigen-decomposition in countably many eigen-spaces that are finite-dimensional with the exception of the null-space. For further details we refer to [35]. For the moment, the main consequence we draw from Theorem 2.4.1 is that the FDOT forward model is well-posed.

The next result specifies the dependency of \mathcal{H} with respect to the heterogeneity parameter.

Proposition 2.4.2. *Let $q \in L^2(\Omega)$. The operator*

$$\mathcal{H}(\cdot; q) : P_1 \rightarrow L^2(\Omega) \tag{2.28}$$

$$\kappa^2 \mapsto \mathcal{H}(\kappa^2; q) \tag{2.29}$$

is Frchet differentiable with respect to κ^2 and we have

$$\frac{\partial \mathcal{H}}{\partial \kappa^2} \delta \kappa^2 = -\mathcal{H}(\kappa^2; u \delta \kappa^2) \tag{2.30}$$

where $u = \mathcal{H}(\kappa^2; q)$.

Proposition 2.4.2 provides a closed-form expression for the derivative of the forward model. This turns out to be useful when optimization methods requiring the gradient are employed. In terms of computations, evaluating the gradient has the cost of solving a diffusion equation.

For characterizing the sampling process and solving linear least-squares problems, we need the adjoint of the propagation equation.

Proposition 2.4.3. *Let $\kappa^2 \in P_1$. The Helmholtz operator $\mathcal{H}(\kappa^2; \cdot) : L^2(\Omega) \rightarrow L^2(\Omega)$ is self-adjoint and positive-definite.*

Proof : Let $f, g \in L^2(\Omega)$, and define $u = \mathcal{H}f$ and $v = \mathcal{H}g$. Using Green's

identity we have

$$\begin{aligned}
 \langle \mathcal{H}f, g \rangle_{L^2(\Omega)} &= \int_{\Omega} u (-\nabla^2 v + \kappa^2 v) \, d\mathbf{r} \\
 &= \int_{\Omega} (-\nabla^2 u + \kappa^2 u) v \, d\mathbf{r} + \int_{\partial\Omega} \mathbf{n} \cdot \nabla u v - \mathbf{n} \cdot \nabla v u \, d\sigma \\
 &= \int_{\Omega} (-\nabla^2 u + \kappa^2 u) v \, d\mathbf{r} = \langle f, \mathcal{H}g \rangle_{L^2(\Omega)},
 \end{aligned}$$

which shows that \mathcal{H} is self-adjoint. Another application of Green's identity leads to

$$\begin{aligned}
 \langle \mathcal{H}f, f \rangle_{L^2(\Omega)} &= \int_{\Omega} u (-\nabla^2 u + \kappa^2 u) \, d\mathbf{r} \\
 &= \int_{\Omega} \nabla u \cdot \nabla u + \kappa^2 u^2 \, d\mathbf{r} - \int_{\partial\Omega} \mathbf{n} \cdot \nabla u u \, d\sigma \\
 &= \int_{\Omega} \nabla u \cdot \nabla u + \kappa^2 u^2 \, d\mathbf{r} + \int_{\partial\Omega} \frac{u^2}{\beta} \, d\sigma,
 \end{aligned}$$

which shows that \mathcal{H} is positive-definite \square

Definition 2.4.4. We call Green's function of the Helmholtz equation the solution $g(\mathbf{r}', \mathbf{r})$ of the following equation

$$-\nabla^2 g + \kappa^2 g = \delta(\mathbf{r} - \mathbf{r}'), \quad \text{in } \Omega \quad (2.31)$$

$$g + \beta \mathbf{n} \cdot \nabla g = 0, \quad \text{on } \partial\Omega \quad (2.32)$$

The function g satisfies

$$u = \mathcal{H}(\kappa^2; q) \Leftrightarrow u(\mathbf{r}) = \int_{\Omega} g(\mathbf{r}', \mathbf{r}) q(\mathbf{r}') \, d\mathbf{r}' \quad (2.33)$$

The Green's function enables one to express explicitly the solution of the Helmholtz equation. Unfortunately there are no closed-form expressions for Green's functions in the general case. Yet, the expression is known in the case of an infinite, homogeneous medium ($\Omega = \mathbb{R}^3$, boundary conditions are 0 at infinity).

Proposition 2.4.5. *Assume κ^2 is constant and that $\Omega = \mathbb{R}^3$. The Green's function of the Helmholtz equation is*

$$\frac{1}{4\pi} \frac{e^{-\kappa|\mathbf{r}|}}{|\mathbf{r}|} \quad (2.34)$$

with Fourier transform

$$\frac{1}{|\mathbf{w}|^2 + \kappa^2} \quad (2.35)$$

These expressions ascertain the low-pass behaviour of the forward model and its highly inhomogeneous sensitivities. We recall that in the homogeneous case, we have $\kappa = \sqrt{\mu_a/D}$, which relates the exponential decay of the Green's function to the optical coefficients.

2.4.2 Some properties of the forward model

We use the above properties of the Helmholtz operator to derive properties of the forward model \mathcal{F} . We consider the case of a single source-term q to keep the notation simple. Recall that we have

$$\mathcal{F} : (\kappa^2, c) \mapsto \begin{pmatrix} \mathcal{F}_x(\kappa^2) \\ \mathcal{F}_m(\kappa^2, c) \end{pmatrix} = \begin{pmatrix} \mathcal{S}_q \mathcal{H}(\kappa^2; q) \\ \mathcal{S}_q \mathcal{H}(\kappa^2; c u_x) \end{pmatrix}$$

with $u_x = \mathcal{H}(\kappa^2; q)$, and

$$\mathcal{S}_q : u \mapsto \begin{pmatrix} \langle f_1, u \rangle_{L^2(\Omega)} \\ \vdots \\ \langle f_{N_q}, u \rangle_{L^2(\Omega)} \end{pmatrix}$$

It is clear that \mathcal{S}_q is linear and bounded, which leads to the following result

Property 2.4.6. *Let $(\kappa_0^2, c_0) \in P$. The following properties hold :*

- *The mappings $\kappa^2 \mapsto \mathcal{F}_x(\kappa^2)$ and $\kappa^2 \mapsto \mathcal{F}_m(\kappa^2, c_0)$ are Frchet-differentiable. The Frchet-derivatives follow from proposition 2.4.2.*

- The mapping $c \mapsto \mathcal{F}_m(\kappa_0^2, c)$ is linear, bounded and compact.

We now discuss the structure of the linear operator $\mathcal{F}_m(\kappa_0^2, \cdot)$, which will have implications on the numerical implementation. One point of view is to consider \mathcal{F}_m as the cascaded system of Helmholtz equations (2.6) and (2.7). In this case, evaluating \mathcal{F}_m requires first to solve for $u_x = \mathcal{H}(\kappa^2, q)$, then to solve for $u_m = \mathcal{H}(\kappa^2, c u_x)$, and finally to sample the fields using \mathcal{S}_q . We now present another point of view on \mathcal{F}_m . Let us assume a point source located in $\mathbf{r}_s : q(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_s)$. Let us also assume a single point detector located in $\mathbf{r}_d : f(\mathbf{r}) = \delta(\mathbf{r} - \mathbf{r}_d)$. In this case, the sampling operation is simply a point evaluation of the emission field in \mathbf{r}_d . Using definition 2.4.4, we see that $u_x(\mathbf{r}) = g(\mathbf{r}_s, \mathbf{r})$, and the measurement produced by this model is given by

$$\begin{aligned} y &= \int_{\partial\Omega} d\mathbf{r} \delta(\mathbf{r} - \mathbf{r}_d) \int_{\Omega} d\mathbf{r}' g(\mathbf{r}', \mathbf{r}) c(\mathbf{r}') g(\mathbf{r}_s, \mathbf{r}') \\ &= \int_{\partial\Omega} d\mathbf{r}' g(\mathbf{r}', \mathbf{r}_d) c(\mathbf{r}') g(\mathbf{r}_s, \mathbf{r}') \\ &= \int_{\partial\Omega} d\mathbf{r}' g(\mathbf{r}_d, \mathbf{r}') c(\mathbf{r}') g(\mathbf{r}_s, \mathbf{r}') \end{aligned} \quad (2.36)$$

where we use the principle of inverse propagation of light to write $g(\mathbf{r}', \mathbf{r}_d) = g(\mathbf{r}_d, \mathbf{r}')$. Equality (2.36) was observed early on [2, 28]. It reveals the structure of the forward model. Instead of solving cascaded Helmholtz equations, one solves an equation on the source side, one equation on the sampling side, and then performs the inner-product with the fluorophore parameter c . The following proposition generalizes this result for arbitrary source and sampling functions.

Proposition 2.4.7. *Let $\kappa^2 \in P_1$. Let $q \in L^2(\Omega)$ be a source function, and $(f_1, \dots, f_{N_d}) \in (L^2(\Omega))^{N_d}$ be sampling functions. Then, the linear operator $\mathcal{F}_m(\kappa^2, \cdot) : L^2(\Omega) \rightarrow \mathbb{R}^{N_d}$ is given by*

$$c \mapsto \mathcal{F}_m(\kappa^2, c) = \begin{pmatrix} \langle U_1 V, c \rangle_{L^2(\Omega)} \\ \vdots \\ \langle U_{N_d} V, c \rangle_{L^2(\Omega)} \end{pmatrix} \quad (2.37)$$

where $V = \mathcal{H}(\kappa^2; q)$, and $U_i = \mathcal{H}(\kappa^2; f_i)$, $i = 1..N_d$.

Proof : Use the definition of $\mathcal{F}_m(\kappa^2, \cdot)$ and the fact that $\mathcal{H}(\kappa^2; \cdot)$ is self-adjoint

□

Remark 1 : this result extends to boundary source terms and sampling at the expense of additional mathematical technicalities and heavier notation.

Remark 2 : the Frchet derivatives $\frac{\partial \mathcal{F}_x}{\partial \kappa^2}$ and $\frac{\partial \mathcal{F}_m}{\partial \kappa^2}$ have the same structure as \mathcal{F}_m and can be expressed in a similar manner.

2.5 FDOT inverse problem, regularization and variational approach

The FDOT inverse problem is to solve the equation

$$\mathcal{F}(\kappa^2, c) = \tilde{\mathbf{y}} \quad (2.38)$$

which we can specify in terms of excitation and emission measurements as follows

$$\begin{pmatrix} \mathcal{F}_x(\kappa^2) \\ \mathcal{F}_m(\kappa^2, c) \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{y}}_x \\ \tilde{\mathbf{y}}_m \end{pmatrix}$$

The solution $(\hat{\kappa}^2, \hat{c})$ of this equation yields a reconstruction of heterogeneity and fluorophore parameters. We note that the sub-problem $\mathcal{F}_x(\kappa^2) = \tilde{\mathbf{y}}_x$ is the non-linear inverse-problem corresponding to Diffuse Optical Tomography. In FDOT, one is ultimately interested in the fluorophore concentration, and thus in the fluorophore coefficient c , rather than in the heterogeneity parameter κ^2 . We will therefore limit the discussion to the linear inverse-problem $\mathcal{F}_m(\kappa^2, c) = \tilde{\mathbf{y}}_m$ where κ^2 is assumed to be known. One could think of it as being estimated beforehand, for instance. In addition, what we discuss for the linear problem can be extended to the non-linear problem by linearization using proposition 2.4.2.

2.5.1 Ill-posedness

We have seen in 2.4.6 that $\mathcal{F}_m(\kappa^2, \cdot)$ is a compact linear operator. Equation $\mathcal{F}_m(\kappa^2, c) = \tilde{\mathbf{y}}_m$ is therefore a *Fredholm equation of the first kind*. The theory of compact operators and Fredholm equations is well-understood; see for example references [35, 36]. In particular, it is known that compact operators cannot have a

2.5 FDOT inverse problem, regularization and variational approach 29

bounded inverse, which means that the Fredholm equation is ill-posed. Concretely, small perturbations on the measurements can lead to large variations of the inverse (if it exists), rendering meaningless any attempt to use it directly for estimating the unknown parameter.

Of course, in our case, $\mathcal{F}_m : L^2(\Omega) \rightarrow \mathbb{R}^N$ has a non-empty null-space, and is not invertible. Besides, measurement non-idealities may cause the observations $\tilde{\mathbf{y}}_m$ to lie outside of the range of \mathcal{F}_m . A natural way of generalizing the inverse is to consider a least-squares solution (which always exists here because $\text{Range}(\mathcal{F}_m)$ is finite dimensional). This leads to an estimate \hat{c} that minimizes the misfit :

$$\|\tilde{\mathbf{y}}_x - \mathcal{F}_m(\kappa^2, \hat{c})\|_2 \leq \|\tilde{\mathbf{y}}_x - \mathcal{F}_m(\kappa^2, c)\|_2, \quad \forall c \quad (2.39)$$

Uniqueness of that estimate is guaranteed by selecting the element of minimum norm amongst all the minimizers (this additional constraint critical here since \mathcal{F}_m has a non-empty null-space). Interestingly, the operator defined in this way is linear. It is called the *pseudo-inverse*, or *Penrose-Moore generalized inverse* \mathcal{F}_m^\dagger of \mathcal{F}_m .

Unfortunately, the pseudo-inverse performs poorly in practice. The following theorem [37], which is an infinite-dimensional generalization of the *singular values decomposition*, enables one to clarify the situation.

Theorem 2.5.1. *Let H_1 and H_2 be two Hilbert spaces. Let $\mathcal{Q} : H_1 \rightarrow H_2$ be a compact linear operator. There exists a countable set of triples $\{u_i, v_i, s_i\}$ satisfying the following properties :*

1. (v_j) is an orthonormal basis of $(\text{Ker}(\mathcal{Q}))^\perp$,
2. (u_j) is an orthonormal basis of $\text{cl}(\text{Range}(\mathcal{Q}))$,
3. s_j are positive and in non-increasing order : $s_1 \geq s_2 \geq \dots \geq 0$.
4. $\forall j, \mathcal{Q}v_j = s_j u_j$ and $\mathcal{Q}^*u_j = s_j v_j$
5. if $\text{Range}(\mathcal{Q})$ is infinite-dimensional, $s_j \rightarrow 0$.

In addition, \mathcal{Q}^\dagger is defined on $\text{Range}(\mathcal{Q}) + (\text{Range}(\mathcal{Q}))^\perp$ and satisfies

$$\mathcal{Q}^\dagger y = \sum_j \frac{\langle u_j, y \rangle_{H_2}}{s_j} v_j \quad (2.40)$$

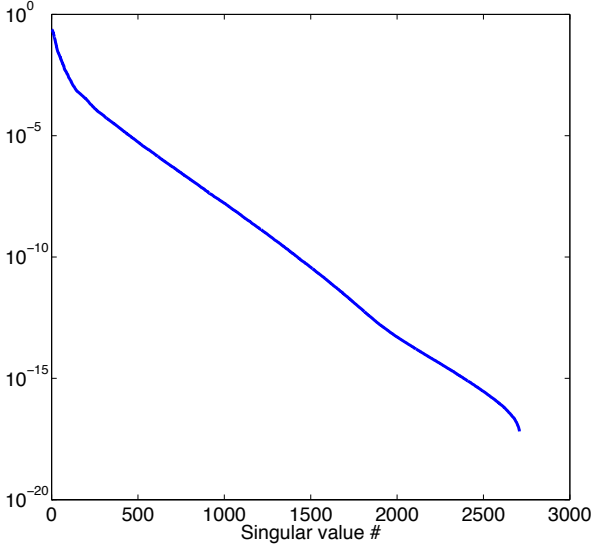


Figure 2.2: Singular values of the discrete forward model for a two dimensional setup.

We note that computing the pseudo-inverse requires inverting arbitrarily small singular values s_j when $\text{Range}(\mathcal{Q})$ is infinite-dimensional (and it is proven that \mathcal{Q}^\dagger cannot be bounded if $\text{Range}(\mathcal{Q})$ is infinite-dimensional). Although $\text{Range}(\mathcal{F}_m)$ is finite-dimensional, we observe the same situation in practice [38]. Singular values of \mathcal{F}_m cluster at 0, which leads to instability of the pseudo-inverse. We understand this intuitively by noting that \mathcal{F}_m is a sampled version of the infinite-dimensional compact operator \mathcal{K} (see Fig 2.1) and inherits its properties as the number of samples increases. We illustrate this effect in Fig 2.2 where we display the singular values of the discrete forward model for a two dimensional setup; in this case we are able to store the forward model matrix in memory.

2.5.2 Regularization, Variational reconstruction approach

We have seen in the previous paragraph that \mathcal{F}_m^\dagger performs poorly because inverting small singular values makes it sensitive to small perturbations of the corresponding components. This phenomenon is exacerbated by the fact that noise (measurement non-idealities) usually sits on the components associated to small singular values. Indeed, it is observed in practice that singular vector corresponding to small singular values exhibit a strong oscillatory behavior akin to noise. Therefore, the exact minimization of the misfit propagates and amplifies noise during the estimation. This is referred to as *overfitting*.

We would like to replace \mathcal{F}_m^\dagger by another mapping that prevents overfitting, but remains close to \mathcal{F}_m^\dagger to preserve the information brought by the forward model. This is done by considering a family of continuous mappings \mathcal{R}_λ , $\lambda > 0$ which approximates \mathcal{F}_m^\dagger in the sense that $\mathcal{R}_\lambda(\mathbf{y}_m) \rightarrow \mathcal{F}_m^\dagger \mathbf{y}_m$, $\lambda \rightarrow 0$. Parameter λ enables one to control the degree of overfitting. We achieve this by introducing a potential function $\Psi(\mathbf{p})$, and defining the cost function

$$J_\lambda(c) = \frac{1}{2} \|\mathbf{y}_m - \mathcal{F}_m c\|_2^2 + \lambda \Psi(c), \quad (2.41)$$

Then, we pose

$$\mathcal{R}_\lambda(\mathbf{y}_m) = \arg \min_c J_\lambda(c) \quad (2.42)$$

For the potential functions Ψ used in practice, one may prove that \mathcal{R}_λ is continuous on \mathbf{y}_m and satisfies the prescribed approximation property. This approach was first introduced by Tikhonov for approximating Fredholm equations of first kind [36]. Tikhonov used $\Psi(c) = \frac{1}{2} \|\mathcal{R}c\|_2^2$, where \mathcal{R} is a suitable linear operator. We call *variational reconstruction approach* an estimation method based on optimizing a cost function of the type of (2.41). The reconstruction algorithms developed on this thesis are based on this formulation. Note that the same framework can be applied to non-linear forward models.

Since \mathcal{F}_m is linear, taking Ψ a convex function results in a convex objective J_λ and is beneficial for numerical algorithms. In addition to numerical considerations, the choice of Ψ reflects the *a priori* information that we want to include in the estimation. In this work, we investigate potential functions inducing *sparsity* in the estimate. Also note that the choice of the data-fitting term is not restricted to least-squares. Other metrics can be used.

The variational approach can be derived from a constrained optimization point of view. Let us assume that we know the noise level ϵ of our measurement system, that is to say,

$$\|\mathbf{y}_m - \mathcal{F}_m \hat{c}\|_2^2 \leq \epsilon^2 \quad (2.43)$$

where \hat{c} is the optimal parameter. Since several candidate solutions may verify this criterion, we use the function Ψ to discriminate. This leads to the constrained optimization problem

$$\arg \min_c \Psi(c) \quad \text{subject to} \quad \|\mathbf{y}_m - \mathcal{F}_m c\|_2^2 \leq \epsilon^2 \quad (2.44)$$

By applying the Kuhn-Tucker conditions, we see that the estimate minimizes a function of the form (2.41), λ being a Lagrange multiplier. We emphasize that λ and ϵ are related through the constraint.

2.5.3 Regularization and statistical inference

Rather than using a deterministic *a priori* in the form of a potential function, one can use statistical prior knowledge about the parameter. Such knowledge is given by a *prior probability distribution*, and employed to construct statistical estimators. The *maximum a posterior* (MAP) estimate is the maximizer of the posterior distribution, given the measurements. It leads to an optimization problem of the form (2.41), and is therefore nothing but a statistical interpretation of the regularization approach presented above. Within the Bayesian statistics framework, one may also consider the *Minimum Mean-Square Estimator (MMSE)*, given by the conditional mean. In practice, computing the MMSE requires performing an integration over a high-dimensional space (as many dimensions as unknown variables), and is therefore intractable. However it is interesting to note that because it is an integral, it is a continuous function of the measurements. It is therefore a good candidate for regularization.

2.6 Remarks

2.6.1 Ill-posedness and uniqueness

In section 2.5 we gave an abstract description of the ill-posedness of the FDOT reconstruction problem, which became apparent with the instability of the direct inversion. A related question is that of the uniqueness of the solution if it exists, independently of stability issues. For the non-linear DOT problem, Arridge et al. exhibit distinct configurations of D and μ_a that result in the same DOT steady-state measurements [29]. Although no such result is known for FDOT at the moment, Herve et al. conducted a similar study, by constructing synthetic phantoms with different compositions for which steady-state measurements are numerically indistinguishable [30].

Given these non-uniqueness results, the question remains open to find a set of conditions on the parameters under which uniqueness is restored. Harrach partially answered this question for the steady-state DOT problem by producing such conditions [39]. We note, however, that Harrach's result holds for ideal measurements and does not guarantee stable recovery. Nevertheless, it gives insights for the design of regularization policies.

2.6.2 Partial differential equations and coefficient identification

In this chapter, we have described FDOT reconstruction abstractly as an inverse problem, in the context of Fredholm equations. However, it belongs to a more restricted class of inverse problems, namely *coefficient identification* in a *partial differential equation* (PDE). The underlying PDE defines the structure of the problem, and finer properties may be derived. A number of mathematical developments were made on this topic, mainly covered by the fields of *optimal control* and *PDE constrained optimization* (more recent and very active). There exists an abundant literature in these fields, see for instance [40].

2.6.3 Optimize-discretize vs. discretize-optimize

The variational approach presented here formulates reconstruction as an optimization problem (and similarly for optimal control and PDE constrained optimization).

However, this optimization problem was formulated in function spaces. Actual reconstruction, on the other hand, will be carried out in a computer on a discrete version of the model. Therefore, one has to decide when to discretize the problem.

In the optimize-discretize paradigm, the rule is to derive optimality conditions in the continuous domain and discretize later. This is essentially the mathematician's approach to the problem. Staying in the continuous domain enables one to prove results, and more importantly, to employ *adaptive* discretization schemes. Alternatively, one could start by discretizing, and deal with a discrete problem from then on. This is termed discretize-optimize. As a rule of thumb, provided that the same discretization scheme is employed, both approaches are rigorously equivalent as far as algorithms are concerned.

With the exception of a few works [41], most reconstruction procedures for FDOT proposed in the literature are based on the discretize-optimize paradigm.

2.6.4 Other regularization methods

Before concluding this part on regularization methods, we shortly mention truncated iterative methods. Since the pseudo-inverse is unstable and leads to overfitting, the idea is to stop an iterative method that computes the pseudo-inverse before convergence, in such a way that the model is fitted but not the noise. We are not using these methods in this work and refer to [42, 36] for further details.

Chapter 3

Numerical model for FDOT

THIS chapter deals with the design and implementation of an efficient discrete model for FDOT. In this regard we focus on the three following aspects. First, the discretization of the propagation equation. Then, the design of a fast algorithm that evaluates the forward model based on the discrete propagation operator. And finally, the construction of appropriate discrete regularization potentials which are used in the inversion algorithm.

We start from the observation that the parameters estimated in the FDOT reconstruction are the coefficients governing the propagation equation (the source term, and the heterogeneity parameter). Because our regularization policies impose the shape the reconstructed coefficients, it is reasonable to solve the propagation equation using a method that is compatible with the properties we try to prescribe for the coefficients. Sparsity-inducing regularization typically results in discontinuous coefficients. Most works on FDOT rely on the finite elements method (FEM) for discretizing the diffusion equation. Because the FEM is based on polynomial interpolation, it is not the best option for dealing with discontinuous coefficients; it does not lead to a satisfactory solution in our case. Here, we propose a novel *Finite Volumes Method* (FVM) for solving the diffusion equation in the context of FDOT. The FVM is well suited to diffusion equations with discontinuous coefficients. Surprisingly, the FVM has hardly been used in FDOT. We are only aware of single work that describes such an approach [43]. Our method has the advantage of being able to handle arbitrary tetrahedral meshes, without additional geometrical

computations (such as Vorono cells, for instance).

We provide implementation guidelines that result in a fast discrete propagation model. By combining this efficient propagation model, and the structure of the FDOT forward model studied in Chapter 2, we obtain fast algorithms for computing the forward model. This is essential for iterative reconstruction procedures where the forward model needs to be evaluated multiple times. In addition, our algorithms are general and can handle arbitrary source and sampling functions. It is the method that we implemented, and that we used for our reconstruction experiments presented in the latter chapters.

For the reconstruction procedure, it is equally important to design adequate regularization potentials. With this in mind, we propose a consistent treatment of the regularization, following the same discretization principles as those that are used for the propagation equation. In particular, we derive a consistent discretization of the total variation regularization potential.

3.1 Notation

We use lowercase bold font to denote vectors in a finite-dimensional space; \mathbf{u} for instance. By convention, we assume column vectors. Similarly, functions defined between finite-dimensional spaces are denoted with uppercase bold font; $\mathbf{F}(\mathbf{u})$ for instance. For linear functions (matrices), we omit the brackets; as an example, $\mathbf{F}\mathbf{u}$. We use the notation $\mathbf{u} \odot \mathbf{v}$ for the vector resulting of the component-wise product of vectors \mathbf{u} and \mathbf{v} . For a vector \mathbf{u} , we denote $\text{diag}(\mathbf{u})$ the diagonal matrix with diagonal given by \mathbf{u} . Note that $\text{diag}(\mathbf{u})\mathbf{v} = \mathbf{u} \odot \mathbf{v}$.

3.2 The discrete variational reconstruction method

In the variational framework, our reconstruction procedure is to minimize a regularized functional of the form

$$J_\lambda(\mathbf{p}) = \frac{1}{2} \|\tilde{\mathbf{y}} - \mathcal{F}\mathbf{p}\|_2^2 + \lambda\Psi(\mathbf{p}) \quad (3.1)$$

where $\mathbf{p} = (\kappa^2, c)$ is the unknown parameter vector. We employ the discretize-optimize paradigm, meaning that we first discretize the functional J_λ , and deal with the resulting discrete problem. This implies discretizing the spaces involved

in the inverse-problem (see Fig 2.1); namely, the parameter space P , the solution space U , the source space Q and the sampling space S . The discrete source and sampling spaces need to be adapted to the acquisition device, whereas we can choose the discrete parameter and solution spaces. This is done by constructing discretization grids and related finite-dimensional approximation spaces $P_h \subset P$ and $U_h \subset U$, where h denotes the characteristic size of the grid. In a similar way we define the space $Q_h \subset Q$. Then, we obtain the discrete propagation operator $\mathbf{K} : P_h \times Q_h \rightarrow U_h$ by discretizing the Helmholtz operator (see (2.15)). Since the Helmholtz operator is central to the propagation model, its discretization is a key step. Here, discretization refers not only to the construction of approximation spaces, but also to the computational algorithm that evaluates numerically the discrete Helmholtz operator. Finally the discrete forward model \mathbf{F} follows by composing \mathbf{K} with the discrete sampling operator \mathbf{S} . A discrete regularization potential needs to be constructed as well; it is a function $P_h \rightarrow \mathbb{R}$.

We recall that \mathcal{F} has the form

$$\mathcal{F} : (\kappa^2, c) \mapsto \begin{pmatrix} \mathcal{F}_x(\kappa^2) \\ \mathcal{F}_m(\kappa^2, c) \end{pmatrix} \quad (3.2)$$

and that the observation is $\mathbf{y} = (\mathbf{y}_x, \mathbf{y}_m)$. Therefore, the functional J_λ can be expressed as

$$J_\lambda(\kappa^2, c) = \frac{1}{2} \|\mathbf{y}_x - \mathcal{F}_x(\kappa^2)\|_2^2 + \frac{1}{2} \|\mathbf{y}_m - \mathcal{F}_m(\kappa^2, c)\|_2^2 + \lambda \Psi(\kappa^2, c) \quad (3.3)$$

Although (κ^2, c) can be estimated jointly using (3.3), we choose to decouple the problem into the non-linear estimation of κ^2 using \mathcal{F}_x and \mathbf{y}_x alone, and the linear estimation of c using \mathcal{F}_m and \mathbf{y}_m alone (considering κ^2 fixed). We end up with the two estimation sub-problems

$$\hat{\kappa}^2 = \arg \min_{\kappa^2} J_x(\kappa^2) \quad \text{where} \quad J_x(\kappa^2) = \frac{1}{2} \|\mathbf{y}_x - \mathcal{F}_x(\kappa^2)\|_2^2 + \lambda \Psi_x(\kappa^2), \quad (3.4)$$

$$\hat{c} = \arg \min_c J_m(c) \quad \text{where} \quad J_m(c) = \frac{1}{2} \|\mathbf{y}_m - \mathcal{F}_m \beta\|_2^2 + \lambda \Psi_m(c), \quad (3.5)$$

that we solve sequentially. Note that problem (3.4) corresponds to DOT.

Using the notation introduced above, we have to solve the corresponding finite-dimensional optimization problems

$$\hat{\mathbf{k}} = \arg \min_{\mathbf{k}} \frac{1}{2} \|\mathbf{y}_x - \mathbf{F}_x(\mathbf{k})\|_2^2 + \lambda \Psi_x(\mathbf{k}), \quad (3.6)$$

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c}} \frac{1}{2} \|\mathbf{y}_m - \mathbf{F}_m \mathbf{c}\|_2^2 + \lambda \Psi_m(\mathbf{c}), \quad (3.7)$$

where \mathbf{k} (resp. \mathbf{c}) is the discrete counterpart of κ^2 (resp. c).

3.3 Discretization grid

Discretization grids need to be adapted to the objects they are meant to capture. The diffusion approximation is not valid at scales smaller than μ'_s , thus, the grid chosen for Q_h , P_h and U_h need not be finer. Besides, diffuse fields do not exhibit strong variations, and therefore rather coarse grids suffice. Since the free-space Green's function is $\exp(-\kappa r)/r$, the relative variation of the solution over a distance dr is approximately κdr . As a rule of thumb, by choosing a grid size $\propto \kappa^{-1}$ for U_h , we get a grid adapted to the solution variations.

Although an ideal discretization would use different grids for Q_h , P_h and U_h , we used the same for all spaces, as it makes the implementation simpler. Moreover, the acquisition system employed in this work returns a mesh of the imaged object along with the measurements. That mesh specifies the geometry of the object and is satisfactory as far as grid-size is concerned. It is therefore natural to use it for discretizing. This mesh is composed of tetrahedra only.

The single grid employed in this work is denoted $\mathcal{T} = (\mathcal{T}_h, \mathcal{T}_f)$. It is an arbitrary tetrahedral grid in the sense that there are no constraints on the aspect ratio of the tetrahedra. \mathcal{T}_h is the set of tetrahedra, called *cells*, making up the mesh. \mathcal{T}_f is the set of triangles, called *faces*, forming the boundary of the cells. \mathcal{T}_f encompasses all triangles, whether they are *inner faces* (inside Ω), or *outer faces* (on $\partial\Omega$). Each inner face $f \in \mathcal{T}_f$ belongs to two tetrahedra, but it is counted only once, and has an arbitrary orientation defined by its normal vector \mathbf{n}_f . For outer faces, it is assumed that \mathbf{n}_f points outside Ω .

Given this grid we define the space V_h that consists of the functions that are constant on every cell (piecewise-constant functions). For P_h , we will use elements of V_h that are bounded from below by a strictly positive constant. V_h is interesting

if we wish to apply regularization potentials favoring parameter distributions which may have discontinuities (for instance piecewise-constant functions). We also have $V_h \subset L^2(\Omega)$, so that V_h is also used for Q_h and U_h .

3.4 Discrete Helmholtz operator

The Helmholtz operator corresponds to solving an elliptic partial differential equation. A number of discretization methods are available and choosing the most appropriate one is not an easy task. In this work we chose the *Finite-Volumes Method* (FVM). It is economical in computations and accurate for diffusion equations because it enforces the continuity equation cell-wise. In addition, the FVM is based on the discretization space V_h (functions constant on every cell). As mentioned above, the Helmholtz equation is discretized on the grid \mathcal{T} .

3.4.1 A Finite-Volumes Method for the Helmholtz equation

The scheme we present here is a variation of a general method call the *Mimetic Finite Difference Method* due to Hyman et al. [44]. We introduce a difference in the handling of the boundary conditions. The Mimetic Finite Difference Method is very similar to a FVM, but with an original way of defining the flux operator.

The diffusion equation is reformulated as a first-order system :

$$\begin{cases} \nabla \cdot J + \kappa^2 u = q, & \Omega \\ J = -\nabla u, & \Omega \\ u + \beta \mathbf{n} \cdot \nabla u = 0, & \partial\Omega \end{cases} \quad (3.8)$$

We recall that the first equation is a *continuity equation* (similar to a mass balance) : at each point the change in u is due to a source term q , absorption in the medium m , and incoming flux J . The second equation describes the *flux operator* (Fourier's law) relating the flux to the gradient.

Equation (3.8) involves scalar-valued and vector-valued functions. We discretize scalar-valued functions at the cell centers (equivalent to using V_h for discretization space), and vector-valued functions at the face centers. The scheme is illustrated in two dimensions in Fig 3.1. Specifically, our discrete unknowns are u_e , $e \in \mathcal{T}_h$ the values of u at the cell centers, and $J_f = J \cdot \mathbf{n}_f$, $f \in \mathcal{T}_f$ the fluxes at the

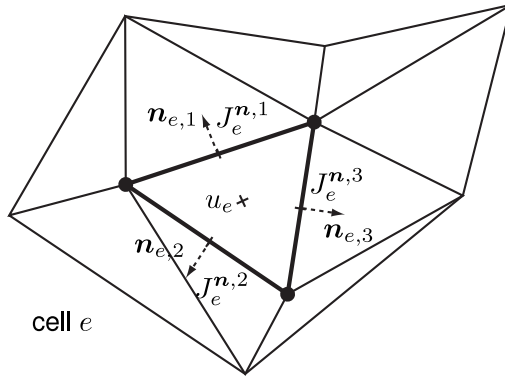


Figure 3.1: Representation of an arbitrary tetrahedral discretization in two dimensions. Discretization of scalar and vector variables in cell e : u_e is the value of the scalar variable u at the cell center, $J_e^{n,i}$, $i = 1..3$ are the normal fluxes of the vector variable J at the face centers.

face centers in the direction normal to the face. We also employ the notation $J_e^{n,i}$, $i \in [1..4]$, $e \in \mathcal{T}_h$ to denote fluxes at the face centers of cell e in the direction normal to the face, and oriented outward with respect to the cell. It is understood that these values are the same as J_f for the corresponding face, up to a sign change. Also, κ^2 and the source term q are discretized at the cell level.

Integrating the continuity equation over each cell and using the divergence theorem yields

$$\int_{\partial e} (J \cdot \mathbf{n}) d\sigma + \int_e \kappa^2 u d\mathbf{r} = \int_e q d\mathbf{r}, \quad e \in \mathcal{T}_h \quad (3.9)$$

$$(3.10)$$

We obtain a consistent discretization of these integrals using our discrete unknowns

$$(J_e^{n,1} a_e^1 + J_e^{n,2} a_e^2 + J_e^{n,3} a_e^3 + J_e^{n,4} a_e^4) + \kappa_e^2 u_e V_e = q_e V_e, \quad e \in \mathcal{T}_h \quad (3.11)$$

where V_e is the volume of cell e , and a_e^i , $i = 1..4$ the areas of its faces. Conservation is imposed cell-wise, which is one of the main advantages of the finite-volumes methods compared to other methods when dealing with continuity equations.

It remains to discretize the flux operator to close the system of equation. Let u be the solution of the equation. Then, $\forall J$ we have

$$\int_{\Omega} \nabla \cdot J u d\mathbf{r} = \int_{\partial\Omega} (J \cdot \mathbf{n}) u d\sigma - \int_{\Omega} J \cdot \nabla u d\mathbf{r} \quad (3.12)$$

$$= \int_{\partial\Omega} \beta (J \cdot \mathbf{n}) (-\nabla u \cdot \mathbf{n}) d\sigma + \int_{\Omega} J \cdot (-\nabla u) d\mathbf{r} \quad (3.13)$$

where we used the boundary condition in the last equation. This is where the method presented here differs from [44]. Here we do not to specify the values of u on $\partial\Omega$ (which translate to additional variables on outer faces). Otherwise, the same reasoning is employed.

Now we define the two inner products

$$\langle f, g \rangle_1 = \int_{\Omega} f g d\mathbf{r} \quad (3.14)$$

$$\langle F, G \rangle_2 = \int_{\partial\Omega} \beta (F \cdot \mathbf{n}) (G \cdot \mathbf{n}) ds + \int_{\Omega} F \cdot G d\mathbf{r} \quad (3.15)$$

Note that $\langle \cdot, \cdot \rangle_1$ applies to scalar-valued functions, whereas $\langle \cdot, \cdot \rangle_2$ applies to vector-valued functions. With this definition, $\nabla \cdot$ and $-\nabla$ are adjoint of each other. We are going to discretize the divergence $\nabla \cdot$, as well as the two inner-products $\langle \cdot, \cdot \rangle_1$ and $\langle \cdot, \cdot \rangle_2$. Then, similarly to the continuous case, we will define the discrete flux operator as the adjoint of the discrete divergence operator.

Using V_e and a_e^i , $e \in \mathcal{T}$, integrals $\int_{\Omega} f g \, d\mathbf{r}$ and $\int_{\partial\Omega} \beta (F \cdot \mathbf{n})(G \cdot \mathbf{n}) \, d\sigma$ are discretized straightforwardly by the midpoint formula. For $\int_{\Omega} F \cdot G \, d\mathbf{r}$ a little bit of work needs to be done because our discrete vector-valued functions are defined at face centers and not cell centers. Since each summit of a tetrahedral cell is common to three faces, the vector values at face centers determine a unique vector at that summit. Then, by linear interpolation, we obtain vector values at cell centers. This completes the discretization of the two inner-products. In matrix form, the discrete inner-products are denoted \mathbf{W}_1 and \mathbf{W}_2 . Note that \mathbf{W}_1 is a diagonal matrix, while \mathbf{W}_2 is a sparse symmetric matrix but not diagonal. In a practical implementation these matrices are computed once and for all.

Finally, we deal with the discrete divergence operator. It maps vector face centered variables to scalar cell centered variables. The well-know gauss theorem applied to cell $e \in \mathcal{T}_h$ reads

$$\int_e \nabla \cdot J \, d\mathbf{r} = \int_{\partial e} J \cdot \mathbf{n} \, d\sigma$$

which yields the desired discrete divergence operator, again by applying the midpoint formula to both sides of this equation. Similarly to \mathbf{W}_1 and \mathbf{W}_2 , the discrete divergence operator is computed once and for all and stored in memory. Since it is a local operator, it results in a sparse matrix.

We can describe the discrete operators so obtained with a matrix formulation. Let us denote by \mathbf{G} the discrete flux operator, and by \mathbf{div} the discrete divergence. By definition of the adjoint we have :

$$(\mathbf{G}\mathbf{u})^\top \mathbf{W}_2 \mathbf{f} = \mathbf{u}^\top \mathbf{W}_1 (\mathbf{div} \mathbf{f}), \quad \forall \mathbf{u}, \mathbf{f}$$

where \mathbf{u} is a cell centered scalar variable, and \mathbf{f} is a face centered vector variable. This leads to the following expression of the discrete gradient operator

$$\mathbf{G} = \mathbf{W}_2^{-1} \mathbf{div}^\top \mathbf{W}_1 \tag{3.16}$$

Interestingly \mathbf{G} is not a local operator in this formulation because \mathbf{W}_2^{-1} is not sparse. Since \mathbf{W}_2 can become ill-conditioned for arbitrary meshes, some care has to be taken for solving the resulting system.

Let \mathbf{k} (resp. \mathbf{q}) be the cell centered variable discretizing the heterogeneity coefficient κ^2 (resp. the source term q). Let \mathbf{u} be the cell centered variable discretizing the fluence rate u and \mathbf{j} the face centered variable discretizing the flux J . Then, our discretization of (3.8) reads

$$\mathbf{W}_1 \operatorname{div} \mathbf{j} + \mathbf{W}_1 \operatorname{diag}(\mathbf{k}) \mathbf{u} = \mathbf{W}_1 \mathbf{q} \quad (3.17)$$

$$\mathbf{j} = \mathbf{G} \mathbf{u} \quad (3.18)$$

The discrete Helmholtz operator inputs \mathbf{q} , \mathbf{k} and outputs \mathbf{u} , which requires the solution of (3.17), (3.18). In the sequel it will be convenient to have \mathbf{H} denote the **inverse** of the discrete Helmholtz operator. With this notation system (3.17), (3.18) is expressed in the compact form

$$\mathbf{H} \mathbf{u} = \mathbf{q} \quad (3.19)$$

Dependency on \mathbf{k} is implicit here, and we expressed only the linear dependency of \mathbf{u} and \mathbf{q} . Note that matrix \mathbf{H} is symmetric.

3.4.2 Numerical implementation

The linear equations (3.17) and (3.18) defining the discrete Helmholtz operator can be written as the following saddle-point system

$$\begin{pmatrix} \mathbf{W}_1 \operatorname{diag}(\mathbf{k}) & \mathbf{W}_1 \operatorname{div} \\ \operatorname{div}^\top \mathbf{W}_1 & -\mathbf{W}_2 \end{pmatrix} \begin{pmatrix} \mathbf{u} \\ \mathbf{j} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1 \mathbf{q} \\ 0 \end{pmatrix} \quad (3.20)$$

Large and sparse linear systems of equations such as (3.20) are usually handled using iterative methods. The method of choice is the *pre-conditioned conjugate gradient* (PCG) when a good pre-conditioner is available. In our case, we do not have a pre-conditioner for (3.20), so trying to solve it directly may lead to poor performance. Another approach is to use the fact that $\mathbf{W}_1 \operatorname{diag}(\mathbf{k})$, and \mathbf{W}_2 are invertible and to perform Gauss elimination either on \mathbf{j} or \mathbf{u} . The resulting systems are called the Schur complements. Eliminating \mathbf{j} leads to

$$(\mathbf{W}_1 \operatorname{div} \mathbf{W}_2^{-1} \operatorname{div}^\top \mathbf{W}_1 + \mathbf{W}_1 \operatorname{diag}(\mathbf{k})) \mathbf{u} = \mathbf{W}_1 \mathbf{q} \quad (3.21)$$

Solving this equation yields \mathbf{u} directly, but this approach is impractical because \mathbf{W}_2^{-1} is not known explicitly, and not sparse. It would result in nested iterations. Instead, let us try to eliminate \mathbf{u} . We obtain

$$(\mathbf{W}_2 + \mathbf{div}^\top \mathbf{W}_1 \text{diag}(\mathbf{k}^{-1}) \mathbf{div}) \mathbf{j} = \mathbf{div}^\top \mathbf{W}_1 \text{diag}(\mathbf{k}^{-1}) \mathbf{q} \quad (3.22)$$

We know explicitly the matrix $\mathbf{W}_2 + \mathbf{div}^\top \mathbf{W}_1 \text{diag}(\mathbf{k}^{-1}) \mathbf{div}$ because $\mathbf{D}_\mathbf{k}$ is diagonal. In addition, it is a sparse matrix. Based on this observation, we have an efficient method to compute \mathbf{j} , which easily yields \mathbf{u} .

3.4.3 Multiple right-hand-sides

If the linear system (3.22) has to be solved one time only, iterative methods such as PCG are interesting. However, concrete imaging problems involves multiple source and the Helmholtz operator needs to be evaluated as many times (see for instance Algorithms 1 to 5). As a consequence, (3.22) has to be solved for multiple right-hand-sides. In this case, iterative methods become too costly. Recent development of direct linear solvers allow for a better approach to this multiple right-hand-sides problem. The idea is to compute and store a sparse factorization of \mathbf{H} , and to use a parallel direct solver for the system. Since the factorization is computed only once, this solution is computationally efficient. In this work we used the MUMPS implementation [45, 46].

3.4.4 Accuracy of the Finite-Volumes Method

A second-order convergence rate was observed for the Mimetic Finite Difference Method [47]. It is worth noting that this rates applies to convergence in $L^2(\Omega)$, and $L^\infty(\Omega)$. For our implementation, we verified experimentally second-order convergence in a 3D case with right-hand-sides presenting various oscillatory behaviours.

3.4.5 Sensitivity relations

Here we give the discrete counterpart of the result in (2.4.2). This yields a closed-form expression for the gradient of the discrete forward model with respect to the heterogeneity parameter. It is useful for optimization algorithms. Let us differen-

tiated with respect to \mathbf{u} , \mathbf{k} and \mathbf{q} . We have

$$(\mathbf{W}_1 \operatorname{div} \mathbf{W}_2^{-1} \operatorname{div}^\top \mathbf{W}_1 + \mathbf{W}_1 \operatorname{diag}(\mathbf{k})) \, d\mathbf{u} + \mathbf{W}_1 \operatorname{diag}(\mathbf{u}) \, d\mathbf{k} - \mathbf{W}_1 \, d\mathbf{q} = 0 \quad (3.23)$$

and therefore, using notation (3.19),

$$\mathbf{g} = \frac{\partial \mathbf{u}}{\partial \mathbf{k}} \mathbf{f} \quad \Leftrightarrow \quad \mathbf{H} \mathbf{g} = -\mathbf{u} \odot \mathbf{f} \quad (3.24)$$

which means that $\frac{\partial \mathbf{u}}{\partial \mathbf{k}}$ is the discrete Helmholtz operator composed with the point-wise multiplication by $-\mathbf{u}$, similarly to what we have in the continuous case.

3.5 Discrete forward model

3.5.1 Sampling

Similarly to the continuous case where the sampling was obtained through inner-products with sampling functions, the sampling of discrete field \mathbf{u} is computed by taking a number of inner products with vectors $\mathbf{s} \in V_h : \mathbf{y} = \mathbf{s}^\top \mathbf{u}$.

3.5.2 Structure of the discrete forward model

Here we express the structure of the fluorescence component of the forward model (\mathbf{F}_m) and give the discrete counterpart of proposition (2.4.7). Two possible implementations of the forward model will follow. Let us consider the case of a single source \mathbf{q} and a single sampling vector \mathbf{s} . Given expression (2.21) of the forward model, computing $\mathbf{F}_m \mathbf{c}$ is naturally done as follows

1. solve $\mathbf{H} \mathbf{u}_x = \mathbf{q}$
2. solve $\mathbf{H} \mathbf{u}_m = \mathbf{u}_x \odot \mathbf{c}$
3. compute $\mathbf{y} = \mathbf{s}^\top \mathbf{u}_m$

Thus we have

$$\mathbf{y} = \mathbf{s}^\top (\mathbf{H}^{-1} (\mathbf{c} \odot \mathbf{H}^{-1} \mathbf{q})) \quad (3.25)$$

$$= (\mathbf{H}^{-1} \mathbf{s})^\top (\mathbf{c} \odot \mathbf{H}^{-1} \mathbf{q}) \quad (3.26)$$

$$= (\mathbf{H}^{-1} \mathbf{s} \odot \mathbf{H}^{-1} \mathbf{q})^\top \mathbf{c} \quad (3.27)$$

where we used the symmetry of \mathbf{H} . Based on this observation, an alternative approach for evaluating $\mathbf{F}_m \mathbf{c}$ is

1. solve $\mathbf{H} \mathbf{u}_q = \mathbf{q}$
2. solve $\mathbf{H} \mathbf{u}_s = \mathbf{s}$
3. compute $\mathbf{y} = (\mathbf{u}_s \odot \mathbf{u}_q)^\top \mathbf{c}$

The advantage of this approach over the direct approach is that Steps 1 and 2 can be precomputed, and thus, a repeated evaluation of \mathbf{F}_m would only require to compute inner-product in step 3, whereas the direct approach requires to solve the linear system in step 2 every time.

The excitation part of the forward model does not exhibit a particular structure. Expression (2.21) gives a direct procedure to evaluate $\mathbf{F}_x(\mathbf{k})$:

1. assemble matrix $\mathbf{H}(\mathbf{k})$
2. solve $\mathbf{H}(\mathbf{k}) \mathbf{u}_x = \mathbf{q}$
3. compute $\mathbf{y} = \mathbf{s}^\top \mathbf{u}_x$

However, proposition (2.4.2) and equation (3.24) show that $\frac{\partial \mathbf{F}_x}{\partial \mathbf{k}}$ has the same structure as \mathbf{F}_m . Thus the two implementation choice presented above can be employed to compute the derivative $\frac{\partial \mathbf{F}_x}{\partial \mathbf{k}}$ which is required in optimization algorithms.

3.5.3 Multiple sources and implementation

As in the continuous case, we specify the notation employed for multiple sources in a tomographic setup. We call N_s the number of sources, and \mathbf{q}_i , $i = 1..N_s$ the source vectors. To each source \mathbf{q}_i , we associate N_d^i sampling vectors \mathbf{s}_j^i , $j = 1..N_d^i$.

Algorithm 1 Forward operator $(\mathbf{F}_x(\mathbf{k}), \mathbf{F}_m \mathbf{b})$, direct method.

Input : \mathbf{k} , \mathbf{c} , $(\mathbf{q}_i)_{i=1..N_s}$, $(\mathbf{S}_i)_{i=1..N_s}$

Output : $\mathbf{F}_x(\mathbf{k})$ and
 $\mathbf{F}_m \mathbf{c}$

$\mathbf{H} \leftarrow \mathbf{H}(\mathbf{k})$

for $i = 1 \rightarrow N_s$ **do**

$\mathbf{u}_x^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{q}_i$

$\mathbf{y}_x^i \leftarrow \mathbf{S}^i \mathbf{u}_x^i$

$\mathbf{u}_m^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{u}_x^i \odot \mathbf{c}$

$\mathbf{y}_m^i \leftarrow \mathbf{S}^i \mathbf{u}_m^i$

end for

$\mathbf{F}_x(\mathbf{k}) \leftarrow ((\mathbf{y}_x^1)^\top, \dots, (\mathbf{y}_x^{N_s})^\top)^\top$

$\mathbf{F}_m \mathbf{c} \leftarrow ((\mathbf{y}_m^1)^\top, \dots, (\mathbf{y}_m^{N_s})^\top)^\top$

Note : The for loop can be replaced by a parallel implementation.

Algorithm 2 Backward operator \mathbf{F}_m^\top , direct method.

Input : \mathbf{H} , \mathbf{y} , $(\mathbf{q}_i)_{i=1..N_s}$, $(\mathbf{S}_i)_{i=1..N_s}$

Output : $\mathbf{v} = \mathbf{F}_m^\top \mathbf{y}$

$\mathbf{v} \leftarrow 0$

for $i = 1 \rightarrow N_s$ **do**

$\mathbf{u}_x^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{q}_i$

$\mathbf{u}_m^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = (\mathbf{S}^i)^\top \mathbf{y}^i$

$\mathbf{v} \leftarrow \mathbf{v} + \mathbf{u}_x^i \odot \mathbf{u}_m^i$

end for

Note 1 : Vectors $(\mathbf{u}_x^i)_{i=1..N_s}$ can be precomputed.

Note 2 : The for loop can be replaced by a parallel implementation.

Algorithm 3 Forward-backward operator $\mathbf{F}_m^\top \mathbf{F}_m$, direct method.

Input : \mathbf{H} , \mathbf{c} , $(\mathbf{q}_i)_{i=1..N_s}$, $(\mathbf{S}_i)_{i=1..N_s}$

Output : $\mathbf{v} = \mathbf{F}_m^\top \mathbf{F}_m \mathbf{c}$

$\mathbf{v} \leftarrow 0$

for $i = 1 \rightarrow N_s$ **do**

$\mathbf{u}_x^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{q}_i$

$\mathbf{u}_m^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{u}_x^i \odot \mathbf{c}$

$\mathbf{u}'_m \leftarrow$ solution of $\mathbf{H} \mathbf{u} = (\mathbf{S}^i)^\top \mathbf{S}^i \mathbf{u}_m^i$

$\mathbf{v} \leftarrow \mathbf{v} + \mathbf{u}_x^i \odot \mathbf{u}'_m$

end for

Note 1 : Vectors $(\mathbf{u}_x^i)_{i=1..N_s}$ can be precomputed.

Note 2 : The for loop can be replaced by a parallel implementation.

Algorithm 4 Forward operator \mathbf{F}_m in sampling form.

Input : \mathbf{H} , \mathbf{c} , $(\mathbf{q}_i)_{i=1..N_s}$, $(\mathbf{S}_i)_{i=1..N_d}$

Output : $\mathbf{y} = \mathbf{F}_m \mathbf{b}$

// Precompute

for $i = 1 \rightarrow N_s$ **do**

$\mathbf{u}_x^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{q}_i$

end for

for $i = 1 \rightarrow N_d$ **do**

$\mathbf{u}_m^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{s}_i$

end for

//

for all valid source-sampling vector combination (i, j) **do**

$\mathbf{y}^{(i,j)} \leftarrow (\mathbf{u}_x^i \odot \mathbf{u}_m^j)^\top \mathbf{c}$

end for

Note 1 : All for loops can be replaced by a parallel implementation.

Algorithm 5 Backward operator \mathbf{F}_m^\top in sampling form.

Input : \mathbf{H} , \mathbf{y} , $(\mathbf{q}_i)_{i=1..N_s}$, $(\mathbf{s}_i)_{i=1..N_d}$

Output : $\mathbf{v} = \mathbf{F}_m^\top \mathbf{y}$

// Precompute

$\mathbf{H} \leftarrow \mathbf{S}_1 \operatorname{div} \mathbf{S}_2^{-1} \operatorname{div}^\top \mathbf{S}_1 + \mathbf{S}_1 \mathbf{D}_k$

for $i = 1 \rightarrow N_s$ **do**

$\mathbf{u}_x^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{q}_i$

end for

for $i = 1 \rightarrow N_d$ **do**

$\mathbf{u}_m^i \leftarrow$ solution of $\mathbf{H} \mathbf{u} = \mathbf{s}_i$

end for

//

$\mathbf{v} \leftarrow 0$

for $i = 1 \rightarrow N$ **do**

for all valid source-sampling vector combination (k, l) **do**

$v_i \leftarrow v_i + (\mathbf{u}_x^k)_i \mathbf{y}^{(k,l)} (\mathbf{u}_m^l)_i$

end for

end for

Note 1 : N denotes the number of cells, $\mathbf{v} = (v_1, \dots, v_N)^\top$.

Note 2 : All for loops can be replaced by a parallel implementation.

These vectors are stacked in a sampling matrix \mathbf{S}_i :

$$\mathbf{S}^i = \begin{pmatrix} (\mathbf{s}_1^i)^\top \\ \vdots \\ (\mathbf{s}_{N_d^i}^i)^\top \end{pmatrix} \quad (3.28)$$

As an alternative notation, we pose $N_d = \sum_{i=1}^{N_s} N_d^i$ the total number of sampling vectors, and enumerate them \mathbf{s}_j , $j = 1..N_d$. The partial forward models for source i are expressed as follows

$$\mathbf{y}_x^i = \mathbf{F}_x^i(\mathbf{k}) = \mathbf{S}^i \mathbf{u}_x^i \quad \text{where} \quad \mathbf{H} \mathbf{u}_x^i = \mathbf{q}_i \quad (3.29)$$

$$\mathbf{y}_m^i = \mathbf{F}_m^i \mathbf{c} = \mathbf{S}^i \mathbf{u}_m^i \quad \text{where} \quad \mathbf{H} \mathbf{u}_m^i = \mathbf{u}_x^i \odot \mathbf{c} \quad (3.30)$$

By stacking these partial forward models, we obtain

$$\mathbf{F}_x(\mathbf{k}) = \begin{pmatrix} \mathbf{F}_x^1(\mathbf{k}) \\ \vdots \\ \mathbf{F}_x^{N_s}(\mathbf{k}) \end{pmatrix} \quad \text{and} \quad \mathbf{F}_m = \begin{pmatrix} \mathbf{F}_m^1 \\ \vdots \\ \mathbf{F}_m^{N_s} \end{pmatrix}. \quad (3.31)$$

We assume that the data vectors \mathbf{y}_x and \mathbf{y}_m are obtained by a similar stacking of partial data vectors for each source. In \mathbf{y}_x (resp. \mathbf{y}_m) each component corresponds to a combination source-sampling vector. We denote $\mathbf{y}_x^{(i,j)}$ (resp. $\mathbf{y}_m^{(i,j)}$) the component for source i and sampling vector j (provided this combination is valid).

In this setting, the data consistency terms are expressed

$$d_x(\mathbf{k}) = \frac{1}{2} \|\mathbf{y}_x - \mathbf{F}_x(\mathbf{k})\|_2^2 = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{y}_x^i - \mathbf{F}_x^i(\mathbf{k})\|_2^2 \quad (3.32)$$

$$d_m(\mathbf{b}) = \frac{1}{2} \|\mathbf{y}_m - \mathbf{F}_m \mathbf{c}\|_2^2 = \frac{1}{2} \sum_{i=1}^{N_s} \|\mathbf{y}_m^i - \mathbf{F}_m^i \mathbf{c}\|_2^2 \quad (3.33)$$

Most optimization algorithms require the gradient of d_x and d_m , either for computing an update, or in a necessary optimality condition. We have for $i = 1..N_s$

$$\nabla d_x^i(\mathbf{k}) = \left(\frac{\partial \mathbf{F}_x^i}{\partial \mathbf{k}} \right)^\top (\mathbf{F}_x^i(\mathbf{k}) - \mathbf{y}_x^i) \quad (3.34)$$

$$\nabla d_m^i(\mathbf{c}) = (\mathbf{F}_m^i)^\top (\mathbf{F}_m^i \mathbf{c} - \mathbf{y}_m^i) \quad (3.35)$$

Besides, in modified Newton methods, such as the Gauss-Newton method, or the Marquardt-Levenberg method, the non-linear least-squares term d_x is iteratively approximated with the following linear least-squares term

$$\tilde{d}_x^i(\delta \mathbf{k}) = \frac{1}{2} \left\| \frac{\partial \mathbf{F}_x^i}{\partial \mathbf{k}} \delta \mathbf{k} + \mathbf{F}_x(\mathbf{k}) - \mathbf{y}_x \right\|_2^2 \quad (3.36)$$

The bottom line is that, in order to apply optimization algorithms, we need to be able to *evaluate* the following operators

$$\mathbf{F}_m, \mathbf{F}_m^\top, \mathbf{F}_m^\top \mathbf{F}_m, \mathbf{F}_x(\mathbf{k}), \left(\frac{\partial \mathbf{F}_x^i}{\partial \mathbf{k}} \right)^\top, \text{ and } \left(\frac{\partial \mathbf{F}_x^i}{\partial \mathbf{k}} \right)^\top \frac{\partial \mathbf{F}_x^i}{\partial \mathbf{k}} \quad (3.37)$$

We emphasize that most iterative methods require only evaluation of the linear operators, without requiring their storage in memory. In Algorithms 1 to 5 we propose two implementation strategies for \mathbf{F}_m , \mathbf{F}_m^\top and $\mathbf{F}_m^\top \mathbf{F}_m$, based on the ideas developed in the previous paragraph. We name them "direct approach", and "forward model in sampling form". What is presented for \mathbf{F}_m can be transposed directly to $\frac{\partial \mathbf{F}_x}{\partial \mathbf{k}}$ since they have the same structure.

3.5.4 Performance

Let us comment on the performance of each strategy. It is essentially the number of Helmholtz equations solved that governs the computing time need to evaluate the model. The main advantage of the sampling form is that once the fields \mathbf{u}_x^i and \mathbf{u}_m^i have been computed, evaluation of the model is very fast; it consists in computing some inner-products, which is easily parallelized. The computing cost is not far off from that of a matrix multiplication. Note that even if the fields \mathbf{u}_x^i and \mathbf{u}_m^i are stored, it does not mean that the entire matrix \mathbf{F}_m is stored. We only store the fields and the combinations source-sampling vectors. When \mathbf{k} (or equivalently κ^2) is fixed and the model has to be evaluated multiple times, it is the most efficient method because it requires to solve the Helmholtz equation many times less than the direct method. However, if the fields \mathbf{u}_x^i and \mathbf{u}_m^i have to be re-evaluated repeatedly (as for instance in the non-linear estimation of \mathbf{k}) the best approach will depend on the numbers N_s and N_d . In the direct approach, we need to solve $n_1 = 2N_s$ Helmholtz equation, while in the sampling form, $n_2 = N_s + N_d$.

The number n_2 can further be decreased if we employ the same source-vectors and sampling-vectors as is the case for point sources and point detectors at the same location.

3.5.5 Misfit rescaling

In practical computations we do not use directly the misfit expression (3.33), because of the large range in the magnitude distribution of the components of \mathbf{y}_m and $\mathbf{F}_m \mathbf{c}$. Instead we use a rescaled misfit given by

$$\frac{1}{2} \|\mathbf{y}_m - \mathbf{F}_m \mathbf{c}\|_{\mathbf{W}}^2, \quad \text{where} \quad \|\mathbf{x}\|_{\mathbf{W}}^2 = \mathbf{x}^\top \mathbf{W} \mathbf{x} \quad (3.38)$$

and \mathbf{W} is a positive-definite diagonal matrix. As in the Rytov expansion [5], or the Born normalization [12], we found that $\mathbf{W} = \text{diag}(1/\mathbf{y}_x)$ is a good choice. Note, however, that we employ \mathbf{W} for making the optimization problem better scaled, and not in the context of the Rytov expansion, or Born normalization. An intuitive reason why $1/\mathbf{y}_x$ is an appropriate rescaling is that \mathbf{y}_x and \mathbf{y}_m both result from propagation of light in the object and are therefore correlated.

3.6 Regularization of the discrete problem

For simplicity of notation, we express the cost functional (3.7) in the form

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \psi(\mathbf{x}) \quad (3.39)$$

where \mathbf{K} is the forward model matrix.

3.6.1 General considerations

In the previous section we derived an implementation of the matrix \mathbf{K} that is consistent with the continuous forward model of the measurement setup. Similarly, we expect the penalty function $\psi(\mathbf{x})$ to be reasonable with respect to the continuous object approximated by \mathbf{x} . As an example, let us consider $\mathbf{x} \in \mathbb{R}^N$, and $\psi(\mathbf{x}) = \|\mathbf{x}\|_2^2/2$. Vector \mathbf{x} represents an element of $x_h \in V_h$, that is, a piecewise constant function with value x_i on cell number i of \mathcal{T}_h . Since cells may have different sizes,

$\psi(\mathbf{x}) = \|\mathbf{x}\|_2^2/2$ is not a reasonable choice because it penalizes equally coefficients corresponding to large cells or small cells. Clearly the appropriate choice in this case would be to use the continuous definition

$$\int_{\Omega} (x_h(\mathbf{r}))^2 d\mathbf{r} = \sum_{e \in \mathcal{T}_h} V_e \mathbf{x}_e^2 \quad (3.40)$$

For penalty functionals defined in the continuous domain, we employ discretization methods similar to those presented in Sections 2 and 3 to derive a penalty function. In the general case, let v_1, \dots, v_N be the basis of V_h composed of indicator functions of cells. Let $x_h \in V_h$. Expansion

$$x_h = \sum_{i=1}^n x_i v_i \quad (3.41)$$

is consistent with the FVM presented in the Section 2 since x_i is the value of x_h on cell i . Basis vectors v_1, \dots, v_N are un-normalized but orthogonal. We define a class of consistent regularizers by applying the penalty to the normalized coefficient with the formula

$$\Psi(x_h) = \psi(\mathbf{x}) = \psi(\|v_1\|_2 x_1, \dots, \|v_N\|_2 x_N) = \psi(V_1 x_1, \dots, V_N x_N) \quad (3.42)$$

Example : The two following penalty function induce sparsity of coefficients, and group of coefficients, respectively. We will investigate them in the next chapters. For sparsity of coefficients we consider

$$\Psi(x_h) = \psi(\mathbf{x}) = \sum_{i=1}^N V_i |x_i| \quad (3.43)$$

For group sparsity, we introduce the $\ell_{2,1}$ -norm of a compound vector. Let $\mathbf{z} = (\mathbf{z}_1^\top, \dots, \mathbf{z}_M^\top)^\top$ be a compound vector. We have

$$\|\mathbf{z}\|_{2,1} = \sum_{i=1}^M \|\mathbf{z}^i\|_2 = \sum_{i=1}^M \left(\sum_{j=1}^{N_i} (z_j^i)^2 \right)^{\frac{1}{2}} \quad (3.44)$$

It yields

$$\Psi(x_h) = \psi(\mathbf{x}) = \sum_{i=1}^M \left(\sum_{j=1}^{N_i} (V_i^j)^2 (x_j^i)^2 \right)^{\frac{1}{2}} \quad (3.45)$$

An alternative which leads to a simpler optimization problem is

$$\psi_1(\mathbf{x}) = \sum_{i=1}^M \|v_i\|_2 \|x^i\|_2. \quad (3.46)$$

It satisfies $\psi(\mathbf{x}) \leq \psi_1(\mathbf{x})$ (Cauchy-Schwarz inequality).

3.6.2 Discrete gradient and application to regularization

In this paragraph, we derive a cell centered discretization of two important continuous-domain regularizers :

$$|u|_{H^1} = \int_{\Omega} \|\nabla u\|_2^2 \, d\mathbf{r} \quad (3.47)$$

$$|u|_{\text{TV}} = \int_{\Omega} \|\nabla u\|_2 \, d\mathbf{r} \quad (3.48)$$

The first penalty function leads to Tikhonov regularization, while the second leads to *total variation* (TV) regularization. In both cases the gradient operator is involved, and therefore we need to define a discrete gradient operator. With a cell centered discretization on an arbitrary tetrahedral mesh, it is a non-trivial task. For a uniform square grid (pixels), finite-difference formulae are available. When a nodal, instead of cell-based, discretization is used, one can use interpolation and quadrature formulae. Unfortunately, these solutions are not satisfactory in our case. We propose a simple method for defining a discrete gradient operator, and use it to discretize the two above penalty functionals.

Discrete gradient

We use a similar reasoning as in the FVM of Section 3.4.1 to derive our gradient. The idea is to define it by duality using the following relation, valid $\forall V \in H_0^1(\Omega)^n$

$$\int_{\Omega} (\nabla u) \cdot V \, d\mathbf{r} = - \int_{\Omega} u (\nabla \cdot V) \, d\mathbf{r} \quad (3.49)$$

By discretizing the divergence operator as well as the scalar and vector inner products we obtain the gradient as the adjoint of the divergence. We already constructed such a gradient operator in (3.16). Unfortunately, this expression has two major drawback for us. The first one is that it gives us a value of the normal component of the gradient at each cell, and not the magnitude. Second, it involves \mathbf{W}_2^{-1} which is unknown and not sparse. Hence it would impractical to use this method for the reason we discussed in the numerical implementation of the Helmholtz operator.

We address these two issues as we show in Fig 3.2. We replace the scalar unknowns $J_e^{n,i}$, $i = 1..4$ corresponding to the normal components of the flux J by vector unknowns J_e^i , $i = 1..4$ that are the values of J at the face centers. In this way, we compute one flux value per cell by linear interpolation. A discrete inner product for vector variables is obtained using the midpoint formula, which results in a matrix \mathbf{W}_3 . The discrete inner-product \mathbf{W}_1 is unchanged, and we define the discrete divergence \mathbf{div} for the new discrete vector fields in a way similar to what we did for the FVM. Formula (3.16) remains valid if we replace \mathbf{W}_2 with \mathbf{W}_3 . The resulting expression is

$$\mathbf{G} = \mathbf{W}_3^{-1} \mathbf{div}^T \mathbf{W}_1 \quad (3.50)$$

It is usable in practice because matrix \mathbf{W}_3 is diagonal, contrary to \mathbf{W}_2 .

3.6.3 Application to Tikhonov regularization

Using the framework presented above we can discretize the Tikhonov regularization potential

$$\Psi(u) = \int_{\Omega} \|\nabla u\|_2^2 \, d\mathbf{r}. \quad (3.51)$$

We obtain

$$\psi(\mathbf{u}) = \mathbf{u}^T \mathbf{W}_1^T \mathbf{div} \mathbf{W}_3^{-1} \mathbf{div}^T \mathbf{W}_1^T \mathbf{u}. \quad (3.52)$$

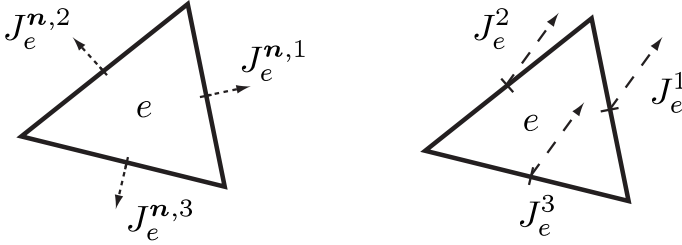


Figure 3.2: Replacing the normal component variables with vector flux variables at the face centers.

3.6.4 Application to total-variation

Total variation (TV) is a popular regularizer for reconstructing piecewise-constant solutions with jumps. Thus, it does not suffer from over-smoothing artifacts introduced by quadratic regularizers. Some care has to be taken for discretizing TV. Indeed, expression (3.48) is not defined as such for functions with jumps (because the gradient is not defined at the discontinuities). The correct definition is variational

$$|u|_{\text{TV}} = \sup_{v \in C} \int_{\Omega} u (\nabla \cdot v) \, d\mathbf{r} \quad \text{with} \quad C = \{v \in C_0^1(\Omega); \|v(x)\|_2 \leq 1\} \quad (3.53)$$

Integrating by parts shows that this formula is equivalent to (3.48) when $u \in C^1(\Omega)$. Using a density argument, and we can extend (3.53) for elements V_h .

With the discrete gradient operator described above we obtain the following consistent discrete TV operator :

$$|\mathbf{u}|_{\text{TV}} = \sup_{\|\mathbf{V}\|_{2,\infty} \leq 1} \mathbf{u}^T \mathbf{W}_1 \mathbf{div} \mathbf{V} = \|\mathbf{div}^T \mathbf{u}\|_{2,1} \quad (3.54)$$

Note the similarity of this expression with the one valid for regular grids.

Chapter 4

General ℓ_p -regularization for Fluorescence Diffuse Optical Tomography

THE¹ starting point of this work is the observation that a number biological mechanisms studied using FDOT exhibit some kind of *sparsity* pattern. For instance, the fluorescence signal obtained when imaging protease activity in early stage of cancer originates from a localized area surrounding the tumor [24]. In this case the signal is spatially sparse; only a few pixels are activated. On the other hand, one might also encounter cases where the distribution of fluorescent molecules is spread more evenly, but restricted to specific anatomical structures; images of inflammation and bone metabolism fall into that category. Resulting images are characterized by sharp transitions at the boundaries between regions of presence and regions of absence of fluorophore. There, it is the variations, or equivalently the gradient, that are limited to a confined region, and therefore sparse. Our purpose in this work is to harness this inherent sparsity in order to improve FDOT reconstruction. We consider sparsity as a form of *a priori* knowledge that we include in the inversion algorithm.

¹Part of this chapter is based on our published paper [48].

Our reconstruction algorithms are derived in the variational framework presented in Chapter 2, whereby sparsity is imposed by means of a suitable regularization potential. In this work we deal with the linear FDOT inverse problem described in Chapter 3 by (3.7). It is noteworthy that conventional regularization methods for linear inverse problems rely on quadratic regularization potentials, or equivalently a ℓ_2 -norm penalty. The resulting reconstruction algorithms are linear. Although linear methods are appealing due to their low computational cost, they are not suited to induce sparsity in the reconstructed images. To understand this point, one may consider the *Truncated Singular Values Decomposition* (TVSD) method, which is one of the prominent linear regularization methods. There, one uses only the data components corresponding to the largest singular values of the forward model to reconstruct the signal. Because large singular values are associated to low-frequency singular-vectors, the resulting reconstruction contains only low frequencies. Clearly, such reconstructions are not adapted to represent sparse signals. This *smoothing* behaviour of TSVD is also observed for ℓ_2 -norm regularization. Sparsity-inducing potentials therefore correspond to *non-linear* reconstruction algorithms, despite the fact that the inverse problem is a linear one.

Amongst all sparsity-promoting regularization potentials we focus on those that are based on the ℓ_p -norm, for $p \in [1, 2]$. Specifically, we consider the family of functions of the form $\psi(\mathbf{x}) = \|\mathbf{R}\mathbf{x}\|_p^p$, where \mathbf{R} is a suitable linear operator chosen according to some *a priori* information. Here p might be understood as a sparsity index. For $p = 2$ we obtain ℓ_2 -norm penalty where no sparsity is induced, while taking p closer to 1 favors more and more sparsity. The limiting case $p = 1$ imposes as much sparsity as possible, while retaining a convex regularization potential.

Our contribution is to propose state-of-the-art algorithms implementing a ℓ_p -norm penalty, which enables one to incorporate a sparsity *a priori* in FDOT reconstruction. The performance of the proposed algorithms relies on the fast implementation of the forward model presented in Chapter 3, and on convex optimization methods tailored to ℓ_p -norm penalization. Several methods are presented, in order to deal with the specifics of the potential functions in the family.

This chapter is organized as follows. First, we introduce the family of ℓ_p potentials used in this work. Then, we review convex optimization techniques needed in the reconstruction algorithm. Our intent is to convey an intuitive understanding of each method and its domain of application. Emphasis is put on design principles rather than mathematical analysis of the methods. After that comes an experimental part in which the proposed algorithm was tested on synthetic and phantom

data. This part includes a comparison of ℓ_1 regularization versus ℓ_2 regularization. There, our goal is to investigate the improvement of the proposed method over linear methods when the sparsity *a priori* is verified.

4.1 Notation

We make extensive use of the notion of ℓ_p -norm. We recall that the ℓ_p -norm of a vector $\mathbf{x} = (x_1, \dots, x_N)$ is denoted $\|\mathbf{x}\|_p$, and defined as follows

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^N |x_i|^p \right)^{\frac{1}{p}}. \quad (4.1)$$

The cases we use the most in practice are the ℓ_0 , ℓ_1 and ℓ_2 norms :

$$\|\mathbf{x}\|_0 = \max_{i=1..N} |x_i|, \quad (4.2)$$

$$\|\mathbf{x}\|_1 = \sum_{i=1}^N |x_i|, \quad (4.3)$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^N x_i^2}. \quad (4.4)$$

Finally, note that ℓ_p -norm penalization is quite a misnomer because it usually refers to the penalization of $\|\mathbf{x}\|_p^p$.

4.2 Sparsity-inducing regularization potentials

The connection between sparsity and ℓ_1 -norm penalization was observed early on [49, 50]. Recently, with the fast development of compressed-sensing and its applications, the need for algorithms computing sparse solutions of underdetermined linear systems led to the consideration of optimization problems of the form

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_0 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b} \quad (4.5)$$

60 General ℓ_p -regularization for Fluorescence Diffuse Optical Tomography

Unfortunately, problem (4.5) is combinatorial and computationally intractable (NP-hard [51]). In practice, an approximate solution is obtained by solving a convex relaxation based on the ℓ_1 -norm :

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{A}\mathbf{x} = \mathbf{b}. \quad (4.6)$$

Problem (4.6) is convex and can be solved efficiently. The hope is that the solutions of (4.5) and (4.6) coincide; see for instance [52] for sufficient conditions on \mathbf{A} guaranteeing the equivalence.

In the context of FDOT, we are interested in sparsity as a regularization criterion; it is some *a priori* knowledge that we have on the solution. In its simplest form, we include the sparsity *a priori* in the variational reconstruction approach (2.41) by considering $\psi(\mathbf{x}) = \|\mathbf{x}\|_1$. It leads to the following regularized least-squares problem

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \quad (4.7)$$

The sparsity-inducing effect of the ℓ_1 penalty can be understood intuitively by comparing the ℓ_1 and ℓ_2 unit-balls represented Fig 4.1 for the 2D case. For a given ℓ_2 -norm value, we observe that the ℓ_1 -norm has a higher value on vectors composed of many small coefficients than vectors composed of few large coefficients. This sparsity-inducing property was formalized and extended to more general functions in [53]. Along the same line of idea, we observe empirically that the convex potentials $\psi(\mathbf{x}) = \|\mathbf{x}\|_p^p$, $1 \leq p \leq 2$ lead to sparser and sparser solutions as $p \rightarrow 1$. Their advantage is to be differentiable, which is not the case of $\|\mathbf{x}\|_1$.

The cost function in (4.7) is useful to express sparsity of the solution vector $\hat{\mathbf{x}}$ itself. However, one might be interested in capturing sparsity occurring in another basis, or more generally, after a linear transformation (the gradient for instance). Besides, we saw in Section 3.6 that one desirable property of the discrete regularizer is to be consistent with the continuous object represented by the discretization. This is typically not the case of $\|\mathbf{x}\|_1$. These issues are addressed by considering potential functions of the $\psi(\mathbf{x}) = \|\mathbf{R}\mathbf{x}\|_1$, where \mathbf{R} is a suitable linear operator. There, it is sparsity in the range of \mathbf{R} that is promoted. One last capability we wish for our potential is to be able to deal with vector-valued operators. The main representative is the gradient operator and its discrete counterpart. Let \mathbf{R} be a vector-valued operator. We denote N_R the dimension of the space in which \mathbf{R}

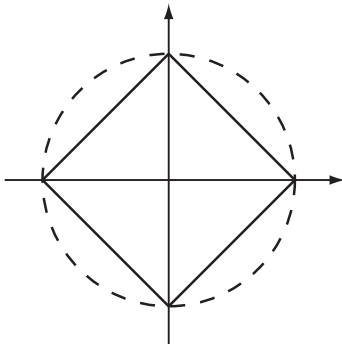


Figure 4.1: ℓ_2 unit-ball (dashed lined) and ℓ_1 unit-ball

takes its values. For instance, if $\mathbf{R} = \nabla$, we have $N_R = 3$ in three dimensions. We call r the number of vector values returned by R . Taking again the example $\mathbf{R} = \nabla$ and assuming a square pixels grid, one can define a gradient value at each pixel using a finite-difference formula. Thus, r is the number of pixels in the grid. In summary, $\mathbf{R}\mathbf{x}$ is composed of r sub-vectors of size N_R . By convention we will assume that $\mathbf{R}\mathbf{x}$ is obtained by stacking these r sub-vectors and denote

$$\mathbf{R}\mathbf{x} = \begin{pmatrix} (\mathbf{R}\mathbf{x})_1 \\ \vdots \\ (\mathbf{R}\mathbf{x})_r \end{pmatrix} \quad (4.8)$$

where $(\mathbf{R}\mathbf{x})_i, i = 1..r$ are the sub-vectors of size N_R .

This leads us to the general sparsity-inducing potential that we consider in this work. Let $p \in [1, 2]$.

$$\psi(\mathbf{x}) = \frac{1}{p} \sum_{i=1}^r \|(\mathbf{R}\mathbf{x})_i\|_2^p \quad (4.9)$$

Taking $\mathbf{R} = \mathbf{I}$ and $p = 1$, we see that (4.9) reduces to $\|\mathbf{x}\|_1$. Taking $\mathbf{R} = \nabla$ and $p = 1$, it reduces to TV. Although the case $p = 2$ does not correspond to sparsity-promoting potentials we include it in order to complete our family of regularizers.

4.3 Iterative optimization methods for sparsity-inducing regularization

In this section, we describe state-of-the-art iterative methods for solving the problem

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \frac{\lambda}{p} \sum_{i=1}^r \|(\mathbf{R}\mathbf{x})_i\|_2^p \quad \text{where } 1 \leq p \leq 2. \quad (4.10)$$

Problem (4.10) is a convex optimization problem, and the minimum is attained provided $\text{Ker}\mathbf{K} \cap \text{Ker}\mathbf{R} = \{0\}$.

For $p = 2$, the solution of (4.10) is obtained by solving the linear system

$$(\mathbf{K}^\top \mathbf{K} + \lambda \mathbf{R}^\top \mathbf{R}) \mathbf{x} = \mathbf{K}^\top \mathbf{y} \quad (4.11)$$

Iterative linear solvers such as the *Preconditioned Conjugate Gradient* can be employed to solve the system. For $p < 2$ the solution of (4.10) is non-linear, but the linear solver is often a building block in non-linear techniques. As a remark, there is generally no satisfactory preconditioning technique for system (4.11) when \mathbf{K} corresponds to an inverse problem

One challenge in solving (4.10) is the large scale of the problem : a typical three-dimensional problem will involve 10^4 up to 10^5 variables. Another challenge when $p = 1$ is that of non-differentiability of the objective function. Indeed, the function $\|(\mathbf{R}\mathbf{x})_i\|_2$ has "kinks" at the points where $(\mathbf{R}\mathbf{x})_i = 0$. Since these kinks occur only for $p = 1$ and in localized places, a pragmatic approach is to bypass the problem, either by using $p > 1$ close to 1, or by replacing $\psi(\mathbf{x})$ with an approximation where the kinks have been smoothed out. With this goal, we introduce the Huber potential :

Definition 4.3.1. We call *Huber potential* the function $h_{p,\epsilon} : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ defined by

$$h_{p,\epsilon}(x) = \begin{cases} \frac{x^p}{p}, & \text{when } x > \epsilon \\ \frac{\epsilon^{p-2}}{2} x^2 + \left(1 - \frac{p}{2}\right) \frac{\epsilon^p}{p}, & \text{when } x \leq \epsilon \end{cases} \quad (4.12)$$

4.3 Iterative optimization methods for sparsity-inducing regularization 63

Function $h_{p,\epsilon}$ is differentiable, with continuous derivative. Using $h_{p,\epsilon}$ we obtain a "smoothed" version of (4.10);

$$\arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^r h_{p,\epsilon}(\|(\mathbf{R}\mathbf{x})_i\|_2) \quad (4.13)$$

In practice, using (4.10) or (4.13) produces similar results provided ϵ is small enough.

4.3.1 Limitations of conventional methods

For $p > 1$ the objective in (4.10) is convex and C^∞ . Therefore standard optimization methods such as gradient descent, Newton's method, and their variants are applicable. For $p = 1$, points of non-differentiability are localized. Hence, sub-gradient methods are expected to be mostly equivalent to gradient methods in terms of iterates and rate of convergence. Alternatively, one can use an approximation based on the Huber potential. In practice, however, the performance of these methods is poor.

The behaviour of gradient-descent-type methods is illustrated on two simple cases in Figs 4.2 and 4.3. In Fig 4.2 we show iterates of the subgradient method for optimizing $f(\mathbf{x}) = \|\mathbf{x}\|_1$, while in Fig 4.3 we show iterates for the function $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \|\mathbf{x}\|_1$. We observe the same behaviour in both cases. The abrupt changes in the gradient due to the non-differentiability points cause the methods to oscillate and converge slowly to the minimizer.

Regarding Newton's method, it results in much more expensive iterations. In addition, when $p = 1$, the absence of curvature in the function ψ (its Hessian is 0) makes it inapplicable.

Much better convergence rates are achieved if we harness the additive structure of the objective.

4.3.2 Optimization transfer methods

In optimization transfer methods iterates are obtained by the minimization (or partial minimization) of a sequence of surrogate functions that upper-bound the objective. The surrogate functions are assumed to be "easier" functions to minimize. The objective in (4.10) is the sum of a misfit term and a regularization term. With

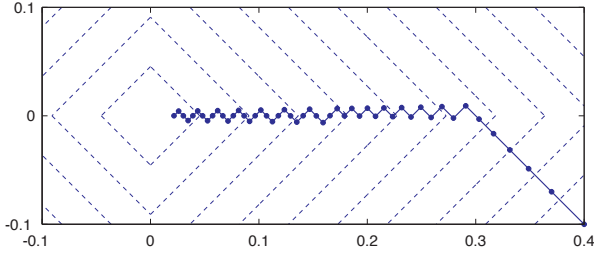


Figure 4.2: Subgradient method applied to minimization of $f(\mathbf{x}) = \|\mathbf{x}\|_1$.

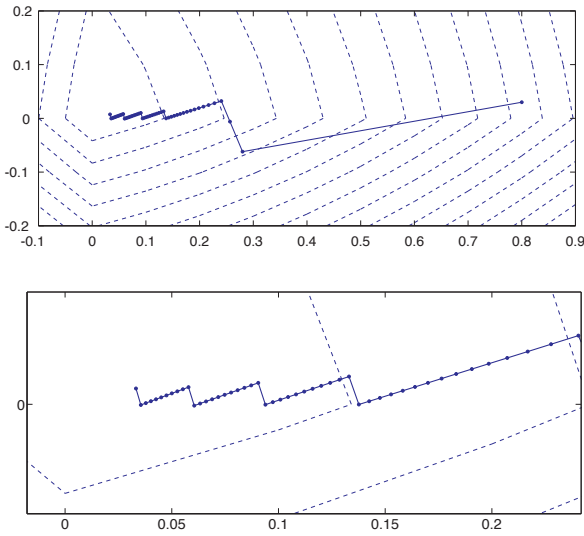


Figure 4.3: Subgradient method applied to minimization of $f(\mathbf{x}) = \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 + \|\mathbf{x}\|_1$, for $\mathbf{y} = (1/2, 4/5)$.

4.3 Iterative optimization methods for sparsity-inducing regularization 65

this structure in mind, one may derive a surrogate function by upper-bounding the misfit term or the regularization term.

Upper-bounding the potential function

The most intuitive way of tackling problem (4.10) is perhaps the *iteratively re-weighted norm method* (IRNM) [54]. The idea here is that when $(\mathbf{R}\mathbf{x})_i \neq 0$, then we have

$$\|(\mathbf{R}\mathbf{x})_i\|_2^p = \|(\mathbf{R}\mathbf{x})_i\|_2^{p-2} \|(\mathbf{R}\mathbf{x})_i\|_2^2. \quad (4.14)$$

This suggests an iteration based on weighted ℓ_2 -norms :

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \frac{\lambda}{2} \sum_{i=1}^r w_{k,i} \|(\mathbf{R}\mathbf{x})_i\|_2^2 \quad \text{where} \quad w_{k,i} = \frac{2}{p} \|(\mathbf{R}\mathbf{x}^k)_i\|_2^{p-2} \quad (4.15)$$

Each iterate is computed by solving a linear system with an iterative method such as PCG. Clearly, formulation of (4.15) is problematic when $(\mathbf{R}\mathbf{x})_i = 0$ because $w_{k,i}$ is undefined in this case. In practice, this problem is addressed by introducing a (small) threshold value ϵ and treating differently the weights $w_{k,i} < \epsilon$. In [55] the authors study the case $\mathbf{R} = \mathbf{I}$, and propose to set to zero in subsequent iterations the coefficients smaller than ϵ . In [54], weights $w_{k,i}$ smaller than ϵ are simply replaced by ϵ in (4.15). Either way, due to the presence of this threshold parameter ϵ , the convergence of these iterations to a solution (4.10) is not established.

Since the solution of the non-smooth problem (4.10) with a re-weighted norm approach is elusive, a more pragmatic point of view is to use this method for the smooth problem (4.13). It is based on the following property.

Property 4.3.2. *Let $p \in [1, 2]$. Let $\alpha \geq \epsilon > 0$. Let $f_\alpha : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ the quadratic function defined by*

$$f_\alpha(x) = \frac{\alpha^{p-2}}{2} x^2 + \left(1 - \frac{p}{2}\right) \frac{\alpha^p}{p} \quad (4.16)$$

We have the following inequality :

$$\frac{x^p}{p} \leq h_{p,\epsilon}(x) \leq f_\alpha(x), \quad \forall x \geq 0. \quad (4.17)$$

where $h_{p,\epsilon}$ is the Huber potential defined in (4.12).

66 General ℓ_p -regularization for Fluorescence Diffuse Optical Tomography

Starting from this observation, the regularization term of the objective in (4.13) is upper-bounded using a quadratic function, resulting in a sequence of surrogate functions. We propose the following variation of (4.15).

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^r f_{w_{k,i}}(\|\mathbf{R}\mathbf{x}\|_2) \quad (4.18)$$

$$\text{with } w_{k,i} = \begin{cases} \|\mathbf{R}\mathbf{x}^k\|_2, & \|\mathbf{R}\mathbf{x}^k\|_2 \geq \epsilon \\ \epsilon, & \text{otherwise} \end{cases} \quad (4.19)$$

This strategy is therefore an optimization transfer method for problem (4.13).

Each iteration consists in solving the following linear system :

$$(\mathbf{K}^\top \mathbf{K} + \lambda \mathbf{R}^\top \mathbf{W}_k \mathbf{R}) \mathbf{x} = \mathbf{K}^\top \mathbf{y} \quad \text{where } \mathbf{W}_k = \text{diag} \left(w_{k,i}^{p-2} \right) \quad (4.20)$$

System (4.20) is one Schur complement of the saddle-point system

$$\begin{pmatrix} \mathbf{R}^\top \mathbf{W}_k \mathbf{R} & \mathbf{K}^\top \\ \mathbf{K} & -\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{u} \end{pmatrix} = \begin{pmatrix} \mathbf{K}^\top \mathbf{y} \\ 0 \end{pmatrix} \quad (4.21)$$

Provided the matrix $\mathbf{R}^\top \mathbf{W}_k \mathbf{R}$ is invertible (which is the case if and only if $\text{Ker} \mathbf{R} = \{0\}$), one can use the second Schur complement of (4.21)

$$\left(\mathbf{K} (\mathbf{R}^\top \mathbf{W}_k \mathbf{R})^{-1} \mathbf{K}^\top + \mathbf{I} \right) \mathbf{u} = \mathbf{K} (\mathbf{R}^\top \mathbf{W}_k \mathbf{R})^{-1} \mathbf{K}^\top \mathbf{y} \quad (4.22)$$

Formulation (4.22) is mostly useful when \mathbf{R} is a diagonal matrix. For instance, if $\mathbf{R} = \mathbf{I}$, (4.22) reduces to

$$(\mathbf{K} \mathbf{W}_k^{-1} \mathbf{K}^\top + \mathbf{I}) \mathbf{u} = \mathbf{K} \mathbf{W}_k^{-1} \mathbf{K}^\top \mathbf{y} \quad (4.23)$$

Implementation-wise, the main advantages of (4.22) are first that $\mathbf{K} \mathbf{W}_k^{-1} \mathbf{K}^\top$ is a smaller matrix than $\mathbf{K}^\top \mathbf{K}$ for underdetermined problems, which results in a smaller system to solve. And second, the matrix \mathbf{W}_k^{-1} does not contain the inverses of the regularization values, which might lead to a better condition-number of the system. In the end, the best strategy should be determined for the problem at hand.

Remark : The *iteratively re-weighted norm method* presented here is sometimes confused with the *Iteratively Re-weighted Least-Squares* (IRLS) [56] algorithm, also

4.3 Iterative optimization methods for sparsity-inducing regularization 67

called FOCal Underdetermined System Solver (FOCUSS) [57]. Although they are based on somewhat similar re-weighting procedures, these algorithms were not designed to solve the same problem. For IRLS, the problem is to compute

$$\min \|\mathbf{x}\|_1 \quad \text{subject to} \quad \mathbf{Ax} = \mathbf{b} \quad (4.24)$$

The difference is in the equality constraint, which makes IRLS unadapted to our case (non-ideal model). Contrary to the method presented above, IRLS solves the non-smooth problem [56].

Upper-bounding the misfit term

One can think of deriving a surrogate function by upper-bounding the misfit-term in (4.10). Since the misfit-term is a quadratic function, one can obtain a surrogate using simple algebraic manipulations :

$$\frac{1}{2} \|\mathbf{Kx} - \mathbf{y}\|_2^2 = \frac{1}{2} \|\mathbf{K}(\mathbf{x} - \mathbf{x}^k) + \mathbf{Kx}^k - \mathbf{y}\|_2^2 \quad (4.25)$$

$$= \frac{1}{2} (\mathbf{x} - \mathbf{x}^k)^\top \mathbf{K}^\top \mathbf{K} (\mathbf{x} - \mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^\top \mathbf{K}^\top (\mathbf{Kx}^k - \mathbf{y}) + \frac{1}{2} \|\mathbf{Kx}^k - \mathbf{y}\|_2^2 \quad (4.26)$$

$$\leq \frac{\alpha}{2} (\mathbf{x} - \mathbf{x}^k)^\top (\mathbf{x} - \mathbf{x}^k) + (\mathbf{x} - \mathbf{x}^k)^\top \mathbf{K}^\top (\mathbf{Kx}^k - \mathbf{y}) + \frac{1}{2} \|\mathbf{Kx}^k - \mathbf{y}\|_2^2 \quad (4.27)$$

$$= \frac{\alpha}{2} \left\| \mathbf{x} - \mathbf{x}^k + \frac{1}{\alpha} \mathbf{K}^\top (\mathbf{Kx}^k - \mathbf{y}) \right\|_2^2 - \frac{1}{2\alpha} \|\mathbf{K}^\top (\mathbf{Kx}^k - \mathbf{y})\|_2^2 + \frac{1}{2} \|\mathbf{Kx}^k - \mathbf{y}\|_2^2 \quad (4.28)$$

where $\|\mathbf{K}^\top \mathbf{K}\|_2 \leq \alpha$.

This upper-bound suggests the following iteration, known as the *Iterated Soft-Thresholding Algorithm (ISTA)* [58, 59]:

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{\alpha}{2} \left\| \mathbf{x} - \left(\mathbf{x}^k - \frac{1}{\alpha} \mathbf{K}^\top (\mathbf{Kx}^k - \mathbf{y}) \right) \right\|_2^2 + \lambda \psi(\mathbf{x}) \quad (4.29)$$

We can decompose ISTA iterations in two steps :

$$\mathbf{x}^{k+1/2} = \mathbf{x}^k - \frac{1}{\alpha} \mathbf{K}^\top (\mathbf{Kx}^k - \mathbf{y}) \quad (4.30)$$

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{\alpha}{2} \left\| \mathbf{x} - \mathbf{x}^{k+1/2} \right\|_2^2 + \lambda \psi(\mathbf{x}) \quad (4.31)$$

which reveals that it is composed of a gradient descent step on the misfit term, followed by an implicit step corresponding to a de-noising problem (the forward model is taken to be \mathbf{I}). We note that ISTA makes no differentiability assumption on the regularization potential. The only restriction is to be able to compute step (4.31) for a given ϕ potential. Interestingly, this remark may be formalized and ISTA can be related to a more general class of algorithms in convex optimization by introducing the *proximal map*.

Definition 4.3.3. [60] Let $\psi : \mathbb{R}^N \rightarrow \mathbb{R}$ a convex function bounded from below. We call *proximal map* of ψ and denote prox_ψ , or $(I + \partial\psi)^{-1}$, the mapping defined as follows

$$\text{prox}_\psi(\mathbf{x}) = (I + \partial\psi)^{-1}(\mathbf{x}) = \arg \min_{\mathbf{y} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \psi(\mathbf{y}) \quad (4.32)$$

Let us try to get an intuitive understanding of the proximal map. Assume ϕ is differentiable. Then, the optimality condition reads

$$\mathbf{y} = \text{prox}_{\tau\psi}(\mathbf{x}) \iff \mathbf{y} = \mathbf{x} - \tau \nabla\psi(\mathbf{y}) \quad (4.33)$$

Therefore, applying the proximal map is nothing but doing an *implicit* gradient-descent step. One advantage of the proximal map is that it is defined for non-differentiable functions. Hence we can think of it as the equivalent of a gradient descent that is valid for non-differentiable functions. In addition, the use of an implicit step brings some robustness to the algorithm.

Using the proximal map, and posing $f(\mathbf{x}) = \frac{1}{2} \|\mathbf{K}\mathbf{x} - \mathbf{y}\|_2^2$, we reformulate (4.29) :

$$\mathbf{x}^{k+1/2} = \mathbf{x}^k - \frac{1}{\alpha} \nabla f(\mathbf{x}^k) \quad (4.34)$$

$$\mathbf{x}^{k+1} = \left(I + \frac{\lambda}{\alpha} \partial\psi \right)^{-1} \left(\mathbf{x}^{k+1/2} \right) \quad (4.35)$$

Expressed in this form, we identify the ISTA iteration (4.29) with the *Forward-Backward Splitting* (FBS) algorithm applied to the objective function $J(\mathbf{x}) = f(\mathbf{x}) + \lambda\psi(\mathbf{x})$. The FBS algorithm is well-understood theoretically and was shown to converge under mild assumptions [61]. More generally, FBS belongs a larger class of proximal-based algorithms referred to as *splitting methods* in the field of convex

4.3 Iterative optimization methods for sparsity-inducing regularization 69

optimization [62]. These algorithms are specifically designed to handle objective functions presenting an additive structure. FBS is the most adapted splitting approach in our case.

The practical performance of the FBS algorithm depends on our ability to evaluate the proximal map efficiently. Ideally it is known in closed-form. For $\psi(\mathbf{x}) = \frac{1}{p} \sum_{i=1}^r \|(\mathbf{R}\mathbf{x})_i\|_2^p$, this is unfortunately the case only when $\mathbf{R} = \mathbf{I}$ (or equivalently, using a change of variable, when \mathbf{R} is invertible). In this case, the proximal map is separable, and the expression follows from the classical result recalled in Proposition 4.3.4. Otherwise, efficient iterative methods need to be used, and the FBS algorithm will contain inner iterations.

Proposition 4.3.4. *Let $\phi_p(\mathbf{x}) = \frac{1}{p} \|\mathbf{x}\|_2^p$ and $\lambda > 0$. For $p > 1$ we have*

$$(I + \lambda \partial \phi_p)^{-1}(\mathbf{x}) = w\mathbf{x}, \quad \text{where } w > 0 \text{ satisfies } w + \lambda \|\mathbf{x}\|_2^{p-2} w^{p-1} = 1. \quad (4.36)$$

In the case $p = 1$,

$$(I + \lambda \partial \phi_1)^{-1}(\mathbf{x}) = \max\left(0, 1 - \frac{\lambda}{\|\mathbf{x}\|_2}\right) \mathbf{x}. \quad (4.37)$$

Operator $(I + \lambda \partial \phi_1)^{-1}(\mathbf{x})$ is commonly known as the *soft-thresholding* operator. In a practical implementation, evaluating w in (4.36) is done by means of a lookup table and interpolation. Proposition 4.3.4 gives the proximal map when $r = 1$. It immediately extends to $r > 1$ using the separability of the function $\frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \frac{\lambda}{p} \sum_{i=1}^r \|\mathbf{x}_i\|_2^p$.

The proximal map is also available for Huber potentials $h_{p,\epsilon}(\|\mathbf{x}\|_2)$; this is interesting if one needs to compare ISTA to the re-weighting method.

Proposition 4.3.5. *Let $\lambda > 0$ and $\epsilon > 0$. For all $p \geq 1$, we have*

$$(I + \lambda \partial h_{p,\epsilon})^{-1}(\mathbf{x}) = \begin{cases} w\mathbf{x}, & \|\mathbf{x}\|_2 \geq \epsilon + \lambda \epsilon^{p-1} \\ \frac{1}{1 + \lambda \epsilon^{p-2}} \mathbf{x}, & \|\mathbf{x}\|_2 \leq \epsilon + \lambda \epsilon^{p-1} \end{cases} \quad (4.38)$$

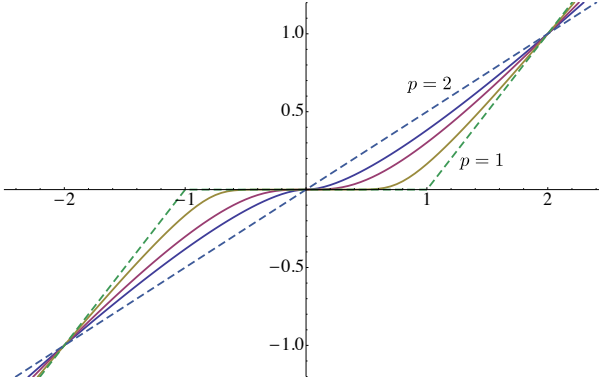


Figure 4.4: Proximal map of $\psi(\mathbf{x}) = \frac{1}{p}\|\mathbf{x}\|_p^p$ in the 1D case, for $p = 1, 1.1, 1.3, 1.5, 2$.

where $w > 0$ satisfies $w + \lambda \|\mathbf{x}\|_2^{p-2} w^{p-1} = 1$.
 In particular, for $p = 1$, we have

$$(I + \lambda \partial h_{1,\epsilon})^{-1}(\mathbf{x}) = \begin{cases} \left(1 - \frac{\lambda}{\|\mathbf{x}\|_2}\right) \mathbf{x}, & \|\mathbf{x}\|_2 \geq \epsilon + \lambda \\ \frac{1}{1 + \lambda/\epsilon} \mathbf{x}, & \|\mathbf{x}\|_2 \leq \epsilon + \lambda \end{cases} \quad (4.39)$$

In Fig 4.4, we represent the proximal map of $\psi(\mathbf{x}) = \frac{1}{p}\|\mathbf{x}\|_p^p$ in the 1D case, for varying values of p .

Even in the most favorable case where a closed form of $\text{prox}_{\lambda\psi}$ is known, we observe a rather poor convergence rate of ISTA in practice, which results in a slow reconstruction algorithm. Successful attempts to address this issue have been made by using tighter upper-bounds instead of (4.28). Acceleration was obtained for imaging problems such as deconvolution [63, 64] or parallel MRI [65]. Nevertheless, the derivation of such bounds is forward model specific and cumbersome for FDOT. A more generic acceleration strategy is to employ multi-step techniques. Here the overall structure of the algorithm remains the same, and the idea is to use not only the last iterate but the entire history to compute the next iterate. Resulting

state-of-the-art methods are *Fast ISTA* (FISTA) [66] and NESTA [67]. We found FISTA to be the most efficient reconstruction method in practice.

4.3.3 Dual and primal-dual methods

In this sub-section we present methods whereby problem (4.10) is transformed in an equivalent but distinct problem, which is then solved instead of the initial one. The hope is that this so-called *dual problem* is easier to solve than the initial *primal problem*. One may also consider *primal-dual* methods that alternate between the two problems. We consider methods in which the dual problem is formulated based on convex duality, or Lagrange duality.

Convex-duality-based methods

The fundamental observation here is the following property.

Property 4.3.6. *Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then,*

$$\frac{1}{p} \|\mathbf{x}\|_2^p = \max_{\mathbf{u} \in \mathbb{R}^N} \mathbf{u}^\top \mathbf{x} - \frac{1}{q} \|\mathbf{x}\|_2^q \quad (4.40)$$

We also have, in the case $p = 1$,

$$\|\mathbf{x}\|_2 = \max_{\mathbf{u} \in \mathbb{R}^N} \mathbf{u}^\top \mathbf{x} - \delta_B(\mathbf{u}) \quad (4.41)$$

where δ_B is the characteristic function of the set $B = \{\mathbf{x} \in \mathbb{R}^N ; \|\mathbf{x}\|_2 \leq 1\}$, that is,

$$\delta_B(\mathbf{x}) = \begin{cases} 0 & \mathbf{x} \in B \\ +\infty & \text{otherwise} \end{cases} \quad (4.42)$$

Expressions in Property 4.3.6 are not without bearing resemblance with the one used for total-variation; they are in fact its discrete counterpart. Functions $\frac{1}{p} \|\mathbf{x}\|_2^p$ (resp. $\|\mathbf{x}\|_2$) and $\frac{1}{q} \|\mathbf{x}\|_2^q$ (resp. $\delta_B(\mathbf{x})$) are said to be convex duals (or Fenchel duals) of each other. For clarity of exposition we will treat only the case $p = 1$, but what

72 General ℓ_p -regularization for Fluorescence Diffuse Optical Tomography

is said extends directly to $p > 1$ using the property above. We also consider $r = 1$ in (4.10). The case $r > 1$ follows from.

Let us pose some notations. Let $J(\mathbf{x})$ be the primal objective :

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \|\mathbf{R}\mathbf{x}\|_2, \quad (4.43)$$

and

$$L(\mathbf{x}, \mathbf{u}) = \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \mathbf{u}^\top \mathbf{R}\mathbf{x} - \delta_B(\mathbf{u}). \quad (4.44)$$

Our original (or primal) problem is

$$\min_{\mathbf{x} \in \mathbb{R}^N} J(\mathbf{x}) \quad (4.45)$$

Based on Property 4.3.6 we have

$$J(\mathbf{x}) = \max_{\mathbf{u} \in \mathbb{R}^M} L(\mathbf{x}, \mathbf{u}) \quad (4.46)$$

and therefore problem (4.45) is equivalent to the primal-dual problem

$$\min_{\mathbf{x} \in \mathbb{R}^N} \max_{\mathbf{u} \in \mathbb{R}^M} L(\mathbf{x}, \mathbf{u}) \quad (4.47)$$

Variable \mathbf{u} is referred to as the dual variable. It is natural to try exchanging min and max in this expression, which leads the dual objective $K(\mathbf{u})$:

$$K(\mathbf{u}) = \min_{\mathbf{x} \in \mathbb{R}^N} L(\mathbf{x}, \mathbf{u}) \quad (4.48)$$

By definition, we have $\forall \mathbf{x}, \mathbf{u}$,

$$K(\mathbf{u}) \leq L(\mathbf{x}, \mathbf{u}) \leq J(\mathbf{x}) \quad (4.49)$$

and thus, in our setting,

$$\underbrace{\max_{\mathbf{u} \in \mathbb{R}^M} K(\mathbf{u})}_{\text{dual problem}} \leq \underbrace{\min_{\mathbf{x} \in \mathbb{R}^N} \max_{\mathbf{u} \in \mathbb{R}^M} L(\mathbf{x}, \mathbf{u})}_{\text{primal-dual problem}} = \underbrace{\min_{\mathbf{x} \in \mathbb{R}^N} J(\mathbf{x})}_{\text{primal problem}} \quad (4.50)$$

4.3 Iterative optimization methods for sparsity-inducing regularization 73

Inequality (4.50) is valid for any function $L(\mathbf{x}, \mathbf{u})$, and it is remarkable that provided $\text{Ker}\mathbf{K} \cap \text{Ker}\mathbf{R} = \{0\}$ in our problem, we have both equality (which means, that min and max can be exchanged) and attainment : $\exists (\mathbf{x}^*, \mathbf{u}^*)$ such that

$$K(\mathbf{u}^*) = L(\mathbf{x}^*, \mathbf{u}^*) = J(\mathbf{x}^*) \quad (4.51)$$

Thus, $(\mathbf{x}^*, \mathbf{u}^*)$ satisfies $\forall \mathbf{x}, \mathbf{u}$,

$$L(\mathbf{x}^*, \mathbf{u}) \leq L(\mathbf{x}, \mathbf{u}) \leq L(\mathbf{x}, \mathbf{u}^*) \quad (4.52)$$

Point $(\mathbf{x}^*, \mathbf{u}^*)$ is called a *saddle-point*. Note that there might not be unicity of a saddle-point in our problem.

Solving the dual problem

We note that the dual function K is concave, and therefore convex optimization methods are well-suited to carry out the minimization of $-K$. Unfortunately, one easily convinces oneself that the dual objective is inaccessible if the forward model matrix \mathbf{K} corresponds the FDOT forward model operator. Indeed, evaluating the dual objective requires inverting the matrix $\mathbf{K}^T\mathbf{K}$, which is not possible. However, the dual problem may still be of some use when matrix \mathbf{K} is invertible, and specifically when $\mathbf{K} = \mathbf{I}$. In this case the primal objective is

$$J(\mathbf{x}) = \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{R}\mathbf{x}\|_2, \quad (4.53)$$

and thus, the primal problem corresponds to evaluating the proximal map. Using some algebraic manipulations, one checks that the corresponding dual objective $K(\mathbf{u})$ is

$$K(\mathbf{u}) = -\frac{1}{2} \left\| \mathbf{R}^T \mathbf{u} - \frac{\mathbf{y}}{\lambda} \right\|_2^2 - \delta_B(\mathbf{x}), \quad (4.54)$$

and that a primal solution \mathbf{x}^* is derived from a dual solution \mathbf{u}^* by $\mathbf{x}^* = \mathbf{y} - \lambda \mathbf{R}^T \mathbf{u}^*$. It is noteworthy that the dual problem is a constrained quadratic minimization problem with simple constraints :

$$\min_{\mathbf{u}} \frac{1}{2} \left\| \mathbf{R}^T \mathbf{u} - \frac{\mathbf{y}}{\lambda} \right\|_2^2 \quad \text{subject to} \quad \|\mathbf{u}\|_2 \leq 1. \quad (4.55)$$

This problem is handled efficiently by standard methods, and thus we have a reasonable way of evaluating the proximal operator when no closed-form is available. It is of particular interest for implementing iterative thresholding algorithms.

Solving the primal-dual problem

Solving the primal-dual problem requires a saddle-point search method. This is made easier by the fact that L is convex-concave. The optimality condition is

$$\begin{cases} 0 \in \partial_{\mathbf{x}}L(\mathbf{x}^*, \mathbf{u}^*) \\ 0 \in -\partial_{\mathbf{u}}L(\mathbf{x}^*, \mathbf{u}^*) \end{cases} \quad (4.56)$$

which is equivalent to

$$\begin{cases} \mathbf{x}^* \in \mathbf{x}^* + \tau \partial_{\mathbf{x}}L(\mathbf{x}^*, \mathbf{u}^*) \\ \mathbf{u}^* \in \mathbf{u}^* - \sigma \partial_{\mathbf{u}}L(\mathbf{x}^*, \mathbf{u}^*) \end{cases} \Leftrightarrow \begin{cases} \mathbf{x}^* = (I + \tau \partial_{\mathbf{x}}L)^{-1}(\mathbf{x}^*, \mathbf{u}^*) \\ \mathbf{u}^* = (I - \sigma \partial_{\mathbf{u}}L)^{-1}(\mathbf{x}^*, \mathbf{u}^*) \end{cases} \quad (4.57)$$

for $\tau, \sigma > 0$. This suggest a fixed-point iteration of the form

$$\begin{cases} \mathbf{x}^{k+1} = (I + \tau \partial_{\mathbf{x}}L)^{-1}(\mathbf{x}^k, \mathbf{u}^k) \\ \mathbf{u}^{k+1} = (I - \sigma \partial_{\mathbf{u}}L)^{-1}(\mathbf{x}^{k+1}, \mathbf{u}^k) \end{cases} \quad (4.58)$$

This iteration is known in the optimization community as the *Primal-Dual Hybrid Gradient* (PDHG) algorithm [68].

Concretely the resulting reconstruction algorithm alternates between the following steps :

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \frac{1}{2\tau} \|\mathbf{x} - \mathbf{x}^k\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \mathbf{x}^\top \mathbf{R}^\top \mathbf{u}^k \quad (4.59)$$

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \frac{1}{2\sigma} \|\mathbf{u} - \mathbf{u}^k\|_2^2 - \lambda \mathbf{u}^\top \mathbf{R}\mathbf{x}^{k+1} + \delta_B(\mathbf{u}) \quad (4.60)$$

$$= \Pi_B (\mathbf{u}^k + \lambda \sigma \mathbf{R}\mathbf{x}^{k+1}) \quad (4.61)$$

where $\Pi_B(\mathbf{u})$ denotes the orthogonal projection on the ball B . The first step is a linear estimation step. It is the most costly, since it consists in solving a system involving matrix $\mathbf{K}^\top \mathbf{K} + \tau^{-1} \mathbf{I}$. Implementation remarks made for the re-weighted methods apply in this case as well. Computing $\Pi_B(\mathbf{u})$ corresponds to evaluating a proximal map and is inexpensive here.

Remark : Dual and primal-dual methods are readily extended to the Huber potentials $h_{p,\epsilon}$ based on the same arguments.

Lagrange-duality-based methods

Methods in this class are obtained by transforming the initial unconstrained problem into a constrained problem. The underlying design principle is again to perform a splitting into simpler sub-problems. The following constrained problem is equivalent to (4.10) :

$$\min_{\mathbf{x} \in \mathbb{R}^N} \quad \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \frac{\lambda}{p} \sum_{i=1}^r \|\mathbf{u}_i\|_2^p \quad \text{subject to} \quad (\mathbf{R}\mathbf{x})_i = \mathbf{u}_i \quad (4.62)$$

We introduce the *Augmented Lagrangian* function $L(\mathbf{x}, \mathbf{u}, \mathbf{p})$:

$$L(\mathbf{x}, \mathbf{u}, \mathbf{p}) = \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \frac{\lambda}{p} \sum_{i=1}^r \|\mathbf{u}_i\|_2^p + \mathbf{p}^\top (\mathbf{R}\mathbf{x} - \mathbf{u}) + \frac{c}{2} \|\mathbf{R}\mathbf{x} - \mathbf{u}\|_2^2 \quad (4.63)$$

with parameter $c > 0$. As its name indicates, the Augmented Lagrangian function is the Lagrangian function to which a quadratic penalty of the constraints was added. We note that L is a saddle-function.

In the *Method of Multipliers* [69] ones solves the saddle-point problem using iteration

$$(\mathbf{x}^{k+1}, \mathbf{u}^{k+1}) = \arg \min_{(\mathbf{x}, \mathbf{u})} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \frac{\lambda}{p} \sum_{i=1}^r \|\mathbf{u}_i\|_2^p + (\mathbf{p}^k)^\top (\mathbf{R}\mathbf{x} - \mathbf{u}) + \frac{c}{2} \|\mathbf{R}\mathbf{x} - \mathbf{u}\|_2^2 \quad (4.64)$$

$$\mathbf{p}^{k+1} = \mathbf{p}^k + c (\mathbf{R}\mathbf{x}^{k+1} - \mathbf{u}^{k+1}) \quad (4.65)$$

Generally, update step (4.64) is as difficult to compute as the original problem. A possible remedy to this situation is to decouple (4.64) in two minimization problems by fixing alternatively one of the variables. This is known as the *Alternating Directions Method of Multipliers* (ADMM) :

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + (\mathbf{p}^k)^\top (\mathbf{R}\mathbf{x} - \mathbf{u}^k) + \frac{c}{2} \|\mathbf{R}\mathbf{x} - \mathbf{u}^k\|_2^2 \quad (4.66)$$

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \frac{c}{2} \|\mathbf{u} - \mathbf{R}\mathbf{x}^k\|_2^2 + (\mathbf{p}^k)^\top (\mathbf{R}\mathbf{x}^k - \mathbf{u}) + \frac{\lambda}{p} \sum_{i=1}^r \|\mathbf{u}_i\|_2^p \quad (4.67)$$

$$\mathbf{p}^{k+1} = \mathbf{p}^k + c (\mathbf{R}\mathbf{x}^k - \mathbf{u}^k) \quad (4.68)$$

Similarly to other methods, we obtain an algorithm that alternates between solving quadratic problems and proximal maps. An implementation advantage is that proximal maps are known in closed form.

4.3.4 Practical performance on FDOT reconstruction

Apart from IRNM, the optimization methods described above have been analyzed. Proofs of convergence, and convergence rates, are available under a number of more or less strict assumptions; see above references. Without entering into details, we mention that in the most favorable case (a well-conditioned matrix \mathbf{K} , for instance $\mathbf{K} = \mathbf{I}$) the fastest algorithms achieve linear convergence rate. In the general case we can only expect sub-linear convergence. When the convergence rate is sub-linear, the performance in early iterations is crucial. In addition, since a number of implementation choices in each method are regularization dependent, it is difficult to single out one dominating method in terms of execution time. The choice of which algorithm to use will therefore be driven by practical comparisons.

In order to give an idea of the practical performance, we present a short case study. We simulate a FDOT setup with point sources, point detectors, and 27360 source-detector pairs. The simulated object is a cylinder with diameter of 25mm, and a height of 40mm. It contains fluorescent rods of diameter 5mm, and height 10mm. The rods are situated 6.5mm away from the boundary, and approximately 8mm from each other. We use a mesh made of 29719 tetrahedra for reconstruction. The discrete forward model was presented in Chapter 2. Reconstruction is made using total variation as a regularizer. We implemented the methods presented above and run the resulting algorithms for a duration of three hours each.

Figure 4.5 shows the objective value as function of number of iterations. The per-iteration execution time varies greatly depending on the method which explains the difference in iteration numbers. FISTA and PDHG iterations are inexpensive compared to IRNM and ADMM. This not surprising since each iteration of the latter methods requires to solve a linear problem. With the exception of ADMM, all the methods have converged. We note that IRNM stabilized at a higher objective value than FISTA and PDHG. The reason for this is that we do not perform the full inversion of the linear system in IRNM. Rather, we use PCG and stop the algorithm after a fixed number of iterations. In practice, this is mandatory because the linear systems become highly ill-conditioned as the solution becomes sparser. Thus, the number of PCG iterations needed for full inversion is unaffordable.

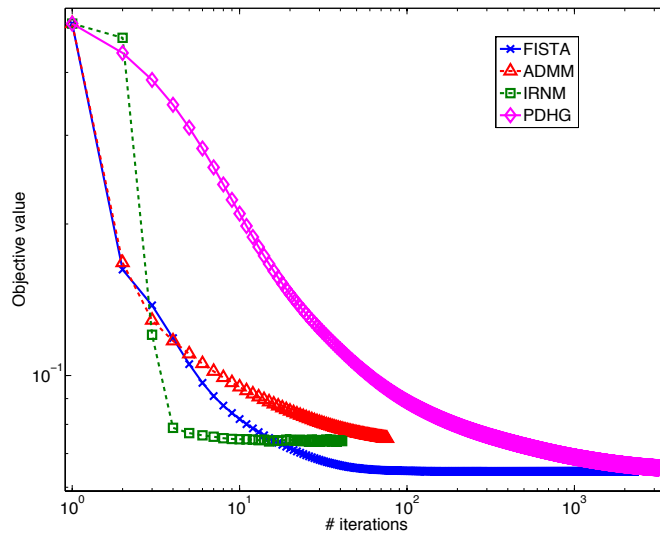


Figure 4.5: Total variation regularization. Evolution of the objective value as a function of iterations for different methods.

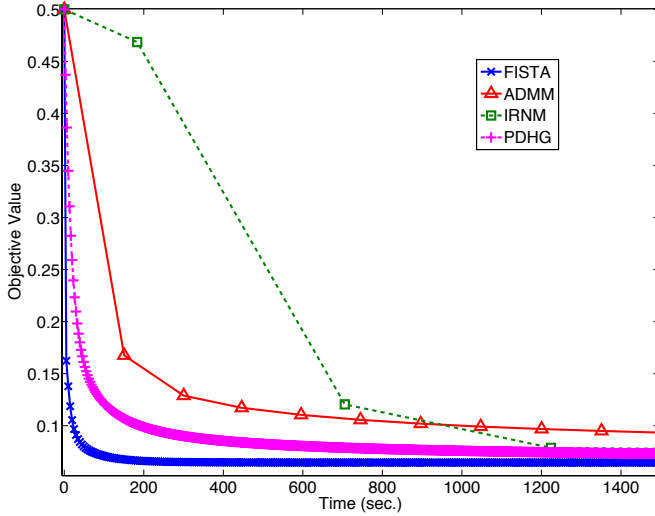


Figure 4.6: Total variation regularization. Evolution of the objective value as a function of time for different methods.

The evolution of the objective as a function of time over a window of 25 minutes is presented in Fig. 4.6. With an execution time budget of 25mins, one could consider FISTA, PDHG or IRNM. For total variation regularization, the best performance is achieved by FISTA which stabilizes after about 300s. As a matter of comparison, ℓ_2 regularization with $\mathbf{R} = \nabla$ requires around 240s. It is noteworthy that with efficient algorithms sparsity-inducing regularization can be performed at a computational cost that is not far off from that of linear methods.

Table 4.1: Optical coefficients for the phantoms used in the experiments.

Wavelength	Absorption (mm^{-1})	Reduced scattering (mm^{-1})
Excitation $\lambda^{\text{in}} = 655\text{nm}$	0.018	1.68
Emission $\lambda^{\text{out}} = 702\text{nm}$	0.017	1.66

4.4 Reconstruction experiments : methods

4.4.1 Phantom experiments

The setup employed to acquire FMT data is depicted in Fig. 1.1. We use a non-contact, trans-illumination setup in the continuous wave mode. The beam of a diode laser emitting at 655nm is focused onto the surface of a cylindrical phantom containing either one or two fluorescent inclusions. A sensitive CCD camera (iKon-M, Andor Technology, Belfast, Northern Ireland), cooled to -60°C to reduce dark-counts, records the light emitted from the opposite side of the phantom. An interference filter is placed in front of the camera objective ($f/1.8$, $f = 50\text{mm}$, Linos AG, Goettingen, Germany) to block the light emitted by the laser. In order to further reduce the detection of spurious signal from the laser, a clean-up interference filter is placed between the laser diode and the focusing lens. To obtain a complete data set, the sample is rotated with respect to the laser/camera reference and images are taken every 10° .

Two cylindrically shaped phantoms made of silicon with diameters of 35mm and 25mm respectively were used in the experiments. The 25mm phantom contained a cylindrical hole (4mm in diameter) parallel to the symmetry axis and the larger phantom contained two holes (4mm). Otherwise, the cylinders were homogeneous. Absorption and scattering coefficients were adapted to be similar to those of biological tissue by adding the required amounts of india ink as an absorber and titanium oxide as a scatterer. The values of the optical coefficients are shown in table 4.1. The holes were filled with an aqueous solution of Alexa Fluor 680 (Invitrogen AG, Basel, Switzerland). The absorption and scattering coefficients of this solution were adapted to those of the silicon by adding the adequate amounts of india ink and Intralipid.

We adjusted the noise level by changing the integration time and laser power in-

cident on the phantoms. The laser power was adapted by means of OD filters placed in the laser beam. Computing the SNR required a reference measurement, for which the noise was negligible. For each experiment, such a reference was obtained using a long integration time and high laser power. The reference measurement was then fitted to the noisy measurements with least squares to yield $\hat{\mathbf{M}}$. Note that in this methodology, measurements that are obtained with a long integration time and high laser power are essentially considered as noise-free. Therefore, we do not give an SNR value in that case.

4.4.2 Simulations

In simulation we modeled the geometry employed in the phantom experiments. The fluence rate of the excitation laser and the concentration were always set to one. Poisson noise was added to the simulated measurements in the following manner. Denoting $\hat{\mathbf{m}}$ the noise-free measurement vector output by the forward model, we generate the noisy vector

$$\mathbf{m} \sim \frac{1}{\gamma} \mathcal{P} \{ \gamma \hat{\mathbf{m}} \} \quad (4.69)$$

where γ is a parameter used to adjust the noise level, and $\mathcal{P} \{ \mathbf{x} \}$ is the poisson distribution with mean \mathbf{x} . We assess the noise level with the signal to noise ratio, $\text{SNR} = 10 \log_{10} (\|\hat{\mathbf{m}}\|_2^2 / \|\hat{\mathbf{m}} - \mathbf{m}\|_2^2)$. Because the forward model is linear, the SNR and the CNR are invariant to scaling, working with unit excitation and concentration is a valid approach.

4.4.3 Contrast-to-noise ratio

Whenever possible, we assessed reconstruction quality by visual inspection. Nevertheless, this approach has two drawbacks. First, visual inspection is subject to personal appreciation. Second, in experiments where several parameters are varied, inspecting all the reconstructed images becomes intractable. For large studies, one would rather employ a criterion that can be computed automatically on all reconstructions.

The *contrast-to-noise ratio* (CNR) is a measure that indicates if a localized feature in an image is well discernable or lost in the image noise[70]. In [71] the authors use the CNR to assess the performance of a FDOT reconstruction system.

The CNR is defined as the image contrast between a feature that is to be detected and the background, divided by a measure of the image noise. Specifically,

$$\text{CNR} = \frac{\mu_{\text{ROI}} - \mu_{\text{BCK}}}{(w_{\text{ROI}} \sigma_{\text{ROI}}^2 + w_{\text{BCK}} \sigma_{\text{BCK}}^2)^{1/2}} \quad (4.70)$$

where μ_{ROI} and μ_{BCK} are the mean concentration values in the ROI (the region of interest to which the feature is confined) and background respectively, σ_{ROI} and σ_{BCK} are the variances, and w_{ROI} and w_{BCK} are weighting factors.

In our CNR computations, the ROI was defined by the actual fluorophore inclusions. We treated as background the signal in the complement of the ROI. The weights w_{ROI} and w_{BCK} were set to the fraction of area occupied by the ROI and the background respectively.

4.4.4 Reconstruction algorithm

For the reconstruction results presented in this chapter we employed the IRNM algorithm of Section 4.3.2 applied with a 2D forward model. Use of a 2D model is justified by the fact that all the phantoms considered were cylindrical and invariant by translation along the symmetry axis. We set the number of IRNM iterations to five, because we observed no further improvement of the cost after that in practice. We also confirmed this fact by visual inspection of the results. The reconstruction grids were restricted to the region enclosed 1.5mm away from the boundary of the sample. For instance, in the case of a disk sample with diameter 25mm, the grid would overlay the 23mm diameter disk that is centered on the sample. This restriction helps handling boundary artifacts. It is reasonable since the diffusion approximation is not valid close the boundaries, rendering reconstructions at the boundary meaningless.

In practice a crucial issue is the selection of the regularization parameter λ . Although it was done either by visual inspection, or based on the CNR in these results, we observed that the *L-curve* method [72] is an efficient approach in our problem. The L-curve is the curve obtained by plotting the penalty value at the estimate, against the misfit value at the estimate, for a varying regularization parameter. In practice we observed the expected L-shape of this curve. At the corner of the curve, changing the parameter the ratio between deterioration and improvement of misfit and penalty is best. Thus, one selects the regularization parameter

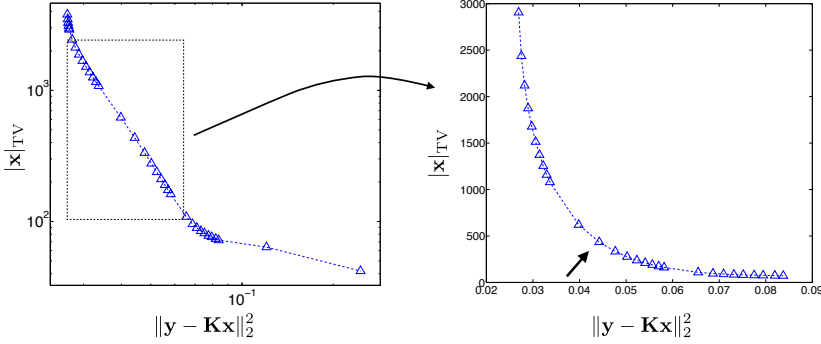


Figure 4.7: L-curve obtained for total variation regularization. The same problem as in the case study of section 4.3.4 (Fig 4.5) was used.

by identifying this corner. Fig 4.7 displays the L-curve obtained for the case study of section 4.3.4; the regularization potential is total variation.

4.5 Reconstruction experiments : results

We refer to Section 4.4 for a description of experimental conditions. Here we present and discuss reconstruction results obtained with ℓ_2 and ℓ_1 penalties (corresponding to $p = 2$ and $p = 1$, respectively). We recall that ℓ_2 potentials lead to linear reconstruction algorithms and do not induce sparsity. We consider regularization operators $\mathbf{R} = \mathbf{I}$ and $\mathbf{R} = \nabla$. The combination of ℓ_1 with \mathbf{I} corresponds to an *a priori* of spatial sparsity; that is, a localized fluorophore distribution. The combination of ℓ_1 with ∇ corresponds to total-variation, *i.e.* the *a priori* that the fluorophore distribution is piecewise-constant.

4.5.1 Experiment 1 : synthetic data, a single inclusion.

Here we present a representative reconstruction of simulated data. The investigated sample was a cylinder with a diameter of 25mm, containing a single cylindrical

inclusion with a diameter of 4mm. The inclusion was parallel to the symmetry axis of the phantom, and centered 7.5mm away from the axis. Poisson noise was added to obtain a SNR of 15dB. The regularization parameter λ was chosen by visual inspection, independently for the different methods. Note that the so-obtained values of λ corresponded to reconstructions with CNR close to the highest values (as a function of λ).

Figures 4.8 and 4.9 display the reconstructions. The values of CNR for these reconstructions are given in Table 4.2. Regardless the regularization operator, ℓ_1 reconstructions show less artifacts and better localization accuracy. As we notice on the cross-sections in Figs 4.8-(3) and 4.9-(3), the ℓ_2 reconstructions suffer from the expected *over-smoothing* effect. This effect is not present in ℓ_1 reconstructions. On Fig 4.8-(3), we observe that for $\mathbf{R} = \mathbf{I}$, the ℓ_1 reconstruction found a marker distribution that is smaller than the true distribution. This is expected because the algorithm induces spatial sparsity. In that case, it did select a single basis function, which is the one that explains best the observed data. The overall superiority of the ℓ_1 method is confirmed by the CNR values of Table 4.2.

Finally, results of Figures 4.8 and 4.9 show that quantification is not directly achieved by the optimization algorithm, even in simulation. A calibration step is necessary; see experiment 6.

4.5.2 Experiment 2 : phantom data, a single inclusion.

This experiment confirms on data acquired on phantom, the results obtained on synthetic data in Experiment 1. The geometry of the phantom is the same as in Experiment 1. It is a cylinder with diameter 25mm, containing a 4mm cylindrical inclusion centered 7.5mm away from the axis. The fluorophore concentration is 100nM. The data was acquired using an integration time of 86ms, and laser power of 0.3mW. Note that these experimental conditions are acceptable for *in-vivo* imaging. The noise-level was high, with a computed SNR of 8.7dB. Similarly to above, the regularization parameter λ was chosen by visual inspection for the reconstruction. We performed reconstructions for $\mathbf{R} = \mathbf{I}$ and $\mathbf{R} = \nabla$. CNR values are given in Table 4.2.

Figure 4.10 displays the reconstruction obtained with $\mathbf{R} = \mathbf{I}$ (in arbitrary units). Both methods located accurately the marker, but the ℓ_1 image contains less artifacts.

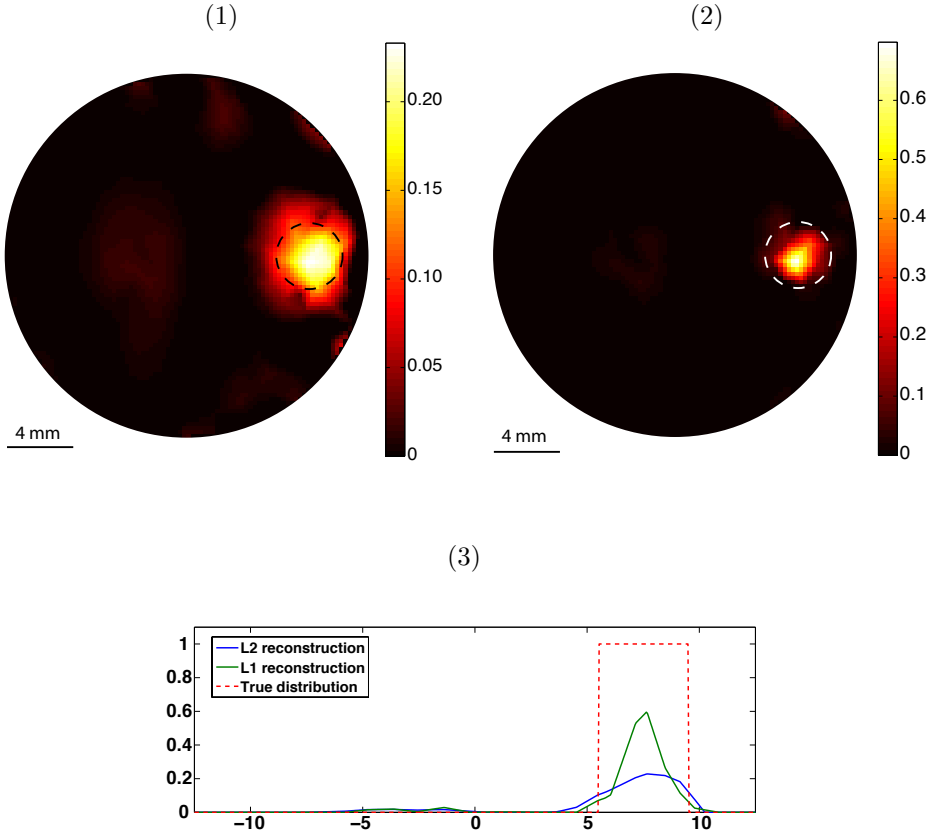


Figure 4.8: Experiment 1 : synthetic data, one inclusion with diameter 4mm in a cylinder with diameter 25mm. Regularization operator $\mathbf{R} = \mathbf{I}$. (1) ℓ_2 regularization; (2) ℓ_1 regularization; (3) Cross-sections of the reconstructions along a horizontal line.

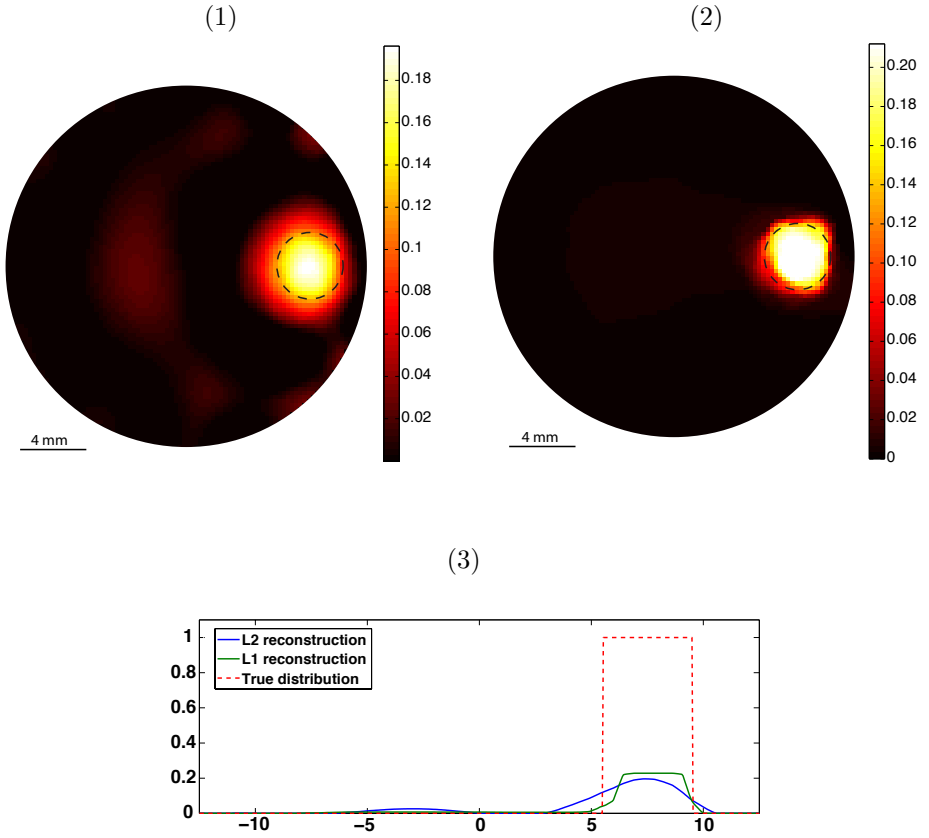


Figure 4.9: Experiment 1 : synthetic data, one inclusion with diameter 4mm in a cylinder with diameter 25mm. Regularization operator $\mathbf{R} = \nabla$. (1) ℓ_2 regularization; (2) ℓ_1 regularization; (3) Cross-sections of the reconstructions along a horizontal line.

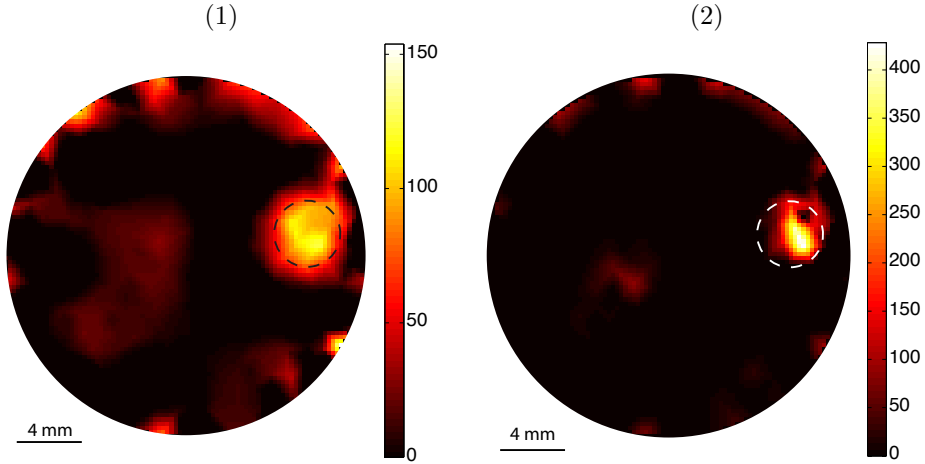


Figure 4.10: Experiment 2 : data acquired on phantom. One inclusion with diameter 4mm in a cylinder with diameter 25mm. (1) $\mathcal{R} = \mathbf{I}$, ℓ_2 -regularization; (2) $\mathcal{R} = \mathbf{I}$, ℓ_1 -regularization. Arbitrary units.

Similar to what was observed in simulation in Experiment 1, the ℓ_1 reconstruction is more localized than the true inclusion, and the reconstructed value is higher than the value obtained with ℓ_2 . ℓ_1 regularization with $\mathcal{R} = \mathbf{I}$ has favored a single basis function with a high weight to explain the data. Although the values obtained with ℓ_1 and ℓ_2 differ, the average values inside the inclusion do correspond. We found 119 for ℓ_1 and 124 for ℓ_2 . This indicates that the two methods reconstruct the same quantity of fluorophore.

4.5.3 Experiment 3 : phantom data, two inclusions.

There, we test the reconstruction algorithm on data acquired on a phantom with two inclusions. The phantom is a cylinder with a diameter of 35mm. It contains two inclusions with diameter 4mm, centered 10mm away from the axis. The distance between the centers of the two inclusions is 15mm. The phantom is larger than in

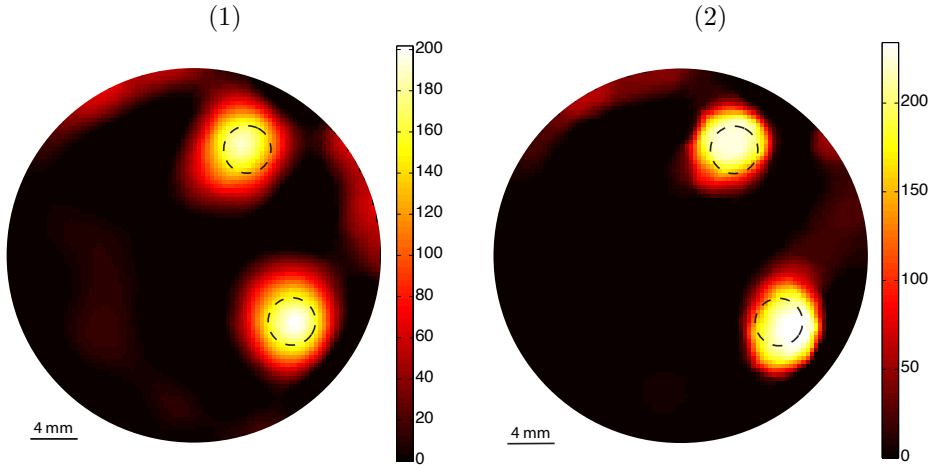


Figure 4.11: Experiment 3 : data acquired on phantom. Two inclusions with diameter 4mm in a cylinder with diameter 35mm. (1) $\mathbf{R} = \nabla$, ℓ_2 -regularization; (2) $\mathbf{R} = \nabla$, ℓ_1 -regularization. Arbitrary units.

the two previous experiments, and the marker inclusions are buried deeper inside. The fluorophore concentration is 200nM in both inclusions. The integration time is 1s, and the laser power 30mW. Again, these values are compatible with *in-vivo* imaging. Since the integration time and laser power are high enough, this data set is considered to be noise-free. We performed reconstructions for $\mathbf{R} = \mathbf{I}$ and $\mathbf{R} = \nabla$, and chose the regularization parameter visually. CNR values are given in Table 4.2.

Reconstructions for $\mathbf{R} = \nabla$ are shown in Figure 4.11. In this case, both methods yield comparable reconstructions. The ℓ_1 method, however, yields slightly less artifacts. In addition, the ℓ_1 image is composed of sharper transitions, between constant regions. This leads to a better CNR value.

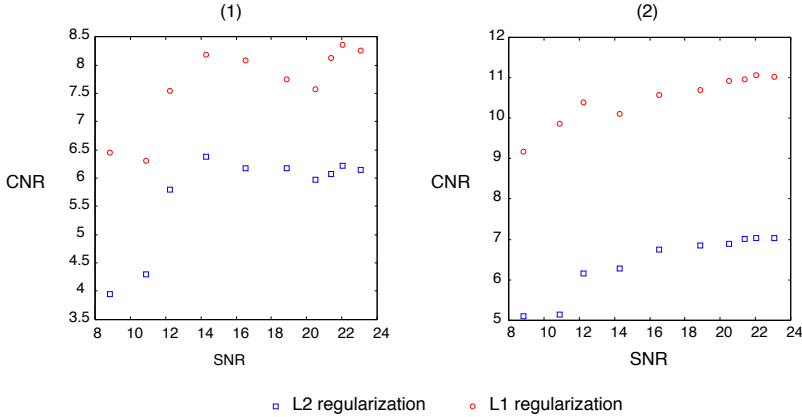


Figure 4.12: Experiment 4. Results obtained with measured data. CNR as the function of the noise level on the measurements. (1) $\mathbf{R} = \nabla$. (2) $\mathbf{R} = \mathbf{I}$.

4.5.4 Experiment 4 : CNR study, phantom data, a single inclusion.

Here, we investigate the robustness of the reconstruction to acquisition noise. The CNR presented is used as a performance criterion. The phantom is the same as in Experiments 1 and 2. The fluorophore concentration is 100nM. We vary the SNR of the measurements by changing the integration time and laser power. The integration time ranges from 0.13s to 1.53s, and the laser power is either 0.3mW, or 3mW. The regularization parameter is set such as to maximize the CNR.

The results of this experiment are presented in Figure 4.12. We notice that the ℓ_1 CNR is consistently above the corresponding ℓ_2 CNR, confirming the trend observed in the previous experiments. This supports the adequacy of ℓ_1 regularization for reconstructing localized inclusions.

Table 4.2: CNR values for reconstructions obtained in Experiments 1, 2 and 3

Experiment	CNR			
	$\mathbf{R} = \mathbf{I}$		$\mathbf{R} = \nabla$	
	ℓ_2	ℓ_1	ℓ_2	ℓ_1
Experiment 1	7.4	8.7	7.7	11.2
Experiment 2	6.4	8.1	6.2	10.1
Experiment 3	6.6	8.9	7.7	9

4.5.5 Experiment 5 : CNR study, synthetic data, a single inclusion of varying size.

The protocol of Experiment 4 was reproduced in simulation. In simulation, we varied the input SNR by changing the parameter γ of Section 4.4.2. In addition, we repeated the experiment for various diameters of the inclusion : 2mm, 4mm and 6mm. This experiment enabled to confront simulation and experimental results, as well as to test the algorithm on two other inclusion sizes.

The results are displayed in Figures 4.13 and 4.14. These simulations are in agreement with Experiment 4 : ℓ_1 achieves best CNRs. As expected, \mathbf{I} is more efficient than ∇ on small inclusions (2mm), whereas ∇ performs better on large inclusions (6mm).

4.5.6 Experiment 6 : Quantification, synthetic data

To further address the quantification issue, we simulate again the setup of Experiments 1 and 2. We vary the concentration (in arbitrary units) from 1 to 10, and evaluate the maximum of the reconstructed concentration. The reconstructions are performed using $\mathbf{R} = \mathbf{I}$. The SNR of the synthetic data is 20dB. The results are presented in Figure 4.15. We see that both ℓ_2 and ℓ_1 estimates are linear functions of the marker concentration. From the linear trend, we may infer that an appropriate calibration of the system would enable the proposed reconstruction method to quantify the fluorophore.

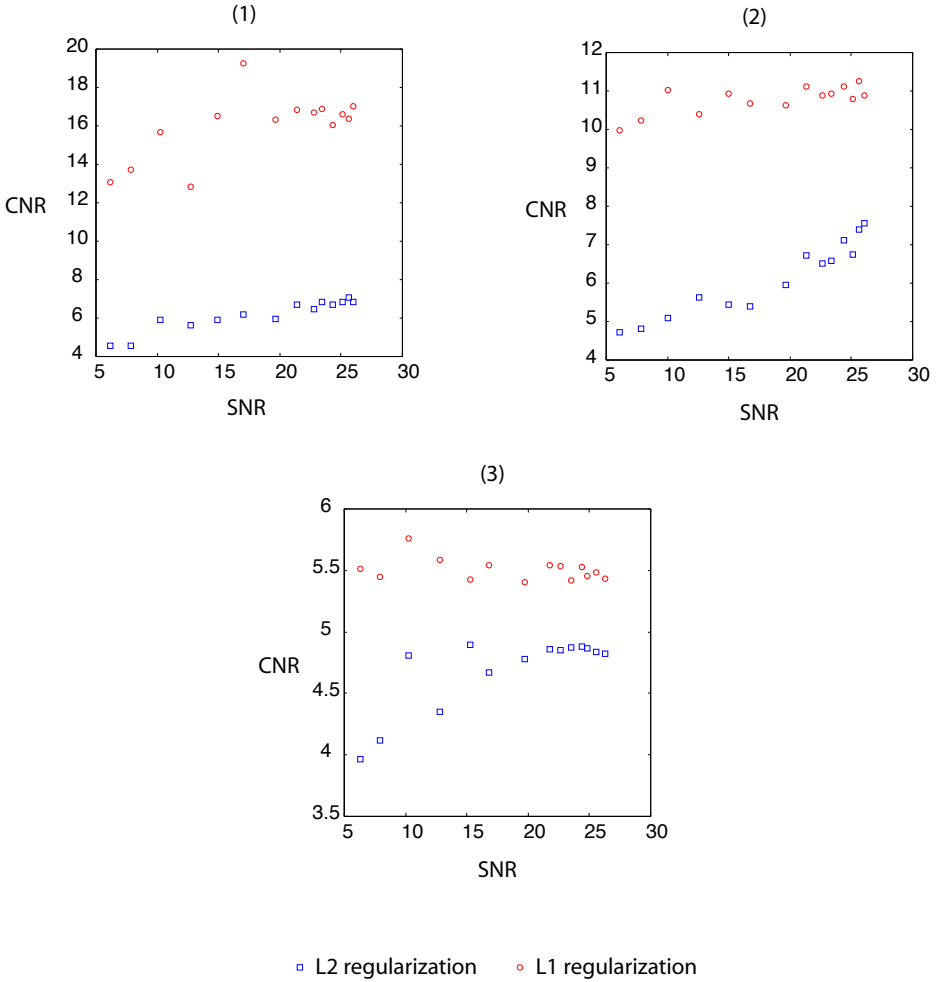


Figure 4.13: Experiment 5. Results obtained for synthetic data with $\mathbf{R} = \mathbf{I}$. CNR as the function of the noise level on the measurements. (1) fluorophore inclusion with 2mm diameter. (2) fluorophore inclusion with 4mm diameter. (3) fluorophore inclusion with 6mm diameter.

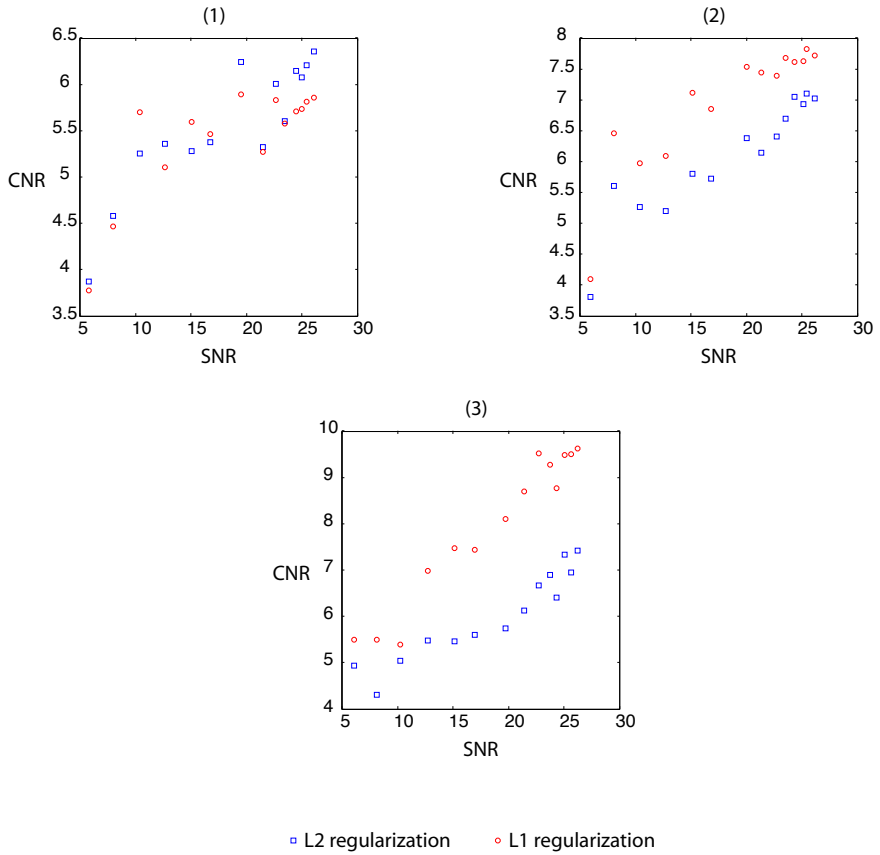


Figure 4.14: Experiment 5. Results obtained for synthetic data with $\mathbf{R} = \nabla$. CNR as the function of the noise level on the measurements. (1) fluorophore inclusion with 2mm diameter. (2) fluorophore inclusion with 4mm diameter. (3) fluorophore inclusion with 6mm diameter. .

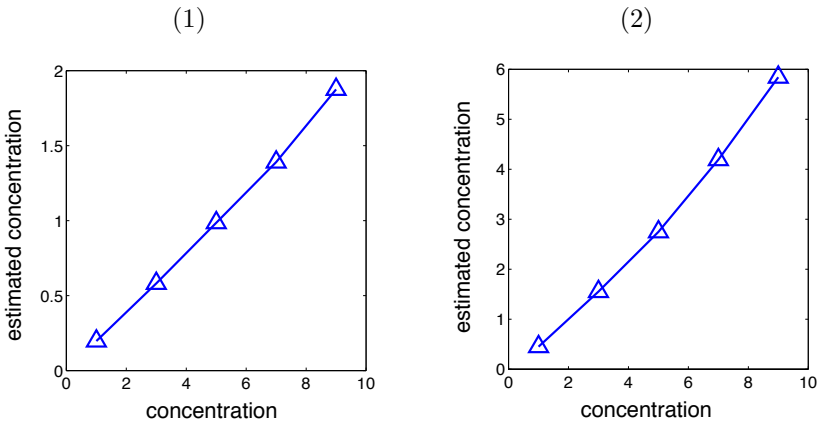


Figure 4.15: Experiment 6 : maximum value of the reconstructed concentration as a function of the true concentration; reconstructions obtained using $\mathbf{R} = \mathbf{I}$. (1) for ℓ_2 regularization; (2) for ℓ_1 regularization. Synthetic data, arbitrary units.

Chapter 5

Sparsity-driven reconstruction for FDOT with anatomical priors

I^N¹ this chapter we propose a method based on $\ell_{2,1}$ -mixed-norm penalization for incorporating a structural prior in FDOT image reconstruction. The effect of $\ell_{2,1}$ -mixed-norm penalization is twofold : first, a sparsifying effect which isolates few anatomical regions where the fluorescent probe has accumulated, and second, a regularization effect inside the selected anatomical regions. After formulating the reconstruction in a variational framework, we derive a practical algorithm based on the discrete forward model presented in Chapter 3 and tailored convex optimization methods according to the strategies presented in Chapter 4. The proposed method includes as particular cases other sparsity-promoting regularization methods such as ℓ_1 -norm penalization and total variation penalization. Results on synthetic and experimental data are presented.

¹This chapter is based on our published paper [73].

5.1 Introduction

Recent developments of multi-modal imaging systems have enabled to couple structural (usually high-resolution) and functional (low-resolution) imaging modalities; the most noticeable examples are CT-PET [74] and MRI-PET [75]. In the field of diffuse optical imaging, hybrid CT-FDOT systems have been designed [76, 77, 78, 79, 80], and the combination MRI-FDOT is also being explored [81, 82, 83]. Not only do these systems make it possible to register anatomic and functional image, but structural information can also be used to improve the accuracy of the FDOT reconstruction. In the context of fluorescence tomography, anatomical priors were either used to enhance the forward modeling, or directly in the reconstruction procedure. In [84] the authors report that improving the forward model with structural information leads to more accurate images. On the reconstruction side, the structural *a priori* knowledge often takes form of a labeling of the pixels, derived from a segmentation of the anatomical image. That type of labeling was used in previous works to design edge-preserving regularization [85], hierarchical Bayesian models [86], and space-varying quadratic regularization [87, 88, 6, 33], which helped improving reconstruction quality; resolution in particular.

In this contribution we propose to incorporate structural information into the reconstruction procedure of FDOT by means of a suitable sparsity-promoting regularization functional. We put the emphasis on group-sparsity, *i. e.* sparsity between groups of coefficients, rather than sparsity of the coefficients themselves. To that end, we introduce the $\ell_{2,1}$ -mixed-norm as a regularizer for FDOT reconstruction, and use it to incorporate anatomical information. The properties of the $\ell_{2,1}$ -mixed-norm lead to an inversion method that automatically selects the fluorescent parts of the anatomy, and focusses the reconstruction on these areas. Concretely the reconstruction algorithm consists in solving a convex optimization problem of the type that was present in Chapter 4. We focus on proximal-based methods presented in Section 4.3.2. To further speed up the algorithm, we provide closed form expressions of the proximal map that replace inner iterations when the regularization term belongs to a set of instances often encountered in practice. Finally, we show that our $\ell_{2,1}$ -mixed-norm framework includes as particular cases a number of sparsity-promoting regularizers such as ℓ_1 -norm and total variation (TV) semi-norm.

This chapter is organized as follows. In Section 5.2 we introduce the (2,1)-mixed-norm and explain how we use it to incorporate structural *a priori* knowledge.

After that, in Section 5.3 we present the reconstruction. And finally, in Section 5.4 we present results and compare the proposed method to existing regularization methods including *a priori* knowledge.

5.2 Group sparsity, anatomical *a priori*, and $\ell_{2,1}$ penalty

We assume that the structural *a priori* information comes in the form of a labeling of pixels in the reconstructed image. The labeling defines image segments that correspond to anatomical features; organs for instance. Such a labeling is typically obtained by segmenting another high-resolution structural image of the same specimen. Instances of labelings (partitioning of the domain into regions/segments), are presented in Section 5.4, Fig 5.1-(f), 5.2-(c) and 5.3-(e). Our present goal is to design an algorithm that uses the structural information given by the labeling as a soft-prior. This means that the reconstruction will not be restricted to specific regions; rather, all pixels are admissible, but incur a penalty related to which segment they belong.

We now introduce the $\ell_{2,1}$ -mixed-norm. Let $\mathbf{x} \in \mathbb{R}^N$ be a vector representing our image. We assume that the image is partitioned into S segments. We have $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_S)$, where each of the sub-vectors \mathbf{x}_i , $i \in [1..S]$ corresponds to the pixels in a segment. We will also employ the notation $\mathbf{x} = (\mathbf{x}_i)_{1 \leq i \leq S}$ to denote such a compound vector. The $\ell_{2,1}$ -mixed-norm of \mathbf{x} is defined as follows :

$$\|\mathbf{x}\|_{2,1} = \sum_{i=1}^S \|\mathbf{x}_i\|_2 \quad (5.1)$$

Notice that the labeling of the pixels is implicit in this notation. Observe that the $\ell_{2,1}$ -mixed-norm is a distance measure that applies ℓ_2 -norms inside segments and a ℓ_1 -norm across segments. This suggests the idea of an algorithm based on $\ell_{2,1}$ penalization. Similarly to ℓ_1 -norm penalization which promotes sparsity, we expect $\ell_{2,1}$ penalization to enforce sparsity across segments while retaining ℓ_2 regularization inside a segment. Formulated differently, we expect an algorithm based on $\ell_{2,1}$ penalization to isolate few anatomical groups, and to apply ℓ_2 regularization inside these groups. In such an algorithm, ℓ_2 penalization operates mainly on those pixels

contained in the selected segments which is likely to lead to a higher reconstruction accuracy.

The use of $\|\mathbf{x}\|_{2,1}$ on its own might not be entirely satisfactory. In particular it does not enforce smoothness constraints. We can therefore extend the idea to penalization of the more general term $\|\mathbf{R}\mathbf{x}\|_{2,1}$, where \mathbf{R} is a linear operator. Penalizing $\|\nabla\mathbf{x}\|_{2,1}$, for instance, would yield a solution with few segments having a high gradient, and constant (most likely 0) in the others.

Let us mention two special cases $\ell_{2,1}$ that we already encountered in Chapter 3 :

1. choosing \mathbf{R} to be the identity, and one group per pixel, leads to $\|\mathbf{R}\mathbf{x}\|_{2,1} = \|\mathbf{x}\|_1$; $\ell_{2,1}$ -regularization then corresponds ℓ_1 -regularization.
2. choosing $\mathbf{R} = \nabla$ and the groups corresponding to $(\partial_x, \partial_y, \partial_z)$ at each pixel, we obtain $\|\mathbf{R}\mathbf{x}\|_{2,1} = \|\mathbf{x}\|_{\text{TV}}$; $\ell_{2,1}$ -regularization is TV regularization in that case.

Although it is natural to choose a labeling that is provided by a prior anatomy-based segmentation, one could consider some variations of the method. For instance, several anatomical parts could be merged under the same label, as it would be the case if one wanted to mark the two lungs at the same time. The labeling is another degree of freedom of the method for defining regularization policies.

5.3 Reconstruction algorithm

By comparing expressions (4.9) and (5.1) we see that $\ell_{2,1}$ potentials are closely related to the potentials presented in Chapter 4 (setting $p = 1$). Expression (5.1) extends (4.9) in the sense that the sub-vectors $\mathbf{x}_1, \dots, \mathbf{x}_S$ are not assumed to be of the same size. Iterative methods presented in Section 4.3 are straightforwardly adapted to handle the $\ell_{2,1}$ -mixed-norm.

For reconstruction, we consider the following variational method, which is a generalization of the method described in (4.10) :

$$\arg \min_{\mathbf{x} \in C} \frac{1}{2} \|\mathbf{y} - \mathbf{K}\mathbf{x}\|_2^2 + \lambda \|\mathbf{R}\mathbf{x}\|_{2,1}, \quad (5.2)$$

where we have introduced a convex set of constraints C . We are interested in imposing simple constraints such as positivity ($C = \mathbb{R}_+^N$), or box constraints ($C =$

Table 5.1: Proximity mappings for $\ell_{2,1}$ regularization

Regularizer $\psi(\mathbf{x})$	Constraints set C	Proximal map $\text{prox}_{\psi}(\mathbf{x})$
$\lambda\ \mathbf{x}\ _{2,1}$	\mathbb{R}^N	$\left(\mathbf{x}_i \max \left(0, 1 - \frac{\lambda}{\ \mathbf{x}_i\ _2} \right) \right)_{i=1..S}$
$\lambda\ \mathbf{W}\mathbf{x}\ _{2,1}$	\mathbb{R}^N	$\left(\mathbf{x}_i \max \left(0, 1 - \frac{\lambda w_i}{\ \mathbf{x}_i\ _2} \right) \right)_{i=1..S}$
$\lambda\ \mathbf{x}\ _{2,1}$	\mathbb{R}_+^N	result of Proposition B.0.1
$\lambda\ \mathbf{x}\ _{2,1}$	$[0, \mathbf{b}[$	result of Proposition B.0.2
$\lambda\ \mathbf{R}\mathbf{x}\ _{2,1}$	Arbitrary convex set	No analytical expression, use an iterative algorithm.

Note : here $\mathbf{W} = \text{diag}(\mathbf{w}_1, \dots, \mathbf{w}_S)$ is a weighting matrix with one weight for each segment, and $\mathbf{b} > 0$.

$[0, \mathbf{b}[$). In this work, we implemented the proximal-based methods discussed in Section 4.3, in our reconstruction algorithm. The constraint set C is incorporated in these methods by replacing the unconstrained proximal map (4.32) by a constrained proximal map :

$$\text{prox}_{\psi}(\mathbf{x}) = \arg \min_{\mathbf{y} \in C} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \psi(\mathbf{y}) \quad (5.3)$$

As we mentioned, proximal-based methods are most effective when the proximal map is known in closed form. Using Proposition 4.3.4 in Chapter 4, and Propositions B.0.1 and B.0.2 in Appendix B, we obtain expressions summarized in Table 5.1. Interestingly it covers most of the cases encountered in practice.

5.4 Results

We present five experiments. The first one is made on synthetic data, while the last four are based on experimental data acquired on phantoms. In each experiment the algorithm is run using a range of regularization parameters, and the best reconstruction is selected based on visual assessment. This operation is repeated in

order to refine the parameter value. Reconstruction quality is also evaluated using contrast-to-noise ratio.

5.4.1 Experiment 1 : two dimensions, synthetic data, multiple inclusions

In this experiment we consider a cylindrical shape geometry with radius 12.5mm. The cross-section is shown in Fig 5.1-(f). There are four fluorophore inclusions, each with unit concentration (in arbitrary units). They are outlined by dashed circles. Otherwise, the cylinder is homogeneous with $\mu_a = 0.175\text{cm}^{-1}$ and $\mu'_s = 16.5\text{cm}^{-1}$. It is partitioned into seven regions (including background). The regions are displayed in grey levels. We assume that the fluorophore distribution is invariant by translation along the axis of the cylinder, which enables us to employ a 2D forward model in order to alleviate the computational burden. A triangular reconstruction grid with 1mm cell size is used. Whether it be simulated or experimental, the acquisition setup considered in this work is transillumination. For each acquisition, a point source (of the excitation light field) is generated by a laser beam at the surface of the measured object. An acquisition consists in sampling the fluorescence light field at several points on the surface opposite to the source position. These points are referred to as detectors in the following. In the present experiment we simulate 36 sources spaced every 10 degrees. For each source the light field is sampled at 90 detectors spaced every 2 degrees. Lastly, the inclusions are positioned rather close to the boundary, but some of them are also close to each other.

Figs 5.1-(a) to (e) display the reconstructions obtained with various regularizers. Fig 1-(f) presents the regions (or labelling) used as structural *a priori*. We differentiate separate compartments by employing distinct grey levels. In order to test the robustness of the scheme, we define more regions than inclusions, with various sizes compared to the inclusion size. We present results for $\|\nabla\mathbf{x}\|_2^2$, $\|\cdot\|_1$, total variation, $\|\mathbf{W}\mathbf{x}\|_{2,1}$ and $\|\mathbf{W}\nabla\mathbf{x}\|_{2,1}$. The matrix \mathbf{W} , which reflects our prior knowledge as well, is a weighting matrix allocating a weight of 1 inside the regions, and 2 in the background.

We notice that the locations of the reconstructed inclusions are correct for all methods. However, the three methods that do not include any structural *a priori* are unable to resolve the four inclusions. On the contrary, all inclusions are accurately recovered when structural information is incorporated. Also, recovered values are more accurate when a structural *a priori* is employed. Contrary to ℓ_2

and total variation regularizations, which lead to fluorophore distributions spread over large regions, the $\ell_{2,1}$ reconstructions are well localized in space. This enables the algorithm to recover accurate concentration values. The result obtained with ℓ_1 shares the same property. Overall, it is the $\|\mathbf{W}\nabla\mathbf{x}\|_{2,1}$ that yields the best result; the solution is smooth and accurate.

5.4.2 Experiment 2 : three dimensions, experimental data, two inclusions, accurate structural information

Here, we consider a cylindrical phantom with radius 12.5mm and height 50mm. The cylinder is equipped with two 3mm diameter longitudinal holes that can receive capillaries filled with an aqueous fluorophore solution; Alexa Fluor 680 (Invitrogen, 5791 Van Allen Way, Carlsbad, California, USA) in this case. Except for the two holes, the phantom is homogeneous with constant optical coefficients. It is made of silicon mixed with india ink and titanium oxide in order to match absorption and scattering coefficients of biological tissue. We have $\mu_a = 0.2\text{cm}^{-1}$ and $\mu'_s = 15\text{cm}^{-1}$.

Prior to optical measurements, the phantom was imaged with a micro-CT system, which enables us to outline its inner structure. Although the structure is very simple in that case, it is a ground truth to validate FDOT reconstruction, and a good starting point to define a structural *a priori*. A three dimensional reconstruction mesh is defined around the two inclusions. We use a mesh resolution of about 1mm in xy and 0.7mm in z , which results in a mesh with 46349 tetrahedra. The micro-CT image of the phantom with the two inclusions is displayed in Fig 5.2-(a) and (b). Data is acquired with a laser power of 4mW and an integration time of 200ms. The object was sampled for 220 source positions spaced every 18 degrees around the rotation axis, and 2mm in z .

The goal of this experiment is to assess the performance of $\ell_{2,1}$ regularization, in a scenario where the structural *a-priori* is close to the fluorophore distribution. For that, we partition the phantom into 5 regions : four cylinders, and the background. This situation is illustrated in Fig 5.2. Note that the background is also an admissible reconstruction region for the $\ell_{2,1}$ algorithm. Among the four cylinders, two are enclosing tightly the inclusions, and the other two are empty. The two empty regions are placed to demonstrate the behavior of the algorithm in presence of spurious regions. We compare the performance of $\ell_{2,1}$ regularization with three other methods based on ℓ_2 regularization that we will denote QR0, QR1 and QR2 (for Quadratic Regularization). QR0 is the standard Tikhonov regularization that

Table 5.2: CNR for Experiment 2

QR0 method	QR1 method	QR2 method	$\ell_{2,1}$ method
4.03	7.05	20.74	20.75

does not include any structural *a-priori*. For QR1, the structural *a-priori* is exploited by restricting the ℓ_2 reconstruction to those mesh-nodes contained in the four cylinders. Similarly, QR2 is restricted to the two cylinders that coincide with the inclusions; this is a very strong *a priori*.

Results for this experiment are presented in Fig 5.3. An xy cross-section through the regions employed is displayed in grey levels in Fig 5.3-(e). Note that the regularization parameter was tuned separately for each method. We see on Fig 5.3-(a) that QR0 locates the fluorophore correctly but is unable to resolve the two inclusions. On the contrary, QR1 and QR2 take advantage of the structural information to resolve the sources. QR1 however, wrongly reconstructs some fluorophore in the two empty regions, as we can see on Fig 5.3-(b). As expected, the result of QR2 (Fig 5.3-(c)) is almost perfect, since, in that case, the reconstruction is restricted to a region that coincide with the actual fluorophore distribution. Fig 5.3-(d) displays the result obtained with $\ell_{2,1}$ regularization. We use the weighted $\ell_{2,1}$ -norm with a weight of 1 for the four cylinder regions, and 10 outside. This choice of weights expresses that the fluorophore has more chance to be in the four cylinders than in the background. The latter is therefore penalized, but still an admissible region. As we can see on Fig 5.3-(d), the reconstruction is as good as with QR2 (strongest *a priori*). Contrary to what happened with QR1, the $\ell_{2,1}$ method is not hindered by the presence of the extra regions. Since there is a good correspondence between the *a priori* and the inclusions, this is what we expect. Indeed, the effect of the $\ell_{2,1}$ term is to select a few active regions for reconstruction, while setting the solution to zero elsewhere. In the present case, some regions enclose tightly the actual fluorophore distribution. The algorithm has selected these regions, which leads to an accurate reconstruction. These observations are confirmed by the CNR results compiled in Table 5.2.

Table 5.3: CNR for Experiment 3

QR method	$\ell_{2,1}$ method
7.05	9.4

5.4.3 Experiment 3 : three dimensions, experimental data, two inclusions, inaccurate structural information

The setup employed here is same as for Experiment 2. The difference is in the structural *a priori* that we define. In Experiment 2 the *a priori* contained two segments coinciding with the inclusions. At present we study the behavior of $\ell_{2,1}$ regularization in the presence of less accurate prior knowledge. For this we segment the phantom in three regions : one large ellipse containing the first inclusion, one smaller cylindrical region containing the second inclusion, and the background. The xy cross-section through the regions is presented in grey levels in Fig 5.4-(c). There is a close correspondence between structural information and fluorophore distribution for one inclusion, while the second one is embedded in a much larger region. Again we compare $\ell_{2,1}$ regularization with ℓ_2 regularization including structural knowledge. We denote by QR a ℓ_2 based reconstruction restricted to the elliptical and cylindrical regions.

The results are presented in Fig 5.4-(a) and (b), for QR and $\ell_{2,1}$, respectively. CNR is also computed and shown in Table 5.3. We observe that $\ell_{2,1}$ outperforms slightly QR in terms of localization. Indeed, the inclusion placed in the accurate region is perfectly recovered by both methods, while $\ell_{2,1}$ yields a reconstruction more concentrated around the inclusion location in the less informative elliptical region. Overall, this result leads to think that $\ell_{2,1}$ penalty allows to exploit structural information at best. If the *a priori* is very informative about the fluorophore distribution, then we get an accurate reconstruction, and otherwise, we revert to a ℓ_2 -like behaviour.

5.4.4 Experiment 4 : three dimensions, experimental data, acquired *a priori* information

In the previous two experiments we employed arbitrary structural information so as to demonstrate the response of the proposed method to different *a priori* accuracies.

Table 5.4: CNR for Experiment 4

QR method	$\ell_{2,1}$ method
9.2	13.4

Now we present an experiment with a heterogeneous phantom, and we use structural information obtained from a segmentation of the CT image. Figure 5.5 shows the geometry of the measure object. The phantom has a cylindrical shape of the same size as the one used in experiments 2 and 3. In Fig 5.5-(a) we display the reconstruction mesh registered with the CT image. Then, Fig 5.5-(b) shows the xy cross-section through the CT image. We see that the cylinder is equipped with five holes; four small ones (3mm diameter), and a bigger one (5mm diameter). We realize a segmentation of the CT image to get the structural *a priori*. This results in the partition represented schematically in Fig 5.5-(c) : five regions corresponding to the holes represented in grey levels, numbered 1 to 5 on the figure, and the background. Optical properties in the phantom are inhomogeneous. Holes number 1 and 2 contain a solution with scattering and absorption five times higher than for the rest of the cylinder; otherwise, the optical coefficients have the same values as in experiments 2 and 3. A capillary filled with fluorophore (same characteristics as above) is inserted in hole number 1.

Figure 5.6 presents the reconstruction results. Similar to experiment 3, we compare ℓ_2 regularization restricted to regions 1 to 5 (QR method) to $\ell_{2,1}$ regularization using the segmentation (regions 1 to 5, and background) as *a priori*. The results are consistent with observations made in experiments 2 and 3. QR tends to reconstruct fluorophore wrongly over several holes. By contrast $\ell_{2,1}$ selects the correct region of the *a priori* for reconstructing, leading to a more faithful reconstruction. We display the CNRs for these reconstructions in Table 5.4, confirming visual inspection. This demonstrates the efficiency of the proposed method in a non-homogeneous phantom with a priori knowledge acquired from the CT image.

5.4.5 Experiment 5 : three dimensions, experimental data, quantification

Now we perform a quantification experiment to demonstrate the ability of the $\ell_{2,1}$ method to quantify the fluorescent probe. The same is also done for ℓ_2 regularization, as a control. In this case, we employ a cylindrical phantom with a single hole. Otherwise, the characteristics of the phantom remain the same as in the previous experiments. A capillary of known volume is filled with a fluorophore solution of concentration varying from 10^{-7}M to $8.92 \times 10^{-6}\text{M}$, and inserted into the hole. Then, the reconstructed fluorophore concentration is integrated over the domain, and monitored as a function of the actual concentration. We use a structural *a priori* composed of two regions : one is slightly larger than the hole containing the capillary, and the other one is the background. The region containing the hole is used as admissible region for the ℓ_2 regularization, similarly to experiments 2 and 4.

In Fig 5.7 we display the total reconstructed fluorescence as a function of concentration for ℓ_2 and $\ell_{2,1}$. The observed responses are linear, from which we can infer that an appropriate calibration would enable to quantify fluorophore concentration.

5.5 Conclusion

In this contribution we have shown how a sparsity promoting technique can be employed to incorporate structural priors in FDOT imaging. We presented a novel regularization scheme based on $\ell_{2,1}$ penalization, and proposed a practical numerical reconstruction method. We further refined the method by deriving analytical formulae for the proximal maps required in several configurations of interest. The proposed regularization method was demonstrated on synthetic and experimental data. These experiments lead us to think that a structural prior combined with the proposed method helps improving the reconstruction in terms of localization, and contrast. We observed that $\ell_{2,1}$ regularization systematically yields more realistic reconstructions of fluorophore distribution compared to ℓ_2 with hard constraints, when the *a priori* contains only partial information about shape and number of inclusions. In particular, artifacts are reduced and CNR is increased. Lastly, our experiments suggest that for uninformative *a priori*, both regularizations lead to similar results. This agrees with the intuition that if we use a single segment of large

size, the mixed-norm-based method reverts back to ℓ_2 regularization. At the other extreme (one region per node of the mesh), it is equivalent to a sparsity-promoting ℓ_1 regularization.

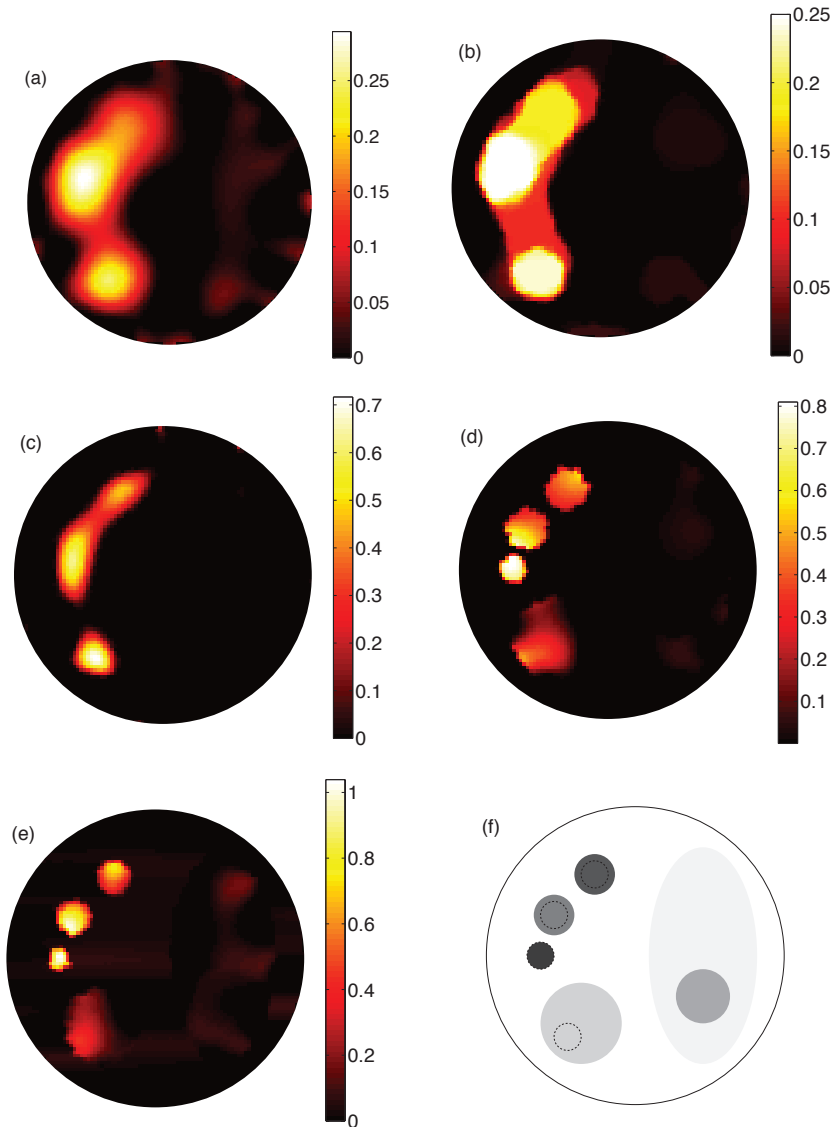


Figure 5.1: Cylinder geometry. Results of reconstruction for different regularizations : (a) Tikhonov ($\|\nabla \mathbf{x}\|_2^2$), (b) total variation, (c) ℓ_1 -norm, (d) Weighted $\ell_{2,1}$ -mixed-norm $\|\mathbf{W}\mathbf{x}\|_{2,1}$, (e) Weighted $\ell_{2,1}$ -mixed norm $\|\mathbf{W}\nabla \mathbf{x}\|_{2,1}$. Figure (f) shows the structural information (regions/labeling) in grey levels, and the fluorophore inclusions in dashed circles.

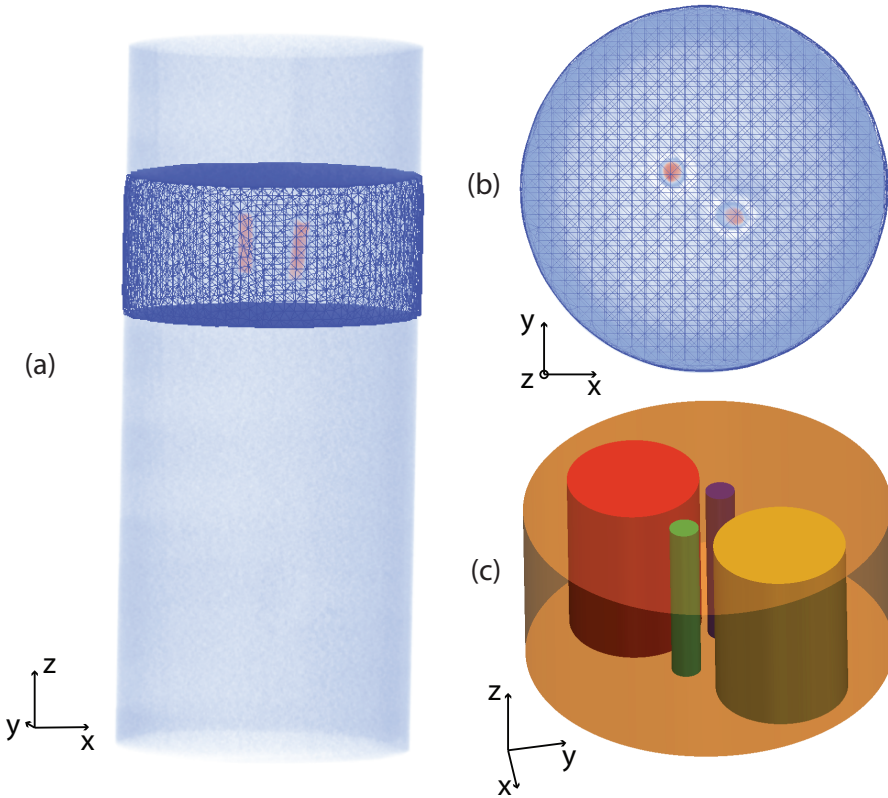


Figure 5.2: Geometry employed in Experiment 2. (a) and (b) registration of reconstruction mesh and CT image of the phantom. The capillaries containing the fluorophore appear in red on this image. (c) structural prior used in Experiment 2.

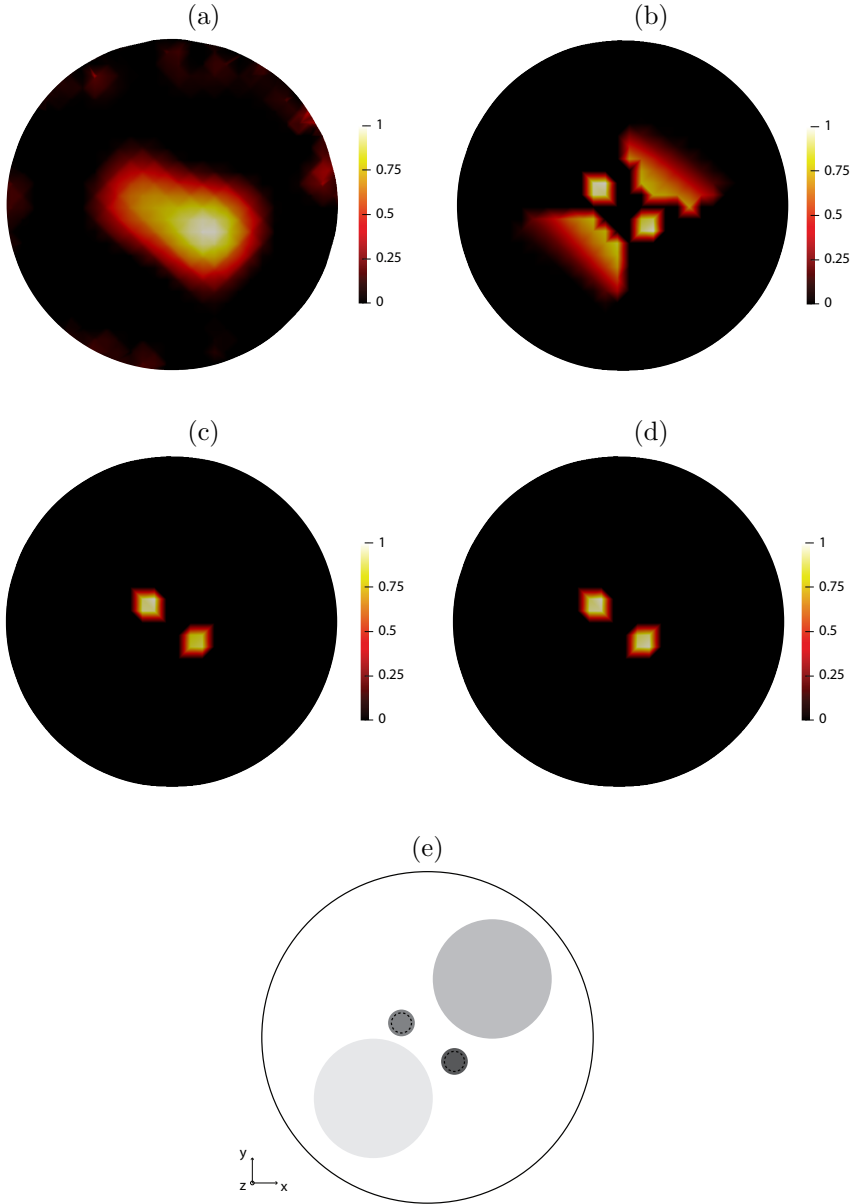


Figure 5.3: Results for Experiment 2, normalized concentration. (a) QR0, standard ℓ_2 regularization, (b) QR1, ℓ_2 regularization restricted to the four cylindrical regions of the structural *a priori*, (c) QR2, ℓ_2 regularization restricted to the two cylindrical regions containing the inclusions, (d) $\ell_{2,1}$ regularization, (e) xy cross-section through the regions used as structural *a priori* in grey levels, and sources in dashed cricles.

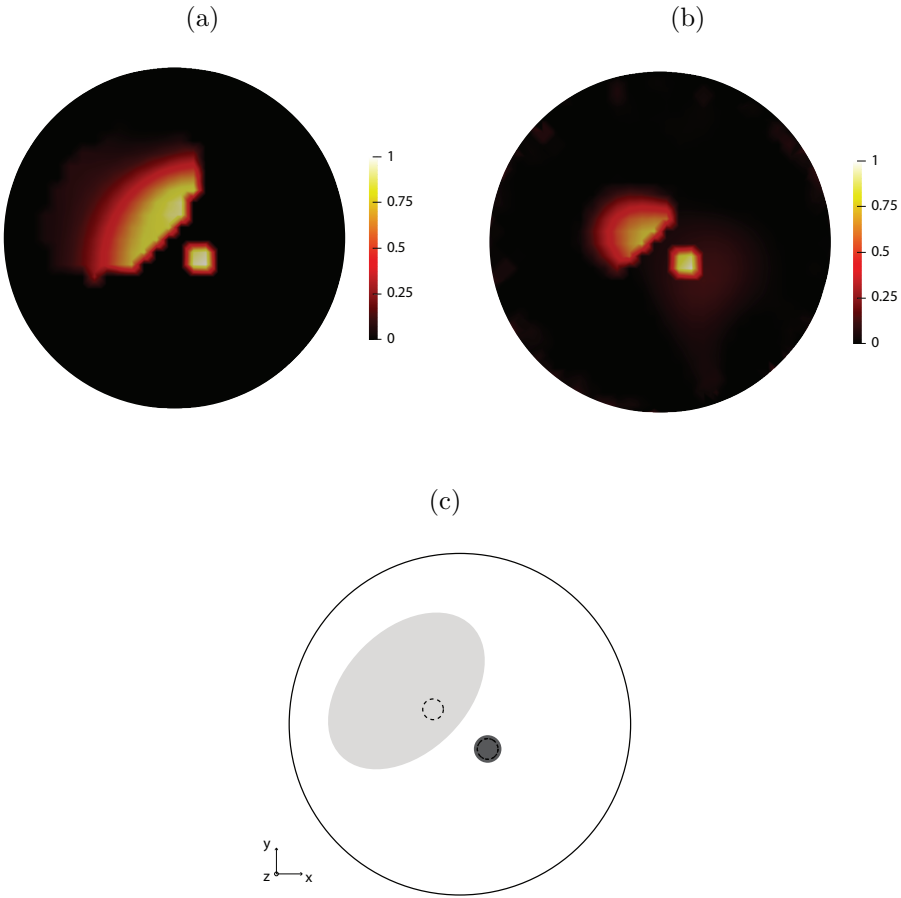


Figure 5.4: Results for Experiment 3, normalized concentration. (a) QR, l_2 regularization restricted to the two segments containing the fluorophore, (b) $l_{2,1}$ regularization, (c) xy cross-section through the regions used as structural *a priori* in grey levels, and sources in dashed circles.

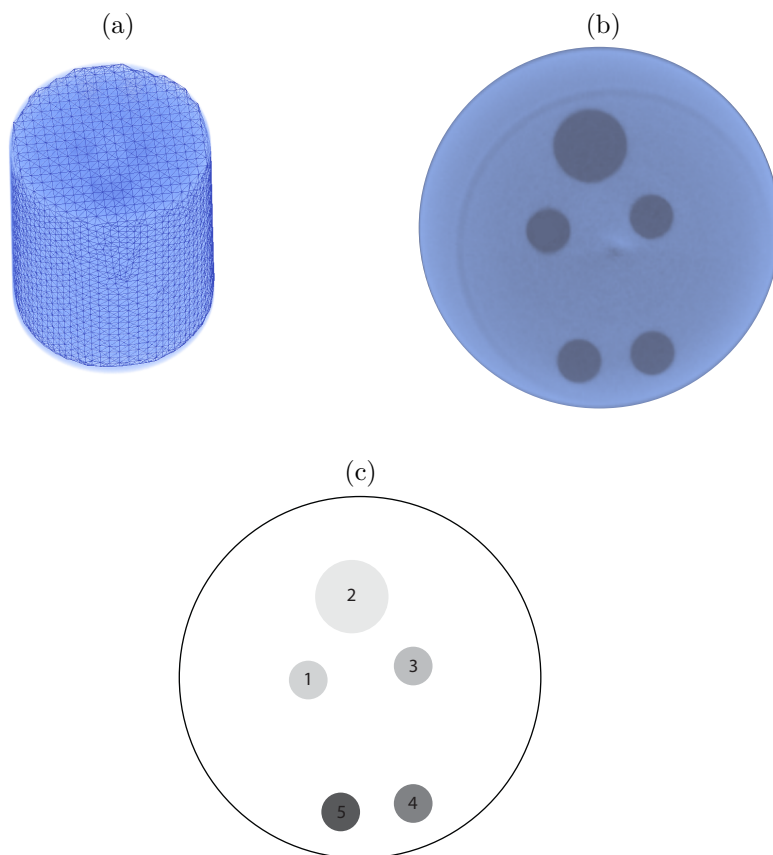


Figure 5.5: Geometry employed in Experiment 4. (a) Registration of reconstruction mesh and CT image of the phantom. (b) xy cross-section through the CT image. (c) Schematic xy cross-section through the structural *a priori* obtained after segmentation of the CT image. Each region is identified by a number and a grey level.

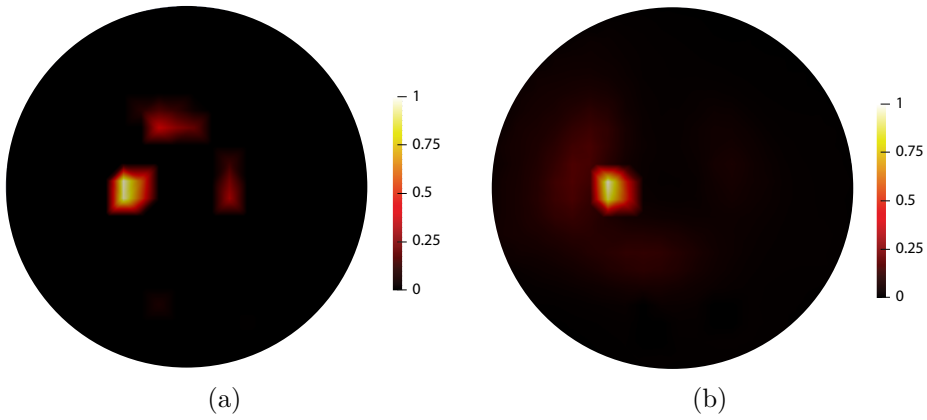


Figure 5.6: Results for Experiment 4, normalized concentration. (a) QR, ℓ_2 -regularized reconstruction restricted to regions 1-5, (b) $\ell_{2,1}$ regularization.

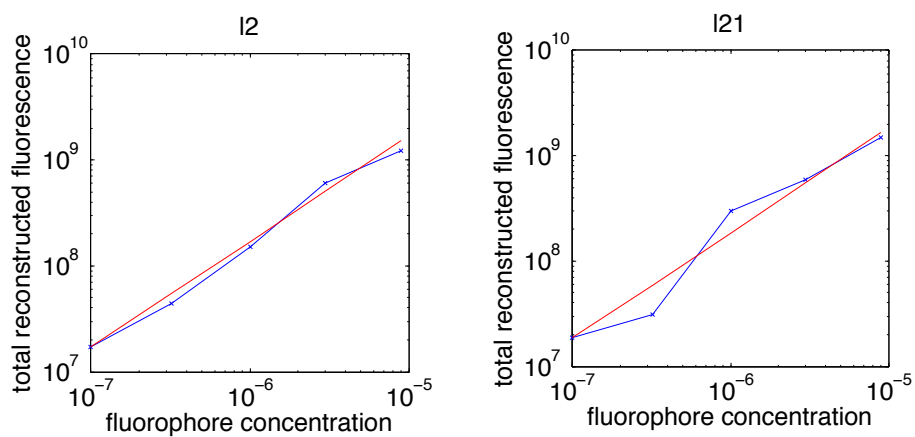


Figure 5.7: Results for Experiment 5 : quantification. In blue, the total reconstructed fluorescence as function of fluorophore concentration (in M). In red, ideal linear response.

Chapter 6

Conclusion

6.1 Summary

We have described an integrated numerical framework for sparsity-inducing reconstruction methods in FDOT. The first idea was to develop a discrete forward model that is consistent with the regularization potentials employed in the inversion algorithm. To that end, we proposed a discretization strategy that handles rigorously propagation equation and regularization on the same grid. The proposed method accommodates arbitrary source and sampling patterns. We showed that our discrete model leads to a computationally efficient implementation.

Then, we proposed sparsity as a new form of regularization for FDOT reconstruction. We demonstrated on phantom results that a sparsity *a priori* enhances the reconstruction of signals that are localized in space, or piecewise-constant. Specifically, our observations indicate a better localization accuracy, and higher robustness to measurement noise. We expect a similar trend for *in-vivo* reconstructions, which is encouraging since in many applications the fluorescent molecules tend to accumulate in a limited area surrounding their target. We also provided efficient algorithms implementing sparsity-inducing regularization at a computational cost that is similar to that of conventional regularization techniques.

Finally, we derived a method to include anatomical information in the regularization term of the FDOT reconstruction algorithm. This was found to im-

prove the localization accuracy. One feature of the proposed method is to use the sparsity-inducing formulation that we developed. As a result, we have a computationally efficient algorithm that incorporates the new information. In this formulation, anatomical information is employed as a *soft prior*, which is an advantage in complex imaging scenarios where this information might be approximative. Using phantom experiments, we compared the performance of the proposed method with another method based on hard constraints. Our results show a systematic improvement in terms of artifacts and contrast.

6.2 Outlook for future research

The research presented here opens several interesting avenues for future investigations. Some of them are listed below.

***In-vivo* validation** The next step is to extend our results to *in vivo* data, and to confirm the improvement observed on phantom experiments.

Quantitative imaging Sparsity-promoting regularization results in non-linear algorithms. This is a critical issue for quantitative imaging. We observed that optimal selection of the regularization parameter somewhat restores the linear dependency. A better understanding is needed, and possibly, correction methods. Such a study should be conducted in parallel with *in-vivo* experiments.

Regularization parameter selection The problem is common to all regularization methods. However, encouraging results were obtained for better conditioned inverse problems such as de-noising, and deconvolution.

Forward model acceleration Evaluating the forward model requires solving the propagation equation as many times as the number of sources and detectors. Using stochastic methods, it might be possible to reduce the number of solutions required, while retaining an accurate forward model.

Preconditioning techniques The lack of satisfactory preconditioner is a handicap for linear reconstruction methods. In principle, the reconstruction time achieved with linear methods could be much reduced, which would have repercussions on non-linear methods as well.

Noise and signal models We showed that non-linear estimators based on convex objective functions could be computed in reasonable time. This suggests that fast algorithms may be found for application-specific noise and signal models.

Experiment design The design of source and sampling patterns that would make the inverse problem better conditioned.

Appendix A

Derivation of the mathematical models for photon migration in tissue

For the sake of completeness we include the derivation of the two prominent models of photon migration in tissue. We start with the transport equation, and then proceed with the diffusion approximation. We take the point of view of photon transport, rather than luminous energy transport as in the body of the text, which is why we change the notation in this appendix. The derivations presented here are adapted from the reference books [9, 8]. We start this appendix by recalling some identities useful for deriving the diffusion approximation.

A.1 Some moments on the unit sphere

We denote

$$\hat{\mathbf{s}} = \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}, \tag{A.1}$$

a vector on the unit sphere of \mathbb{R}^3 . Note that up to a normalization factor, the spherical harmonic of order 0 is the constant 1, and the spherical harmonics of order 1 are the three components s_1, s_2, s_3 .

We denote 4π the unit-sphere, and $2\pi_+$ the half unit-sphere $z > 0$. $\delta_{i,j}$ is the Kronecker symbol. We have the following moments :

$$\begin{aligned} \int_{4\pi} d\hat{s} &= 4\pi \\ \int_{4\pi} \hat{s} d\hat{s} &= \mathbf{0} \\ \int_{2\pi_+} s_i d\hat{s} &= 0, \quad i = 1, 2 \\ \int_{4\pi} s_i s_j d\hat{s} &= \frac{4\pi}{3} \delta_{i,j}, \quad i, j = 1..3 \\ \int_{2\pi_+} s_3 d\hat{s} &= \pi \\ \int_{2\pi_+} s_i^2 d\hat{s} &= \frac{2\pi}{3} \delta_{i,j}, \quad i = 1..3 \\ \int_{4\pi} s_i s_j s_k d\hat{s} &= 0, \quad i, j, k = 1..3. \end{aligned}$$

A.2 Derivation of the transport equation

The photon transport equation is a general model for photon migration in random media. In order to establish this equation, one need to precise :

- variables for the photon state,
- mechanisms controlling the photon state dynamics,
- variables for the photon population.

In this model, photons are seen as particles whose state are represented by position \mathbf{r} , direction of propagation \hat{s} , and time of observation t . We assume all

photons to have the same velocity c [ms^{-1}], which is the average velocity in the tissue under investigation. As far as photon propagation in tissue is concerned, other variables such as quantum state, polarization, or coherency can be safely ignored.

As photons travel in tissue, they are subjected to interactions with matter, which causes their state to evolve. Matter-photon interactions in tissue are of two types: absorption, and scattering. Scattering refers to a change in photon direction, and is typically due to local variations of refractive index in tissue. In transport models, scattering and absorption properties of tissue are described in terms of *cross-sections*. This leads us to define scattering and absorption cross-sections, σ_s [m^{-1}] and σ_a [m^{-1}], respectively. One can give two meanings to these quantities. On the one hand they represent an interaction probability per unit path length. On the other hand, their inverses give the *mean free-path*, that is, the mean distance traveled by a photon before interacting. It is useful to notice that $c\sigma_a$ (resp. $c\sigma_s$) is the frequency with which a photon undergoes absorption (resp. scattering) in tissue. A finer description of scattering events is obtained using the differential scattering cross-section $\sigma_s^d(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ [$\text{m}^{-1} \text{sr}^{-1}$], which corresponds to the probability per unit length, per unit of solid angle, for a photon traveling in direction $\hat{\mathbf{s}}$ to be scattered in direction $\hat{\mathbf{s}}'$. In tissue, scatterers' orientations are essentially random. Therefore, one can assume that σ_s^d depends only on $\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}'$, that is, on the angle between $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}'$.

We now introduce quantities that will help us describe the photon population. First, we define the *angular photon density* $\phi(\mathbf{r}, \hat{\mathbf{s}}, t)$ [$\# \text{m}^{-3} \text{sr}^{-1}$], which is the number of photons per unit of volume, per unit of solid angle, at point \mathbf{r} , at time t , traveling in direction $\hat{\mathbf{s}}$. This is a scalar quantity. For expressing photon displacement, it is useful to define the *angular photon current density* $\mathbf{j}(\mathbf{r}, \hat{\mathbf{s}}, t) = c\phi(\mathbf{r}, \hat{\mathbf{s}}, t)\hat{\mathbf{s}}$ [$\# \text{m}^{-2} \text{sr}^{-1} \text{s}^{-1}$]. For an elementary surface $d\vec{A}$, $\mathbf{j}(\mathbf{r}, \hat{\mathbf{s}}, t) \cdot d\vec{A}$ is the number of photons traveling in the direction $\hat{\mathbf{s}}$, and flowing through $d\vec{A}$ per second.

Let us consider an arbitrary volume Ω , with boundary $\partial\Omega$. We are going to make a balance of $\phi(\mathbf{r}, \hat{\mathbf{s}}, t)$ inside Ω . The variation in number of photons inside Ω at time t , that travel in direction $\hat{\mathbf{s}}$ is

$$\frac{\partial}{\partial t} \int_{\Omega} \phi(\mathbf{r}, \hat{\mathbf{s}}, t) d\mathbf{r} \quad (\text{A.2})$$

Over time, photons are added or subtracted from this number due to

1. loss to absorption and scattering,

$$- \int_{\Omega} c(\sigma_a + \sigma_s)\phi(\mathbf{r}, \hat{\mathbf{s}}, t) \, d\mathbf{r} \quad (\text{A.3})$$

2. photons traveling in some other direction $\hat{\mathbf{s}}'$ and being scattered in direction $\hat{\mathbf{s}}$ (in-scattering term),

$$\int_{\Omega} \int_{4\pi} c\sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}})\phi(\mathbf{r}, \hat{\mathbf{s}}', t) \, d\hat{\mathbf{s}}' \, d\mathbf{r} \quad (\text{A.4})$$

3. source terms denoted $q(\mathbf{r}, \hat{\mathbf{s}}, t)$ [$\# \text{ m}^{-3} \text{ sr}^{-1} \text{ s}^{-1}$],

$$\int_{\Omega} q(\mathbf{r}, \hat{\mathbf{s}}, t) \, d\mathbf{r} \quad (\text{A.5})$$

4. photons leaving or entering Ω through the interface (streaming term),

$$- \int_{\partial\Omega} \mathbf{j}(\mathbf{r}, \hat{\mathbf{s}}, t) \cdot d\vec{S} \quad (\text{A.6})$$

Since Ω is arbitrary, and using Stokes' theorem on the boundary integral we obtain the photon transport equation

$$\begin{aligned} \frac{\partial}{\partial t}\phi(\mathbf{r}, \hat{\mathbf{s}}, t) + c\hat{\mathbf{s}} \cdot \nabla_{\mathbf{r}}\phi(\mathbf{r}, \hat{\mathbf{s}}, t) + c(\sigma_a + \sigma_s)\phi(\mathbf{r}, \hat{\mathbf{s}}, t) \\ = \int_{4\pi} c\sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}})\phi(\mathbf{r}, \hat{\mathbf{s}}', t) \, d\hat{\mathbf{s}}' + q(\mathbf{r}, \hat{\mathbf{s}}, t) \end{aligned} \quad (\text{A.7})$$

A.3 Derivation of the diffusion equation

Here the idea is to seek an approximation to the previous model that gets rid of the angular dependency. To that end, we integrate the transport equation over the

angle variable $\hat{\mathbf{s}}$. We get

$$\begin{aligned}
 & \frac{\partial}{\partial t} \rho(\mathbf{r}, t) + \nabla \cdot \mathbf{J}(\mathbf{r}, t) + c(\sigma_a + \sigma_s) \rho(\mathbf{r}, t) \\
 &= \int_{4\pi} \int_{4\pi} c \sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) \phi(\mathbf{r}, \hat{\mathbf{s}}', t) d\hat{\mathbf{s}}' d\hat{\mathbf{s}} + Q(\mathbf{r}, t) \\
 &= \int_{4\pi} \left(\int_{4\pi} \sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) d\hat{\mathbf{s}} \right) c \phi(\mathbf{r}, \hat{\mathbf{s}}', t) d\hat{\mathbf{s}}' + Q(\mathbf{r}, t) \\
 &= c \sigma_s \rho(\mathbf{r}, t) + Q(\mathbf{r}, t)
 \end{aligned}$$

where we defined

$$\rho(\mathbf{r}, t) = \int_{4\pi} \phi(\mathbf{r}, \hat{\mathbf{s}}, t) d\hat{\mathbf{s}} \quad (\text{A.8})$$

$$\mathbf{J}(\mathbf{r}, t) = \int_{4\pi} c \phi(\mathbf{r}, \hat{\mathbf{s}}, t) \hat{\mathbf{s}} d\hat{\mathbf{s}} = \int_{4\pi} \mathbf{j}(\mathbf{r}, \hat{\mathbf{s}}, t) d\hat{\mathbf{s}} \quad (\text{A.9})$$

$$Q(\mathbf{r}, t) = \int_{4\pi} q(\mathbf{r}, \hat{\mathbf{s}}, t) d\hat{\mathbf{s}} \quad (\text{A.10})$$

Finally we are left with the following continuity equation

$$\frac{\partial}{\partial t} \rho(\mathbf{r}, t) + \nabla \cdot \mathbf{J}(\mathbf{r}, t) + c \sigma_a \rho(\mathbf{r}, t) = Q(\mathbf{r}, t) \quad (\text{A.11})$$

In this equation $\rho(\mathbf{r}, t)$ [$\# \text{ m}^{-3}$] is the *photon density*, $\mathbf{J}(\mathbf{r}, t)$ [$\# \text{ m}^{-2} \text{ s}^{-1}$] the *photon current density*, and $Q(\mathbf{r}, t)$ [$\# \text{ m}^{-3} \text{ s}^{-1}$] a source term. Note that $\mathbf{J}(\mathbf{r}, t)$ is a flux density in the physical sense because for any element of surface $d\vec{A}$, $\mathbf{J}(\mathbf{r}, t) \cdot d\vec{A}$ is the number of photons passing through $d\vec{A}$ per second. At this point, we see that equation (A.11) is not closed because it involves two unknowns; ρ and \mathbf{J} . It is well-known that we obtain a diffusion equation by assuming the flux density to be proportional to the gradient of the density : $\mathbf{J}(\mathbf{r}, t) = -\alpha \nabla \rho(\mathbf{r}, t)$ (Fick's law). Nevertheless, it is interesting to derive this result directly from the transport equation in order to identify the appropriate constant α , as well as the approximations involved.

In order to get another equation satisfied by $\mathbf{J}(\mathbf{r}, t)$, we multiply the transport

equation by $\hat{\mathbf{s}}$ and integrate over the angle variable. This yields

$$\begin{aligned}
 & \frac{1}{c} \frac{\partial}{\partial t} \mathbf{J}(\mathbf{r}, t) + \int_{4\pi} c \hat{\mathbf{s}} \nabla \cdot (\hat{\mathbf{s}} \phi(\mathbf{r}, \hat{\mathbf{s}}, t)) d\hat{\mathbf{s}} + (\sigma_a + \sigma_s) \mathbf{J}(\mathbf{r}, t) \\
 &= \int_{4\pi} \int_{4\pi} c \hat{\mathbf{s}} \sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) \phi(\mathbf{r}, \hat{\mathbf{s}}', t) d\hat{\mathbf{s}}' d\hat{\mathbf{s}} + \mathbf{Q}_1(\mathbf{r}, t) \\
 &= \int_{4\pi} \left(\int_{4\pi} \hat{\mathbf{s}} \sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) d\hat{\mathbf{s}} \right) c \phi(\mathbf{r}, \hat{\mathbf{s}}', t) d\hat{\mathbf{s}}' + \mathbf{Q}_1(\mathbf{r}, t) \tag{A.12}
 \end{aligned}$$

with

$$\mathbf{Q}_1(\mathbf{r}, t) = \int_{4\pi} q(\mathbf{r}, \hat{\mathbf{s}}, t) \hat{\mathbf{s}} d\hat{\mathbf{s}} \tag{A.13}$$

It remains to deal with the streaming and in-scattering terms. Defining g as the average cosine of the scattering angle :

$$g = \left(\int_{4\pi} (\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) \sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) d\hat{\mathbf{s}} \right) / \sigma_s, \tag{A.14}$$

we have

$$\int_{4\pi} \hat{\mathbf{s}} \sigma_s^d(\hat{\mathbf{s}}' \cdot \hat{\mathbf{s}}) d\hat{\mathbf{s}} = g \sigma_s \hat{\mathbf{s}}' \tag{A.15}$$

Second, we pose

$$\mathbf{J}_2(\mathbf{r}, t) = \int_{4\pi} c \hat{\mathbf{s}} \otimes \hat{\mathbf{s}} \phi(\mathbf{r}, \hat{\mathbf{s}}, t) d\hat{\mathbf{s}} \tag{A.16}$$

and abusing the notation ∇ we obtain finally another continuity equation satisfied by $\mathbf{J}(\mathbf{r}, t)$

$$\frac{1}{c} \frac{\partial}{\partial t} \mathbf{J}(\mathbf{r}, t) + \nabla \cdot \mathbf{J}_2(\mathbf{r}, t) + (\sigma_a + \sigma_s') \mathbf{J}(\mathbf{r}, t) = \mathbf{Q}_1(\mathbf{r}, t) \tag{A.17}$$

where $\sigma_s' = (1 - g)\sigma_s$ is the *reduced scattering coefficient*. The resulting system of equations is still not closed because the in-scattering term has generated a new

unknown \mathbf{J}_2 . For closing the system we make the assumption of weak angular dependence of $\phi(\mathbf{r}, \hat{\mathbf{s}}, t)$:

$$\phi(\mathbf{r}, \hat{\mathbf{s}}, t) \approx \frac{1}{4\pi} \rho(\mathbf{r}, t) + \frac{3}{4\pi} \mathbf{J}(\mathbf{r}, t) \cdot \hat{\mathbf{s}} \quad (\text{A.18})$$

One can check that this expression is consistent with our previous definition of ρ and \mathbf{J} . With this approximation we can evaluate the streaming term in (A.21)

$$\nabla \cdot \mathbf{J}_2(\mathbf{r}, t) = \int_{4\pi} c \hat{\mathbf{s}} \nabla \cdot (\hat{\mathbf{s}} \phi(\mathbf{r}, \hat{\mathbf{s}}, t)) d\hat{\mathbf{s}} \quad (\text{A.19})$$

$$= \frac{1}{3} c \nabla \rho(\mathbf{r}, t) \quad (\text{A.20})$$

Equation (A.21) becomes

$$\frac{1}{c} \frac{\partial}{\partial t} \mathbf{J}(\mathbf{r}, t) + c \nabla \rho(\mathbf{r}, t) + (\sigma_a + \sigma'_s) \mathbf{J}(\mathbf{r}, t) = \mathbf{Q}_1(\mathbf{r}, t) \quad (\text{A.21})$$

In the end, we have obtained the closed system

$$\begin{aligned} \frac{\partial}{\partial t} \rho(\mathbf{r}, t) + \nabla \cdot \mathbf{J}(\mathbf{r}, t) + c \sigma_a \rho(\mathbf{r}, t) &= Q(\mathbf{r}, t) \\ \frac{1}{c} \frac{\partial}{\partial t} \mathbf{J}(\mathbf{r}, t) + c \nabla \rho(\mathbf{r}, t) + (\sigma_a + \sigma'_s) \mathbf{J}(\mathbf{r}, t) &= \mathbf{Q}_1(\mathbf{r}, t) \end{aligned} \quad (\text{A.22})$$

This system is known as the P_1 approximation of the transport equation because it can also be obtained by truncating the spherical harmonic expansion of the angular part of $\phi(\mathbf{r}, \hat{\mathbf{s}}, t)$ at the first order. To obtain the diffusion approximation, we need to add two more assumptions :

- isotropy of the sources. This leads to $\mathbf{Q}_1(\mathbf{r}, t) = 0$.
- the relative variation of the photon current in time is negligible compared to the interaction frequency

$$\frac{1}{\mathbf{J}(\mathbf{r}, t)} \frac{\partial}{\partial t} \mathbf{J}(\mathbf{r}, t) \ll c(\sigma_a + \sigma'_s) \quad (\text{A.23})$$

Note that this assumption is void in steady state, which is the case we will consider in the sequel.

124 Derivation of the mathematical models for photon migration in tissue

From these assumptions we get the expected Fick's relation $\mathbf{J}(\mathbf{r}, t) = -D c \nabla \rho(\mathbf{r}, t)$, where we define the *diffusion coefficient*

$$D = \frac{1}{3(\sigma_a + \sigma'_s)} \quad (\text{A.24})$$

Replacing in the photon continuity equation (A.11) we obtain the *diffusion approximation*

$$\frac{\partial}{\partial t} \rho(\mathbf{r}, t) - \nabla \cdot (D c \nabla \rho(\mathbf{r}, t)) + c \sigma_a \rho(\mathbf{r}, t) = Q(\mathbf{r}, t) \quad (\text{A.25})$$

This equation is usually written for the *photon flux* $u(\mathbf{r}, t) = c \rho(\mathbf{r}, t)$ [$\# \text{ m}^{-2} \text{ s}^{-1}$]. Note here that this quantity is not a flux in the physical sense, like \mathbf{J} is. Rather, its physical interpretation is the number of photons flowing through an infinitesimal sphere per second at each point. Quantity u is equivalent to the *fluence rate* encountered in radiative transfer. Formulating the equation (A.25), in terms of u we obtain the usual diffusion equation employed in FDOT

$$\frac{1}{c} \partial_t u - \nabla \cdot (D \nabla u) + \mu_a u = q \quad (\text{A.26})$$

Appendix B

Constrained proximal mappings for the $\ell_{2,1}$ -norm

In this appendix we derive the proximal mappings associated to the $\ell_{2,1}$ -norm for positivity constraints and box constraints. Since the proximal map is separable, we consider the case of a single group. Unless otherwise mentioned, inequalities between vectors are understood coordinate-wise in this section.

Proposition B.0.1. (*Proximal map with positivity constraints*)

Let $\mathbf{y} \in \mathbb{R}^n$, and $\tau > 0$. Assume without loss of generality that $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$, where $\mathbf{y}_1 \leq 0$ and $\mathbf{y}_2 > 0$. The solution of the constrained denoising problem

$$\arg \min_{\mathbf{x} \geq 0} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \tau \|\mathbf{x}\|_2 \tag{B.1}$$

is $\mathbf{x}^* = (0, \mathbf{x}_2^*)$, where

$$\mathbf{x}_2^* = \mathbf{y}_2 \max \left(0, 1 - \frac{\tau}{\|\mathbf{y}_2\|_2} \right) \tag{B.2}$$

Proof : Clearly, $\forall \mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2) \geq 0$,

$$\frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \tau \|\mathbf{x}\|_2 = \frac{1}{2} \|\mathbf{y}_1 - \mathbf{x}_1\|_2^2 + \frac{1}{2} \|\mathbf{y}_2 - \mathbf{x}_2\|_2^2 + \tau (\|\mathbf{x}_1\|_2^2 + \|\mathbf{x}_2\|_2^2)^{1/2} \quad (\text{B.3})$$

$$\geq \frac{1}{2} \|\mathbf{y}_1\|_2^2 + \frac{1}{2} \|\mathbf{y}_2 - \mathbf{x}_2\|_2^2 + \tau \|\mathbf{x}_2\|_2 \quad (\text{B.4})$$

Thus we only need to solve the problem

$$\arg \min_{\mathbf{x}_2 \geq 0} \frac{1}{2} \|\mathbf{y}_2 - \mathbf{x}_2\|_2^2 + \tau \|\mathbf{x}_2\|_2 \quad (\text{B.5})$$

whose unconstrained solution is known and satisfies the positivity constraint \square

Proposition B.0.2. (*Proximal map with box constraints*)

Let $\mathbf{b}, \mathbf{y} \in \mathbb{R}^n$ be two vectors with strictly positive components, and $\tau > 0$. The solution \mathbf{x}^* of the constrained denoising problem

$$\arg \min_{\mathbf{x} \in [0, \mathbf{b}]} \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \tau \|\mathbf{x}\|_2 \quad (\text{B.6})$$

is $\mathbf{x}^* = 0$ if $\|\mathbf{y}\|_2 \leq \tau$. Otherwise, it is obtained as follows.

Define the sets:

$$I_1 = \{i \in [1..n]; b_i < y_i \leq b_i + \tau\} \quad (\text{B.7})$$

$$I_2 = \{i \in [1..n]; y_i > b_i + \tau\} \quad (\text{B.8})$$

$$I_3 = \{i \in [1..n]; y_i \leq b_i\} \quad (\text{B.9})$$

Let m be the cardinal of I_1 . Without loss of generality we assume that $I_1 = [1..m]$.

Define $p_i = \frac{b_i}{y_i - b_i}$, $i \in I_1$ and assume that I_1 is ordered such that the sequence

$(p_i)_{i \in [1..m]}$ is non decreasing. Furthermore, pose $p_0 = 0$, and $p_{m+1} = +\infty$. Define the sequence of functions $(\phi_k(p))_{k \in [0..m]}$

$$\phi_k(p) = \frac{1}{p^2} \left(\sum_{i \in I_1, i \leq k} b_i^2 + \sum_{j \in I_2} b_j^2 \right) + \frac{1}{(p+1)^2} \left(\sum_{i \in I_1, i > k} y_i^2 + \sum_{j \in I_3} y_j^2 \right) \quad (\text{B.10})$$

There exists a unique $l \in [0, m]$, such that the equation $\phi_l(p) = \tau^2$ has a unique solution p^* in the interval $]p_l, p_{l+1}]$, and \mathbf{x}^* has components

$$x_i = \begin{cases} \frac{b_i}{p^*} & \text{if } (i \in I_2) \text{ or } (i \in I_1, i \leq l) \\ y_i \left(1 - \frac{1}{p^* + 1}\right) & \text{if } (i \in I_3) \text{ or } (i \in I_1, i > l) \end{cases} \quad (\text{B.11})$$

Proof : The dual problem is given by

$$\min_{\|\mathbf{z}\|_2 \leq \tau} h(\mathbf{z}) \stackrel{\text{def}}{=} -\frac{1}{2} \|\mathbf{y} - \mathbf{z} - P_{\mathcal{C}}(\mathbf{y} - \mathbf{z})\|_2^2 + \frac{1}{2} \|\mathbf{y} - \mathbf{z}\|_2^2 - \frac{1}{2} \|\mathbf{y}\|_2^2 \quad (\text{B.12})$$

where $\mathcal{C} = [0, \mathbf{b}]$ and $P_{\mathcal{C}}$ is the orthonormal projection on \mathcal{C} . We have $\nabla g(\mathbf{z}) = -P_{\mathcal{C}}(\mathbf{y} - \mathbf{z})$. Notice that the minimizer of the primal problem is given by $\mathbf{x}^* = P_{\mathcal{C}}(\mathbf{y} - \mathbf{z}^*)$. Since h is convex and the constraints are convex, the KKT conditions are necessary and sufficient. The KKT system reads :

$$\begin{aligned} \frac{1}{2} \|\mathbf{z}^*\|_2^2 - \frac{\tau^2}{2} &\leq 0 \\ p^* &\geq 0 \\ p^* \left(\frac{1}{2} \|\mathbf{z}^*\|_2^2 - \frac{\tau^2}{2} \right) &= 0 \\ -P_{\mathcal{C}}(\mathbf{y} - \mathbf{z}^*) + p^* \mathbf{z}^* &= 0 \end{aligned} \quad (\text{B.13})$$

We have $p^* = 0 \Leftrightarrow P_{\mathcal{C}}(\mathbf{y} - \mathbf{z}^*) = 0 \Leftrightarrow \mathbf{y} \leq \mathbf{z}^* \Rightarrow \|\mathbf{y}\|_2 \leq \tau$. Conversely if $\|\mathbf{y}\|_2 \leq \tau$, then clearly $\mathbf{z}^* = \mathbf{y}$ is a minimizer of g (even unconstrained), which implies $p^* = 0$. In that case, $\mathbf{x}^* = 0$.

Let us consider the case $p^* > 0$. This leads to $P_{\mathcal{C}}(\mathbf{y} - \mathbf{z}^*) = p^* \mathbf{z}^*$. From this we deduce that $\forall i \in [1..n]$, $z_i^* \geq 0$. For all $i \in [1..n]$, we have

$$(P_{\mathcal{C}}(\mathbf{y} - \mathbf{z}^*))_i = \begin{cases} b_i & \text{when } y_i - z_i^* > b_i \\ y_i - z_i^* & \text{when } y_i - z_i^* \leq b_i \end{cases} \quad (\text{B.14})$$

Let us denote $K_1 = \{i \in [1..n]; y_i - z_i^* > b_i\}$, $K_2 = \{i \in [1..n]; y_i - z_i^* \leq b_i\}$. We clearly have that $I_2 \subset K_1$ and $I_3 \subset K_2$. From now on, we assume that $i \in I_1$, and we pose $p_i = \frac{b_i}{y_i - b_i}$. Using the fact that $P_{\mathcal{C}}(\mathbf{y} - \mathbf{z}^*) = p^* \mathbf{z}^*$, we get (see drawing

of the two functions, or do the equivalences) that

$$i \in I_1 \cap K_1 \Leftrightarrow \begin{cases} p^* > p_i \\ p^* z_i = b_i \end{cases} \quad (\text{B.15})$$

$$i \in I_1 \cap K_2 \Leftrightarrow \begin{cases} p^* \leq p_i \\ (p^* + 1)z_i = y_i \end{cases} \quad (\text{B.16})$$

The number p^* belongs to a unique interval $]p_l, p_{l+1}]$. This is equivalent to saying that $I_1 \cap K_1 = [1..l]$ and $I_1 \cap K_2 = [l+1..m]$. This implies that $\tau^2 = \|\mathbf{z}^*\|_2^2 = \phi_l(p^*)$. Conversely, the functions $\phi_k(p)$ are strictly decreasing in p , from $+\infty$ to 0. Thus, $\forall k \in [0..m]$, the equation $\phi_k(p) = \tau^2$ has a unique solution. Notice that $\forall k \in [1..m]$, $\frac{b_k}{p_k} = \frac{y_k}{p_k + 1}$ which implies that $\forall k \in [0, m-1]$, $\phi_k(p_{k+1}) = \phi_{k+1}(p_{k+1})$. It is therefore the case for a unique function $\phi_l(p)$, that the equation $\phi_l(p) = \tau^2$ has a solution in $]p_l, p_{l+1}]$. Taking the corresponding p^* and \mathbf{z}^* , we obtain a solution of the minimization problem \square

Bibliography

- [1] M. Cutler, “Transillumination of breasts,” *Journal of the American Medical Association*, vol. 93, no. 25, pp. 1991, 1929.
- [2] S. R. Arridge, M. Schweiger, M. Hiraoka, and D. T. Delpy, “A finite element approach for modeling photon transport in tissue,” *Medical Physics*, vol. 20, no. 2, pp. 299–309, 1993.
- [3] MA. OLeary, DA. Boas, XD. Li, B. Chance, and AG. Yodh, “Fluorescence lifetime imaging in turbid media,” *Opt. Letters*, vol. 21, no. 2, pp. 158–160, JAN 15 1996.
- [4] V. Ntziachristos, C. Tung, C. Bremer, and R. Weissleder, “Fluorescence molecular tomography resolves protease activity in vivo,” *Nature Medicine*, vol. 8, no. 7, pp. 757–761, 2002.
- [5] L. Herv, A. Koenig, A. Da Silva, M. Berger, J. Boutet, J. M. Dinten, P. Peltié, and P. Rizo, “Noncontact fluorescence diffuse optical tomography of heterogeneous media,” *Appl. Opt.*, vol. 46, no. 22, pp. 4896–4906, 2007.
- [6] S. C. Davis, H. Dehghani, J. Wang, S. Jiang, B. W. Pogue, and K. D. Paulsen, “Image-guided diffuse optical fluorescence tomography implemented with laplacian-type regularization,” *Opt. Express*, vol. 15, no. 7, pp. 4066–4082, 2007.
- [7] P. Mohajerani, A. A. Eftekhari, J. Huang, and A. Adibi, “Optimal sparse solution for fluorescent diffuse optical tomography : theory and phantom experimental results,” *Appl. Opt.*, vol. 46, no. 10, pp. 1679–1685, 2007.

-
- [8] A. Ishimaru, *Wave propagation and Scattering in Random Media*, New York: Academic, 1978.
- [9] J. J. Duderstadt and L. J. Hamilton, *Nuclear Reactor Analysis*, John Wiley & Sons, 1976.
- [10] A. D. Klose, V. Ntziachristos, and A. H. Hielscher, “The inverse source problem based on the radiative transfer equation in optical molecular imaging,” *Journal of Computational Physics*, vol. 202, no. 1, pp. 323 – 345, 2005.
- [11] R. C. Haskell, L. O. Svaasand, T. Tsay, T. Feng, M. S. McAdams, and B. J. Tromberg, “Boundary conditions for the diffusion equation in radiative transfer,” *J. Opt. Soc. Am. A*, vol. 11, no. 10, pp. 2727, 1994.
- [12] V. Ntziachristos and R. Weissleder, “Experimental three-dimensional fluorescence reconstruction of diffuse media by use of a normalized born approximation,” *Opt. Letters*, vol. 26, no. 12, pp. 893–895, JUN 15 2001.
- [13] C. H. Schmitz, M. Lcker, J. M. Lasker, A. H. Hielscher, and R. L. Barbour, “Instrumentation for fast functional optical tomography,” *Rev. Sci. Instr.*, vol. 73, no. 2, pp. 429–439, 2002.
- [14] J. C. Hebden, S. R. Arridge, and D. T. Delpy, “Optical imaging in medicine: I. experimental techniques,” *Physics in Medicine and Biology*, vol. 42, no. 5, pp. 825–840, 1997.
- [15] E. M. C. Hillman, J. C. Hebden, F. E. W. Schmidt, S. R. Arridge, M. Schweiger, H. Dehghani, and D. T. Delpy, “Calibration techniques and dataarticle extraction for time-resolved optical tomography,” *Rev. Sci. Instr.*, vol. 71, no. 9, pp. 3415–3427, 2000.
- [16] T. H. Pham, O. Coquoz, J. B. Fishkin, E. Anderson, and B. J. Tromberg, “Broad bandwidth frequency domain instrument for quantitative tissue optical spectroscopy,” *Rev. Sci. Instr.*, vol. 71, no. 6, pp. 2500–2513, 2000.
- [17] G. Yu, T. Durduran, D. Furuya, J. H. Greenberg, and A. G. Yodh, “Frequency-domain multiplexing system for in vivo diffuse light measurements of rapid cerebral hemodynamics,” *Appl. Opt.*, vol. 42, no. 16, pp. 2931–2939, Jun 2003.

- [18] V. Ntziachristos, J. Ripoll, and R. Weissleder, "Would near-infrared fluorescence signals propagate through large human organs for clinical studies?," *Opt. Lett.*, vol. 27, no. 5, pp. 333–335, Mar 2002.
- [19] J. P. Culver, R. Choe, M. J. Holboke, L. Zubkov, T. Durduran, A. Slemple, V. Ntziachristos, B. Chance, and A. G. Yodh, "Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: Evaluation of a hybrid requery domain/continuous wave clinical system for breast imaging," *Medical Physics*, vol. 30, no. 2, pp. 235–247, 2003.
- [20] D. Grosenick, K. T. Moesta, M. Mller, J. Mucke, H. Wabnitz, B. Gebauer, C. Stroszczyński, B. Wassermann, P. M Schlag, and H. Rinneberg, "Time-domain scanning optical mammography: I. recording and assessment of mammograms of 154 patients," *Physics in Medicine and Biology*, vol. 50, no. 11, pp. 2429, 2005.
- [21] A. Villringer and C. Britton, "Non-invasive optical spectroscopy and imaging of human brain function," *Trends Neurosci.*, vol. 20, no. 10, pp. 435–442, Oct 1997.
- [22] D. K. Joseph, T. J. Huppert, M. A. Franceschini, and D. A. Boas, "Diffuse optical tomography system to image brain activation with improved spatial resolution and validation with functional magnetic resonance imaging," *Appl. Opt.*, vol. 45, no. 31, pp. 8142–8151, Nov 2006.
- [23] G. Gratton, M. Fabiani, P. M. Corballis, D. C. Hood, M. R. Goodman-Wood, J. Hirsch, K. Kim, D. Friedman, and E. Gratton, "Fast and localized event-related optical signals (eros) in the human occipital cortex: comparisons with the visual evoked potential and fmri.," *NeuroImage*, vol. 6, no. 3, pp. 168–180, 1997.
- [24] M. J. Niedre, R. H. de Kleine, E. Aikawa., D. G. Kirsch, R. Weissleder, and V. Ntziachristos, "Early photon tomography allows fluorescence detection of lung carcinomas and disease progression in mice in vivo," *Proceedings of the National Academy of Sciences*, vol. 105, no. 49, pp. 19126–19131, 2008.
- [25] J. Haller, D. Hyde, N. Deliolanis, R. de Kleine, M. Niedre, and V. Ntziachristos, "Visualization of pulmonary inflammation using noninvasive fluorescence molecular imaging," *J. Appl. Physiol.*, vol. 104, no. 3, pp. 795–802, 2008.

-
- [26] K. M. Kozloff, R. Weissleder, and U. Mahmood, "Noninvasive optical detection of bone mineral," *Journal of Bone and Mineral Research*, vol. 22, no. 8, pp. 1208–1216, 2007.
- [27] A. Corlu, R. Choe, T. Durduran, M. A. Rosen, M. Schweiger, S. R. Arridge, M. D. Schnall, and A. G. Yodh, "Three-dimensional in vivo fluorescence diffuse optical tomography of breast cancer in humans," *Opt. Express*, vol. 15, no. 11, pp. 6696–6716, 2007.
- [28] S. R. Arridge, "Optical tomography in medical imaging," *Inverse Problems*, vol. 15, no. 2, pp. R41–R93, 1999.
- [29] S. R. Arridge and W. R. B. Lionheart, "Nonuniqueness in diffusion-based optical tomography," *Opt. Lett.*, vol. 23, no. 11, pp. 882–884, Jun 1998.
- [30] L. Herv, A. Koenig, and J.-M. Dinten, "Non-uniqueness in fluorescence-enhanced continuous wave diffuse optical tomography," *Journal of Optics*, vol. 13, no. 1, pp. 015702, 2011.
- [31] Y. Yao, Y. Wang, Y. Pei, W. Zhu, and R. L. Barbour, "Frequency-domain optical imaging of absorption and scattering distributions by a born iterative method," *J. Opt. Soc. Am. A*, vol. 14, no. 1, pp. 325–342, 1997.
- [32] R. B. Schulz, J. Ripoll, and V. Ntziachristos, "Experimental fluorescence tomography of tissues with noncontact measurements," *IEEE Transactions on Medical Imaging*, vol. 23, no. 4, pp. 492–500, Apr. 2004.
- [33] D. Hyde, E. L. Miller, D. H. Brooks, and V. Ntziachristos, "Data specific spatially varying regularization for multimodal fluorescence molecular tomography," *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 365–374, 2010.
- [34] A. Soubret, J. Ripoll, and V. Ntziachristos, "Accuracy of fluorescent tomography in the presence of heterogeneities : study of the normalized born ratio," *Medical Imaging, IEEE Transactions on*, vol. 24, no. 10, pp. 1377–1386, Oct. 2005.
- [35] C. W. Groetsch, *Elements of applicable functional analysis*, New York: Marcel Dekker, 1980.

-
- [36] H. W. Engl, M. Hanke, and A. Neubauer, *Regularization of inverse problems*, Kluwer Academic Publishers, Dordrecht, 1996.
- [37] C. R. Vogel, *Computational methods for inverse problems*, SIAM, 2002.
- [38] E. E. Graves, J. P. Culver, J. Ripoll, R. Weissleder, and V. Ntziachristos, “Singular-value analysis and optimization of experimental parameters in fluorescence molecular tomography,” *J. Opt. Soc. Am. A*, vol. 21, no. 2, pp. 231–241, 2004.
- [39] B. Harrach, “On uniqueness in diffuse optical tomography,” *Inverse Problems*, vol. 25, no. 5, pp. 055010, 2009.
- [40] J. L. Lions, *Optimal control of systems governed by partial differential equations*, Dunod, Paris, 1968.
- [41] W. Bangerth and A. Joshi, “Adaptive finite element methods for the solution of inverse problems in optical tomography,” *Inverse Problems*, vol. 24, no. 3, pp. 034011, 2008.
- [42] C. C. Paige and M. A. Saunders, “Lsqr: An algorithm for sparse linear equations and sparse least squares,” *ACM Trans. Math. Softw.*, vol. 8, pp. 43–71, March 1982.
- [43] L. Lecordier, L. Hervé, J.-M. Dinten, and F. Peyrin, “A finite volume method for fluorescence diffuse optical tomography: Influence on forward model and reconstruction,” *Biomedical Optics*, 2010.
- [44] J. Hyman, J. Morel, M. Shashkov, and S. Steinberg, “Mimetic finite difference methods for diffusion equations,” *Computational Geosciences*, vol. 6, pp. 333–352, 2002.
- [45] P. R. Amestoy, I. S. Duff, J. Koster, and J.-Y. L’Excellent, “A fully asynchronous multifrontal solver using distributed dynamic scheduling,” *SIAM Journal of Matrix Analysis and Applications*, vol. 23, pp. 15–41, 2001.
- [46] P. R. Amestoy, A. Guermouche, J.-Y. L’Excellent, and S. Pralet, “Hybrid scheduling for the parallel solution of linear systems,” *Parallel Computing*, vol. 32, pp. 136–156, 2006.

-
- [47] J. Hyman, M. Shashkov, and S. Steinberg, “The numerical solution of diffusion problems in strongly heterogeneous non-isotropic materials,” *Journal of Computational Physics*, vol. 132, no. 1, pp. 130 – 148, 1997.
- [48] J.-C. Baritau, K. Hassler, and M. Unser, “An efficient numerical method for general ℓ_p regularization in fluorescence molecular tomography,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 4, pp. 1075–1087, April 2010.
- [49] B. Logan, *Properties of high-pass signals*, PhD thesis, Columbia University, New York, 1965.
- [50] D. L. Donoho and B. F. Logan, “Signal recovery and the large sieve,” *SIAM Journal of Applied Mathematics*, vol. 52, no. 2, pp. 577–591, 1992.
- [51] S. G. Mallat and Z. Zhifeng, “Matching pursuits with time-frequency dictionaries,” *IEEE Transactions on Signal Processing*, vol. 41, no. 12, pp. 3397–3415, dec 1993.
- [52] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Transactions on Information Theory*, vol. 49, no. 12, pp. 3320 – 3325, dec. 2003.
- [53] M. Nikolova, “Weakly constrained minimization: Application to the estimation of images and signals involving constant regions,” *Journal of Mathematical Imaging and Vision*, vol. 21, pp. 155–175, 2004.
- [54] P. Rodriguez and B. Wohlberg, “Efficient minimization method for a generalized total variation functional,” *IEEE Transactions on Image Processing*, vol. 18, no. 2, pp. 322 –332, feb. 2009.
- [55] 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*, vol. 16, no. 12, pp. 2980 –2991, dec. 2007.
- [56] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gntnk, “Iteratively reweighted least squares minimization for sparse recovery,” *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

-
- [57] I. F. Gorodnitsky and B. D. Rao, “Sparse signal reconstruction from limited data using focuss: a re-weighted minimum norm algorithm,” *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 600–616, mar 1997.
- [58] M. A. T. Figueiredo and R. D. Nowak, “An em algorithm for wavelet-based image restoration,” *IEEE Transactions on Image Processing*, vol. 12, no. 8, pp. 906–916, aug. 2003.
- [59] 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*, vol. 57, no. 11, pp. 1413–1457, 2004.
- [60] J. J. Moreau, “Proximitéet dualité dans un espace hilbertien,” *Bull. Soc. Math. France*, vol. 93, pp. 273–299, 1965.
- [61] P. L. Combettes and V. R. Wajs, “Signal recovery by proximal forward-backward splitting,” *Multiscale Modeling & Simulation*, vol. 4, no. 4, pp. 1168–1200, 2005.
- [62] P. L. Combettes and J.-C. Pesquet, *Proximal Splitting Methods in Signal Processing*, Springer New York, 2011.
- [63] C. Vonesch and M M. Unser, “A fast thresholded landweber algorithm for wavelet-regularized multidimensional deconvolution,” *IEEE Transactions on Image Processing*, vol. 17, no. 4, pp. 539–549, 2008.
- [64] I. Bayram and I. W. Selesnick, “A subband adaptive iterative shrinkage/thresholding algorithm,” *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1131–1143, 2010.
- [65] M. Guerquin-Kern, M. Häberlin, K. P. Pruessmann, and M. Unser, “A fast wavelet-based reconstruction method for magnetic resonance imaging,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 9, pp. 1649–1660, 2011.
- [66] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences*, vol. 2, pp. 183–202, 2009.
- [67] Yu. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical Programming*, vol. 103, pp. 127–152, 2005.

- [68] A. Chambolle and T. Pock, “A first-order primal-dual algorithm for convex problems with applications to imaging,” *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.
- [69] D. P. Bertsekas, “Multiplier methods: A survey,” *Automatica*, vol. 12, no. 2, pp. 133–145, 1976.
- [70] M. Rudin, *Molecular imaging : basic principles and applications in biomedical research*, Imperial College Press, 2005.
- [71] S. C. Davis, B. W. Pogue, H. Dehghani, and K. D. Paulsen, “Contrast-detail analysis characterizing diffuse optical fluorescence tomography image reconstruction,” *Journal of Biomedical Optics*, vol. 10, no. 5, pp. 050501, 2005.
- [72] P. C. Hansen and D. P. O’Leary, “The use of the l-curve in the regularization of discrete ill-posed problems,” *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993.
- [73] J.-C. Baritiaux, K. Hassler, M. Bucher, S. Sanyal, and M. Unser, “Sparsity-driven reconstruction for fdot with anatomical priors,” *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1143–1153, May 2011.
- [74] G. K. von Schulthess, *Clinical Molecular Anatomic Imaging: PET, PET/CT, and SPECT/CT 2nd Edition*, Lippincott Williams & Wilkins, 2007.
- [75] C. Catana, D. Procissi, Y. Wu, M. S. Judenhofer, J. Qi, B. J. Pichler, R. E. Jacobs, and S. R. Cherry, “Simultaneous in vivo positron emission tomography and magnetic resonance imaging,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 10, pp. 3705–3710, 2008.
- [76] D. Hyde, R. de Kleine, S. A. MacLaurin, E. Miller, D. H. Brooks, T. Krucker, and V. Ntziachristos, “Hybrid fnt-ct imaging of amyloid-[beta] plaques in a murine alzheimer’s disease model,” *NeuroImage*, vol. 44, no. 4, pp. 1304 – 1311, 2009.
- [77] D. Kepshire, N. Mincu, M. Hutchins, J. Gruber, H. Dehghani, J. Hypnarowski, F. Leblond, M. Khayat, and B. W. Pogue, “A microcomputed tomography guided fluorescence tomography system for small animal molecular imaging,” *Review of Scientific Instruments*, vol. 80, no. 4, pp. 043701, 2009.

- [78] Y. Lin, H. Yan, O. Nalcioglu, and G. Gulsen, “Quantitative fluorescence tomography with functional and structural a priori information,” *Appl. Opt.*, vol. 48, no. 7, pp. 1328–1336, 2009.
- [79] W. C. Barber, Y. Lin, O. Nalcioglu, J. S. Iwanczyk, N. E. Hartsough, and G. Gulsen, “Combined fluorescence and x-ray tomography for quantitative in vivo detection of fluorophore,” *Technology in Cancer Research & Treatment*, vol. 9, no. 1, pp. 45–51, 2010.
- [80] R. B. Schulz, A. Ale, A. Sarantopoulos, M. Freyer, E. Soehngen, M. Zientkowska, and V. Ntziachristos, “Hybrid system for simultaneous fluorescence and x-ray computed tomography,” *IEEE Transactions on Medical Imaging*, vol. 29, no. 2, pp. 465–473, 2010.
- [81] B. Brooksby, S. D. Jiang, H. Dehghani, B. W. Pogue, K. D. Paulsen, C. Koge, M. Dooley, J. B. Weaver, and S. P. Poplack, “Magnetic resonance-guided near-infrared tomography of the breast,” *Review of Scientific Instruments*, vol. 75, no. 12, pp. 5262–5270, 2004.
- [82] S. C. Davis, B. W. Pogue, R. Springett, C. Leussler, P. Mazurkewitz, S. B. Tuttle, S. L. Gibbs-Strauss, S. S. Jiang, H. Dehghani, and K. D. Paulsen, “Magnetic resonance-coupled fluorescence tomography scanner for molecular imaging of tissue,” *Review of Scientific Instruments*, vol. 79, no. 6, 2008.
- [83] S. C. Davis, K. S. Samkoe, J. A. O’Hara, S. L. Gibbs-Strauss, H. L. Payne, P. J. Hoopes, K. D. Paulsen, and B. W. Pogue, “Mri-coupled fluorescence tomography quantifies egfr activity in brain tumors,” *Academic Radiology*, vol. 17, no. 3, pp. 271 – 276, 2010.
- [84] D. Hyde, R. Schulz, D. Brooks, E. Miller, and V. Ntziachristos, “Performance dependence of hybrid x-ray computed tomography/fluorescence molecular tomography on the optical forward problem,” *J. Opt. Soc. Am. A*, vol. 26, no. 4, pp. 919–923, 2009.
- [85] A. Douiri, M. Schweiger, J. Riley, and S. R. Arridge, “Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information,” *Measurement Science and Technology*, vol. 18, no. 1, pp. 87–95, 2007.

- [86] M. Guven, B. Yazici, X. Intes, and B. Chance, “Diffuse optical tomography with a priori anatomical information,” *Physics in Medicine and Biology*, vol. 50, no. 12, pp. 2837–2858, 2005.
- [87] A. Li, G. Boverman, Y. Zhang, D. Brooks, E. L. Miller, M. E. Kilmer, Q. Zhang, E. M. C. Hillman, and D. A. Boas, “Optimal linear inverse solution with multiple priors in diffuse optical tomography,” *Appl. Opt.*, vol. 44, no. 10, pp. 1948–1956, 2005.
- [88] P. K. Yalavarthy, B. W. Pogue, H. Dehghani, C. M. Carpenter, S. Jiang, and K. D. Paulsen, “Structural information within regularization matrices improves near infrared diffuse optical tomography,” *Opt. Express*, vol. 15, no. 13, pp. 8043–8058, 2007.

Curriculum Vitæ

Curriculum Vitae

Jean-Charles BARITAUX

Personal Details

Nationality: French

Date of birth: 31 Dec 1981

Present appointment

Research assistant at École Polytechnique Fédérale de Lausanne (EPFL), in the Biomedical Imaging Group (Prof. Michael Unser).

Research Interests

Physics	Optics, Wave propagation, Magnetic resonance
Mathematics	Inverse problems, Numerical analysis, Statistical models
Applications	Molecular Imaging, Medical Imaging, Reconstruction methods

Education

- JUNE 2007 PhD in BIOMEDICAL SIGNAL-PROCESSING
 - *present* **École Polytechnique Fédérale de Lausanne (EPFL)**
 Title: “Advanced Reconstruction Algorithms for Fluorescence Diffuse Optical Tomography”
 - Advisor: Prof. Michael UNSER
 In collaboration with the company SCANCO MEDICAL AG
- SEPT 2006 Master’s thesis, **Sun Microsystems Labs**, Menlo Park, CA
- APR 2007 Title: “Multi-terabit switch based on Proximity Communication” - Advisors: Hans EBERLE, and Prof. Jean-Yves LEBOUDEC
 Developed a FPGA interface for a multi-terabit switch prototype based on the *Proximity Communication* interconnect of Sun Labs.
- OCT 2005 Master’s Degree in COMPUTER SCIENCE *with very high honors*
- JULY 2006 **École Polytechnique Fédérale de Lausanne (EPFL)**
- SEPT 2002 Bachelor’s Degree *with very high honors*
- APR 2005 **École Polytechnique de Paris**, Palaiseau, France
 Major : Electrical Engineering, Computer Science

Experience

- SUMMER 2005 | Research Internship, **Philips Research**, Eindhoven
Integration of super-scalar components in the Trimedia processor simulator of Philips, with the goal of deriving code optimization.

Teaching

- 2008 - 2011 | Teaching assistant for the course SIGNALS AND SYSTEMS I & II at EPFL
Exercise sessions, grading, tutorship, exam preparation

Journal Publications

J.-C. Baritoux, K. Hassler, M. Bucher, S. Sanyal, M. Unser, "Sparsity-Driven Reconstruction for FDOT with Anatomical Priors," *IEEE Transactions on Medical Imaging*, vol. 30, no. 5, pp. 1143-1153, May 2011.

J.-C. Baritoux, K. Hassler, M. Unser, "An Efficient Numerical Method for General Lp Regularization in Fluorescence Molecular Tomography," *IEEE Transactions on Medical Imaging*, vol. 29, no. 4, pp. 1075-1087, April 2010.

Conference Proceedings

M. Guerquin-Kern, **J.-C. Baritoux**, M. Unser, "Efficient Image Reconstruction Under Sparsity Constraints with Application to MRI and Bioluminescence Tomography," *Proceedings of the Thirty-Sixth IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'11)*, Prague, pp. 5760-5763.

J.-C. Baritoux, M. Unser, "A primal-dual reconstruction algorithm for fluorescence and bioluminescence tomography", *IEEE International Symposium on Biomedical Imaging, ISBI 2011*, Chicago, pp. 960-963.

J.-C. Baritoux, M. Unser, "A-priori guided reconstruction for FDOT using mixed-norms", *IEEE International Symposium on Biomedical Imaging, ISBI 2010*, Rotterdam, pp. 692-695.

J.-C. Baritoux, M. Guerquin-Kern, M. Unser, "Integrated modeling and reconstruction with sparsity constraints for FDOT", *IEEE International Symposium on Biomedical Imaging, ISBI 2009*, Boston, pp. 173-176.

J.-C. Baritoux, S. C. Sekhar, M. Unser, "A spline-based forward model for Optical Diffuse Tomography", *IEEE International Symposium on Biomedical Imaging, ISBI 2008*, Paris, pp. 384-387.

