ADVANCED REGULARIZATION AND DISCRETIZATION METHODS IN DIFFUSE OPTICAL TOMOGRAPHY

by

WENQI LU

A thesis submitted to the University of Birmingham

For the degree of

DOCTOR OF PHILOSOPHY

School of Computer Science

College of Engineering and Physical Sciences

University of Birmingham

June 2020

Abstract

Diffuse optical tomography (DOT) is an emerging technique that utilizes light in the near infrared spectral region (650−900nm) to measure the optical properties of physiological tissue. Comparing with other imaging modalities, DOT modality is non-invasive and non-ionising. Because of the relatively lower absorption of haemoglobin, water and lipid at the near infrared spectral region, the light is able to propagate several centimeters inside of the tissue without being absolutely absorbed. The transmitted near infrared light is then combined with the image reconstruction algorithm to recover the clinical relevant information inside of the tissue.

Image reconstruction in DOT is a critical problem. The accuracy and precision of diffuse optical imaging rely on the accuracy of image reconstruction. Therefore, it is of great importance to design efficient and effective algorithms for image reconstruction. Image reconstruction has two processes. The process of modelling light propagation in tissues is called the forward problem. A large number of models can be used to predict light propagation within tissues, including stochastic, analytical and numerical models. The process of recovering optical parameters inside of the tissue using the transmitted measurements is called the inverse problem. In this thesis, a number of advanced regularization and discretization methods in diffuse optical tomography are proposed and evaluated on simulated and real experimental data in reconstruction accuracy and efficiency.

In DOT, the number of measurements is significantly fewer than the number of optical parameters to be recovered. Therefore the inverse problem is an ill-posed problem which would suffer from the local minimum trap. Regularization methods are neces-

sary to alleviate the ill-posedness and help to constrain the inverse problem to achieve a plausible solution. In order to alleviate the over-smoothing effect of the popular used Tikhonov regularization, L_1-norm regularization based nonlinear DOT reconstruction for spectrally constrained diffuse optical tomography is proposed. This proposed regularization can reduce crosstalk between chromophores and scatter parameters and maintain image contrast by inducing sparsity. This work investigates multiple algorithms to find the most computational efficient one for solving the proposed regularization methods.

In order to recover non-sparse images where multiple activations or complex injuries happen in the brain, a more general total variation regularization is introduced. The proposed total variation is shown to be able to alleviate the over-smoothing effect of Tikhonov regularization and localize the anomaly by inducing sparsity of the gradient of the solution. A new numerical method called graph-based numerical method is introduced to model unstructured geometries of DOT objects. The new numerical method (discretization method) is compared with the widely used finite element-based (FEM) numerical method and it turns out that the graph-based numerical method is more stable and robust to changes in mesh resolution.

With the advantages discovered on the graph-based numerical method, graph-based numerical method is further applied to model the light propagation inside of the tissue. In this work, two measurement systems are considered: continuous wave (CW) and frequency domain (FD). New formulations of the forward model for CW/FD DOT are proposed and the concepts of differential operators are defined under the nonlocal vector calculus. Extensive numerical experiments on simulated and realistic experimental data validated that the proposed forward models are able to accurately model the light propagation in the medium and are quantitatively comparable with both analytical and FEM forward models. In addition, it is more computational efficient and allows identical implementation for geometries in any dimension.

Acknowledgements

First, I want to express my sincere gratitude to my supervisor Prof. Iain Styles, for his constant support and encouragement over the past four years. He is always kind to share his academic ideas and discuss with me about the research. He also gives me the freedom to do the research that I am interested in. I want to thank my thesis group members, Prof. Ela Claridge and Prof. Hamid Dehghani for their support and advice throughout my PhD research. I am also grateful to my colleagues, Joshua Veesa, Matthew Doulgerakis, Stanislaw Wojtkiewicz, Mario Forcionea and Daniel Lighter, who have given me a great amount of support in academia. I also want to express my gratitude to our project officer Claire Lindow for her help and sharing of life experiences and our BITMAP colleagues for sharing their knowledge from different backgrounds.

Finally I would like to thank my grandparents, Shaochun Lu and Wenjin Song, my parents, Ming Lu and Liane Jiang, my beloved husband Jinming Duan and all lovely members of my family for their extensive support in my life and study.

This work has been in full funded by the European Union Horizon 2020 Marie Sklodowska-Curie BITMAP Innovative Training Network (675332).

Publications

Journal Papers

Wenqi Lu, Daniel Lighter, and Iain B. Styles. "L_1-norm based nonlinear reconstruction improves quantitative accuracy of spectral diffuse optical tomography." Biomedical optics express 9, no. 4 (2018): 1423-1444.

Wenqi Lu, Jinming Duan, David Orive-Miguel, Lionel Herve, and Iain B. Styles. "Graph- and finite element-based total variation models for the inverse problem in diffuse optical tomography." Biomedical Optics Express 10, no. 6 (2019): 2684-2707.

Wenqi Lu, Jinming Duan, Joshua Deepak, and Iain B. Styles. "New nonlocal forward model for diffuse optical tomography." Biomedical Optics Express 10, no. 12 (2019): 6227-6241.

Conference Proceedings

Wenqi Lu, and Iain B. Styles. "Spectrally constrained L1-norm improves quantitative accuracy of diffuse optical tomography." In European Conference on Biomedical Optics, p. 1041211. Optical Society of America, 2017.

Wenqi Lu, and Iain B. Styles. "Nonlocal Differential Operators Improve Image Reconstruction in Diffuse Optical Tomography." In Optical Tomography and Spectroscopy, pp. JTu3A-32. Optical Society of America, 2018.

Wenqi Lu, Jinming Duan, Joshua Deepak Veesa, and Iain B. Styles. "A new forward model for diffuse optical tomography." In European Conference on Biomedical Optics, p. 11074-83. Optical Society of America, 2019.

List of Figures

List of Tables

List of Abbreviations:

MRI	Magnetic Resonance Imaging
CT	Computed Tomography
NMR	Nuclear Magnetic Resonance
DOT	Diffuse Optical Tomography
NIR	Near-infrared
TV	Total Variation
HbO_2	Oxy-haemoglobin
Hb	Deoxy-haemoglobin
H_2O	Water
CW	Continuous Wave
FD	Frequency Domain
TD	Time Domain
RF	Radio-frequency
TPSF	Temporal Point Spread Function
RTE	Radiative Transfer Equation
DE	Diffusion Equation

PDE	Partial Differential Equation
RBC	Robin Boundary Condition
FDM	Finite Difference Method
2D	Two Dimension
3D	Three Dimension
FEM	Finite Element Method
SVD	Singular Value Decomposition
ART	Algebraic Reconstruction Technique
SIRT	Simultaneous Iterative Reconstruction Technique
RSC	Regularization Methods with Sparsity Constraints
RSP-I	Regularization Methods with Spatial Priors
RSP-II	Regularization Methods with Spectral Priors
ROI	Regions of Interest
HbT	Total Hemoglobin Concentration
StO_2	Haemoglobin Oxygen Saturation
DCA	Depth Compensation Algorithm
FISTA	Fast Iterative Shrinkage Thresholding Algorithm
SNR	Signal-to-Noise Ratio
SCDOT	Spectrally Constrained Diffuse Optical Tomography
IRLS	Iteratively Reweighted Least Square Algorithm
ADMM	Alternating Directional Method of Multipliers

AC	Average Contrast
PC	Pearson Correlation
ISTA	Iterative Shrinkage-Thresholding Algorithm
COV	Covariance
MSE	Mean Squared Error
SPM	Statistical Parametric Mapping
CSF	Cerebrospinal Fluid
HD	High Density
SD	Source-detector
LE	Localization Error
TV	Total Variation
PAT	Photoacoustic Tomography
BLT	Bio-luminescence Tomography
FT	Fluorescence Tomography
GTikhonov	Graph based Tikhonov Regularization
A-FETV	Finite Element-based Anisotropic Total Variation Regularization
I-FETV	Finite Element-based Isotropic Total Variation Regularization
A-GTV	Graph-based Anisotropic Total Variation Regularization
I-GTV	Graph-based Isotropic Total Variation Regularization
MBLL	Modified Beer-Lambert Law
NIRS	Near-infrared Spectroscopy

FVM	Finite Volume Method
BEM	Boundary Element Method
GNM	Graph-based Numerical Method
NDE	Nonlocal Diffusion Equation
NBC	Nonlocal Boundary Condition
BF	Boundary Flux
FR	Fluence Rate
NBF	Normalized Boundary Flux
SSIM	Structural Similarity Index
RMSE	Root Mean Square Error
TBI	Traumatic Brain Injury
IC	Incomplete Cholesky
ILU	Incomplete LU
FSAI	Factorized Sparse Approximate Inverse
CG	Conjugate Gradient
GMRES	Generalized Minimal Residual
BiCGSTAB	Biconjugate Gradient Stabilised
ANA	Analytical Solution
GPU	Graphics Processing Unit

Contents

CHAPTER 1

INTRODUCTION

1.1 Background

The electromagnetic spectrum is the range of frequencies of electromagnetic radiation and their respective wavelengths and photon energies. The visible light (400-700 nanometers (nm)) is applied to observe and interpret images. It only represents a very small portion of the electromagnetic spectrum. Other types of light which are imperceptible to human eyes include radio waves, microwaves, infrared radiation, ultraviolet rays, X-rays and gamma rays. The energy of electromagnetic radiation determines its usefulness for diagnostic imaging (Figure 1.1). For example, magnetic resonance imaging (MRI) uses radio frequency electromagnetic radiation as a transmission medium. Ultrasonography uses ultrasound to find a source of a disease or to exclude pathology. Gamma rays and X-rays are capable of penetrating large body tissues due to their extremely short wavelengths. Gamma rays are used in medicine (radiotherapy), industry (sterilization and disinfection) and the nuclear industry while X-rays are normally used for plain film and computed tomography (CT) imaging.

Specifically, X-ray is an electromagnetic radiation wave with wavelength between 0.01nm and 10nm. CT makes use of a series of X-ray images acquired from different angles to produce cross-sectional (tomographic) images of specific areas of a scanned object. However X-ray is an ionizing radiation which is harmful to living tissues [7] and it has low resolution in soft tissues. Therefore, it fails to provide anatomical details in

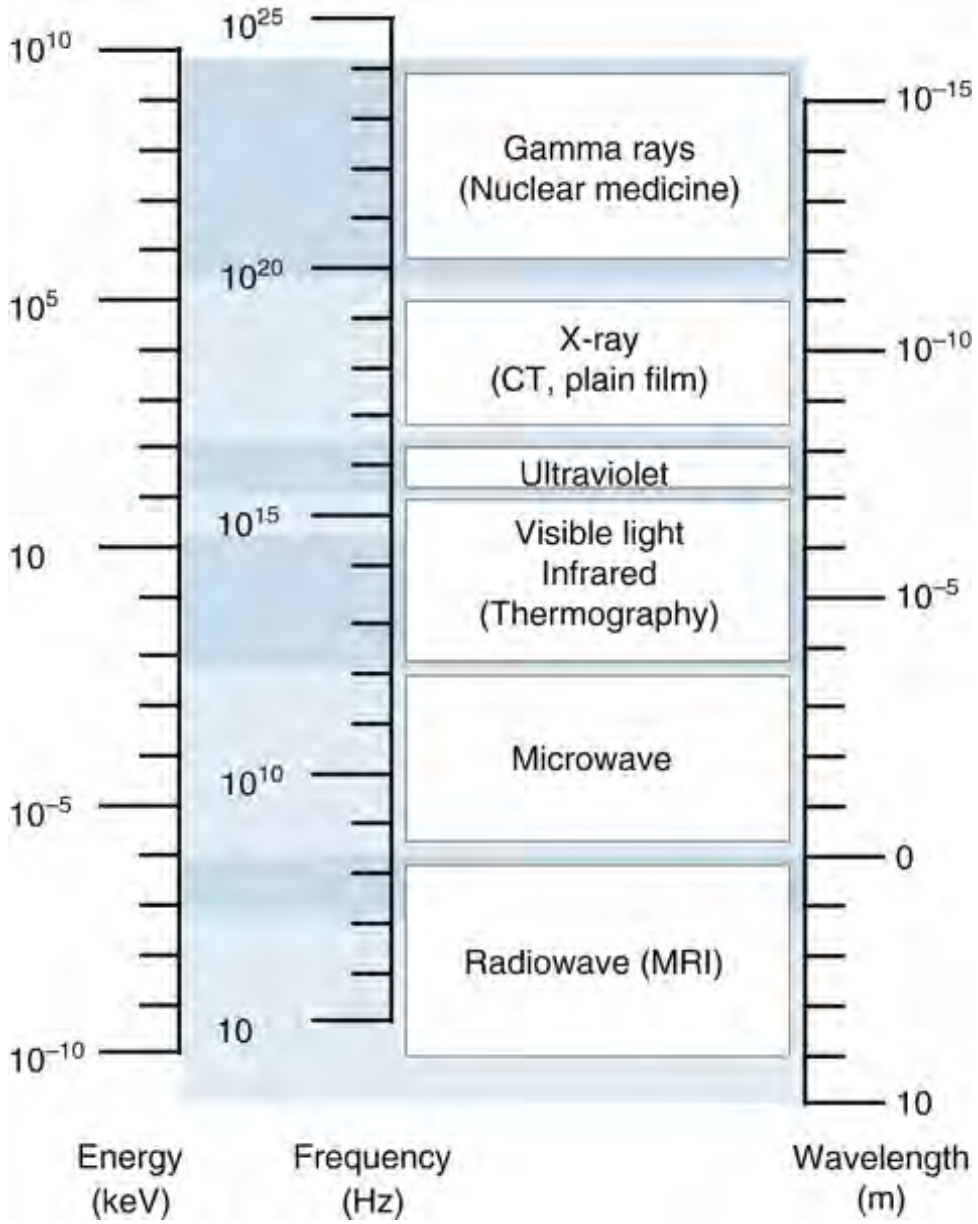


Figure 1.1: The energy, frequency and wavelength of the electromagnetic spectrum and their associated imaging modalities [1].

some brain regions such as the cortex surface. Ultrasound is an imaging technology that uses high-frequency sound waves to characterize tissues. It is real-time, portable and low cost and does not involve ionizing radiation. However, ultrasound images have low spatial resolution especially in deep tissue regions [8, 9]. Therefore, it is difficult to reconstruct structure of tissues behind bone and air in human brain. Magnetic resonance imaging [10, 11] is a medical imaging technique to form pictures of the anatomy and the physiological processes of the body in both health and disease. MRI can recover

the internal structure of the subject based on the nuclear magnetic resonance (NMR) phenomenon. MRI has been applied in the study of the human brain and it can provide millimetre spatial resolution in the whole brain region. In addition, MRI is a non-invasive and non-ionizing imaging modality, which is widely used in hospitals and clinics for medical diagnosis, staging of disease and follow-up. However there may be risks and discomfort associated with MRI scans. Compared with CT scans, MRI scans typically take longer and are louder, and they usually require that the subject enter a narrow, confining tube. In addition, people with some medical implants or other non-removable metal inside the body may be unable to undergo an MRI examination safely.

Diffuse optical tomography (DOT) is an emerging imaging modality which uses near-infrared (NIR) light with wavelengths between 650nm and 900nm. Comparing with the modalities involving ionizing radiations, DOT modality is non-invasive and non-ionising. Typical building costs and the physical size of the device are smaller than in MRI devices and no special environment, e.g. a magnetically shielded room, is needed. The instrumentation can be made portable and suitable for continuous bedside monitoring infant and adult subjects. Even though the spatial resolution and depth sensitivity of DOT is lower than in MRI, the temporal resolution can be higher [12]. During recent years, DOT has many applications such as diagnosing breast cancer [13–15], analyzing functional changes in brain [16–19], and imaging small animals for study of disease and treatment [20, 21].

Image reconstruction in DOT is a critical problem. The accuracy and precision of diffuse optical imaging is related to the accuracy of image reconstruction. Hence, the design of an efficient algorithm for image reconstruction is of great importance. The image reconstruction process in DOT includes modelling the propagation of photons in highly scattering media and reconstructing images out of a finite number of surface measurements. There are a few different models to describe light propagation in tissue-like media. One of the simplest is called the diffusion approximation. The optical properties, the absorption and the scattering coefficients, are the parameters which determine the

light propagation in the diffusion approximation model. When optical properties are known and one is interested in estimating the photon flux at the boundary, the problem is called a forward problem. Likewise, when the photon flux at the boundary is measured and the task is to find the optical properties, the problem is called an inverse problem.

1.2 Goals and contributions

The inverse model in DOT is an ill-posed problem because of the limited number of measurements and diffusive nature of light propagation. Hence unique solution cannot be guaranteed and small errors in the measurements can lead to large errors in the solution. Tikhonov regularization method is normally adopted to overcome issues related to ill-posedness [22, 23]. This method is known to suppress the high-frequency components of the reconstructed image (normally noise) leading to smooth reconstructions. However this has the drawback of being unable to preserve sharp features in the reconstructed images and over-smoothing the results. Therefore the first problem addressed in this thesis is finding suitable regularization methods to recover images with high accuracy and localization. Solving this problem requires quantitative evaluation and comparison of different regularization methods. In this thesis, regularization methods proposed in the DOT community are studied systematically. Previous studies mainly focus on sparsity preserving regularization methods for single wavelength DOT image reconstruction or combining anatomical information from high resolution imaging modalities to guide DOT reconstruction. Borrowing anatomical information from other modalities has been proved to be an efficient strategy for improving the quality of the reconstructed images. However, the segmentation process is prone to errors and can be extremely time consuming. Hence one work in this thesis combines the pure sparsity preserving regularization method (L_1-norm regularization) with the spectral prior to promote localized recovery of chromophore concentration for multiple wavelength DOT image reconstruction. It is the first time that L_1-norm regularization methods and spectrally constrained DOT methods are used together. It gives detailed descriptions of how this can be done, and performs

systematic comparisons of the performance and efficiency of the different methods on both simulated and real data. A method to automatically select the regularization parameters is developed in this work.

The work with sparsity preserving regularization methods is based on the assumption that the optical properties (representing the image) to be reconstructed are spatially sparse. It may tend to over-sparsify the distribution of the optical properties when such an assumption does not hold. For example, in the case of multiple activations or complex injuries in the brain, the features of interest are not spatially localized and the optical properties relative to the background are therefore non-sparse. In order to address this limitation, the second work in this thesis introduces the spatial gradient differential operator to DOT reconstruction. Gradient differential operator can help to impose smoothness to the solution. It can improve the conditioning of the minimization problem, thus enabling a robust numerical solution. Two regularizations are proposed at this stage: Tikhonov regularization with a spatial gradient operator and Total variation (TV) regularization, which uses the L_1-norm of gradients. We introduce finite element and graph-based numerical methods to model unstructured geometries of DOT objects. The discrete differential operators resulting from each representation are defined so that the minimization of the inverse problem associated with the regularization can be carried out on unstructured domains. To the best of our knowledge, this is the first time that graph-based numerical method is introduced to the DOT reconstruction. This work gives a systematic comparison between these two numerical methods. Algorithms based on alternating direction method of multipliers are proposed for each regularized inverse problem. Results of this work determine whether the reconstruction performance is correlated with the discretization method. This work also validates the better performance of the proposed total variation regularization with higher reconstruction accuracy and better localization.

Generation of the forward model which simulates the light propagation is one of the main processes in DOT recovery. Several different models have been applied to

the generation process, and numerical models such as the finite element method are one of the most commonly used forward models for complex subject such as human brain. Under this numerical method, the computational domain is divided into a series of elements (triangles in 2D, tetrahedrain 3D). However, finite element implementations can be difficult and time-consuming, especially when higher-order polynomial basis (shape) functions are used for non-linear interpolation between vertices of high-order elements [24]. Hence, the third work in this thesis is applying the graph-based numerical method which has shown much advantages over the finite element-based numerical method to the forward model. New forward models are proposed in the continuous wave DOT and frequency domain DOT reconstruction respectively. In order to fully leverage the power of graph-based discretization, we use the nonlocal vector calculus in which the differential operators include more pixel information in the domain. Results of the work provide the most efficient solvers for the forward model in the frequency domain DOT measurement system and validate the superior of the proposed forward model in the computational efficiency, light propagation accuracy and simplicity.

The proposed discretization and regularization techniques can be used in many clinical applications. For example, in traumatic brain injury, our proposed methods can be incorporated to help to recover the injury with high accuracy and efficiency. It is very important as these improvements can avoid the delay of the accurate clinical diagnosis and prevent the symptoms from getting worse. In addition, the widely used finite element discretization method in DOT reconstruction is computational expensive. It is problematic, especially for longitudinal as well as large patient population studies. Our proposed graph based discretization method can be used directly to alleviate this drawback.

1.3 Outline of the thesis

This thesis consists of 8 chapters.

Chapter 1 (this chapter) is the introduction of the work undertaken.

Chapter 2 is the introduction of diffuse optical tomography.

Chapter 3 is the detailed implementation of using the finite element method for solving the partial diffusion equation which is the forward model in DOT reconstruction.

Chapter 4 is a literature review of different regularization method and the proposed spectral-L_1 regularization for spectrally constrained diffuse optical property

Chapter 5 is the introduction of the graph-based inverse model for diffuse optical tomography and two new proposed regularization methods.

Chapter 6 is the introduction of the graph-based forward model for continuous-wave diffuse optical tomography and the comparison of the proposed forward model with the popular used finite element-based forward model.

Chapter 7 is the introduction of the graph-based forward model for frequency-domain diffuse optical tomography and the comparison of the proposed forward model with the popular used finite element-based forward model.

Chapter 8 is the summary and conclusion.

CHAPTER 2

DIFFUSE OPTICAL TOMOGRAPHY

Diffuse Optical Tomography (DOT) uses near-infrared (NIR) light in the wavelength range from 650 to 950 nm to reconstruct optical properties of biological tissues. This part of the spectrum is the so-called "NIR window" where chromophore absorption coefficients (oxy-haemoglobin (HbO_2), deoxy-haemoglobin (Hb), water (H_2O) and lipid) are relatively low (Figure 2.1). In addition, H_2O and lipid have lower absorption coefficients than other major chromophores that absorb light and the human tissue consists of mainly water. As a result, NIR light can propagate several centimeters into the tissue without being completely absorbed. The wavelength dependent absorption length in the tissue corresponds to the expectation value of the distance travelled by a photon before it is absorbed. Its reciprocal, the absorption coefficient, is denoted by $\mu_{a,\lambda}$ where λ is the wavelength of light. The propagation of light through tissues is also affected by scattering. Two length scales are important in this context: a rather short "scattering length" which corresponds to the expectation value of the distance travelled by photons before they scatter, and a longer "transport mean-free path" or "random walk step" which corresponds to the typical distance travelled by photons before their direction is randomized. The reciprocal of the photon transport mean-free path is called the reduced scattering coefficient. It is wavelength dependent and is denoted by $\mu'_{s,\lambda}$. $\mu'_{s,\lambda}$ is relative to the scattering coefficient $\mu_{s,\lambda}$ with $\mu'_{s,\lambda} = (1-g)\,\mu_{s,\lambda}$ and g is the anisotropy factor which is the average cosine of the scattering angle. With the optical properties at each wavelength, the chromophore concentrations can be calculated using a constrained linear

relationship fit to the Beer's law relation $\mu_a = \varepsilon C$. ε is the molar absorption spectra of the tissues absorbing chromophores and C is the concentration of these chromophores. Similarly, the spectrum of tissue has been shown to fit well to an empirical approximation to Mie scattering theory [25, 26] given by $\mu'_{s,\lambda} = a\lambda^{-b}$ where a is the scattering amplitude and b represents the scattering power.

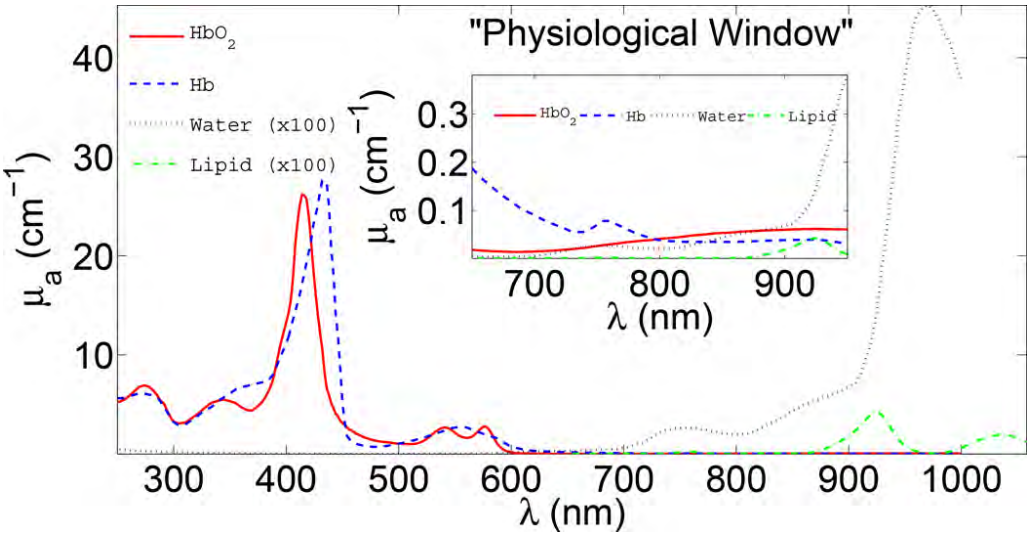


Figure 2.1: A schema of the in vivo NIR window and the absorption spectra of main tissue chromophores are plotted over a large wavelength range. There are clear features in the spectra which enable estimation of chromophore concentrations from diffuse optical measurements at several wavelengths [2].

The imaging process of DOT typically involves injecting NIR light into biological tissue through optical fibres placed on its surface (step 1 in Figure 2.3). The light propagates through the tissue and the spatial distribution of light remitted from the medium's surface is measured for each source fibre (Figure 2.2, step 2 in Figure 2.3). The internal chromophore concentrations and optical properties can be reconstructed directly using a transport-model-based image reconstruction algorithm (step 3 and 4 in Figure 2.3). The mathematical model of photon transport which is built on the absorption and the reduced scattering coefficient of the tissue is called the forward problem. The process in which the optical properties are determined based on given measurements is known as an inverse problem.

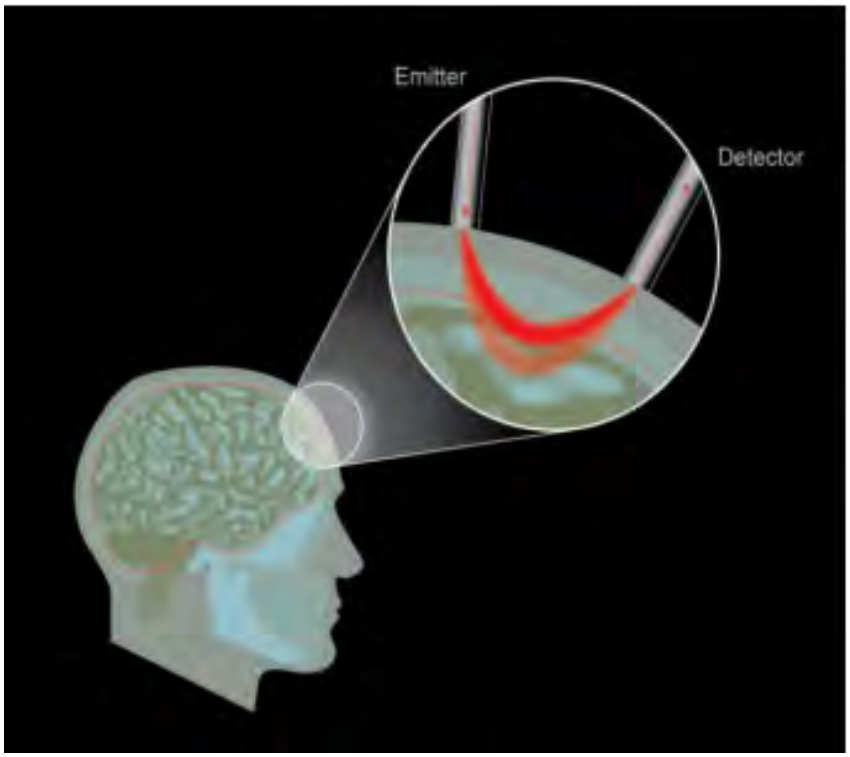


Figure 2.2: NIR light is applied to the surface using source and the reflected light is measured via a detector fibre. The path of light is diffuse and its spectrally varying attenuation provides information about chromophore concentrations in the tissue [3].

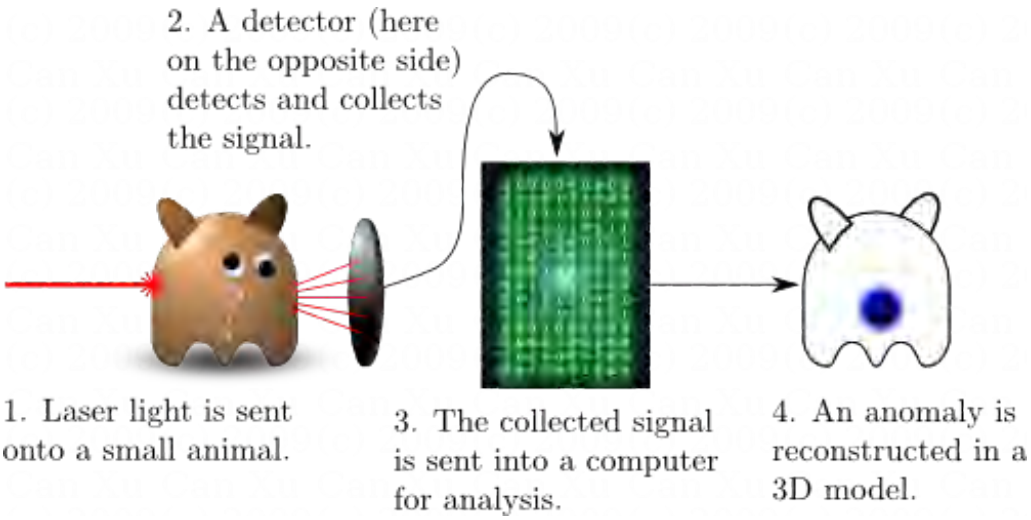


Figure 2.3: Reconstruction process in DOT [4].

2.1 Sources and measurables

There are three types of light sources commonly used in DOT: continuous wave (CW), intensity modulated (FD), and time pulsed (TD).

The simplest source type is CW light, where the intensity remains constant over time [2, 5] (Figure 2.4). This enables fast data acquisition and straightforward data collection and analysis. However, this technique cannot measure the optical pathlength, with the result that instruments of this type do not provide absolute values of chromophore concentrations.

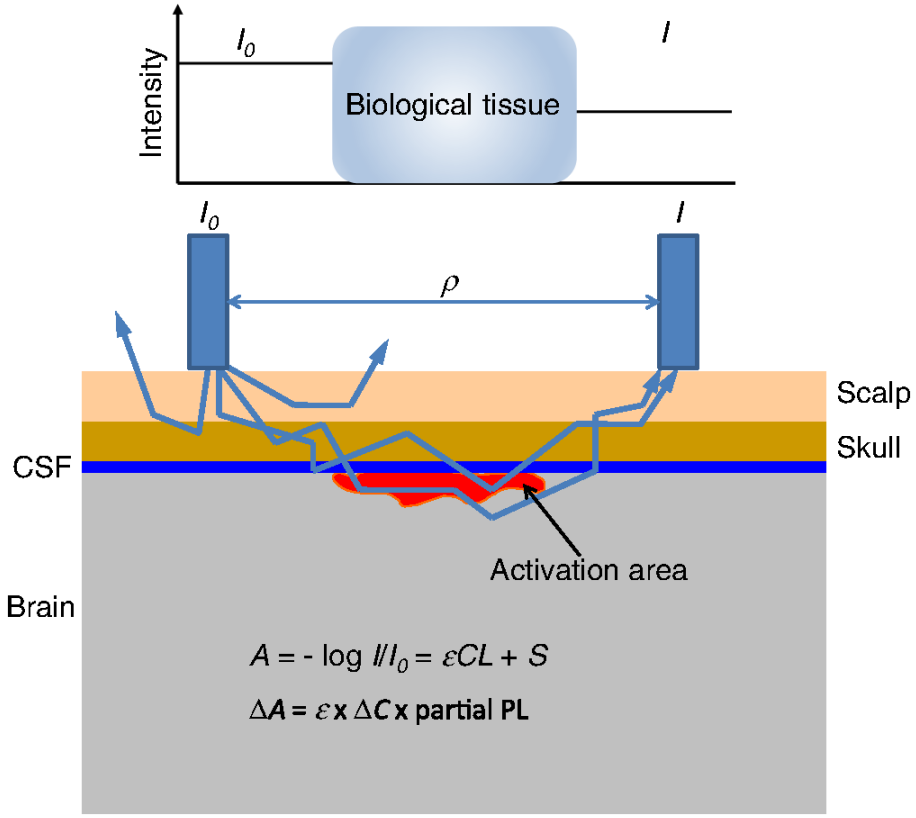


Figure 2.4: Continuous wave (CW) sources [5].

Intensity modulated sources (the frequency-domain technique, FD) are more complex but also give more information about the tissue [27](Figure 2.5). This source emits high frequency (between 100 and 1000 MHz [28]) light via a laser diode driven by a radio-frequency (RF) oscillator. At a given source-detector separation, both the amplitude and phase of the transmitted light are measured. The additional information from the phase, in principle, permits simultaneous determination of absolute values of chromophore concentrations.

Time pulsed light (the time-domain technique, TRS) is related to intensity modulated light via a Fourier transform. It contains the same information content as intensity

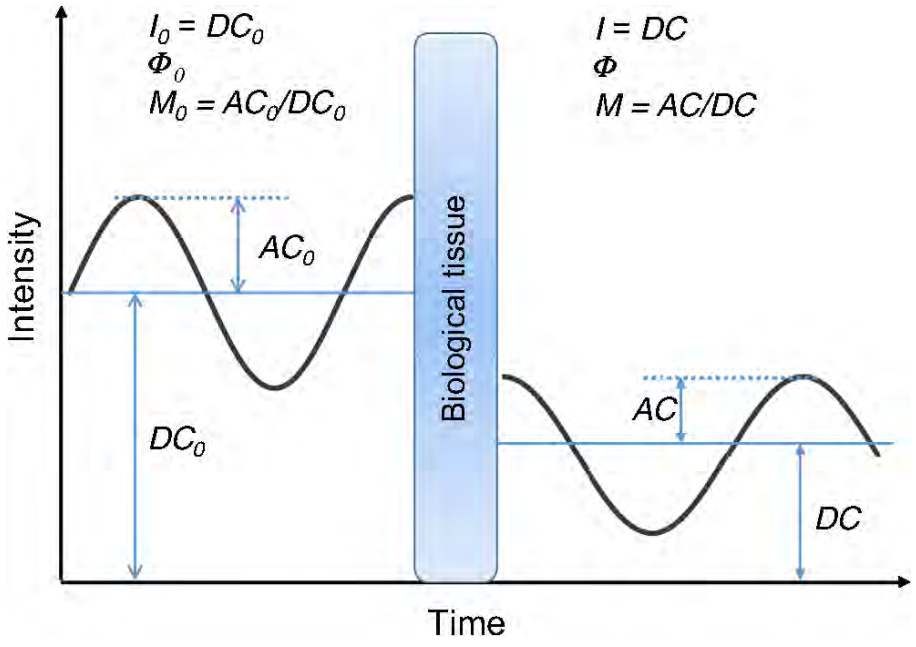


Figure 2.5: Intensity modulated (FD) sources [5].

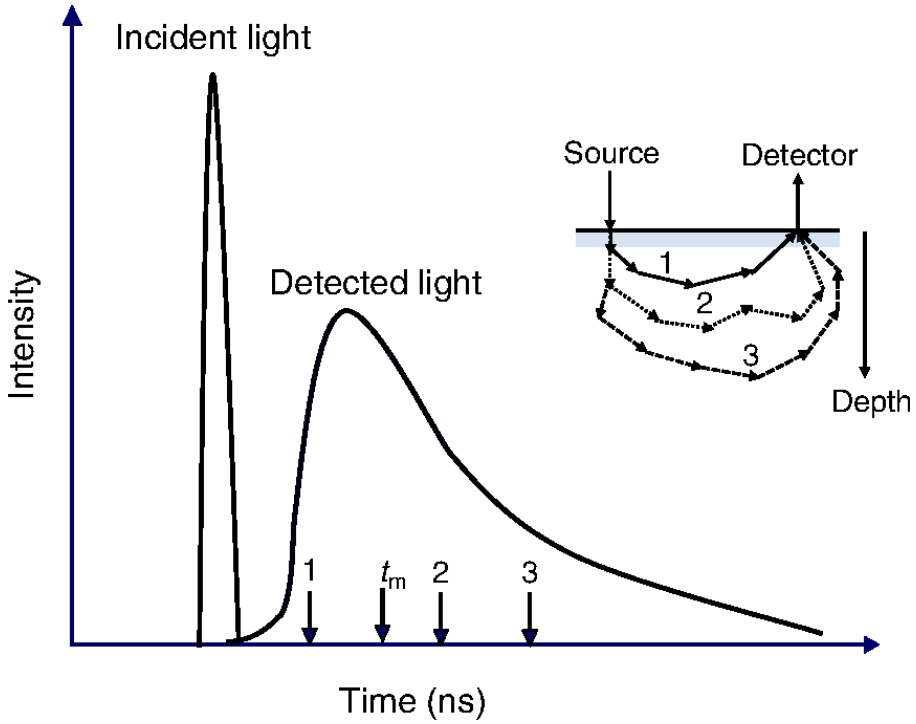


Figure 2.6: Time-resolved (TR) sources [5].

modulated sources scanned over the wide range of modulation frequencies present in the pulse [29](Figure 2.6). In TD measurements, the tissue is irradiated by ultrashort (pico second order) laser pulses, and the intensity of the emerging light is recorded over time to generate a temporal point spread function (TPSF) with picosecond resolution. The TPSF is a delayed, broadened and attenuated version of the light pulse, and normally extends

over several nanoseconds. The TRS data carries information about depth-dependent attenuation based on the correlation of the detection time to the penetration depth of photons. Changes in chromophore concentrations can be more selectively and quantitatively determined by analyzing the TPSF, as the TPSF contains information about the optical properties of the tissue, i.e., the absorption and reduced scattering coefficients. However, TD systems are comparably more expensive than CW and FD systems and require more sophisticated analysis.

FD and TD measurements can be mutually translated using the Fourier (and its inverse) transform and so one TD measurement contains information at all frequencies whereas one FD measurement contains information at only one frequency. Therefore FD measurements have to be taken at several modulation frequencies in order to obtain comparable amount of information to single TD measurement [30].

2.2 Models of photon transport (Forward problem)

When light is interpreted as photons (particles) and its wave nature is neglected, the main interactions between light and biological tissue are absorption and scattering. It is generally accepted that the radiative transfer equation (RTE), which is a conservation equation for photon travelling through a volume medium with absorbers and scatters, accurately describes how light propagates in biological tissues [31–34]. Equation (2.1) shows the FD RTE.

$$\left[\frac{i\omega}{c} + \hat{s}\cdot\nabla + \mu_a + \mu_s\right] I\left(r,\hat{s},\omega\right) = \mu_s \int_{4\pi} P\left(\hat{s},\hat{s}'\right) I\left(r,\hat{s},\omega\right) d\hat{s}' + q\left(r,\hat{s},\omega\right) \tag{2.1}$$

where $I\left(r,\hat{s},\omega\right)$ is the energy radiance (the light intensity) as a function of the position r, angular direction $\hat{s}$, and modulation frequency ω. The parameters μ_a and μ_s are the absorption and scattering coefficients, respectively. c is the light speed in a turbid medium and $q\left(r,\hat{s},\omega\right)$ is light source at position r at frequency ω emitted in direction $\hat{s}$. $P\left(\hat{s},\hat{s}'\right)$ is the scattering phase function, which describes the probability of a photon scattering from the direction $\hat{s}'$ to the direction $\hat{s}$ during a scattering event. The most commonly

used scattering phase function is the Henyey-Greenstein scattering function [35–37] which is defined as

$$P(\cos\beta) = \frac{1-g^2}{2(1+g^2-2g\cos\beta)^{3/2}} \tag{2.2}$$

where β is the angle between the two directions $\hat{s}$ and $\hat{s}'$ and g is the anisotropy factor, which is used to characterize the angular distribution of tissue scattering. In time domain (TD) systems, $i\omega$ is replaced with $\frac{\partial}{\partial t}$ and $I(\mathbf{r}, \Omega, \omega)$ becomes $I(\mathbf{r}, \Omega, t)$.

Many groups [38–40] use the RTE based forward problem for DOT image reconstruction, taking advantage of the accuracy of the RTE for problems where the propagation of photons may not be assumed as diffuse and problems where low scattering may be present. Analytical solutions exist for the RTE only for simple geometries with nearly homogeneous interior objects [41–43]. Although a number of algorithms exist that can seek its numerical solution for more complex inhomogeneous domains [36, 44–47], they are extremely computationally expensive, especially for large 3D volumes. Due to the complexity of the RTE and the limitations of existing algorithms for its solutions, approximate light propagation models are adopted to simplify the RTE, which can be either stochastic or deterministic.

2.2.1 Stochastic models

Stochastic (or statistical) methods predict interactions of each individual photon using either explicit (e.g. Monte Carlo) or implicit methods (e.g. random walk or Markov random field).

2.2.1.1 Monte Carlo methods:

The most commonly used stochastic model in diffuse optics is the Monte Carlo method [48–54]. The transport paths of individual photons are simulated by local values of optical parameters. Photons are followed until they are absorbed by tissue, or escape from the surface of the object contributing to a measurement. Monte Carlo has the advantages of

simulating data with arbitrarily low statistical errors [13], and offering great flexibility in modelling light transport in complex heterogeneous objects. It is however costly in computational time, because a large number of photons need to be simulated so as to acquire meaningful statistics. In order to overcome this disadvantage, various acceleration methods are proposed. Among them, parallel computation is the most preferred method because the movements of photons inside tissue are intrinsically independent.

2.2.1.2 Random walk theory:

Another popular stochastic approach is based on the random walk theory [55, 56], where photon migration is modelled as a series of steps on a discrete cubic lattice. Random walk theory is widely used in time domain imaging systems [57]. It has been applied to calculate the photon travelling time in an absorptive inclusion within a scattering medium [58], to quantify the optical parameters of a breast tumor [59], and to simulate the light diffusion in brain extracellular space [60].

2.2.1.3 Markov random field method:

This is one general stochastic model based on transition probabilities which is proposed by Professor Grunbaum et al. at the University of California at Berkeley [61–63]. With exact values of the probabilities on the boundary, this model can reconstruct the internal transition probabilities in the time independent case. Therefore this model needs noiseless data. Even though this method can lead to an exact reconstruction result, it has never been applied to real data because of the difficulty in relating the essentially topologically invariant analysis to real cases.

2.2.2 Deterministic models

The RTE is a deterministic formulation of radiance transport, from which simpler deterministic models can be derived [64, 65]. If the scattering coefficients dominate over absorption coefficients in tissues and the region of interest is far from the light sources,

light propagation can be modelled by a deterministic diffusion equation (DE) using the P_1 approximation of the RTE (i.e. $I(r, \hat{s}, \omega)$ in equation (2.1) is approximated by the first two terms of a series expansion in spherical harmonics). The DE is able to generate isotropic fluence fields given a distribution of source fibres and the tissue optical properties. Let Ω be the computational domain with the boundary surface Γ. Given the optical parameters (κ, μ_a), DE approximates the photon fluence, also called the photon density, $\Phi(r, \omega) = \int I(r, \Omega, \omega) \mathrm{d}\omega$ by the following elliptic partial differential equation (PDE)

$$-\nabla \cdot (\kappa(r) \nabla \Phi(r, \omega)) + f(r, \omega) \Phi(r, \omega) = q_0(r, \omega) \text{ for } r \in \Omega, \tag{2.3}$$

with

$$f(r, \omega) = \mu_a(r) + \frac{i\omega}{c}. \tag{2.4}$$

Here the source function $q_0(r, \omega)$ is a distributed Gaussian source, matching the intensity profile at the tip of the optical fiber. When the source function is located at the depth of the one transport scattering distance $(1/\mu_s')$ below the boundary, the simulation can accurately reflect the experimental data. $\nabla \cdot (\cdot)$ denotes the differential divergence of a vector function (i.e. $\kappa \nabla \Phi$), while ∇ is the gradient vector. Note that these two differential operators are normally defined based on the classical local vector calculus. $\Phi(r, \omega)$ is the photon fluence rate at position r and modulation frequency ω. It needs to be solved explicitly. The diffusion coefficient $\kappa = 1/(3(\mu_a + \mu_s'))$, where μ_s' is the reduced scattering coefficient. c is defined as c_0/n, where c_0 is the speed of light in vacuum and n is the refractive index. The boundary condition adopted is the Robin boundary condition (RBC), which is used for the case where the fluence at the edge of the tissue exits but does not return [66, 67]. The flux leaving the external boundary is equal to the fluence rate at the boundary weighted by a factor that accounts for the internal reflection of light back into the tissue. This relationship is described in the following equation:

$$2A\boldsymbol{n} \cdot (\kappa(m) \nabla \Phi(m, \omega)) + \Phi(m, \omega) = 0 \quad \text{for } m \in \Gamma, \tag{2.5}$$

where $\boldsymbol{n}$ denotes the outward unit normal on the boundary and A depends on the relative refractive index mismatch between the tissue domain and air. The value of A can be

obtained via Fresnel's law.

The boundary quantity Ψ that is measured at the surface Γ is defined as

$$\Psi(m,\omega) = -\boldsymbol{n} \cdot (\kappa(m)\nabla\Phi(m,\omega)) \quad \text{for } m \in \Gamma. \tag{2.6}$$

Combining equation (2.6) with the Robin boundary condition (2.5) simply gives the form $\Psi(m,\omega) = \Phi(m,\omega)/2A$.

2.2.3 Solution methods for deterministic equations

2.2.3.1 Analytical method:

A general approach for analytically solving the PDE with a source condition is to apply the Green function [65]. The Green function can give the solution for the case where the source is a δ-function. The pulsed sources adopted in DOT are often sufficiently close approximations to δ-function to allow the Green function to be used. However, solutions only exist for simple homogeneous objects [2, 13, 43, 64]. Numerical techniques are required if more complex geometries are modelled.

2.2.3.2 Finite difference method:

The standard numerical technique to solve a PDE is the finite difference method (FDM). FDM is suitable for dealing with the Cartesian grid where each element represents a pixel in two dimensions (2D) or voxel in three dimensions (3D) in the image [68]. Then the differential operators in the problem domain, such as gradient, divergence, Laplacian and curvature, can be discretized straightforwardly using the FDM. However, in biomedical applications, the computational domain (i.e. the multi-layer head) usually has a complex shape. It is non-trivial to represent the complex geometry using a Cartesian grid and therefore FDM is not always practical.

2.2.3.3 Finite element method:

In contrast to FDM, finite element method (FEM) is more flexible for handling objects with complex geometry. FEM works by dividing the computational domain into a series

of elements (triangles in 2D, tetrahedra in 3D). Therefore FEM can impose the boundary conditions very easily and is a common choice to discretize the DE and its RBC. The key principle of FEM as applied to photon transport problems is the reduction of the general problem to that of finding an approximate solution that lies in the vector space spanned by a finite number of basis functions. After that the forward problem is reduced to one of matrix algebra of a finite size which can be solved by efficient techniques. In principle this method can be applied to any PDE model of the transport process. Detailed implementation of FEM for solving the DE and its RBC is described in Section 3.

2.3 Jacobian matrix

The Jacobian matrix (sensitivity matrix) is defined as the change in the surface measurements given a small change in the optical parameters. In the frequency domain, the sensitivity matrix has the form

$$\mathbf{J} = \left(\frac{\partial \Psi (m)}{\partial \kappa (r)} \; \frac{\partial \Psi (m)}{\partial \mu_a (r)} \right). \tag{2.7}$$

There are many different ways to form the Jacobian matrix and one of those is called the direct form or the forward sensitivity analysis [69]. It calculates the changes in the measurement by running the forward problem (Equation (2.3)) multiple times with slightly different parameter values. However, this method is impractical because of the expensive computation time. Instead of perturbing every degree of freedom and calculating the corresponding change in the measurement [70], Arridge [65] presented an efficient way to calculate the sensitivity of the measurement in DOT which is called the adjoint method or the adjoint sensitivity analysis. If we assume the change in the optical parameters at the boundary is close to zero, by applying the Green function method we have the following expression

$$\delta \Psi (m) = - \int_{\Omega} \delta \kappa (r) \nabla G (r) \cdot \nabla \Phi (r) + \delta \mu_a (r) G (r) \Phi (r) \, dr, \tag{2.8}$$

where $\delta\kappa(r)$ and $\delta\mu_a(r)$ are changes in optical parameters and $G(r)$ is the solution of the adjoint equation

$$-\nabla\cdot\kappa(r)\nabla G(r,\omega)+\left(\mu_a(r)-\frac{i\omega}{c(r)}\right)G(r,\omega)=0, \tag{2.9}$$

with the boundary condition:

$$G(m,\omega)+2A\boldsymbol{n}\cdot\kappa(m)\nabla G(m,\omega)=q_0^+(m,\omega), \tag{2.10}$$

where $q_0^+(m,\omega)$ is the adjoint source at the measurement location. Since equation (2.8) is linear in both $\delta\kappa(\mathbf{r})$ and $\delta\mu_a(\mathbf{r})$, the two components in the Jacobian (Equation (2.7)) can be defined as the Fréchet derivatives of the following forms

$$\frac{\partial\Psi(m)}{\partial\kappa(r)}=-\nabla G(r)\cdot\nabla\Phi(r),$$

$$\frac{\partial\Psi(m)}{\partial\mu_a(r)}=-G(r)\Phi(r).$$

Alternatively, instead of considering changes in the intensity directly, the Jacobian can be derived from changes in some transformation of the intensity. For example, by considering the logarithmic intensity, we can obtain the following variant of the Jacobian

$$\tilde{\mathbf{J}}=\left(\frac{\partial(\log\Psi)}{\partial\kappa}\ \frac{\partial(\log\Psi)}{\partial\mu_a}\right), \tag{2.11}$$

where

$$\frac{\partial(\log\Psi)}{\partial\kappa}=\frac{1}{\Psi}\frac{\partial\Psi}{\partial\kappa}=\frac{-\nabla G(r)\cdot\nabla\Phi(r)}{\Psi(m)},$$

$$\frac{\partial(\log\Psi)}{\partial\mu_a}=\frac{1}{\Psi}\frac{\partial\Psi}{\partial\mu_a}=\frac{-G(r)\Phi(r)}{\Psi(m)}.$$

The new two components in equation (2.11) are now normalized by the boundary quantity $\Psi(m)$. Arridge [71] compared the image recovery results from intensity and log intensity measurements, and found that the latter allows for the large dynamic range of the measurement and can thus considerably improve the reconstruction. Later, O'Leary [72] confirmed it and pointed out that the logarithmic intensity is actually equivalent to the Rytov approximation whereas absolute intensity is the Born approximation. Technically,

the logarithmic transformation can be considered as a measurement operator or postprocessing transformation on the obtained boundary data $\Psi(m)$. Arridge [70] has shown that the imaging reconstruction algorithm can be improved by introducing the general measurement operator to the Jacobian calculation.

2.4 Image reconstruction (Inverse problem)

Reconstructing an image representing internal optical properties from a series of boundary measurements is a so-called inverse problem. Depending on the application, the DOT image reconstruction can be approximately divided into two kinds: linear approach and nonlinear iterative approach. DOT reconstruction based on linear approaches is mainly used in neuroimaging studies where brain activity-related optical properties changes are very small. In this case, high-quality images can be recovered by using high-density DOT systems, such as the instruments with spatially overlapping multidistance source-detector arrangements [73,74]. This approach reconstructs the spatiotemporal changes of internal optical properties using temporal derivatives of boundary measurements, which is sufficient for functional neuroimaging studies. In clinical use, quantitative images of optical properties in steady-state are further useful with diagnostic optical imaging. Therefore DOT based on the nonlinear iterative approaches have been developed to enable this.

2.4.1 Linear inverse problem

The linear inverse problem is defined as: given the temporal derivatives of boundary measurements δy, find the spatiotemporal changes of internal optical properties $\delta\mu$ within the computational domain, such that

$$\delta y = \mathbf{J}\delta\mu, \tag{2.12}$$

where Jacobian $\mathbf{J}$ is a linear operator mapping from $\delta\mu$ to δy and $\delta\mu = (\delta\kappa, \delta\mu_a)$.

Equation (2.12) has a unique solution $\delta\mu = \mathbf{J}^{-1}\delta y$ if it is a well-posed system. However, because $\mathbf{J}$ in real applications is most likely non-invertible (or even if it would be invertible, its inversion may be ill-conditioned), equation (2.12) is ill-posed most of the time. In fact, as the data δy contains measurement noise and there may exist modelling errors, instead of directly inverting $\mathbf{J}$ one can seek a solution that minimizes the norm of the residual $e = \delta y - \mathbf{J}\delta\mu$. This leads to the normal equation of the following form

$$\mathbf{J}^*\mathbf{J}\delta\mu = \mathbf{J}^*\delta y. \tag{2.13}$$

Here $\mathbf{J}^*$ is the adjoint operator of $\mathbf{J}$. The solutions of the normal equation (2.13), $\delta\mu = (\mathbf{J}^*\mathbf{J})^{-1}\mathbf{J}^*\delta y$, exist and is unique if $(\mathbf{J}^*\mathbf{J})^{-1}$ exists and continuous. The minimum residual solution $\delta\mu$ is also known as the least square solution. Unfortunately, equation (2.13) may still be a ill-posed problem since $\mathbf{J}^*\mathbf{J}$ may not be invertible or its inversion may be ill-conditioned. To overcome the ill-posedness, Levenberg-Marquardt(LM) approach introduces a diagonal term to stabilize the problem. In this case the update equation is altered to

$$(\mathbf{J}^*\mathbf{J} + \lambda\mathbf{I})\delta\mu = \mathbf{J}^*\delta y. \tag{2.14}$$

In equation (2.14) $\mathbf{I}$ is the identity operator and $\lambda > 0$ is a regularization parameter which is often chosen based on the noise level present in the measurements δy. Note that equation (2.14) has the same solution as the minimization problem with Tikhonov regularization

$$\delta\mu = \arg\min_{\delta\mu} \left\{ \|\delta y - \mathbf{J}\delta\mu\|_2^2 + \lambda\|\delta\mu\|_2^2 \right\}. \tag{2.15}$$

It is possible to show that equation (2.15) has a unique solution $\delta\mu = (\mathbf{J}^*\mathbf{J} + \lambda\mathbf{I})^{-1}\mathbf{J}^*\delta y$ when λ is fixed.

In a more general version of Tikhonov regularization, one can substitute the term $\lambda\|\delta\mu\|_2^2$ in equation (2.15) with $\lambda\|\mathbf{L}\delta\mu\|_2^2$ where $\mathbf{L}$ can be regarded as a highpass operator such as a linear differential operator like gradient or Laplacian. The minimization problem thus becomes

$$\delta\mu = \arg\min_{\delta\mu} \left\{ \|\delta y - \mathbf{J}\delta\mu\|_2^2 + \lambda\|\mathbf{L}\delta\mu\|_2^2 \right\}. \tag{2.16}$$

The new regularization term $\|\mathbf{L}\delta\mu\|_2^2$ can be seen as a prior or a smoothness constraint on $\delta\mu$. It can improve the conditioning of the problem, thus enabling a numerical solution. Now assuming that $\mathbf{L}$ is linear and continuous with adjoint $\mathbf{L}^*$, $\delta\mu$ is the solution of equation (2.16) if and only if

$$(\mathbf{J}^*\mathbf{J} + \lambda\mathbf{L}^*\mathbf{L})\delta\mu = \mathbf{J}^*\delta y. \tag{2.17}$$

2.4.2 Nonlinear iterative inverse problem

Nonlinear iterative inverse problem is to recover the underlying optical properties (κ, μ_a) at each finite element node from the measurements y using a model-based approach. This reconstruction process can be accomplished by matching the experimental measurements y with the model predictions $\Psi(\kappa, \mu_a)$ iteratively in the least squares sense. We note that the general theory of nonlinear inverse problems is however a lot more complicated. We thereby focus only on how to numerically solve the given minimization problem for DOT image reconstruction in this part. One can refer to [64, 65, 69] for detailed discussions on the general theory of inverse problems.

Assume that we want to recover optical properties $\mu = (\kappa, \mu_a)$ from M measurements for a finite element mesh with N nodes. Let $\Psi(\cdot)$ be the nonlinear operator induced from the forward problem (Equation (2.3)) in DOT. This finite-dimensional forward operator explicitly describes the relationship between the measurement and the optical parameters via $y = \Psi(\kappa, \mu_a)$. Instead of adopting the intensity of measurements directly, Arridge [71] suggest using log intensity measurements which has been proved to allow for the large dynamic range of the measurement and can considerably improve the reconstruction. Therefore we transform the relationship into the logarithmic domain as

$$\log y = \log\Psi. \tag{2.18}$$

Given a reasonable guess for (κ, μ_a), one can linearize the nonlinear function $\Psi(\kappa, \mu_a)$ using the first order Taylor expansion as follows

$$\log\Psi \approx \log\Psi^k + \tilde{\mathbf{J}}^k\delta\mu, \tag{2.19}$$

where the superscript k denotes the kth iteration. $\delta\mu_j = (\kappa_j - \kappa_j^k\, \mu_{a_j} - \mu_{a_j}^k)^{\mathrm{T}}$ where $j = 1, ..., N$, is the changes in the optical properties. According to the definition of the Jacobian (Equation (2.11)), $\tilde{\mathbf{J}}^k$ is given by

$$\tilde{\mathbf{J}}_{i,j}^k = \left(\frac{-\nabla G_j^k(r) \cdot \nabla \Phi_j^k(r)}{\Psi_j^k(m)} \quad \frac{-G_j^k(r) \cdot \Phi_j^k(r)}{\Psi_j^k(m)} \right), \tag{2.20}$$

where $i = 1, ..., M$. G_j and Φ_j, defined on each finite element node, are the discrete solutions of the adjoint model in equation (2.9) and equation (2.3) respectively. Substituting equation (2.19) to (2.18) leads to

$$\delta\Psi^k = \tilde{\mathbf{J}}^k \delta\mu, \tag{2.21}$$

where $\delta\Psi^k = \log y - \log\Psi^k$ is the data-model mismatch in the kth iteration. However, the Jacobian is often ill-posed. Strategies that can be employed to invert an ill-posed matrix includes truncated singular value decomposition (SVD), algebraic reconstruction technique (ART), simultaneous iterative reconstruction technique (SIRT), and k-space expansions [69, 75–77]. Another most commonly used approach is the Gauss-Newton algorithm. Let the residual $e = \delta\Psi^k - \tilde{\mathbf{J}}^k \delta\mu$. The optical properties can be found by the classical Gauss-Newton method through minimizing the following L_2-norm of the residual

$$\delta\mu^{k+1} = \arg\min_{\delta\mu} \left\| \delta\Psi^k - \mathbf{J}^k \delta\mu \right\|_2^2 \tag{2.22}$$

Equation (2.22) is in fact the least square formulation of the linear inverse problem (equation (2.12)), and its solution can be written as

$$\delta\mu^{k+1} = \left(\mathbf{J}^{k\mathrm{T}}\mathbf{J}^k\right)^{-1} \mathbf{J}^{k\mathrm{T}} \delta\Psi^k. \tag{2.23}$$

The Gauss-Newton algorithm for approximately solving the DOT inverse problem (equation (2.18)) is therefore iteratively updating Ψ and $\delta\mu$ by using equation (2.19) and equation (2.23) until some convergence condition is satisfied. Unfortunately, there is no guarantee that this process is convergent due to the nonlinearity of equation (2.18). Moreover, if the initial guess is far away from the actual solution, the algorithm is more likely to get stuck into a local minimum which most often is meaningless.

As $\mathbf{J}^{k\mathrm{T}}\mathbf{J}^k$ is usually singular or close to singular, it is difficult to calculate the inversion of $\mathbf{J}^{k\mathrm{T}}\mathbf{J}^k$ in equation (2.23) directly for the Gauss-Newton iteration. Further,

experimental noise in the measurements y tends to lead to reconstruction artefacts. Regularization terms are normally imposed to convert equation (2.22) into a more readily solvable problem. In Section 4, I will give the literature review about the regularization methods proposed in the DOT community and present our new proposed regularization method to alleviate the ill-posedness of the inverse problem.

CHAPTER 3

FINITE ELEMENT METHOD FOR SOLVING THE PARTIAL DIFFERENTIAL EQUATION

In this chapter, we describe how the 2D and 3D partial differential equations in diffuse optical tomography imaging reconstruction can be numerically solved by the Galerkin finite element method. The main idea is to first derive the variational formulation of the original partial differential equation and discretize it with the help of the first order piecewise linear basis functions. Afterwards, the discretized formulation over the whole calculation domain is broken down onto each local element. The Gauss-Legendre quadrature rules are then applied, which allow the integrals over each element to be numerically evaluated efficiently and accurately. Lastly, the computed integrals are assembled element by element, giving rise to a sparse linear equation system which can be solved by direct solvers (e.g. Gaussian elimination) or iterative solvers (e.g. generalized minimal residual method). We first start with the 2D FEM case, followed by the corresponding 3D case in Section 3.4.

3.1 Variational (weak) formulation

To apply the Galerkin finite element method to solve the differential equation (2.3) on irregular geometries, one needs to convert it into a variational (weak) formulation. In

mathematics, variational (weak) formulations permit the transfer of concepts of linear algebra to the analysis of partial differential equations such as equation (2.3). To derive the variational formulation of equation (2.3), we multiply the two sides of the differential equation by a test function $v(x, y)$ and integrate them over the whole computational domain Ω. Note that in a 2D geometry, we use (x, y) to represent the coordinate pair of position r.

$$\begin{aligned}\iint_{\Omega} -\nabla \cdot (\kappa(x,y)\nabla\Phi(x,y))\, v(x,y)\, dxdy + \iint_{\Omega} f(x,y)\,\Phi(x,y)\, v(x,y)\, dxdy \\ = \iint_{\Omega} q_0(x,y)\, v(x,y)\, dxdy.\end{aligned} \tag{3.1}$$

By applying integration by parts to the first term of the left-hand side of equation (3.1) and omitting the function variables (x, y) for brevity, we arrive at

$$\iint_{\Omega} (\kappa\nabla\Phi)\nabla v dxdy - \oint_{\Gamma} \boldsymbol{n} \cdot (\kappa\nabla\Phi) v ds + \iint_{\Omega} f\Phi v dxdy = \iint_{\Omega} q_0 v dxdy. \tag{3.2}$$

Substituting the boundary condition (Equation (2.5)) into equation (3.2), we obtain

$$\iint_{\Omega} (\kappa\nabla\Phi)\nabla v dxdy + \oint_{\Gamma} \frac{\Phi}{2A} v ds + \iint_{\Omega} f\Phi v dxdy = \iint_{\Omega} q_0 v dxdy. \tag{3.3}$$

Equation (3.3) is now the variational formulation of equation (2.3) for $\Phi \in \mathcal{V}$, where $\mathcal{V}$ denotes the vector space of the test function $v = v(x, y)$.

3.2 Discretization using Galerkin finite element method

Given a 2D domain, we discretize the computational domain with a set of triangles or quadrilaterals (rectangles). In this paper, we only focus on the triangular triangulation but the Galerkin method is also applicable to other triangulations such as quadrilateral. Now assume that a 2D finite element domain Ω is divided by M triangular elements $\{T_i\}_{i=1}^{M}$ connected with N nodes. Let $\mathcal{V}_h$ be the 2D vector space of continuous and piecewise linear functions on these triangles. The Galerkin finite element method seeks a solution Φ to equation (3.3), that can be expressed in terms of a set of basis functions $\{\varphi_j\}_{j=1}^{N}$, $\varphi_j \in \mathcal{V}_h$ (also known as shape or trial functions). Specifically, the solution Φ

can be rewritten in the form of

$$\Phi = \sum_{j=1}^{N} \Phi_j \varphi_j, \tag{3.4}$$

where given three nodes of a element T, i.e. $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$ and $v_3 = (x_3, y_3)$, the basis functions φ_j for the three vertices of the triangle can be respectively expressed as

$$\begin{aligned}
\varphi_1(x,y) &= \frac{(x_2y_3 - x_3y_2 + (y_2 - y_3)x + (x_3 - x_2)y)}{2A_T}, \\
\varphi_2(x,y) &= \frac{(x_3y_1 - x_1y_3 + (y_3 - y_1)x + (x_1 - x_3)y)}{2A_T}, \\
\varphi_3(x,y) &= \frac{(x_1y_2 - x_2y_1 + (y_1 - y_2)x + (x_2 - x_1)y)}{2A_T}.
\end{aligned}$$

Here A_T denotes the triangular area of T, which can be evaluated by

$$A_T = \frac{|x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)|}{2}. \tag{3.5}$$

Note that the problem equation (3.3) now reduces down to the calculation of the fluence rate Φ at each node over the whole domain Ω. We thus substitute equation (3.4) into equation (3.3) to obtain

$$\sum_{j=1}^{N} \Phi_j \left(\iint_\Omega \kappa \nabla \varphi_j \nabla v dx dy + \iint_\Omega f \varphi_j v dx dy + \oint_\Gamma \frac{1}{2A} \varphi_j v ds \right) = \iint_\Omega q_0 v dx dy. \tag{3.6}$$

Since $\{\varphi_j\}_{j=1}^N \in \mathcal{V}_h$ are the basis functions of $\mathcal{V}_h$, equation (3.6) is equivalent to

$$\sum_{j=1}^{N} \Phi_j \left(\iint_\Omega \kappa \nabla \varphi_j \nabla \varphi_i dx dy + \iint_\Omega f \varphi_j \varphi_i dx dy + \oint_\Gamma \frac{1}{2A} \varphi_j \varphi_i ds \right) = \iint_\Omega q_0 \varphi_i dx dy. \tag{3.7}$$

Equation (3.7) is a sparse, well-posed quadratic system of N linear equations and N unknowns, of the form

$$\boldsymbol{\mathcal{M}} \Phi = Q. \tag{3.8}$$

Equation (3.8) is normally known as the Galerkin system, where $\Phi = (\Phi_1, \Phi_2, ..., \Phi_N)^T$. The matrix $\boldsymbol{\mathcal{M}} = \mathbf{K} + \mathbf{C} + \mathbf{B}$ the terms of which have elements given by

$$K_{ij} = \iint_\Omega \kappa \nabla \varphi_i \nabla \varphi_j dx dy, \tag{3.9}$$

$$C_{ij} = \iint_\Omega f \varphi_i \varphi_j dx dy, \tag{3.10}$$

$$B_{ij} = \oint_{\Gamma} \frac{1}{2A} \varphi_i \varphi_j ds, \tag{3.11}$$

and the source term $Q = (Q_1, Q_2, ..., Q_N)^T$ has the form of

$$Q_i = \iint_{\Omega} q_0 \varphi_i dx dy. \tag{3.12}$$

In the linear system, $\mathbf{K}$ is the stiffness matrix and $\mathbf{C}$ the mass matrix. Because $\mathbf{B}$ contains the boundary contributions to the linear system matrix, its domain of integration is different from the one of other three matrices $\mathbf{K}$, $\mathbf{C}$ and $\mathbf{Q}$. Note that it is however not straightforward to find values for $\mathbf{K}$, $\mathbf{C}$, $\mathbf{B}$ and $\mathbf{Q}$ over the whole domain Ω directly. Alternatively, these matrices can be evaluated locally on each individual triangle or boundary edge and then assembled to construct the global matrix $\boldsymbol{\mathcal{M}}$. In Section 3.3, we shall introduce the whole procedure in detail.

3.3 Simplex coordinates, integral evaluations and quadrature rules

3.3.1 Simplex coordinates

To evaluate the matrices $\mathbf{K}$, $\mathbf{C}$ and $\mathbf{Q}$ over each triangle, simplex coordinates (also known as area coordinates or barycentric coordinates) are normally used. By choosing this coordinate system, Gauss-Legendre quadrature rules can be conveniently applied to evaluate the domain integrals in these matrices. We note that Gauss-Legendre quadrature allows fewer quadrature points to be used to achieve high accuracy. To develop the simplex coordinate transformation, we consider a triangle T which is defined by three vertices $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$ and $v_3 = (x_3, y_3)$ in the Cartesian coordinate system, as shown in Figure 3.1 left. Any point $r = (x, y)$ located on the triangle can be thereby represented as a weighted sum of the three nodes via

$$r = (1 - \xi - \eta)v_1 + \xi v_2 + \eta v_3. \tag{3.13}$$

Equation (3.13) maps the original triangle T in the Cartesian coordinate system spanned by x axis and y axis onto a standard triangle T_{st} in the simplex coordinate system spanned

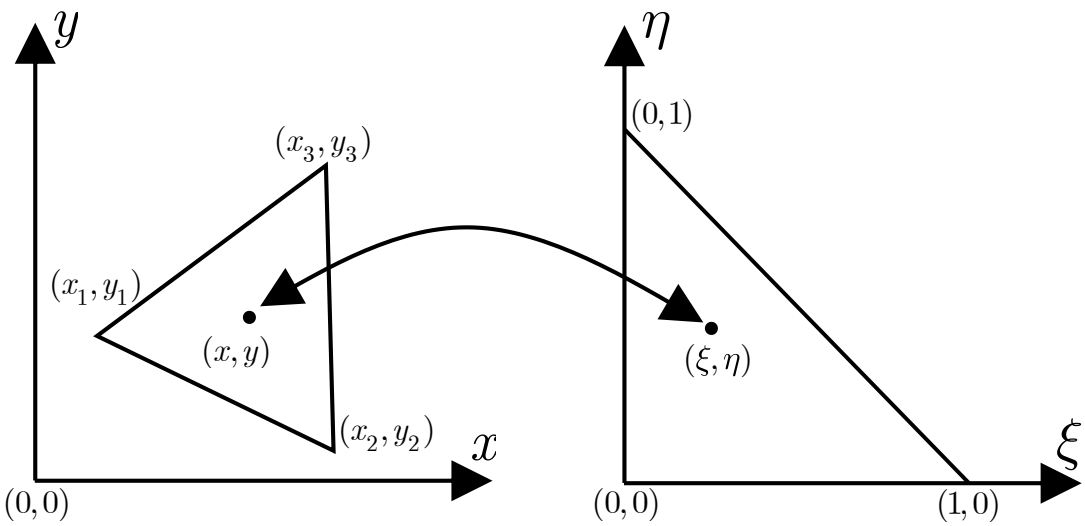


Figure 3.1: The linear transform from an arbitrary triangle T (left) to the standard triangle T_{st} (right), which is defined in the simplex coordinates spanned by ξ axis and η axis.

3.3.2 Integral evaluations

The simplex coordinates now equip us with three new nonzero basis functions over the standard triangle T_{st}

$$\psi_1(\xi, \eta) = 1 - \xi - \eta,$$

$$\psi_2(\xi, \eta) = \xi,$$

$$\psi_3(\xi, \eta) = \eta.$$

With the new coordinate system and basis functions, we can evaluate the integrals over the local triangle as follows

$$K_T = \iint_T \kappa \nabla\varphi_i \nabla\varphi_j dxdy = \nabla\varphi_i \nabla\varphi_j \iint_{T_{st}} \kappa(\xi,\eta) \left|J(\xi,\eta)\right| d\xi d\eta, \tag{3.14}$$

$$C_T = \iint_T f\varphi_i\varphi_j dxdy = \iint_{T_{st}} f(\xi,\eta)\,\psi_i(\xi,\eta)\,\psi_j(\xi,\eta) \left|J(\xi,\eta)\right| d\xi d\eta, \tag{3.15}$$

$$Q_T = \iint_T q_0\varphi_i dxdy = \iint_{T_{st}} q_0(\xi,\eta)\,\psi_i(\xi,\eta) \left|J(\xi,\eta)\right| d\xi d\eta, \tag{3.16}$$

where T is one triangle defined on the whole triangulation Ω, and T_{st} is the standard triangle in the simplex coordinate system mapped from T. As $i,j = 1,2,3$ now, the size of matrices define in equations (3.14), (3.15) and (3.16) is 3×3, 3×3 and 3×1 respectively. Special attention should be paid to equation (3.14). Because the basis functions φ_i are first order piecewise linear (high order basis functions can be also used for higher accuracy but with lower computational efficiency and more complicated implementations), $\nabla\varphi_i\nabla\varphi_j$ in equation (3.14) are simply constant, and so can be taken outside of the integral. Moreover, $J(\xi,\eta)$ is the Jacobian of the transformation, which is of the form

$$J(\xi,\eta) = \left|\frac{\partial(x,y)}{\partial(\xi,\eta)}\right| = \begin{vmatrix} \dfrac{\partial x}{\partial\xi} & \dfrac{\partial x}{\partial\eta} \\ \dfrac{\partial y}{\partial\xi} & \dfrac{\partial y}{\partial\eta} \end{vmatrix} = 2A_T, \tag{3.17}$$

where A_T is the triangular area defined in equation (3.5). Therefore, equations (3.14), (3.15) and (3.16) can be reformulated as

$$K_T = \iint_T \kappa \nabla\varphi_i \nabla\varphi_j dxdy = 2A_T \nabla\varphi_i \nabla\varphi_j \iint_{T_{st}} \kappa(\xi,\eta) d\xi d\eta, \tag{3.18}$$

$$C_T = \iint_T f\varphi_i\varphi_j dxdy = 2A_T \iint_{T_{st}} f(\xi,\eta)\,\psi_i(\xi,\eta)\,\psi_j(\xi,\eta) d\xi d\eta, \tag{3.19}$$

$$Q_T = \iint_T q_0\varphi_i dxdy = 2A_T \iint_{T_{st}} q_0(\xi,\eta)\,\psi_i(\xi,\eta) d\xi d\eta. \tag{3.20}$$

Equation (3.11) must be evaluated separately to the domain integrals for the matrices **K**, **C** and **Q**. First of all, we have to decompose the boundary into *nedges* edges of the triangular elements that lie on it, such that

$$\oint_\Gamma \frac{1}{2A}\varphi_i\varphi_j ds = \sum_{p=1}^{nedges} \int_{L_p} \frac{1}{2A}\varphi_i\varphi_j ds. \tag{3.21}$$

Next, for each edge we consider the two nodes that delimit it: $v_1^{L_p} = (x_1, y_1)$ and $v_2^{L_p} = (x_2, y_2)$. In the same way as we treated, for any point $r^{L_p} = (x, y)$ on the edge L_p, we can transform it into the simplex coordinate system (i.e. interval [0,1]) with

$$r^{L_p} = (1-t)v_1^{L_p} + tv_2^{L_p}. \tag{3.22}$$

Therefore, the new basis functions become

$$\mathcal{L}_1(t) = 1 - t,$$

$$\mathcal{L}_2(t) = t.$$

Next, we can evaluate the integral for each edge separately

$$B_{L_p} = \int_{L_p} \frac{1}{2A} \varphi_i \varphi_j ds = length(L_p) \int_0^1 \frac{1}{2A(t)} \mathcal{L}_i(t) \mathcal{L}_j(t) dt, \tag{3.23}$$

where $i, j = 1, 2$ and equation (3.23) is thus a 2×2 matrix. The 1D two-point quadrature rule can be applied to numerically evaluate this line integral, which is given in Table 3.2 in Section 3.3.3.

3.3.3 Quadrature rules

In order to numerically implement the domain integrals (Equations (3.18), (3.19) and (3.20)), we need to develop quadrature formula over a standard triangle T_{st}, which is given by

$$\iint_{T_{st}} g(\xi, \eta) d\xi d\eta \approx \sum_{i=1}^{N_g} w_i g(\xi_i, \eta_i), \tag{3.24}$$

where N_g is the number of quadrature points used, (ξ_i, η_i) are quadrature points located on a standard triangle T_{st} and w_i are weights (normalized with respect to the original triangular area A_T). Practically, we want to choose points (ξ_i, η_i) and weights w_i so that the quadrature (Equation (3.24)) is accurate and can be also calculated efficiently. In [78], the authors list Gauss-Legendre points and weights for triangles and the degree of the quadrature varies from 1 to 20 (79 quadrature points used for degree 20). Technically, higher degree quadrature involves more points and can lead to higher accuracy, but it is also computationally expensive. For simplicity, we only provide some commonly used

2D quadrature rules using one, three and four quadrature points, which are respectively linear, quadratic and cubic cases. The coordinates of the points and weights are given in Table 3.2, and the corresponding geometry of the points are illustrated in Figure 3.2. Note that quadrature rules for triangular elements are normally not unique. For instance, there exist multiple settings of the quadrature rule for the cubic case.

Table 3.1: Quadrature points and weights corresponding to the geometry in Figure 3.2.

N_g	points	(ξ_i, η_i)	w_i
1	a	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$\frac{1}{2}$
3	a	$\left(0, \frac{1}{2}\right)$	$\frac{1}{6}$
	b	$\left(\frac{1}{2}, 0\right)$	$\frac{1}{6}$
	c	$\left(\frac{1}{2}, \frac{1}{2}\right)$	$\frac{1}{6}$
4	a	$\left(\frac{1}{3}, \frac{1}{3}\right)$	$-\frac{27}{96}$
	b	$\left(\frac{2}{15}, \frac{11}{15}\right)$	$\frac{25}{96}$
	c	$\left(\frac{2}{15}, \frac{2}{15}\right)$	$\frac{25}{96}$
	d	$\left(\frac{11}{15}, \frac{2}{15}\right)$	$\frac{25}{96}$

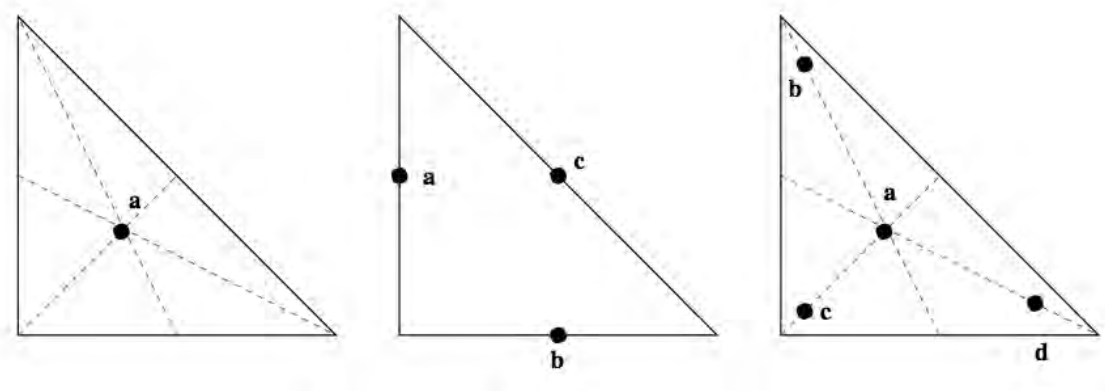


Figure 3.2: A diagram of the 2D quadrature rules with one, three and four quadrature points, respectively.

We are now ready to evaluate equations (3.18), (3.19) and (3.20). Here we take

equation (3.19) as an example, since equations (3.18) and (3.20) are analogous to equation (3.19) and equation (3.19) is more general. The detailed 3×3 matrix form of equation (3.19) can be written as

$$\iint_{T_{st}} f(\xi,\eta)\,\psi_i(\xi,\eta)\,\psi_j(\xi,\eta)d\xi d\eta = \begin{pmatrix} \iint_{T_{st}} f\psi_1\psi_1 d\xi d\eta & \iint_{T_{st}} f\psi_1\psi_2 d\xi d\eta & \iint_{T_{st}} f\psi_1\psi_3 d\xi d\eta \\ \iint_{T_{st}} f\psi_2\psi_1 d\xi d\eta & \iint_{T_{st}} f\psi_2\psi_2 d\xi d\eta & \iint_{T_{st}} f\psi_2\psi_3 d\xi d\eta \\ \iint_{T_{st}} f\psi_3\psi_1 d\xi d\eta & \iint_{T_{st}} f\psi_3\psi_2 d\xi d\eta & \iint_{T_{st}} f\psi_3\varphi_3 d\xi d\eta \end{pmatrix}. \tag{3.25}$$

Let us further use the first component of equation (3.25) as an example. By using the three quadrature points rule in Table 3.1, i.e. $N_g = 3$, $a = (0, \frac{1}{2})$, $b = (\frac{1}{2}, 0)$, $c = (\frac{1}{2}, \frac{1}{2})$ and $w_1 = w_2 = w_3 = \frac{1}{6}$, we have

$$\begin{aligned} \iint_{T_{st}} f\psi_1\psi_1 d\xi d\eta = \frac{1}{6}\Big(& f(0,\frac{1}{2})\psi_1(0,\frac{1}{2})\psi_1(0,\frac{1}{2}) + f(\frac{1}{2},0)\psi_1(\frac{1}{2},0)\psi_1(\frac{1}{2},0) \\ & + f(\frac{1}{2},\frac{1}{2})\psi_1(\frac{1}{2},\frac{1}{2})\psi_1(\frac{1}{2},\frac{1}{2}). \end{aligned} \tag{3.26}$$

As mentioned in Section 3.3.1, the simplex coordinates can perform a linear interpolation of f at points $(0, \frac{1}{2})$, $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{2})$ on the standard triangle T_{st} by only using the values of f at the nodes. The other components in equation (3.25) as well as equation (3.18) and equation (3.20) can be evaluated in the same fashion.

Table 3.2: Quadrature points and weights for the line integral (Equation (3.23)).

N_g	points	t_i	w_i
2	a	0.21132486540519	0.5
	b	0.78867513459481	0.5

Next, we address numerical computation problem for the line integral over interval [0,1]. The two quadrature points rule for the integral is listed in Table 3.2.

Lastly, we present the overall Galerkin finite element algorithm to solve the 2D PDE in equation (2.3)

Algorithm: Galerkin finite element method for equation (2.3)

1: Set up the computational triangular mesh and choose basis functions

2: Compute the matrix $\mathbf{K}$, $\mathbf{C}$, $\mathbf{B}$ and vector $\mathbf{Q}$ in equations (3.9)-(3.12), respectively:

3: • Loop over triangles, for each compute the elemental contributions K_T (3 by 3 matrix) in equation (3.18), C_T (3 by 3 matrix) in (Equation 3.19) and Q_T (3-vector) in equation (3.20)

4: • For this, compute elemental area and use 2D quadrature rule in Table 3.1

5: • Add K_T, C_T and Q_T to $\mathbf{K}$, $\mathbf{C}$ and $\mathbf{Q}$ appropriately

6: • Loop over Neumann boundary edges on Γ, for each compute the boundary contribution B_{L^p} (2 by 2 matrix) in equation (3.23)

7: • For this, compute edge length and use 1D quadrature rule in Table 3.2

8: • Add B_{L^p} to $\mathbf{B}$ appropriately

9: Solve the linear system $\boldsymbol{\mathcal{M}}\Phi = Q$ where $\boldsymbol{\mathcal{M}} = \mathbf{K} + \mathbf{C} + \mathbf{B}$

3.4 3D finite element algorithm

The theory behind 3D FEM is essentially the same as its 2D counterpart. Assume that a 3D finite element domain V is divided by M tetrahedral elements $\{T_i\}_{i=1}^{M}$, with the corresponding N nodes. The 3D Galerkin system has the form of

$$\boldsymbol{\mathcal{M}}\Phi = \mathcal{Q}, \tag{3.27}$$

where $\Phi = (\Phi_1, \Phi_2, ..., \Phi_N)^T$ and $\boldsymbol{\mathcal{M}}$ consists of

$$\mathcal{K}_{ij} = \iiint_V \kappa \nabla\varphi_i \nabla\varphi_j dxdydz, \tag{3.28}$$

$$\mathcal{C}_{ij} = \iiint_V f\varphi_i\varphi_j dxdydz, \tag{3.29}$$

$$\mathcal{B}_{ij} = \iint_\Omega \frac{1}{2A}\varphi_i\varphi_j dxdy. \tag{3.30}$$

The source term $\mathcal{Q} = (\mathcal{Q}_1, \mathcal{Q}_2, ..., \mathcal{Q}_N)^T$ in (3.27) has the form of

$$\mathcal{Q}_i = \iiint_V q_0\varphi_i dxdydz. \tag{3.31}$$

Given the basis functions $\{\varphi_i\}_{i=1}^N$, the 3D solution Φ to the PDE (Equation (2.3)) can be rewritten in the form of

$$\Phi = \sum_{j=1}^{N} \Phi_j \varphi_j, \tag{3.32}$$

where given four vertices of an arbitrary tetrahedron T, i.e. $v_1 = (x_1, y_1, z_1)$, $v_2 = (x_2, y_2, z_2)$, $v_3 = (x_3, y_3, z_3)$ and $v_4 = (x_4, y_4, z_4)$, the basis functions φ_j for the four vertices of the tetrahedron read as

$$\varphi_1(x,y,z) = \frac{1}{6V_T} \begin{vmatrix} 1 & x & y & z \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix},$$

$$\varphi_2(x,y,z) = \frac{1}{6V_T} \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x & y & z \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix},$$

$$\varphi_3(x,y,z) = \frac{1}{6V_T} \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x & y & z \\ 1 & x_4 & y_4 & z_4 \end{vmatrix},$$

$$\varphi_4(x,y,z) = \frac{1}{6V_T} \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x & y & z \end{vmatrix}.$$

Here V_T denotes the volume of the tetrahedron T, which can be evaluated by

$$V_T = \frac{1}{6} \begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}. \tag{3.33}$$

Next, we need simplex coordinates to evaluate volume integrals $\boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{C}}$, and $\boldsymbol{\mathcal{Q}}$ over each tetrahedron. Now consider a tetrahedron T consisting of the four vertices v_1, v_2, v_3 and v_4 in the Cartesian coordinate system. Any point $r = (x, y, z)$ located in the tetrahedron can be represented as a weighted sum of the four nodes via

$$r = (1 - \xi - \eta)v_1 + \xi v_2 + \eta v_3 + \gamma v_4. \tag{3.34}$$

Equation (3.34) represents that the original tetrahedron in the Cartesian coordinate system spanned by x-axis, y-axis and z-axis has been linearly mapped to a standard tetrahedron T_{st} in the simplex coordinate system spanned by ξ-axis, η-axis and γ-axis. See the illustration in Figure 3.3 for details. Since equation (3.34) is the form of a linear transform, the simplex coordinates will allow us to perform a linear interpolation of a function at points in the tetrahedron if the values of the function are known at the nodes.

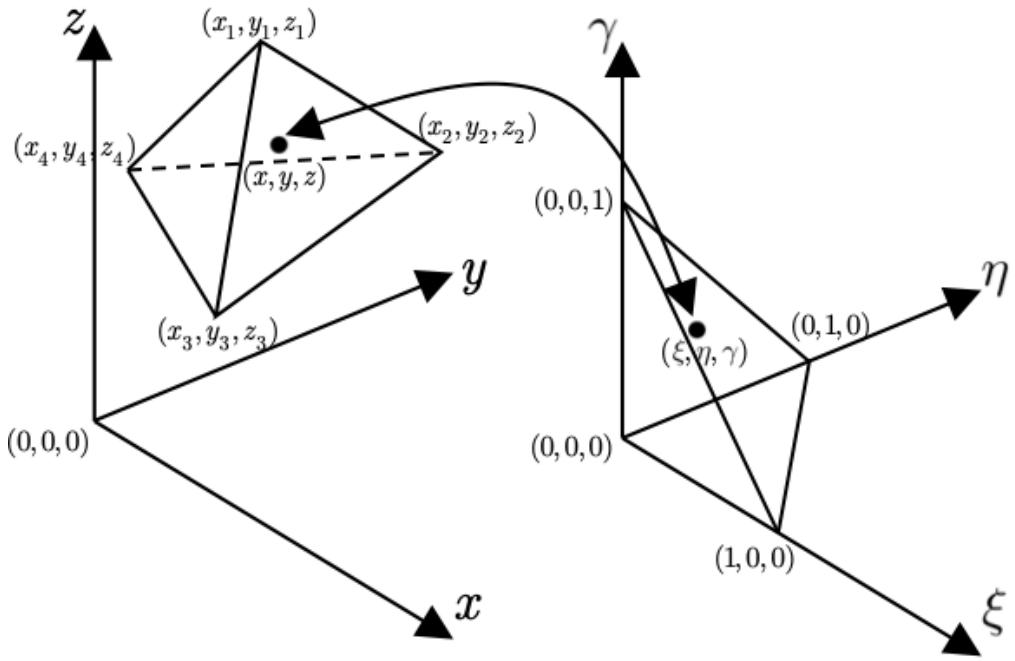


Figure 3.3: The linear transform from an arbitrary tetrahedron T (left) to the standard tetrahedron T_{st} (right), which is defined in the simplex coordinates spanned by ξ-axis, η-axis and γ-axis.

The simplex coordinates now equip us with four new nonzero basis functions over the

standard tetrahedron T_{st}

$$\psi_1(\xi,\eta,\gamma) = 1-\xi-\eta-\gamma,$$

$$\psi_2(\xi,\eta,\gamma) = \xi,$$

$$\psi_3(\xi,\eta,\gamma) = \eta,$$

$$\psi_4(\xi,\eta,\gamma) = \gamma.$$

With the new coordinate system and basis functions, we can evaluate the integrals over each local standard tetrahedron T_{st} in the simplex coordinate system as follows

$$\mathcal{K}_T = \nabla\varphi_i \nabla\varphi_j \iiint_{T_{st}} \kappa(\xi,\eta,\gamma) \left|J(\xi,\eta,\gamma)\right| d\xi d\eta d\gamma, \tag{3.35}$$

$$\mathcal{C}_T = \iiint_{T_{st}} f(\xi,\eta,\gamma)\, \psi_i(\xi,\eta,\gamma)\, \psi_j(\xi,\eta,\gamma) \left|J(\xi,\eta,\gamma)\right| d\xi d\eta d\gamma, \tag{3.36}$$

$$\mathcal{Q}_T = \iiint_{T_{st}} q_0(\xi,\eta,\gamma)\, \psi_i(\xi,\eta,\gamma) \left|J(\xi,\eta,\gamma)\right| d\xi d\eta d\gamma. \tag{3.37}$$

Note that as $i,j = 1,2,3,4$ now, the size of equations (3.35), (3.36) and (3.37) is 4×4, 4×4 and 4×1 respectively. The basis functions φ_i we used in this study are first order piecewise linear, so $\nabla\varphi_i\nabla\varphi_j$ in equation (3.35) always stay constant. $J(\xi,\eta,\gamma)$ is the Jacobian of the following form

$$J(\xi,\eta,\gamma) = \left|\frac{\partial(x,y,z)}{\partial(\xi,\eta,\gamma)}\right| = \begin{vmatrix} \frac{\partial x}{\partial\xi} & \frac{\partial x}{\partial\eta} & \frac{\partial x}{\partial\gamma} \\ \frac{\partial y}{\partial\xi} & \frac{\partial y}{\partial\eta} & \frac{\partial y}{\partial\gamma} \\ \frac{\partial z}{\partial\xi} & \frac{\partial z}{\partial\eta} & \frac{\partial z}{\partial\gamma} \end{vmatrix} = 6V_T, \tag{3.38}$$

where V_T is the tetrahedral volume defined in equation (3.33). We still need to address the integral problem $\mathcal{B}_{ij}$ in equation (3.30) which is on the surface Ω of the whole 3D FEM domain V. Note that this process should be made separately to the volume integrals for the matrices $\boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{Q}}$. First of all we have to decompose the surface into $ntri$ triangles that lie on it, such that

$$\mathcal{B}_{ij} = \iint_{\Omega} \frac{1}{2A}\varphi_i\varphi_j dxdy = \sum_{p=1}^{ntri} \iint_{T_p} \frac{1}{2A}\varphi_i\varphi_j dxdy. \tag{3.39}$$

Note that unless the triangle face is on the boundary surface, this integral is equal to zero.

Next, for each triangle on the boundary surface its integral problem simply goes back to the 2D domain integral evaluation which is evident in Section 3.3.2. We thus can evaluate the integral for each boundary triangle in the simplex coordinate system with the 2D shape functions $\psi_1(\xi,\eta)$, $\psi_2(\xi,\eta)$ and $\psi_3(\xi,\eta)$

$$\mathcal{G}_{T_p} = \iint_{T_p} \frac{1}{2A}\varphi_i\varphi_j dxdy = 2A_{T_p} \iint_{T_{st}} \frac{1}{2A(\xi,\eta)}\psi_i(\xi,\eta)\,\psi_j(\xi,\eta)\,d\xi d\eta. \tag{3.40}$$

Here A_{T_p} is the area of the triangle T_p on the boundary surface, and equation (3.40) is 3×3 matrix due to $i,j = 1,2,3$. The numerical implementation of equation (3.40) is described in Section 3.3.3.

3.4.1 3D quadrature rules

In order to numerically implement the volume integrals (Equations (3.35), (3.36) and (3.37)), we need to develop a quadrature formula over a standard tetrahedron, which is given by

$$\iiint_{T_{st}} g(\xi,\eta,\gamma)\,d\xi d\eta d\gamma \approx \sum_{i=1}^{N_g} w_i g(\xi_i,\eta_i,\gamma_i) \tag{3.41}$$

where N_g is the number of quadrature points used, (ξ_i,η_i,γ_i) are quadrature points located in a standard tetrahedron T_{st} and w_i are weights (normalized with respect to the original tetrahedron volume V_T). In this study, we use the Gauss-Legendre quadrature of degree 4 including 24 quadrature points and weights for our 3D FEM implementation, which is listed in Table 3.3. One can refer to [79] for more options.

Finally, we present the overall Galerkin finite element algorithm to find the 3D solution Φ to the PDE in equation (2.3)

3.5 Challenges

This chapter gives a detailed description about how to use the Galerkin finite element method to solve the DE and its RBC. However, there are some limitations in this dis-

Table 3.3: Quadrature points and weights for the volume integrals (Equations (3.35), (3.36) and (3.37)).

N_g	points	$(\xi_i, \eta_i, \gamma_i)$	w_i
24	a	(0.21460287125915202, 0.21460287125915202, 0.3561913862225439)	6.6537917096945820e-3
	b	(0.21460287125915202, 0.3561913862225439, 0.21460287125915202)	6.6537917096945820e-3
	c	(0.3561913862225439, 0.21460287125915202, 0.21460287125915202)	6.6537917096945820e-3
	d	(0.21460287125915202, 0.21460287125915202, 0.21460287125915202)	6.6537917096945820e-3
	e	(0.040673958534611353, 0.040673958534611353, 0.87797812439616594)	1.6795351758867738e-3
	f	(0.040673958534611353, 0.87797812439616594, 0.040673958534611353)	1.6795351758867738e-3
	g	(0.87797812439616594, 0.040673958534611353, 0.040673958534611353)	1.6795351758867738e-3
	h	(0.04067395853461135, 0.040673958534611353, 0.040673958534611353)	1.6795351758867738e-3
	i	(0.32233789014227551, 0.32233789014227551, 0.0329863295731735)	9.2261969239424536e-3
	j	(0.32233789014227551, 0.0329863295731735, 0.32233789014227551)	9.2261969239424536e-3
	k	(0.0329863295731735, 0.32233789014227551, 0.32233789014227551)	9.2261969239424536e-3
	l	(0.32233789014227551, 0.32233789014227551, 0.32233789014227551)	9.2261969239424536e-3
	m	(0.063661001875017525, 0.26967233145831580, 0.6030056647916492)	8.0357142857142857e-3
	n	(0.063661001875017525, 0.6030056647916492, 0.26967233145831580)	8.0357142857142857e-3
	o	(0.26967233145831580, 0.063661001875017525, 0.6030056647916492)	8.0357142857142857e-3
	p	(0.26967233145831580, 0.6030056647916492, 0.063661001875017525)	8.0357142857142857e-3
	q	(0.6030056647916492, 0.063661001875017525, 0.26967233145831580)	8.0357142857142857e-3
	i	(0.6030056647916492, 0.26967233145831580, 0.063661001875017525)	8.0357142857142857e-3
	s	(0.063661001875017525, 0.063661001875017525, 0.6030056647916492)	8.0357142857142857e-3
	t	(0.063661001875017525, 0.6030056647916492, 0.063661001875017525)	8.0357142857142857e-3
	u	(0.6030056647916492, 0.063661001875017525, 0.063661001875017525)	8.0357142857142857e-3
	v	(0.063661001875017525, 0.063661001875017525, 0.26967233145831580)	8.0357142857142857e-3
	w	(0.063661001875017525, 0.26967233145831580, 0.063661001875017525)	8.0357142857142857e-3
	x	(0.26967233145831580, 0.063661001875017525, 0.063661001875017525)	8.0357142857142857e-3

cretization method.

• Implementation of the FEM is complicated when higher order basis functions are involved. It is because 1) higher degree polynomial basis functions require more operations to compute, 2) the corresponding quadrature rules for numerical integration to compute the matrix for the discrete problem must be of correspondingly higher order,

Algorithm: 3D Galerkin finite element method for equation (2.3)

1: Set up the computational tetrahedral mesh and choose basis functions

2: Compute the matrix $\boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{C}}$, $\boldsymbol{\mathcal{B}}$ and vector $\boldsymbol{\mathcal{Q}}$ in equations (3.28)-(3.31), respectively:

3: • Loop over tetrahedrons, for each compute the elemental contributions $\mathcal{K}_T$ (4 by 4 matrix) in (3.35), $\mathcal{C}_T$ (4 by 4 matrix) in equation (3.36) and $\mathcal{Q}_T$ (4-vector) in equation (3.37)

4: • For this, compute elemental volume and use 3D quadrature rule in Table 3.3

5: • Add $\mathcal{K}_T$, $\mathcal{C}_T$ and $\mathcal{Q}_T$ to $\boldsymbol{\mathcal{K}}$, $\boldsymbol{\mathcal{C}}$ and $\boldsymbol{\mathcal{Q}}$ appropriately

6: • Loop over surface triangles on Ω, for each compute the boundary contribution $\mathcal{B}_{T^p}$ (3 by 3 matrix) in equation (3.40)

7: • For this, compute triangular area and use 2D quadrature rule in Table 3.1

8: • Add $\mathcal{B}_{T_p}$ to $\boldsymbol{\mathcal{B}}$ appropriately

9: Solve the linear system $\boldsymbol{\mathcal{M}}\Phi = \mathcal{Q}$ where $\boldsymbol{\mathcal{M}} = \boldsymbol{\mathcal{K}} + \boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{B}}$

which means more quadrature points at which to evaluate the basis functions, 3) the matrix $\boldsymbol{\mathcal{M}}$ for the discrete problem is denser, which means more computation to solve the linear system [80].

• Even using the first order piecewise linear basis functions, the computational complexity is still high. The 24 quadrature points and weights for the volume integrals as shown in Table 3.3 can complex the FEM implementation and decrease the computational efficiency.

• Under the finite element strategy, triangular elements are used in 2D analysis whereas tetrahedron elements are used in 3D analysis. Therefore, different formulations of basis functions, simplex coordinates and quadrature rules are needed for geometries with different dimensions, leading to different FEM implementations.

• Accurate solution in FE analyses highly relies on the high mesh resolution and refinement. As FEM operates on each local element, high quality meshes are crucial to obtaining accurate results. This is also validated in Chapter 5, where FEM performance is evaluated on the meshes with low and high resolution respectively.

Due to these challenges of FEM, in Section 6, we will present our work on a new discretization approach which is more computationally efficient and allows identical implementation for geometries in different dimensions.

CHAPTER 4

LITERATURE REVIEW AND SPECTRAL-L_1 MODEL FOR DIFFUSE OPTICAL TOMOGRAPHY

Reconstruction of images from DOT measurements is a difficult inverse problem. The limited availability of boundary measurements and the diffusive nature of light propagation in tissue [69, 81] make the problem non-linear and ill-posed. Different distributions of optical parameters can lead to the same boundary measurements. Therefore iterative solutions with effective methods for regularization are necessary to alleviate the ill-posedness and obtain a plausible solution optimized for particular chosen characteristic. In this chapter, we review the regularization methods proposed by the DOT community in the recent two decades and then introduce our proposed regularization methods for improving the DOT reconstruction.

4.1 Literature review of advanced regularization methods in DOT

Reviewing the regularization methods proposed in the DOT community over the past two decades, three categories of regularization methods can be identified: regularization methods with sparsity constraints (RSC), regularization methods with spatial priors (RSP-I), and regularization methods with spectral prior (RSP-II).

4.1.1 Regularization methods with sparsity constraints (RSC)

The simplest regularization applied to alleviate the ill-posed problem is Tikhonov regularization, which was proposed by Pogue et al [82] to reduce high frequency noises in the reconstructed images. Let y and Ψ be the logarithm of the boundary measurements and model based prediction. The corresponding objective function is given as follows

$$\Omega(\mu) = \|y - \Psi(\mu)\|_2^2 + \lambda \|\mu - \mu_0\|_2^2, \tag{4.1}$$

where λ is the regularization parameter and μ_0 is the initial estimate of the optical properties, which in DOT is typically obtained from calibration [83]. μ is the optical properties which need to be recovered. Taking the first order derivative with respect to μ leads to the following first order Maclaurin series expansion

$$\left(\mathbf{J}^{k\mathrm{T}}\mathbf{J}^k + \lambda\mathbf{I}\right)\delta\mu = \mathbf{J}^{k\mathrm{T}}\delta\Psi^{k-1} - \lambda\left(\mu_{k-1} - \mu_0\right). \tag{4.2}$$

where the superscript k denotes the kth iteration. It is easy to see that equation (4.2) can be computed analytically. Tikhonov regularization suppresses the high-frequency components (normally noise) of the reconstructed image leading to smooth reconstructions. Therefore Tikhonov regularization is the most popular approach in the DOT community. However, this regularization has the drawback of being unable to preserve sharp features in the reconstructed images and oversmoothing the results [22]. When the boundary measurements have large dynamic range, Tikhonov regularization leads to artefacts in specific data points near the source-detector locations [23].

Features of interest in DOT, such as tumours in the breast or activations in the brain, are typically spatially localized and in this case the vector corresponding to the difference in the optical properties relative to the background is sparse with only a few non-zero elements [84–87]. Based on this assumption, regularization methods with sparsity constraints (RSC) were proposed to preserve the sparsity of the reconstructed DOT images and improve the accuracy. The typical RSC is L_1-norm regularization $\|\mu - \mu_0\|_1$ which

was first proposed by Cao et al. [88]. Experiments revealed that, compared to Tikhonov regularization, L_1-norm regularization can promote more localized reconstructed results and is able to preserve edge-like features. Both Tikhonov and L_1-norm regularizations can be solved by convex optimization schemes where a unique solution can be guaranteed [87, 89]. A more general L_p-norm ($\|\mu - \mu_0\|^p$ where $0 < p < 1$) regularizations have also been studied for DOT image reconstruction [86, 90] and are known to induce sparsity to the reconstructed image [91]. It was first proposed for time-domain diffuse optical tomography by Okawa et al. [90]. However, in L_p-norm regularization it is difficult to calculate the gradient of the objective function because it is not differentiable everywhere. When calculating the derivative of L_p-norm and $|\delta\mu|$ is small, $|\delta\mu|^{p-1}$ tends to infinity. To avoid this difficulty, Okawa et al. [90] reformulated $|\delta\mu|$ with a parameter z:

$$\delta\mu = |z|^{2/p} \cdot \operatorname{sgn}(z). \tag{4.3}$$

Then the DOT reconstruction with the L_p-norm becomes

$$\Omega(z) = \|y - \Psi(z)\|_2^2 + \lambda \|z\|_2^2. \tag{4.4}$$

Equation (4.4) can be easily solved with the nonlinear conjugate gradient method [92]. Experiments confirmed that L_p-norm regularization can improve the accuracy and preserve sharp features of the solutions. However, with p decreasing, an area with small $\delta\mu$ may disappear because the recovered $\delta\mu$ becomes zero due to the excessive effect of the L_p regularization. The L_p regularization with a small p can strongly localizes the solution, and the size of the reconstructed area decreases with the decrease in p. In addition, L_p-norm regularization is known to be nonconvex meaning that local minima exist [93] and unique solutions cannot be guaranteed.

Theoretically, the sparsest solution should be found in L_0-norm regularization which counts the nonzero elements. However the problem of using exact L_0-norm regularization is an NP-hard problem [94] which is very difficult to solve computationally. Prakash et al. [86] employed an approximation for the L_0-norm, namely smooth L_0-norm, to provide

more accurate sparse reconstruction results. They compared the proposed smooth L_0-norm regularization and other RSCs (L_2-norm, L_1-norm, L_p-norm where p is set as 0.45) on a regular circular geometry. It is illustrated that as p goes to zero, the quantitative accuracy of the reconstruction improves. Smooth L_0-norm performs the best among those RSCs in preserving sharp features. However, it is interesting to notice that, for irregular geometries, L_p-norm with $p = 0.65$ performed superior compared to other RSCs (L_2-norm, L_1-norm, L_0-norm). Therefore for more regular geometry, RSC can provide the best values compared to the traditional L_2-norm based reconstruction method, with p being close to zero providing the highest accuracy. Although L_0-norm regularization is the true sparsity measure, Donoho et al. [95] validated mathematically that L_1-norm regularization is able to give essentially the same sparsity as the L_0-norm regularization.

Even though promising results have been obtained in RSC over the conventional Tikhonov regularization, accurate DOT reconstruction is still challenging. The reasons are: 1) Boundary measurements sometimes involve strong noise which satisfies the sparsity assumption [96]. In this case, true signals are difficult to be differentiated from background noise. 2) RSC is based on the assumption that the underlying image to be recovered is sparse. However when this assumption does not hold, these regularization methods may fail to reconstruct the underlying signal.

4.1.2 Regularization methods with spatial priors (RSP-I)

Regularization methods with spatial priors (RSP-I) are popular in imposing a specific spatial structure to improve quantitative accuracy of reconstructed images. It is first used in the least square scheme where only a data fitting term is included. The most straightforward way is using regions of interest (ROI) segmented from images of other modalities like x-ray, MRI, or ultrasound as spatial priors to limit the DOT image reconstruction [97]. In this case, the Jacobian matrix only needs to be calculated over the ROI area. OLeary [97] describes the method which segments image into different regions. Each region becomes homogeneous and has its own optical property values. As a result,

the number of unknowns is reduced to the number of segmented regions and the speed of image reconstruction is improved. However, these methods rely on an accurate image segmentation. If the segmentation is poor, this can lead to incorrect reconstruction results. In order to overcome this drawback, the recent RSP-I methods [22, 98–101] minimize an objective function that has a data term and a spatial prior based regularization term. This strategy can adjust the influence of the spatial priors by appropriately weighting the regularization term of the objective function.

The simplest RSP-I is the hard prior [98], which applies a matrix transformation to the sensitivity matrix $\mathbf{J}$. This transform matrix contains the region information of each unknown FEM node, transforming the number of unknown parameters to be estimated into the number of regions segmented from MRI images. This transformation significantly reduces the ill-posedness of the inverse problem. However, using the hard prior, each region is constrained to take homogeneous values, resulting in lower spatial information in each region. In order to overcome the drawback, a so-called $\mathbf{L}$ matrix is introduced as a soft prior to the regularization term

$$\Omega(\mu) = \|y - \Psi(\mu)\|_2^2 + \lambda \|\mathbf{L}(\mu - \mu_0)\|_2^2, \tag{4.5}$$

where $\mathbf{L}$ correlated the optical properties of each node to the others in the same region. As such, $\mathbf{L}$ can smooth each region and allow discontinuous changes across different regions. The update equation involving a soft prior is:

$$\left(\mathbf{J}^{k\mathrm{T}}\mathbf{J}^k + \lambda \mathbf{L}^T\mathbf{L}\right)\delta\mu = J^{k\mathrm{T}}\delta\Psi^{k-1} - \lambda \mathbf{L}^T\mathbf{L}(\mu_{k-1} - \mu_0). \tag{4.6}$$

Two forms of the $\mathbf{L}$ matrix discussed in [98] by Brooksby et al. are as follows:

1). Laplacian form [99, 100]

$$L_{i,j} = \begin{cases} 0 & \text{if } i \neq j;\ i \text{ and } j \text{ are not in the same region} \\ -1/N_r & \text{if } i \neq j;\ i \text{ and } j \text{ are in the same region} \\ 1 & \text{if } i = j \end{cases}, \tag{4.7}$$

where N_r is the number of nodes in that region. In this case, $\mathbf{L}^T\mathbf{L}$ approximates a second-order Laplacian smoothing operator within each region. It functionally calculates the average value of the update within a region and allows discontinuity across different regions. This method works well even if there exist errors in the anatomical information.

2). Helmholtz form [22]

$$L_{i,j} = \begin{cases} 0 & \text{if } i \neq j;\ i \text{ and } j \text{ are not in the same region} \\ -1/\left[N_r + (\varsigma h)^2\right] & \text{if } i \neq j;\ i \text{ and } j \text{ are in the same region} \\ 1 & \text{if } i = j \end{cases}, \tag{4.8}$$

where $\varsigma = 1/l$ and l is the largest radius of the region. h is the distance between node i and j. $\mathbf{L}^T\mathbf{L}$ here is a second-order Helmholtz smoothing operator.

These two soft priors (Equations 4.7 and 4.8) and the hard prior were compared in detail in paper [22]. The experiments revealed that the soft priors allow the tissue optical properties to vary within each region, while the hard prior constrains each region to be homogeneous. The hard prior does not perform well when the anatomical information has segmentation boundary errors, while soft-prior based approaches are robust to the boundary errors. Experiments on phantom data reveal that the Helmholtz form (Equation (4.8)) could give a better estimation of scattering coefficient and the Laplacian form (Equation (4.7)) leads to a superior absorption coefficient estimate. Therefore Laplacian form can give the best estimates of total hemoglobin concentration (HbT), haemoglobin oxygen saturation ($StO_2\%$) and water fraction (H_2O) while Helmholtz structure can estimate the scattering power and scattering amplitude better.

Another kind of soft prior is the application of different regularization parameters for different segmented regions [101]. Li et al used an x-ray tomosynthesis volume to segment the breast into two regions and the corresponding objective function is given by (details in [102] part 3.B)

$$\Omega(\mu) = \|y - \Psi(\mu)\|_2^2 + \lambda_1 \|(\mathbf{I} - \mathbf{S})(\mu - \mu_0)\|_2^2 + \lambda_2 \|\mathbf{S}(\mu - \mu_0)\|_2^2, \tag{4.9}$$

where $\mathbf{I}$ is the identity matrix and $\mathbf{S}$ is a diagonal matrix, whose diagonal values describe the hypothesized support of the target indicated by the x-ray image. More specifically, in the discrete form, the ith element of the diagonal is 1 if the ith node is identified as part of the target by the x-ray image and 0 otherwise. λ_1 and λ_2 are the two regularization parameters that control the degree of regularization in the background and target, respectively. Note that, if λ_1 is set equal to λ_2, the formulation is reduced to the conventional Tikhonov regularization scheme. The update equation is

$$\left(\mathbf{J}^{k\mathrm{T}}\mathbf{J}^{k} + \lambda_r \mathbf{I}\right) \Delta\mu = \mathbf{J}^{k\mathrm{T}}\delta^{k-1} - \lambda(r)\left(\mu_{k-1} - \mu_0\right). \tag{4.10}$$

where λ_r is a diagonal matrix with the diagonal elements being λ_1 or λ_2. The advantage of this method is that the influence of the structural information can be adjusted by appropriately weighting the structural component of the objective function. However, it also amplifies the image noise in the designated region [82]. More importantly, the reconstructed area is confined to a volume significantly smaller than the spatial prior collected from other modalities.

Depth information [87] is also understood as a spatial prior, which improves the spatial resolution and depth localization simultaneously in DOT imaging. Depth information is helpful when the number of photons decreases dramatically with the increase in propagation depth. The measurement sensitivity in deep tissue is significantly lower than that in superficial tissue. The lower measurement sensitivity for deeper layers results in poor depth resolution and biases reconstructed images towards the superficial layers. Kavuri et al. [87] combined Depth Compensation Algorithm (DCA) with L_1-norm regularization to improve spatial resolution and depth localization for DOT. DCA adopts a weighted matrix, which provides a pseudo-exponential increase in magnitude with propagation depth, to compensate the sensitivity matrix $\mathbf{J}$ in deeper layers directly. The comprehensive details on DCA can be found in Niu et al. [103].

The RSI-I methods described above can only be solved with an indirect two-step procedure. First, high resolution anatomical images from other modalities are segmented

into a small number of regions. Then, this anatomical information imposes either a hard or a soft constraint on the image reconstruction process. Combining anatomical information from high resolution imaging modalities to guide DOT has been proved to be an efficient strategy for improving the quality of the reconstructed images. However, the segmentation process is prone to errors and can be extremely time consuming. As a solution, a direct reconstruction method is proposed in [104], where the anatomical information is implicitly incorporated into the penalty term without users′ intervention. The matrix $\mathbf{L}$ can be written as

$$L_{i,j} = \begin{cases} 1 & \text{if } i = j \\ -\frac{1}{F_i} exp\left(-\frac{|\gamma_i - \gamma_j|^2}{2\sigma_g}\right) & \text{otherwise} \end{cases}, \tag{4.11}$$

where γ is the grayscale value of anatomical images which corresponds to a particular FEM node. The grayscale values were normalized to the maximum within the image. σ_g is the characteristic grayscale difference over which to apply regularization, and F_i is a normalization factor chosen for each row. Experiments illustrate that the reconstructed results were significantly improved in terms of their qualitative and quantitative accuracy, as well as their robustness. This proposed approach dramatically reduces processing time and expands the potential of multimodal imaging by fully automating the image reconstruction process. Later Althobaiti et al [105] adopted the same technique but using Ultrasound images to regulate the objective function. Improvements in reconstructed lesion shapes and the spatial distribution of absorption maps were clearly observed.

Inspired by RSC and RSP-I, authors in [106–108] proposed a structured sparsity where fewer measurements are required for image reconstruction than the pure RSC method. Chen et al. [96] proposed a clustered sparsity reconstruction method to improve the DOT reconstruction by exploiting the clustered structure. In this work, the clustered sparsity is approximated by grouping each node with its neighbor nodes. With this group setting, non-zero nodes of the reconstructed image will be only in the same groups, leading to the clustered structure of the non-zero pixels. The proposed method utilizes the mixed

$L_{2,1}$-norm regularization, leading to the following objective functional

$$\Omega(\mu) = \|y - \Psi(\mu)\|_2^2 + \lambda \|\mu - \mu_0\|_{2,1}, \tag{4.12}$$

where $\|\mu - \mu_0\|_{2,1} = \Sigma_g \left\| (\mu - \mu_0)_g \right\|_2$, g denotes one of the group described above and $(\mu - \mu_0)_g$ denotes the components in the specific group. This clustered sparsity problem is modelled with convex programming and solved by a new algorithm based on the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) framework [109]. The needed number of measurements becomes $O(K + C\log(M/C))$ where K is the number of nonzero elements, M is the total number of measurements and C denotes the number of clusters with $C \ll K$. In the case where the number of measurement is not sufficient for RSC, it is still possible to achieve successful recovery with clustered sparsity. The proposed method was compared with pure RSCs (L_0, L_p where p is set as 0.5, L_1 and L_2) on human brain activations and found that this method can recover images with the fewest artifacts and best contrast in the expected region. Pure RSCs encourage sparsity and have no other constraints on the locations of the non-zero values, leading to slightly distorted and distributed results.

4.1.3 Regularization methods with spectral priors (RSP-II)

The quantities of interest in DOT experiments are typically chromophore concentrations (mainly HbO_2 and Hb) and scattering parameters (scatter amplitude and scatter power) rather than the absorption and scattering coefficients themselves. Measurements have to be taken at multiple wavelengths in order to provide sufficient information to recover the distributions of these chromophores. There are two main approaches for reconstruction of chromophore images using multiple wavelength measurements: non-spectral methods and spectrally constrained methods. Non-spectral methods (traditional DOT) reconstruct the absorption and scattering coefficients at each wavelength independently and then calculate the chromophore concentrations using Beer's law and scattering parameters using Mie scattering theory [110]. In spectrally constrained DOT (RSP-II), these constraints can be incorporated into the reconstruction directly to estimate oxy-hemoglobin HbO_2,

deoxy-hemoglobin Hb, water, scatter amplitude and scatter power.

This technique was first documented by Corlu et al. [111], wherein absolute concentrations of chromophores are solved iteratively with a minimization function [112]. Compared with non-spectral methods, RSP-II has been shown to be better at suppressing artefacts in the resulting reconstructed images and to reduce crosstalk between chromophores and scatter parameters in breast imaging [100, 113, 114]. After that, Li et al. [115] applied RSP-II to continuous-wave data in functional (also known as temporal or differential) imaging. Experiments revealed that when considering functional imaging, linear DOT algorithms that incorporate spectral priors directly in the image reconstruction can provide a solution that is more stable, albeit with an increase in crosstalk between chromophores. Srinivasan et al. [113] extended the approach to the frequency domain and showed experimental evidence of improved quantification. Intes et al. [116] combined RSP-II with RSP-I to improve the accuracy and quality of NIR images. Analysis of these two techniques (RSP-II and RSP-I) was first provided by Brooksby et al. [100]. It is proved that spatial priors improve image quality by reducing artifacts but does not significantly improve parameter quantification. The spectral prior obtained by including the intrinsic behavior of tissue chromophores and scattering plays a more important role in preserving quantitative parameter estimates. A synergy between these two priors was shown to yield the most accurate characterization of breast tissue properties [100].

In Table 4.1, we give a brief summary of the three regularization categories.

Table 4.1: Summary of regularization methods

Category	Name	Advantage	Disadvantage
Popular used	Tikhonov (L_2-norm) regularization [22, 23, 82]	1. Analytical solution 2. Suppress high-frequency components 3. Convex	1. Over-smooth the results 2. Unable to preserve sharp features
Regularization methods with sparsity constraints (RSC)	L_1-norm regularization [68, 87–89]	1. Promote localized results 2. Keep sharp features 3. Convex	
	L_p-norm regularization [86, 90, 91]	1. Induce sparsity 2. Keep sharp features	1. Nonconvex 2. Small p leads to strong localized results
	L_0-norm regularization [86, 94, 95]	1. Lead to sparsest solution	1. NP-hard problem 2. Give essentially the same sparsity as the L_1-norm regularization
Regularization methods with spatial priors (RSP-I)	Hard prior [98]	1. Significantly reduces the ill-posedness of the inverse problem	1. Low spatial information in each region 2. Suffer from segmentation errors 3. Segmentation process is time consuming
	Soft prior with Laplacian matrix [99, 100]	1. Allow the tissue optical properties to vary within each region 2. More superior absorption coefficient estimate	1. Suffer from segmentation errors 2. Segmentation process is time consuming
	Soft prior with Helmholtz matrix [22]	1. Allow the tissue optical properties to vary within each region 2. Better estimation of scattering coefficient	1. Suffer from segmentation errors 2. Segmentation process is time consuming
	Soft prior using different regularization parameters for different segmented regions [82, 101]	The influence of the structural information can be adjusted by appropriately weighting the structural component of the objective function	1. Amplify the image noise in the designated region 2. Over confined results 3. Suffer from segmentation errors 4. Segmentation process is time consuming
	Depth information [87]	Combine Depth Compensation Algorithm (DCA) with L_1-norm regularization to improve spatial resolution and depth localization	1. Affected by segmentation errors 2. Segmentation process is time consuming
	Direct reconstruction method [104]	No users′ intervention	
Regularization methods with spectral priors (RSP-II)	[100, 111–114]	1. Suppress artefacts in the resulting reconstructed images 2. Reduce crosstalk between chromophores and scatter parameters in absolute imaging	Increase crosstalk between chromophores in functional imaging

4.2 L_1-norm regularization for spectrally constrained diffuse optical tomography

Inspired by the three regularization categories, in this section, I propose an L_1-norm regularization based nonlinear DOT reconstruction method for spectrally constrained diffuse optical tomography (SCDOT). This work combines the L_1-norm regularization(RSC) with RSP-II to promote localized recovery of chromophore concentrations in absolute imaging.

The contents of this chapter were the subject of the following publication.

- **Wenqi Lu, Daniel Lighter, and Iain B. Styles, *L1-norm based nonlinear reconstruction improves quantitative accuracy of spectral diffuse optical tomography*, Biomedical Optics Express 9, 1423-1444 (2018)**

4.2.1 Introduction

In traditional DOT, L_1-norm regularization has only been adopted for single wavelength DOT image reconstruction. However, there are no guarantees that the solutions across multiple wavelengths will be consistent with Beer's law. This would require the assumption that the regularization will have the same sparsifying effect at all wavelengths. This may not necessarily be true, given that signal-to-noise ratio(SNR), scattering etc are different. With the incorporation of spectral priors, spectrally constrained diffuse optical tomography (SCDOT) can be used to constrain the solution space to those solutions that are physically spectrally consistent. Therefore, reconstruction with sparsity and spectral regularization simultaneously applied will constrain the solution space much more reliably than their sequential application. To the best of our knowledge, L_1-norm has not yet been

used in SCDOT image reconstruction. We introduce a novel algorithm, spectral-L_1, which combines the sparsity-preserving advantages of L_1-norm regularization with the spectral priors imposed by coupling optical properties across multiple wavelengths, to solve the inverse problem for image reconstruction in SCDOT. The key advance is to adapt the DOT reconstruction process to incorporate efficient methods for solving each iterative step. These are necessary because the L_1-norm regularization is non differentiable and the update terms in the reconstruction process cannot be computed analytically. We investigate three algorithms for solving the update term: iteratively reweighted least square algorithm (IRLS) [117], alternating directional method of multipliers (ADMM) [118–121] and fast iterative shrinkage-thresholding algorithm (FISTA) [109, 122]. All three methods have been widely used to obtain sparse solutions to linear systems. IRLS and ADMM are second-order algorithms that require explicit inversion of a large matrix; FISTA is a first-order algorithm that does not require explicit matrix inversion, but does require a gradient operator to be constructed.

We adapt the DOT reconstruction process to use these methods for the solution of the update terms. An automated method to automatically select the regularization parameters is developed which is based on the L-curve method but is modified for this use-case. Then we perform a systematic comparison of the different regularization methods (L_1-norm and L_2-norm) and optimization algorithms (IRLS, ADMM and FISTA) on simulated data in two- and three-dimensions. The comparison evaluates the methods on the accuracy of image reconstruction; ability to preserve edges; robustness against noise; and computational efficiency. Comprehensive and robust qualitative and quantitative evaluations are performed to quantitatively compare the reconstruction results using average contrast (AC), Pearson correlation (PC) and peak signal-to-noise ratio (PSNR). To our knowledge, this is the first systematic study in the area of spectral DOT reconstruction to perform such a comprehensive evaluation. We then apply our methods to the reconstruction of functional activations in simulated human brain imaging data using realistic anatomical models and finally evaluate the proposed algorithms using ex-

perimentally acquired data, by imaging a tissue-mimicking, plastic phantom of known optical properties using a multispectral DOT system. The experiments in this work use data obtained from a CW system but the proposed method can be easily applied to FD data with the same implementation.

4.2.2 Theory

Image reconstruction in SCDOT aims to find the tissue composition that best explains the boundary measurements. It typically requires the repeated evaluation of a forward model of light propagation in biological tissues as part of an inverse model that minimizes the difference between the measurements and the model's predicted measurements. In this section, the forward model for CW light propagation is introduced, followed by the spectrally constrained inverse model. The L_1-norm and L_2-norm regularization methods for the inverse problem are described at the end of this section.

4.2.2.1 The forward model

Let $w = 0$ in equation (2.3), the DE in a CW system can be written as

$$-\nabla \cdot \kappa\left(r, \lambda_i\right) \nabla \Phi\left(r, \lambda_i\right) + \mu_a\left(r, \lambda_i\right) \Phi\left(r, \lambda_i\right) = q_0\left(r, \lambda_i\right). \tag{4.13}$$

where λ_i represents the wavelength. $\kappa\left(r, \lambda_i\right)$ and $\mu_a\left(r, \lambda_i\right)$ have specific values at each wavelength with position r. It should be noted that in CW imaging, the value of reduced scattering coefficient $\mu_s'\left(r, \lambda_i\right)$ is not updated by the reconstruction algorithm and is assumed to be a known constant. FEM is used to numerically solve equation (4.13) on a discretized mesh, which has been implemented in several open-source software packages, notably TOAST++ [123] and NIRFAST [110]. In this work, the NIRFAST package is used for all computations.

In CW systems, the tissue absorption coefficient μ_a depends on the concentration of chromophores in the tissue. The relationship between the absorption coefficients at different wavelengths is therefore constrained by the intrinsic absorption properties of

the chromophores via Beer's law. For a dual-wavelength imaging system, and for two chromophores, Beer's law is written in matrix-vector form as

$$\begin{pmatrix} \mu_{a,\lambda_1} \\ \\ \mu_{a,\lambda_2} \end{pmatrix} = \begin{pmatrix} \varepsilon_{c_1,\lambda_1} & \varepsilon_{c_2,\lambda_1} \\ \\ \varepsilon_{c_1,\lambda_2} & \varepsilon_{c_2,\lambda_2} \end{pmatrix} \begin{pmatrix} c_1 \\ \\ c_2 \end{pmatrix}, \tag{4.14}$$

where c_1 and c_2 are chromophore concentrations and λ_1 and λ_2 are two measurement wavelengths. Here c_1 and c_2 correspond to oxy-haemoglobin HbO_2 and deoxy-haemoglobin Hb respectively, with $\lambda_1 = 750$nm and $\lambda_2 = 850$nm. $\varepsilon_{c_i,\lambda_i}$ $(i = 1, 2)$ are the extinction coefficients of the two chromophores at the corresponding wavelength λ_i. The values of $\varepsilon_{c_i,\lambda_i}$ have been documented by Zeff et al (2007) [17].

4.2.2.2 The inverse model for SCDOT image reconstruction

In SCDOT, chromophore concentrations c_1, and c_2 are directly estimated from the boundary measurements in preference to explicitly reconstructing optical properties at each wavelength. The following SCDOT inverse model allows direct estimation of chromophore parameters from two measurement wavelengths (i.e. 750 and 850nm) using some form of iterative procedure. Using a block notation, in which $(\,\vdots\,)$ represents the concatenation of two column vectors, we have:

$$c_1, c_2 = \arg\min_{c_1,c_2} \left\| \begin{pmatrix} y_{\lambda_1} \\ \\ y_{\lambda_2} \end{pmatrix} - \begin{pmatrix} \Psi^k_{\lambda_1} \\ \\ \Psi^k_{\lambda_2} \end{pmatrix} \right\|^2_2, \tag{4.15}$$

where chromophores c_1 and c_2 are the model parameters to be recovered. y_{λ_i} $(i = 1, 2)$ is the measured fluence at the tissue surface and Ψ_{λ_i} is the calculated data using the forward solver. The superscript k denotes the iteration number. Equation (4.15) defines a non-linear least square problem which can be solved via the classical Gauss-Newton method in which the first order Taylor series is used to expand the forward solution Ψ_{λ_i} as

$$\begin{pmatrix} \Psi^k_{\lambda_1} \\ \\ \Psi^k_{\lambda_2} \end{pmatrix} = \begin{pmatrix} \Psi^{k-1}_{\lambda_1} \\ \\ \Psi^{k-1}_{\lambda_2} \end{pmatrix} + \mathbf{J}^{k-1} \begin{pmatrix} c_1^k - c_1^{k-1} \\ \\ c_2^k - c_2^{k-1} \end{pmatrix}, \tag{4.16}$$

in which the spectral Jacobian J relates the changes in boundary data to changes in chromophore concentrations and can be constructed directly with the incorporation of

spectral priors using the adjoint method [70]. Note that when $k = 1$, an initial guess of the chromophore concentrations c_1^0 and c_2^0 is required which can be obtained by a data-calibration procedure explained elsewhere [124]. The spectral Jacobian $\mathbf{J}$ can be derived as [125]:

$$\mathbf{J} = \begin{pmatrix} \frac{\partial \Psi_{\lambda_1}}{\partial c_1} & \frac{\partial \Psi_{\lambda_1}}{\partial c_2} \\ \frac{\partial \Psi_{\lambda_2}}{\partial c_1} & \frac{\partial \Psi_{\lambda_2}}{\partial c_2} \end{pmatrix} = \begin{pmatrix} \frac{\partial \Psi_{\lambda_1}}{\partial \mu_{a,\lambda_1}} \cdot \frac{\partial \mu_{a,\lambda_1}}{\partial c_1} & \frac{\partial \Psi_{\lambda_1}}{\partial \mu_{a,\lambda_1}} \cdot \frac{\partial \mu_{a,\lambda_1}}{\partial c_2} \\ \frac{\partial \Psi_{\lambda_2}}{\partial \mu_{a,\lambda_2}} \cdot \frac{\partial \mu_{a,\lambda_2}}{\partial c_1} & \frac{\partial \Psi_{\lambda_2}}{\partial \mu_{a,\lambda_2}} \cdot \frac{\partial \mu_{a,\lambda_2}}{\partial c_2} \end{pmatrix} = \begin{pmatrix} J_{\lambda_1} \cdot \varepsilon_{c_1,\lambda_1} & J_{\lambda_1} \cdot \varepsilon_{c_2,\lambda_1} \\ J_{\lambda_2} \cdot \varepsilon_{c_1,\lambda_2} & J_{\lambda_2} \cdot \varepsilon_{c_2,\lambda_2} \end{pmatrix}, \tag{4.17}$$

where J_{λ_i} relates the changes in boundary data to changes in the absorption coefficient at wavelength λ_i. The size of $\mathbf{J}$ in this case is the number of wavelengths times the number of measurements per wavelength, by number of finite element nodes times number of chromophore parameters.

Substituting equation (4.16) into equation (4.15) leads to

$$\delta c^k = \arg\min_{\delta c} ||\delta \Psi^{k-1} - \mathbf{J}^{k-1} \delta c||_2^2, \tag{4.18}$$

where δc^k is the change in the chromophore parameters at the k-th iteration and can be written as

$$\delta c^k = \begin{pmatrix} \delta c_1^k \\ \\ \delta c_2^k \end{pmatrix} = \begin{pmatrix} c_1^k - c_1^{k-1} \\ \\ c_2^k - c_2^{k-1} \end{pmatrix}. \tag{4.19}$$

$\delta \Psi$ in equation (4.18) is the data-model mismatch which is given by

$$\delta \Psi^{k-1} = \begin{pmatrix} y_{\lambda_1} - \Psi_{\lambda_1}^{k-1} \\ \\ y_{\lambda_2} - \Psi_{\lambda_2}^{k-1} \end{pmatrix}. \tag{4.20}$$

Minimizing equation (4.18) leads to the normal equations

$$\left(\mathbf{J}^{(k-1)\mathrm{T}} \mathbf{J}^{(k-1)} \right) \delta c^k = \mathbf{J}^{(k-1)\mathrm{T}} \delta \Psi^{k-1}. \tag{4.21}$$

which can be solved to find the update term δc^k using the Gauss-Newton algorithm which is summarized in Algorithm 1.

As discussed in Chapter 2.4.1, it is non-trivial to calculate the inverse of $\mathbf{J}^{(k-1)\mathrm{T}} \mathbf{J}^{(k-1)}$ in equation (4.21) (i.e. step 6 in Algorithm 1) because it is normally singular or close

Algorithm 1: Gauss-Newton Algorithm for Minimizing Equation (4.15)

INPUT: y_{λ_i} $(i = 1, 2)$, $iter$, Tol

Initialize: c_1^0, c_2^0

for $k = 1 : iter$

1: Update μ_{a,λ_i} at each wavelength using Beer's law (Equation (4.14))

2: Update $\Psi_{\lambda_i}^{k-1}$ at each wavelength using the forward model (Equation (4.13))

3: Update $\delta\Psi^{k-1}$ using equation (4.20)

4: Stop if $k = iter$ or $||\delta\Psi^{k-1} - \delta\Psi^{k-2}||_1 \leqslant$ Tol, otherwise go to step 5

5: Update $\mathbf{J}^{k-1}$ using equation (4.17)

6: Update $\delta c^k = (\mathbf{J}^{(k-1)\mathrm{T}}\mathbf{J}^{(k-1)})^{-1}\mathbf{J}^{(k-1)\mathrm{T}}\delta\Psi^{k-1}$

7: Update c_1^k and c_2^k using equation (4.19)

end for

RETURN c_1^k and c_2^k

to singular. Furthermore, experimental noise in the measurements Ψ_{λ_i} tends to lead to reconstruction artefacts if this inversion is computed directly. Therefore regularization is needed to reduce model errors and artefacts caused by measurement noise.

4.2.2.3 The proposed spectral-L_1 inverse model

To convert equation (4.18) into a more readily solvable problem, a Tikhonov (L_2-norm) regularization term is usually introduced into the inverse problem:

$$\delta c^k = \arg\min_{\delta c} \left\{ ||\delta\Psi^{k-1} - \mathbf{J}^{k-1}\delta c||_2^2 + \lambda||\delta c||_2^2 \right\}. \tag{4.22}$$

The regularization parameter λ determines the degree of regularization that will be imposed on the model. This can be solved analytically to give

$$\delta c^k = (\mathbf{J}^{(k-1)\mathrm{T}}\mathbf{J}^{(k-1)} + \lambda\mathbf{I})^{-1}\mathbf{J}^{(k-1)\mathrm{T}}\delta\Psi^{k-1}. \tag{4.23}$$

$\mathbf{I}$ is the identity matrix and its size is the same as that of $\mathbf{J}^{(k-1)\mathrm{T}}\mathbf{J}^{(k-1)}$. The introduction of $\lambda\mathbf{I}$ effectively reduces the condition number of the matrix, thus stabilizing the matrix

inversion. An analytical solution to this problem is possible because equation (4.22) is convex and quadratic which makes L_2-norm regularization an attractive choice for many inverse problems. However, in image reconstruction problems, it tends to over-smooth the result and sharp features such as object boundaries in the reconstructed images are often smeared. Moreover, L_2-norm regularization discourages sparsity, and is not suitable for sparse image reconstruction. In SCDOT image recovery, the perturbation/change δc^k is usually zero or close to zero when the region to be recovered is not in the vicinity of the region of interest. In this case, δc^k is spatially sparse. Recently, L_1-norm regularization has been widely studied because of several useful properties: it is sparsity-promoting, convex, edge-preserving, and is more robust against noise. Therefore, we propose a new inverse model for SCDOT image recovery based on L_1-norm regularization that we refer to as spectral-L_1. This is formulated as

$$\delta c^k = \arg\min_{\delta c} \left\{ ||\delta \Psi^{k-1} - \mathbf{J}^{k-1} \delta c||_2^2 + \lambda ||\delta c||_1 \right\}. \tag{4.24}$$

Although L_1-norm regularization has many advantages over L_2-norm regularization, the L_1-norm is non-differentiable, which makes it difficult to solve equation (4.24). Three candidate algorithms for this task will be investigated in the next section.

4.2.3 Candidate algorithms for solving the proposed spectral-L_1 method

We now consider three algorithms for the solution of equation (4.24): iteratively reweighted least squares (IRLS) [117]; alternating directional method of multipliers (ADMM) [119]; and fast iterative shrinkage-thresholding algorithm (FISTA) [109]. These algorithms will be incorporated into the image reconstruction process by substituting them into step 6 of Algorithm 1, which solves for the update term.

4.2.3.1 IRLS

Instead of solving the L_1-minimization problem directly, IRLS reformulates the problem as a sequence of weighted L_2-minimization problems. Specifically, by introducing a weight matrix $\mathbf{W}$, the L_1-minimization can be converted into finding the optimal solution of the quadratic problem

$$\delta c^k = \arg\min_{\delta c} \left\{ ||\delta\Psi^{k-1} - \mathbf{J}^{k-1}\delta c||_2^2 + \lambda ||\mathbf{W}\delta c||_2^2 \right\}. \tag{4.25}$$

$\mathbf{W}$ is a diagonal matrix with weights, w_s, along its diagonal that are given by

$$w_s = \begin{cases} |\delta c_s^{i-1}|^{-0.5} & \text{if } |\delta c_s^{i-1}| \geqslant \varepsilon \\ \varepsilon^{-1} & \text{if } |\delta c_s^{i-1}| < \varepsilon \end{cases}. \tag{4.26}$$

The superscript i above denotes the i'th IRLS iteration (Algorithm 2), and it should be distinguished from the superscript k denoting the iterations of Algorithm 1. A small positive number $0 < \varepsilon \ll 1$ is used to avoid the possibility of division by zero. It has been suggested that ε should be a series of positive real numbers that decay to zero over iterations [126]. In practice, we have found that using a fixed value in the range $0.001 \leq \varepsilon \leq 0.01$ does not give significantly different results. Equation (4.25) results in the normal equation

$$\left(\mathbf{J}^{(k-1)\mathrm{T}} \mathbf{J}^{(k-1)} + \lambda \mathbf{W}^{\mathrm{T}} \mathbf{W} \right) \delta c = \mathbf{J}^{(k-1)\mathrm{T}} \delta\Psi^{k-1}. \tag{4.27}$$

Equation (4.27) is known as the weighted L_2-minimization scheme. We note that if the diagonal weights w_s are set to 1, the normal equation reduces to the conventional L_2 scheme (Equation (4.23)). The IRLS approach is summarized in Algorithm 2.

The calculation of the elements of $\mathbf{W}$ requires an initial guess for δc for which we use the solution to the L_2-regularized problem (Equation (4.23)). One of the biggest advantages of IRLS is that equation (4.27) has an analytical solution which allows equation (4.25) to be solved exactly, making IRLS almost as easy to implement as the Levenberg-Marquadt scheme. In common with many sparsity-promoting optimization methods, the sparsity level in IRLS is controlled by the regularization parameter λ which must be chosen carefully for each specific problem.

Algorithm 2: Iteratively Reweighted Least Square Algorithm (IRLS)

INPUT: $\delta\Psi^{k-1}$, $\mathbf{J}^{k-1}$, λ, $iter$, Tol

Initialize: Set δc^0 using equation (4.23)

for $i = 1 : iter$

1: Update $\mathbf{W}$ using equation (4.26)

2: Update $\delta c^i = (\mathbf{J}^{(k-1)\mathrm{T}}\mathbf{J}^{(k-1)} + \lambda\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{J}^{(k-1)\mathrm{T}}\delta\Psi^{k-1}$

3: Stop if $i = iter$ or $||\delta c^i - \delta c^{i-1}||_1 \leqslant$ Tol, otherwise go to step 1

end for

RETURN $\delta c^k = \delta c^i$

4.2.3.2 ADMM

ADMM has been widely used to solve optimization problems in machine learning, signal processing, and standard image restoration and reconstruction. This method has become particularly important in the field of variational image processing, which frequently requires the minimization of non-differentiable objectives [118,119]. It has been shown to be able to solve constrained optimization problems effectively and efficiently. The basic idea is to decompose a complex optimization problem into several simpler subproblems, which usually have closed-form solutions [120]. Its simplicity, flexibility, and broad applicability have made it an important part of the modern optimization toolset. To apply ADMM to our spectral-L_1 problem, we first introduce an auxiliary splitting vector variable $\boldsymbol{v}$, an augmented Lagrangian multiplier $\boldsymbol{b}$, and a positive penalty parameter θ, reformulating equation (4.24) as the following unconstrained optimization problem

$$\delta c, \boldsymbol{v}, \boldsymbol{b} = \underset{\delta c, \boldsymbol{v}, \boldsymbol{b}}{\arg\min} \left\{ ||\delta\Psi^{k-1} - \mathbf{J}^{k-1}\delta c||_2^2 + \lambda||\boldsymbol{v}||_1 + \frac{\theta}{2}||\boldsymbol{v} - \delta c - \boldsymbol{b}||_2^2 \right\}. \tag{4.28}$$

This multivariate optimization problem corresponds to a sub-minimization problem with respect to δc, $\boldsymbol{v}$ and $\boldsymbol{b}$, separately. When all the subproblems converge, the solution for δc approximately represents that of equation (4.24). In order to find the minimizers for all of the subproblems, ADMM searches all the saddle points of equation (4.28) by first

fixing the variables ($\boldsymbol{v}$, $\boldsymbol{b}$) and minimizing the subproblem with respect to δc using the following normal equation

$$\left(\mathbf{J}^{(k-1)\mathrm{T}}\mathbf{J}^{(k-1)} + \theta\mathbf{I}\right)\delta c = \mathbf{J}^{(k-1)\mathrm{T}}\delta\Psi^{k-1} + \theta\left(\boldsymbol{v}^{i-1} - \boldsymbol{b}^{i-1}\right). \tag{4.29}$$

By inverting the matrix on the left-hand side of equation (4.29), a unique solution for δc is found. We then fix variables δc and $\boldsymbol{b}$ and set the first order derivative with respect to $\boldsymbol{v}$ to zero. This leads to

$$\lambda\frac{\boldsymbol{v}}{|\boldsymbol{v}|} + \theta\left(\boldsymbol{v} - \delta c^{i} - \boldsymbol{b}^{i-1}\right) = 0, \tag{4.30}$$

which can be solved component-wise using an analytical shrinkage-thresholding method to give

$$\boldsymbol{v}^{i} = \max\left(\left|\delta c^{i} + \boldsymbol{b}^{i-1}\right| - \frac{\lambda}{\theta}, 0\right) \circ sign\left(\delta c^{i} + \boldsymbol{b}^{i-1}\right), \tag{4.31}$$

where $\circ$ and $sign$ symbols denote component-wise multiplication and the signum function, respectively. The last step of ADMM is to update the augmented Lagrangian multiplier b, as $\boldsymbol{b}^{i} = \boldsymbol{b}^{i-1} + \delta c^{i} - \boldsymbol{v}^{i}$. The complete method is presented in Algorithm 3. The key advantage of ADMM is that equations (4.29) and (4.30) have closed-form solutions. We note that the augmented Lagrangian multiplier means that different choices of the penalty parameter θ will provide similar results but with different rates of convergence. In all the experiments we have conducted, we used $\theta = 0.01$ to achieve fast convergence.

4.2.3.3 FISTA

FISTA is an efficient optimization approach that uses the forward-backward splitting technique (FBS) [109,122]. It is an extension of the classical gradient descent method and belongs to the class of first order methods that are a better choice for large-scale problems than second-order methods such as IRLS and ADMM because they do not require the explicit construction of large matrices. Let us consider minimizing the L_1-regularized data fitting energy given by equation (4.24). We begin by analyzing the standard unregularized problem with $\lambda = 0$. Let $\mathrm{F}(\delta c) = ||\delta\Psi^{k-1} - \mathbf{J}^{k-1}\delta c||_2^2$ and $\nabla\mathrm{F}(\delta c) = \mathbf{J}^{(k-1)\mathrm{T}}(\mathbf{J}^{(k-1)}\delta c - \delta\Psi^{k-1})$ denote its gradient. We apply the gradient descent algorithm

$$\delta c^{i} = \delta c^{i-1} - t\nabla\mathrm{F}\left(\delta c^{i-1}\right), \tag{4.32}$$

Algorithm 3: Alternating Directional Method of Multipliers (ADMM)

INPUT: $\delta\Psi^{k-1}$, $\mathbf{J}^{k-1}$, λ, θ, $iter$, Tol

Initialize: $\boldsymbol{v}^0 = 0$, $\boldsymbol{b}^0 = 0$

for $i = 1 : iter$

1: Update $\delta c^i = (\mathbf{J}^{(k-1)\mathrm{T}}\mathbf{J}^{(k-1)} + \theta\mathbf{I})^{-1}(\mathbf{J}^{(k-1)\mathrm{T}}\delta\Psi^{k-1} + \theta(\boldsymbol{v}^{i-1} - \boldsymbol{b}^{i-1}))$

2: Update $\boldsymbol{v}^i$ using equation (4.31)

3: Update $\boldsymbol{b}^i = \boldsymbol{b}^{i-1} + \delta c^i - \boldsymbol{v}^i$

4: Stop if $i = iter$ or $||\delta c^i - \delta c^{i-1}||_1 \leqslant$ Tol, otherwise go to step 1

end for

RETURN $\delta c^k = \delta c^i$

where $t > 0$ is a suitable stepsize which controls how far the iteration moves along the gradient direction in the i'th iteration. The value of t is initialized by estimating the Lipschitz constant $\widetilde{L}$ of $\nabla\mathrm{F}$ as $\widetilde{L} = L(\nabla\mathrm{F})$ and then backtracking rules are adopted to guarantee that the objective has decreased sufficiently [122]. The gradient iteration given by equation (4.32) can be understood as a proximal regularization [127] of the linearized function $\mathrm{F}(\delta c)$ at δc^{i-1}, which corresponds to the following optimization problem:

$$\delta c = \arg\min_{\delta c}\left\{\mathrm{F}\left(\delta c^{i-1}\right) + \nabla\mathrm{F}\left(\delta c^{i-1}\right)\left(\delta c - \delta c^{i-1}\right) + \frac{1}{2t}||\delta c - \delta c^{i-1}||_2^2\right\}. \tag{4.33}$$

Analogously, adopting the same gradient descent idea to solve equation (4.24) with $\lambda \neq 0$ leads to the following minimization problem

$$\delta c = \arg\min_{\delta c}\left\{\frac{1}{2t}||\delta c^{i-1} - t\nabla\mathrm{F}\left(\delta c^{i-1}\right) - \delta c||_2^2 + \lambda||\delta c||_1\right\}. \tag{4.34}$$

Minimizing this results in a formulation similar to equation (4.30) and can be solved in the same way to give

$$\delta c^i = \max\left(\left|\delta c^{i-1} - t\nabla\mathrm{F}\left(\delta c^{i-1}\right)\right| - t\lambda, 0\right) \circ sign\left(\delta c^{i-1} - t\nabla\mathrm{F}\left(\delta c^{i-1}\right)\right). \tag{4.35}$$

The minimizer of the equation (4.24) can then be found by iterating δc^i in equation (4.35) to convergence. In isolation, equation (4.35) is known as the iterative shrinkage-thresholding algorithm (ISTA) [128–133], whose global convergence rate is $O(1/N)$, where

N is the iteration counter. This is improved upon by using a Nesterov-type acceleration technique to obtain faster convergence. In the FISTA algorithm, the iterative shrinkage operator is not used on the value obtained from the previous iteration δc^{i-1}, but rather on a combination of the values from the previous two iterations. Thus, in FISTA, equation (4.35) is replaced with

$$\delta c^i = \max\left(\left|\delta d^i - t\nabla \mathrm{F}\left(\delta d^i\right)\right| - t\lambda, 0\right) \circ sign\left(\delta d^i - t\nabla \mathrm{F}\left(\delta d^i\right)\right), \tag{4.36}$$

where δd^i comes from the prediction procedure given in step 4 of Algorithm 4. This step can help to push the solution to the current iteration further in the direction it moved during the previous iteration, which can significantly improve the computational efficiency. The complete FISTA method is presented in Algorithm 4, where steps 2 and 3 implement the acceleration strategy and can be viewed as an over-relaxation step that improves the global convergence rate from $O(1/N)$ to $O(1/N^2)$.

Algorithm 4: Fast Iterative Shrinkage-Thresholding Algorithm (FISTA)

INPUT: $\delta\Psi^{k-1}$, J^{k-1}, λ, $iter$, Tol

Initialize: Set δc^0 using equation (4.23), $\alpha^0 = 1$, $t < 1/\widetilde{L}$

for $i = 1 : iter$

1: Update δc^i using equation (4.36)

2: Stop if $i = iter$ or the relative residual $\leqslant$ Tol, otherwise go on to step 3

3: Update $\alpha^i = (1 + \sqrt{1 + 4(\alpha^{i-1})^2})/2$

4: Update $\delta d^{i+1} = \delta c^i + \ (\alpha^{i-1} - 1)/(\alpha^i)(\delta c^i - \delta c^{i-1})$

end for

RETURN $\delta c^k = \delta c^i$

4.2.3.4 The spectral-L_1 algorithm

We have introduced three methods for solving the chromophore update terms in SC-DOT with a sparsity enforcing constraint: IRLS, ADMM, and FISTA. These algorithms

replace the single update term of the conventional reconstruction algorithm (step 6 of Algorithm 1). The flow-chart presented in Figure 4.1 shows how these proposed methods are integrated into SCDOT reconstruction for CW imaging. Since the three proposed optimization schemes are themselves iterative, our method contains nested iterations. In

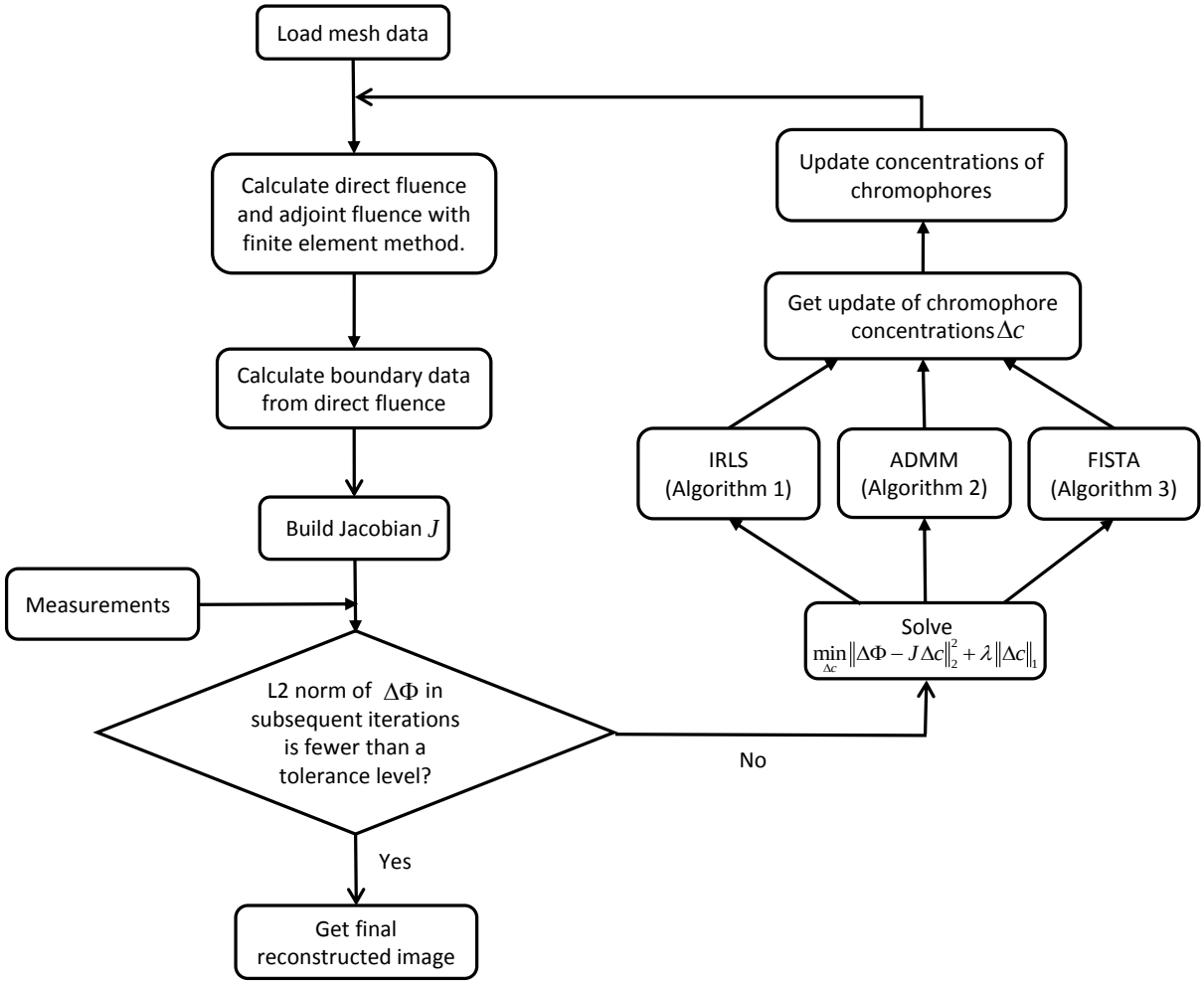


Figure 4.1: Flow chart for SCDOT image reconstruction using the spectral-L_1 model.

4.2.4 Parameter selection

The regularization parameter λ determines the trade-off between the goodness-of-fit of the model to the data, and the strict enforcement of the regularization criteria. An optimal value between the two quantities must therefore be found: if too much regularization is imposed on the model, then it will not fit the data properly; if the regularization

parameter is too small, the fit will be good but the solution will be dominated by data errors and measurement noise (the overfitting regime). There are several methods to find an optimal compromise between these two considerations and the L-curve method is both simple and effective. By plotting the model-data mismatch $||\delta\Psi - \mathbf{J}\delta c||_2^2$ against the model regularization $||\delta c||_2^2$ or $||\delta c||_1$ for a sequence of different λ, a curve which is typically L-shaped can be constructed. Figure 4.2 shows the L-curves obtained from each of the four candidate optimization schemes using the numerical experiments described by Zhan et al [125]. The optimal trade-off occurs at the "elbow" of the L-shaped curve and this can be located by determining the point of maximum curvature of the curve.

Since strong regularization can improve the conditioning of the linear system, we solve the formulations given by equations (4.22) and (4.24) with a relatively large regularization parameter λ and then decrease it gradually by a fixed factor until the curvature of the L-curve starts to decrease. This corner point is considered to be at the optimum value of λ where both the model fit and the regularization function are simultaneously near to their minimum values. In principle, computing the L-curve requires the full image reconstruction process to be run multiple times which is computationally very expensive. We have found that it is sufficient to compute the L-curve for one iteration of the outer loop of Figure 4.1, and then to use the resulting optimal value of λ for the remaining iterations. In addition, in order to avoid the special case where the L-curve does not allows an optimal value of λ to be found by purely numerical means [134, 135], we select a range around the parameter with the highest curvature value. We then adjust the values manually to get the final optimal parameter by visually inspecting the solutions and choosing the one that generates the sparsest solution with a well-defined compact localization. This approximate optimum is then used for subsequent iterations. In Figure 4.2, we note that the choice of algorithm for L_1 regularized reconstruction significantly affects the shape of the L-curve and the optimal value of λ. Comparing with the L-curve generated by IRLS and ADMM, FISTA algorithm has the most L shaped curve which "elbow" gives the largest curvature. It implies that the λ chosen at the "elbow" of the L-curve by FISTA

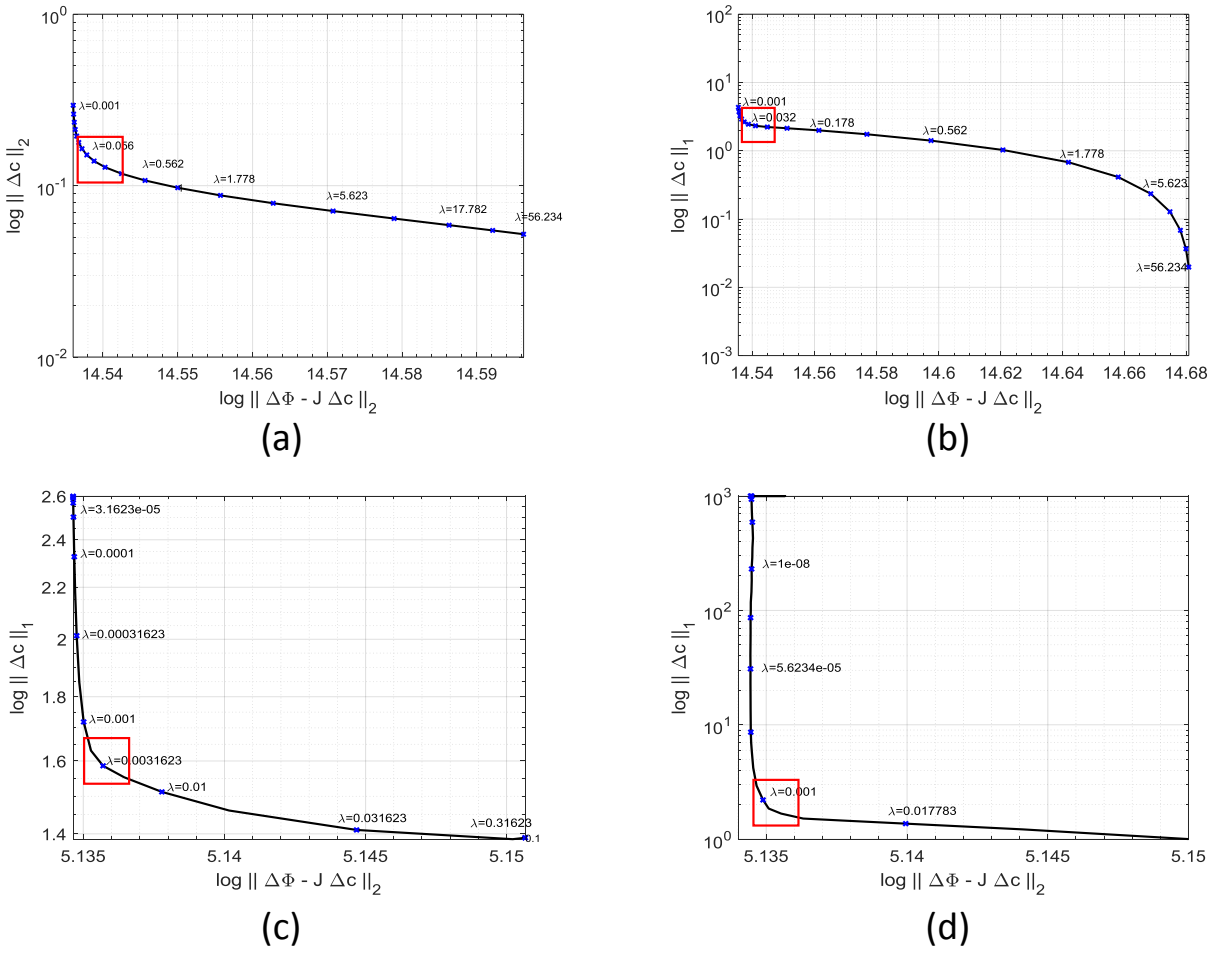


Figure 4.2: L-curves (data fit against model regularization) derived from a synthetic example: a) Tikhonov regularization; b) L_1 regularization using IRLS; c) L_1 regularization using ADMM; d) L_1 regularization using FISTA. The optimal regularization parameter is around the point of maximum curvature (within the red boxes).

In addition to the regularization parameter λ that is common to all three L_1 algorithms, we have considered how to select the other parameters of each method to ensure that our comparison is fair and unbiased. IRLS has one parameter ε and we set this to $0.001 \leq \varepsilon \leq 0.01$ following the recommendations set out by Shaw and Yalavarthy [126]. ADMM has one parameter θ and the use of the augmented Lagrangian multiplier means that different choices of θ provide similar results but lead to different rates of convergence. In all the experiments, θ was set to 0.01 to achieve fast convergence. FISTA has two parameters t and α. t is initialized by estimating the Lipschitz constant and then backtracking rules are adopted to guarantee that the objective has decreased suffi-

ciently [122]. t is therefore updated automatically. α is involved in an over-relaxation step (i.e. step 4 in Algorithm 3) and its update is also automatic (i.e. step 3 in Algorithm 3). The regularization parameter λ is therefore the only parameter that must be optimized for a specific problem.

4.2.5 Experiment setup

We have performed extensive experiments to evaluate the performance of different models and algorithms qualitatively and quantitatively. We first define three evaluation metrics to quantify the quality of the reconstructed images. We then describe simulated numerical experiments, and then real experiments performed on phantom samples. For experiments in which measurement noise was added, ten repeats were performed. In all cases, the forward model was implemented using the NIRFAST package [110] in Matlab R2013a (Mathworks, Natick, USA).

4.2.5.1 Quantitative evaluation metrics

Three quantitative evaluation metrics are considered: the average contrast (AC), Pearson correlation (PC) and peak signal-to-noise ratio (PSNR). Ideally, if the reconstructed image is exactly same as the ground truth image, AC is equal to 1. For PC and PSNR, the recovered image has higher quality if higher PC or PSNR values are obtained.

Average constrast (AC) is based on the mean value of the region of interest and is defined as:

$$\mathrm{AC} = \frac{\sum_{j=1}^{N} c_i^j / N}{\tilde{c}_i} \quad i = 1, 2 \tag{4.37}$$

where c_i^j denotes the recovered values of chromophore i on the finite element node j. N is the number of nodes in the activation region which is selected by thresholding the recovered changes based on 50% of the maximum recovered changes. $\tilde{c}_i$ are the ground truth values of the chromophores in the activation region.

The second evaluation metric PC is given by

$$\mathrm{PC} = \frac{\mathrm{COV}\left(c_i, \tilde{c}_i\right)}{\sigma\left(c_i\right)\sigma\left(\tilde{c}_i\right)} \quad i = 1, 2. \tag{4.38}$$

The numerator is the covariance (COV) of the recovered images with the ground truth and σ indicates standard deviation. The PC is thus a measure of the joint variability of the ground truth image with the reconstructed image.

Finally, PSNR evaluates the difference between the ground truth image and the recovered image. Larger PSNR values means less difference between the target and the recovered image. This measure is defined as follows

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_{c_i}^2}{\text{MSE}} \right) \quad i = 1, 2. \tag{4.39}$$

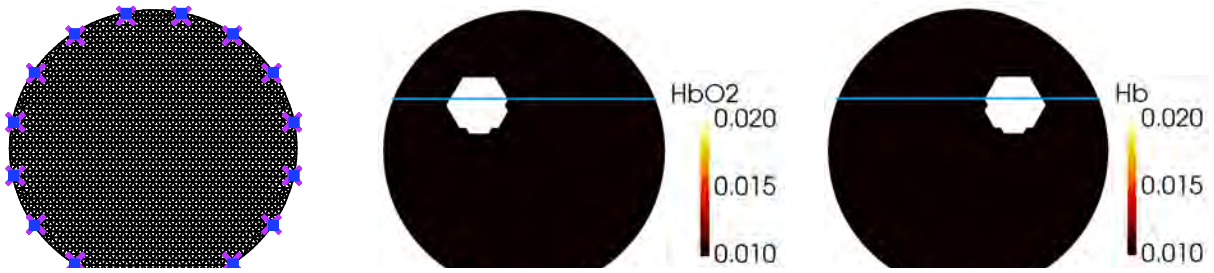
Here, MAX_{c_i} is the maximum pixel value of c_i and MSE is the mean squared error between the reconstructed and ground truth values.

$$\text{MSE} = \frac{1}{N} \sum_{j=1}^{N} \left(c_i^j - \tilde{c}_i^j \right)^2. \tag{4.40}$$

For AC, values closer to 1 indicate better performance. For PC and PSNR, higher values are better.

4.2.5.2 Two dimensional circle model

We first consider a simple simulated 2D circular geometry containing two anomalies (Figure 4.3). The model has a radius of 43mm and is composed of 1785 nodes and 3418 linear triangle elements. Sixteen source-detector fibres are placed equidistant around the external boundary for CW boundary data acquisition. When one fibre as a source is turned on, the rest are used as detectors, leading to 240 total boundary data points per wavelength. All sources were positioned one scattering distance within the outer boundary because the source is assumed to be spherically isotropic. The background chromophore concentrations, c_1 (HbO_2) and c_2 (Hb), are both set to 0.01mM. Two 10mm radius target regions are centred at (-14mm, 14mm) and (14mm, 14mm). The oxyhemoglobin concentration HbO_2 of the left target is set to 0.02mM, and the deoxyhemoglobin Hb of the right target as 0.02mM. Boundary data was generated at 750 and 850nm. To represent various realistic cases, normally distributed randomly generated noise ranging from 0% to 5% at 1% intervals was added to the amplitude of the boundary data.



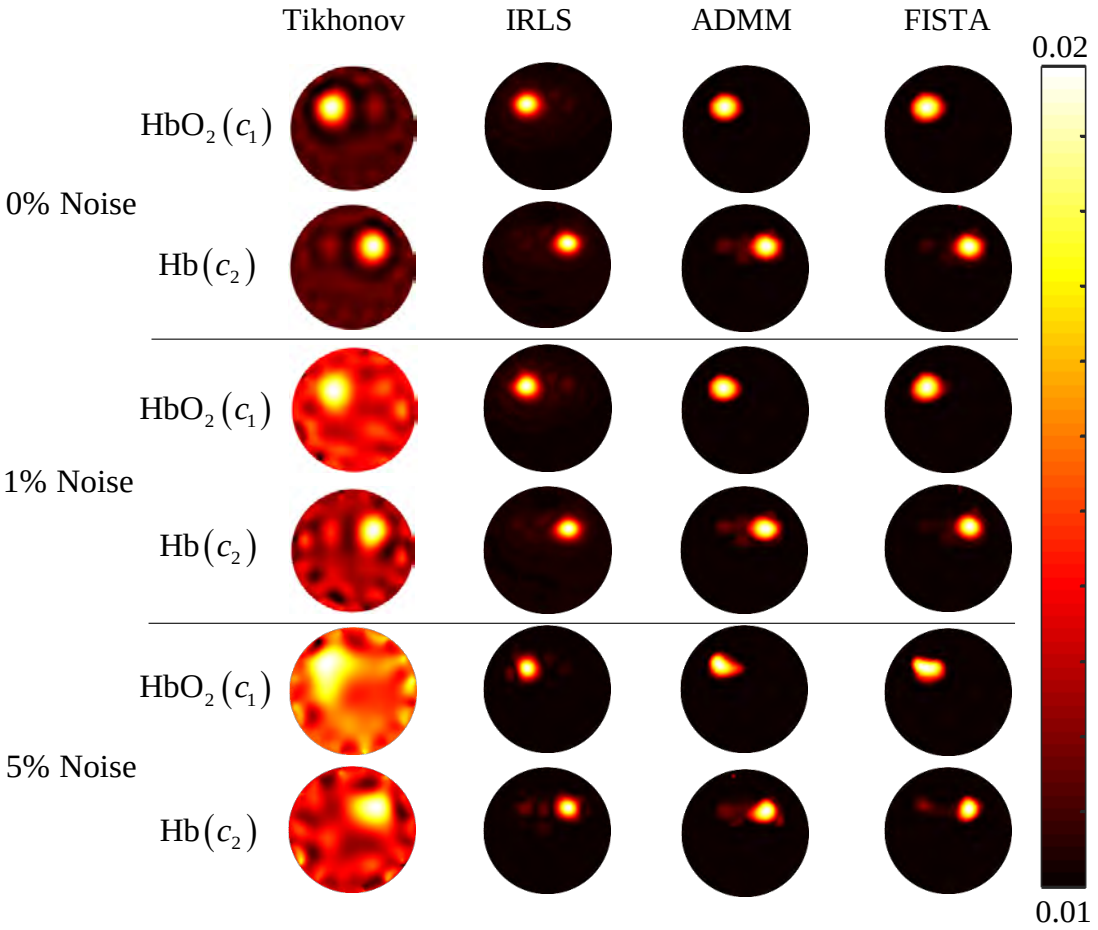


Figure 4.4: Chromophore images reconstruction using Tikhonov and IRLS, ADMM, FISTA (from left to right column) on 0% (top part), 1% (middle part) and 5% (bottom part) noisy data.

In Figure 4.4 we show the chromophore reconstruction results using the Tikhonov model (Equation (4.22)) and the spectral-L_1 model (Equation (4.24)) using the IRLS, ADMM and FISTA methods on 0%, 1% and 5% noisy data respectively. By visual inspection, it is evident that the background of the images recovered by spectral-L_1 is

similar to the target background as shown in Figure 4.3, whilst Tikhonov regularization recovers much higher levels of background signal. This is because spectral-L_1 imposes sparsity on the solution δc in equation (4.24) and this is consistent with low signal in the background area. It is also apparent that Tikhonov regularization performance degrades significantly with increasing noise levels, while spectral-L_1 is consistently more robust even at high levels. In addition, the boundaries of the targets reconstructed by spectral-L_1 are sharper than those recovered by Tikhonov. We observe that the three different L_1 optimization algorithms lead to similar reconstruction results.

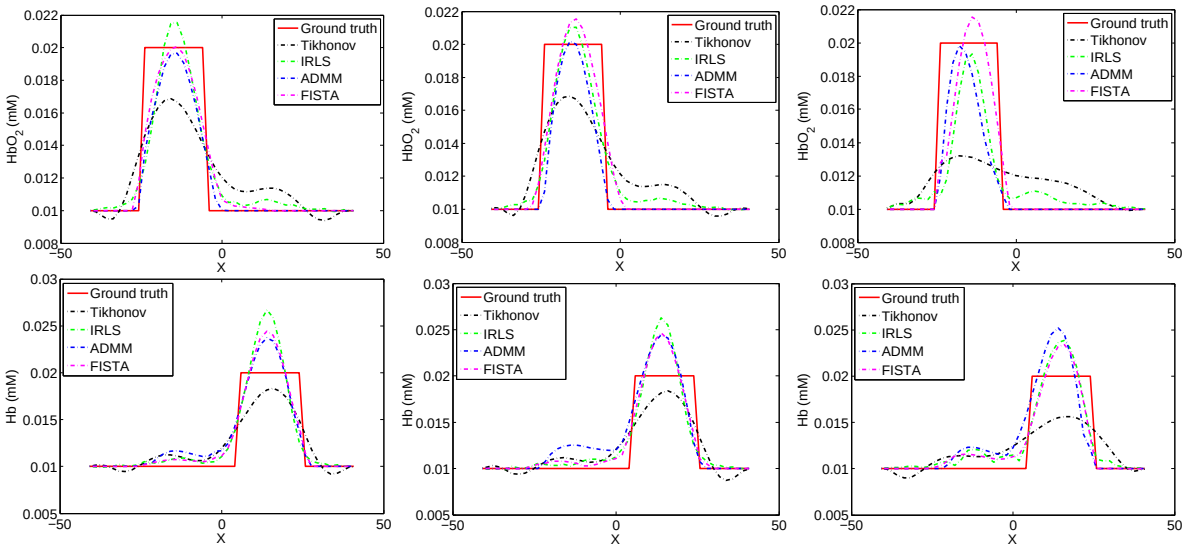


Figure 4.5: 1D cross sections of images recovered in Figure 4.4 along the corresponding light blue lines annotated on the targets in Figure 4.3. Left to right column: 0%, 1% and 5% added Gaussian noise.

The one-dimensional cross-sections of the reconstructed chromophores in Figure 4.5 show that Tikhonov regularization gives lower peak values of oxyhemoglobin and deoxyhemoglobin in comparison to both the ground truth (red) and to spectral-L_1. This is because L_2 over-smooths the results so they tend to have lower contrast. The curves from the three spectral-L_1 variants are seen to have similar shape implying that all three converge towards similar solutions. This is consistent with our visual observation from the reconstructed images in Figure 4.4.

The values of the metrics AC, PC and PSNR are calculated to qualitatively evaluate

the results in Figure 4.4. Specifically, the performance of each method is measured in terms of mean and standard deviation of the corresponding metric over all 6 noise levels. The corresponding boxplots are shown in Figure 4.6 for easy assess of these evaluation metrics numbers. The boxplots shown in Figure 4.6 show the distribution across repeats of the three metrics for reconstructions performed at different noise levels. Tikhonov regularization is seen to have the lowest performance in all cases, while IRLS, ADMM and FISTA are seen to produce more accurate reconstructions based on all computed metrics, and have higher accuracy and lower error rates for both oxyhemoglobin and deoxyhemoglobin. The boxplots also show that there is less variation in performance of each of the three spectral-L_1 methods than is seen with Tikhonov, and that even in the worst case scenario spectral-L_1 yields lower error than the average performance of Tikhonov, showing its robustness against noise. Among the three algorithms, FISTA shows slightly better performance than ADMM and IRLS.

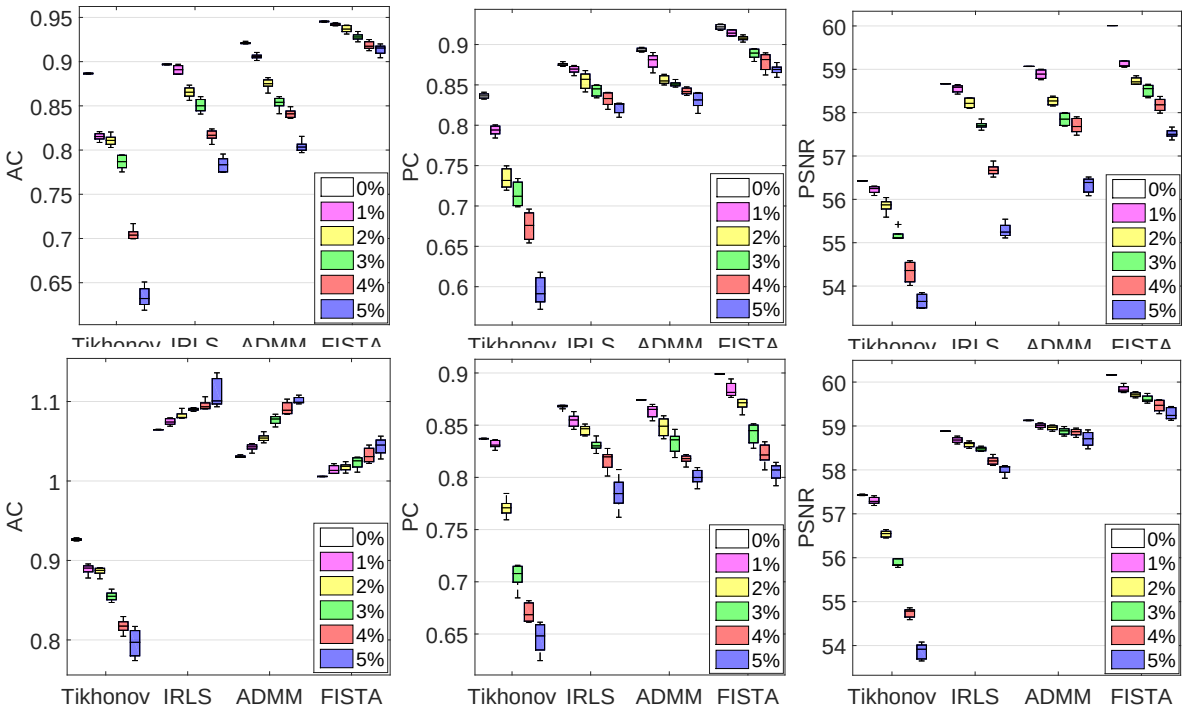


Figure 4.6: Evaluation metrics comparing the performance of different methods on 2D simulated data at six different noise levels. Left to right column: AC index; PC index and PSNR index. The first row gives the results for HbO_2, the second row for Hb.

4.2.5.3 Three dimensional cylinder numerical experiments

We conducted further experiments on a three dimensional (3D) simulated model consisting of one cylinder of radius 43 mm and height 60 mm, composed of 12695 nodes with 63810 tetrahedral elements. Sixteen CW sources and detectors were placed equidistant around the equator (first image in Figure 4.7) giving 240 source-detector pairs per wavelength. The background concentration of HbO_2 and Hb were both set to 0.01mM. The geometric centre of the model is set as (0,0,0). Two 10 mm spherical targets centred on (-25,10,0) and (25,10,0) respectively were added with the chromophore concentrations of HbO_2 and Hb respectively set to 0.02mM (Figure 4.7, middle and right). 0% to 2% in 0.5% intervals Gaussian random noise was added to the boundary measurement data. Higher levels of noise were not used because 2% noise was found to be large enough to reveal differences between methods. We only display the results for 0%, 0.5% and 1% added noise, which is sufficient to show a visual difference. Three evaluation metrics (AC, PC and PSNR) were calculated to assess the reconstruction quality.

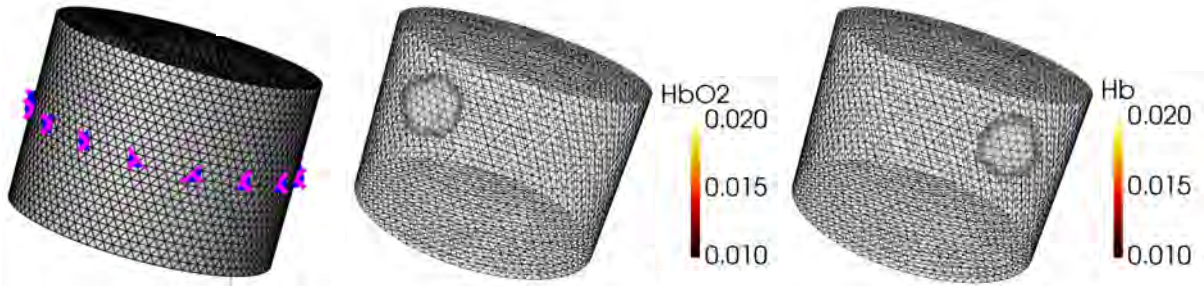


Figure 4.7: Three-dimension mesh used for image simulation. Left to right: source-detector location on the cylinder's equator; HbO_2 inclusion at (-25,10,0); Hb inclusion at (25,10,0).

Reconstructed images of chromophores using Tikhonov for L_2-norm regularization and IRLS, ADMM, FISTA algorithms for L_1-norm regularization are shown in Figure 4.8 with 0% , 0.5% and 1% added Gaussian random noise respectively. In the case of 0% added noise, a region of crosstalk (areas marked with black dotted line) can be easily identified in the images recovered by Tikhonov regularization along with other artefacts.

Reconstructions with L_1-norm regularization do not show evidence of crosstalk, which we believe has been supressed by the sparsity constraint. The results from IRLS, ADMM and FISTA are visually similar, but compared with FISTA, IRLS and ADMM algorithms produce more compact localizations that appear to be somewhat more compact than the ground truth. The images recovered with FISTA appear to be visually closer to the desired results. Increasing levels of noise produce increases in crosstalk and artefacts in Tikhonov regularization whilst L_1-norm regularization is able to continue to reconstruct well-localized and compact inclusions.

Figure 4.9 shows the quantitative evaluation of these experiments. We observe similar results compare to the 2D experiment: all L_1-norm methods are seen to give an average contrast closer to 1, and higher correlations with the ground truth than L_2 regularization.

4.2.5.4 Three dimensional head numerical experiments

We then evaluate our proposed methods using a physically realistic three dimensional head model derived from T1-weighted MPRAGE scans originally acquired by Eggebrecht *et al* [73] that were kindly provided to us by the authors of that work. Following the process described by Wu et al [136], Statistical Parametric Mapping (SPM) software [137] was used to perform a parametric segmentation of the 5 tissue types (scalp, skull, cerebrospinal fluid (CSF), gray matter, white matter) based on the pixel intensity probability function distribution. These five different layers were then further processed in NIRFAST to create masks and layered volumetric FEM meshes.

The mesh is composed of 101046 nodes corresponding to 589658 tetrahedral elements. Each node is labelled by one of the five segmented head tissue types, as shown in Figure 4.10. Chromophore concentrations assigned to each layer are computed from the tissue optical properties at 750nm and 850nm in a previous in vivo study [6] (Table 4.2) using Beer's law and Mie scattering formulae. A high-density (HD) imaging array containing 158 sources and 166 detectors (Figure 4.11) [73] was placed over the whole head, with source-detector (SD) separation distances ranging from 1.3 to 4.8cm. In this study,

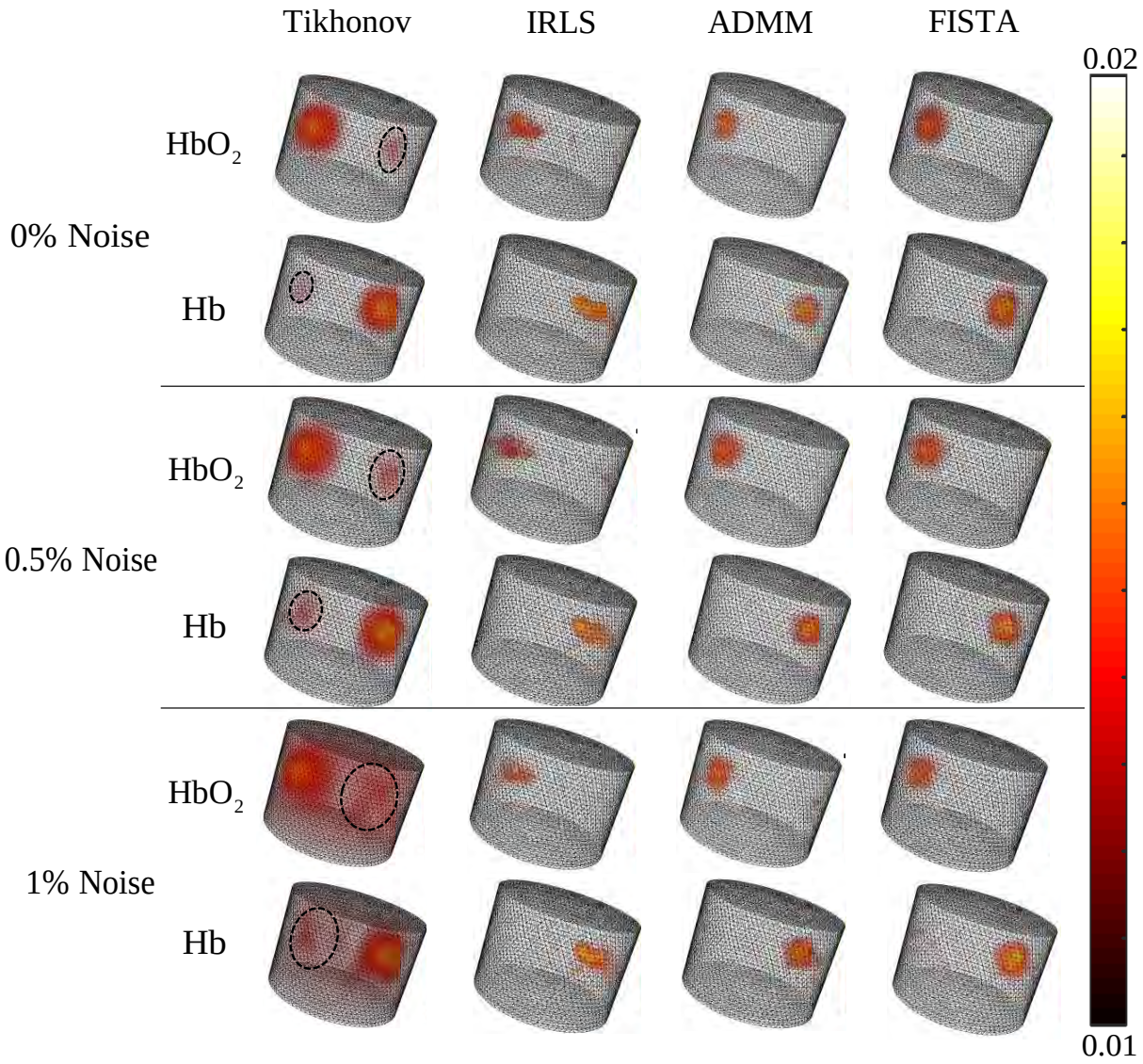


Figure 4.8: Reconstruction of HbO_2 and Hb using (L-R): Tikhonov for L_2-norm regularization; IRLS, ADMM, FISTA algorithms for L_1-norm regularization with different noise levels. Areas marked with black dotted line represent the region of crosstalk between different chromophores.

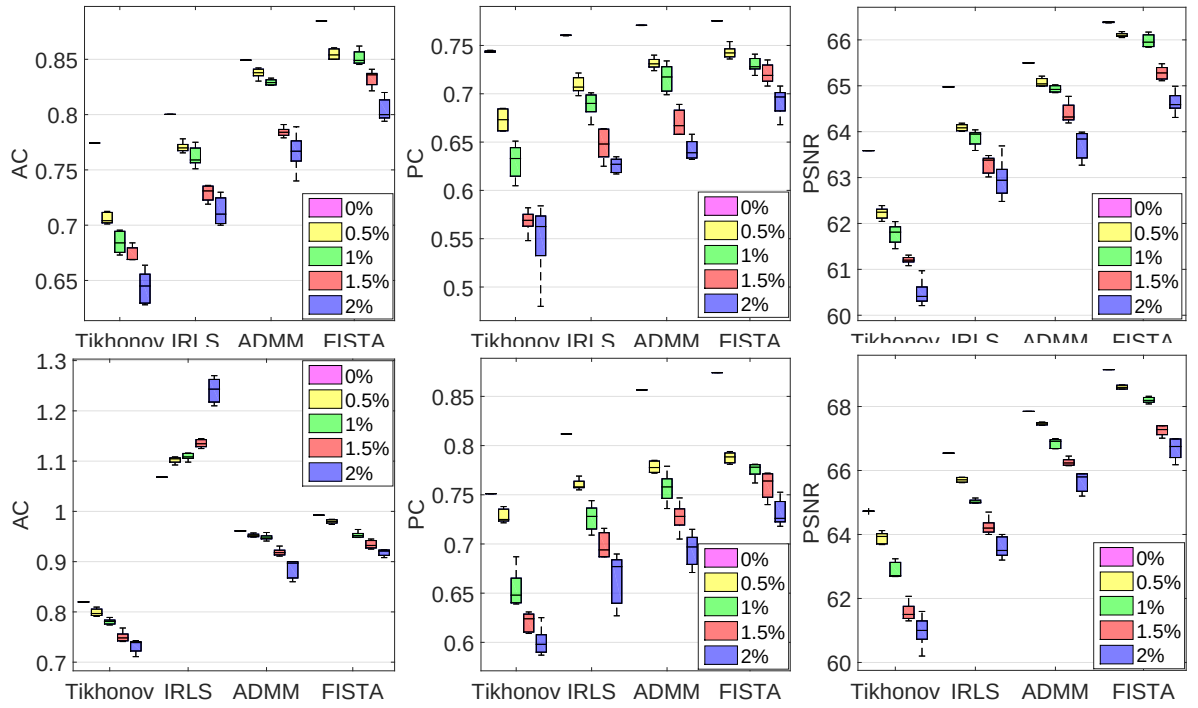


Figure 4.9: Evaluation metrics comparing the performance of different methods on 3D simulated data at five different noise levels. Left to right column: AC index; PC index and PSNR index. The first row gives the results for HbO_2, the second row for Hb.

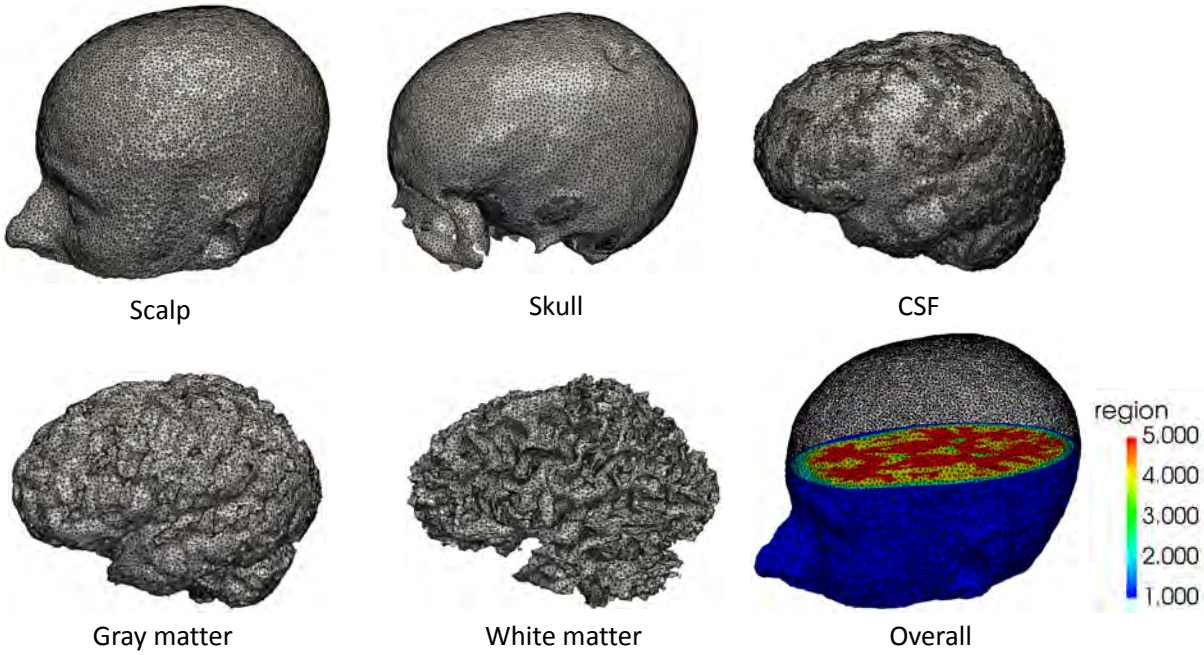


Figure 4.10: Three-dimensional surface mesh for each of the five head layers.

3478 differential measurements per wavelength were used to image the hemodynamic changes in the brain. Two individual activations were simulated in the visual cortex with chromophore concentrations of HbO_2 and Hb respectively increasing to double the

background level in the gray matter (Figure 4.12). Each simulated activation has a radius of 5mm. 0% to 2% distributed Gaussian noise at 0.5% intervals was added to the measurement vector.

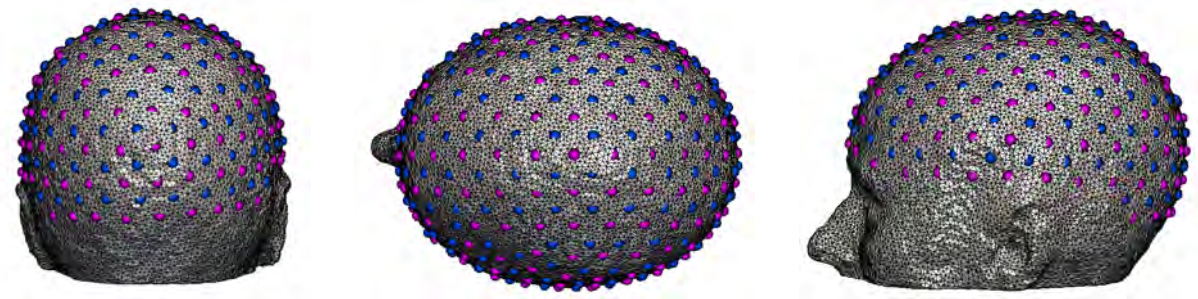


Figure 4.11: Schematic view from three directions showing the distribution of the imaging array with 158 sources (blue circles) and 166 detectors (red circles).

Reconstructed chromophore concentrations of the simulated activation using the Tikhonov model (Equation (4.22)) and the spectral-L_1 model (Equation (4.24)) on noise-free data are displayed in Figure 4.13, while those on data with 1% Gaussian noise are displayed in Figure 4.14. We only show the area with changes in chromophore concentration greater than 0.0001mM. Compared to the spectral-L_1 model, Tikhonov reconstructions have lower image contrast, which can be clearly seen from the first column of Figure 4.13 and Figure 4.14. Some artifacts (areas contained within green ellipses) can be easily observed around the source and detector areas. With increased levels of noise, larger artefacts are seen with Tikhonov regularization and the results are spatially smeared. In contrast, results from the spectral-L_1 model show fewer artifacts in the non-activation area. Higher noise does not noticeably affect the L_1-regularized reconstructions. IRLS produces visually more compact localizations than ADMM and FISTA, whilst ADMM appears to have better sparsity-inducing properties compared with IRLS and FISTA.

Evaluation metrics from these experiments are shown in Figure 4.15. It is clear that the spectral-L_1 model can achieve higher AC, PC and PSNR values than the Tikhonov model which means higher image contrast and accuracy can be achieved with L_1-norm regularization. Although the results of FISTA show more visual artifacts than other L_1-norm methods, it is still able to produce better performance based on the metrics. This

Table 4.2: Head tissue optical property for each of five layers. These numbers in the table are based on [6]. CSF stands for cerebrospinal fluid.

	Scalp	Skull	CSF	Gray Matter	White Matter
c_1 (mM)	0.0575	0.0438	0.011	0.0548	0.0683
c_2 (mM)	0.0313	0.0209	0.0083	0.0354	0.0273
Scattering amplitude	0.53	0.7258	0.3	0.5040	0.8176
Scattering power	1.1599	0.8987	9×10^{-7}	1.7757	1.3048
μ_a (mm^{-1}) at 750nm	0.017	0.012	0.004	0.018	0.017
μ_s' (mm^{-1}) at 750nm	0.74	0.94	0.3	0.84	1.19

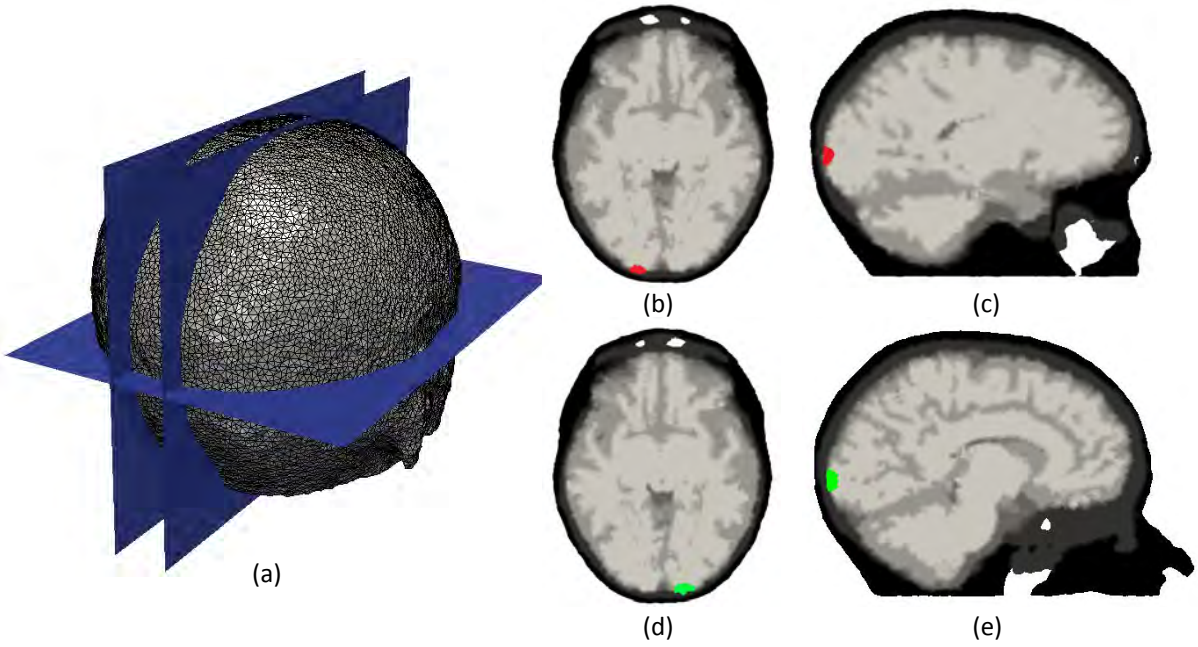


Figure 4.12: Ground-truth image with the activation exists in the gray matter and white matter. (a): Illustration of the overall distribution of slices. (b)-(c): Individual activation is color-coded in red and represents the individual simulation of HbO_2. (d)-(e): Individual activation is color-coded in green and represents the individual simulation of Hb.

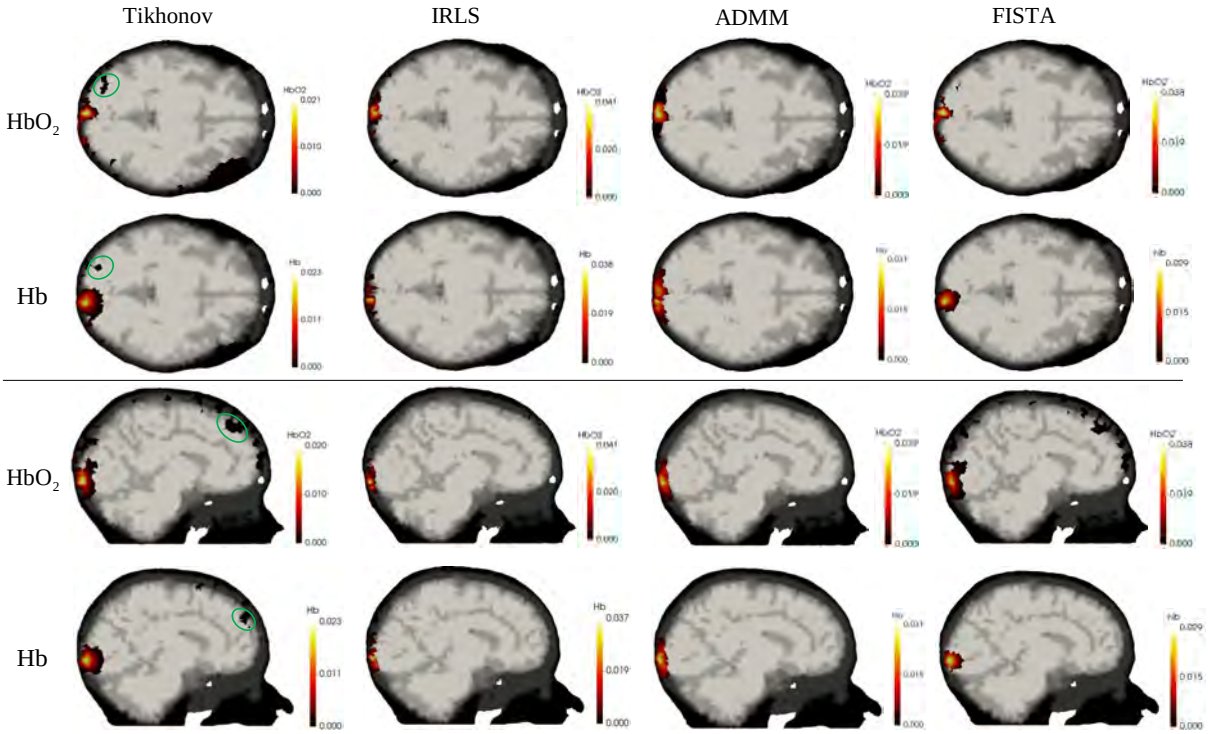


Figure 4.13: The reconstructed image of the change of HbO_2 and Hb in mM with noise-

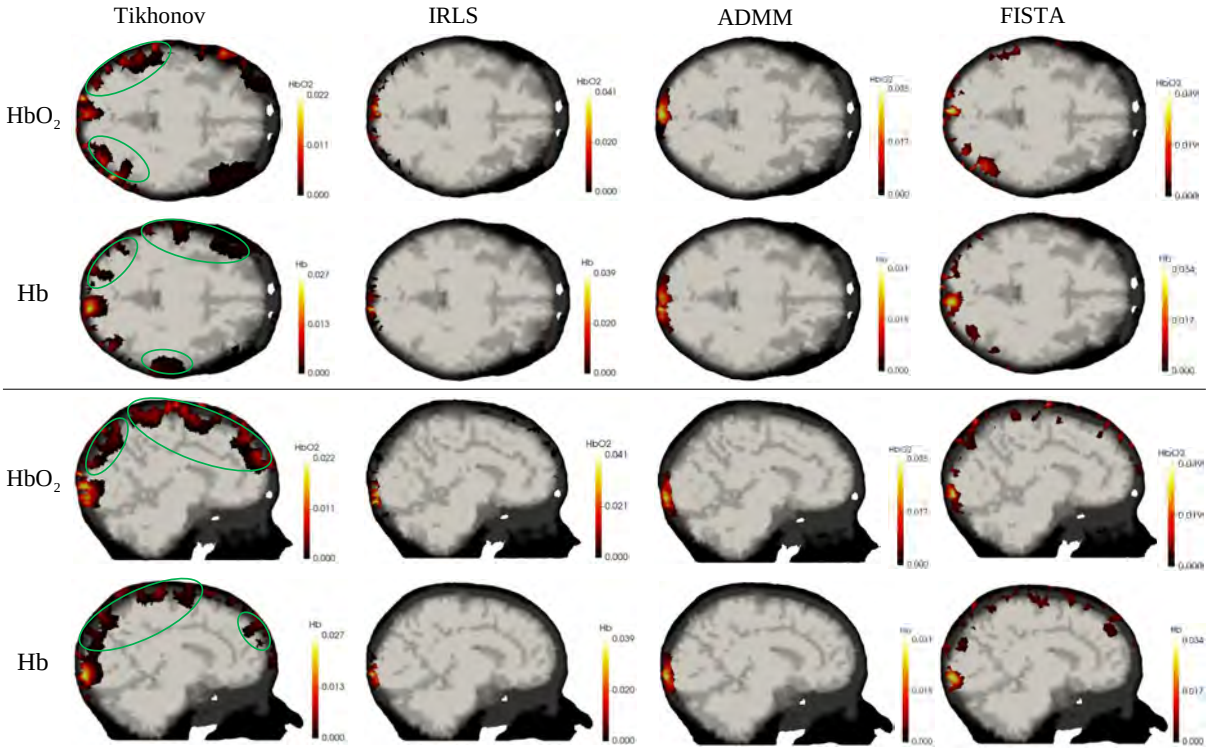


Figure 4.14: The reconstructed image of the change of HbO_2 and Hb in mM with data contaminated by 1% Gaussian noise. Some examples of reconstruction artefacts are highlighted in green ellipses.

is because (i) AC is defined on the activation region which is selected by thresholding the recovered changes based on 50% of the maximum recovered changes, artefacts away from this region do not influence this metric; (ii) By the other metrics (PC and PSNR), the improved ability of FISTA to localize the activation is sufficient to counteract the effect of the artefacts.

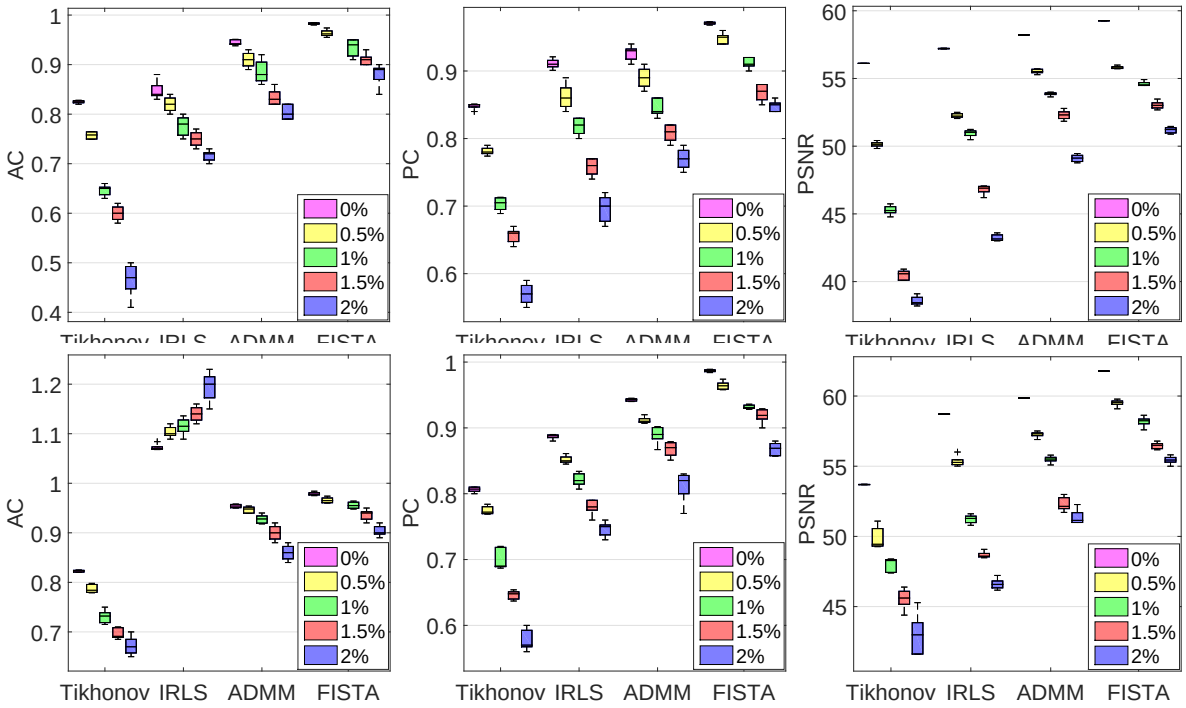


Figure 4.15: Evaluation metrics comparing the performance of different methods on a simulated 3D head model at five different noise levels. Left to right column: AC, PC and PSNR index. The first row gives the results from HbO_2; the second row from Hb.

4.2.5.5 More realistic three dimensional head numerical experiments

Following the proof-of-concept experiments described in Section 4.2.5.4, we extended our analysis to a more realistic case in functional imaging where much smaller changes happened in chromophore concentrations. In the activation area we model a small region with changes in HbO_2 (c_1) of 5μM and Hb (c_2) of -5μM, relative to the background concentrations given in Table 4.2. The mesh is the same as that used in the previous section 4.2.5.4. In line with the expected in vivo performance of imaging systems, 0.12%, 0.15%, 0.41% and 1.42% Gaussian random noise was added to first (13mm), second

(30mm), third (40mm) and fourth (48mm) nearest neighbour measurements to provide realistic data [138].

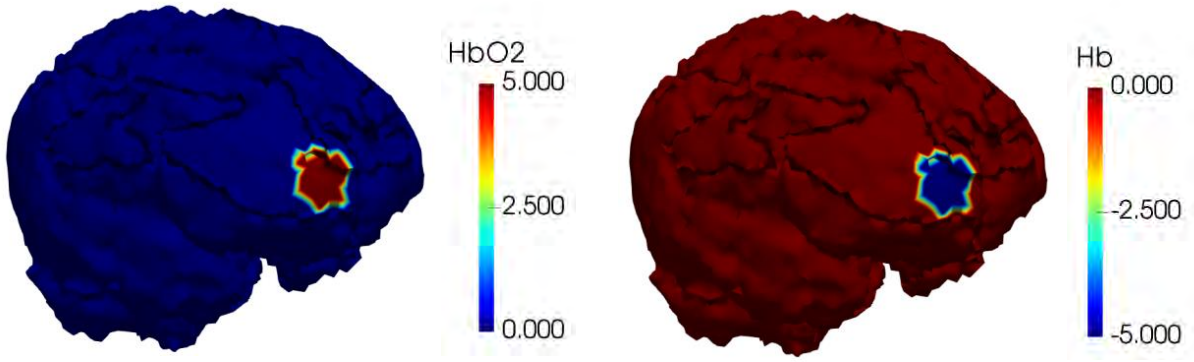


Figure 4.16: Ground-truth image showing the change in chromophore concentration confined to the gray matter.

Reconstruction using the four methods considered here are shown in Figure 4.17 with noise-free and noisy simulated data. With reference to results shown earlier in this chapter, we make a similar observation that in comparison to the ground truth values, results using Tikhonov regularization are visually inferior to those from L_1-norm regularization. With increased noise, Tikhonov regularization performs progressively worse with more artefacts visible in the source-detector areas. L_1-norm regularization induces sparse results with fewer artefacts in non-activated areas. It is noticed that with the same regularization term, different algorithms lead to different reconstructed results. Visual inspection of the results from the three L_1 algorithms suggests that IRLS produces over-sparse reconstructions with strong activations confined to a small area. ADMM and FISTA results are much more visually realistic and they are seen to give higher quantitative accuracy. These differences may result from the different regularization parameter value selected for each algorithm. The regularization parameter λ controls the amount of regularization applied to the model. As can be seen from Figure 4.2, the λ values which meet with the large curvature in the L-curves plotted in different algorithms are different. Therefore different amount of regularization is applied to the model, leading to different reconstruction performance. Because the L-curve by FISTA is the most L shaped, we take the results by FISTA as the best reconstruction results. A quantitative

Table 4.3: Three evaluation metrics for HbO_2 on results by different methods

	Without noise				With noise			
	Tikhonov	IRLS	ADMM	FISTA	Tikhonov	IRLS	ADMM	FISTA
AC	0.72	0.81	0.84	**0.91**	0.67	0.72	0.73	**0.76**
PSNR	73.81	76.48	77.33	**78.55**	38.08	63.79	67.92	**73.43**

evaluation using AC and PSNR is given in Table 4.3 and Table 4.4 and these support the conclusion that even at small changes in chromophore concentration, the spectral-L_1 model can still guarantee higher image contrast and accuracy, with FISTA performing consistently better by all measures (AC closer to 1, higher PSNR).

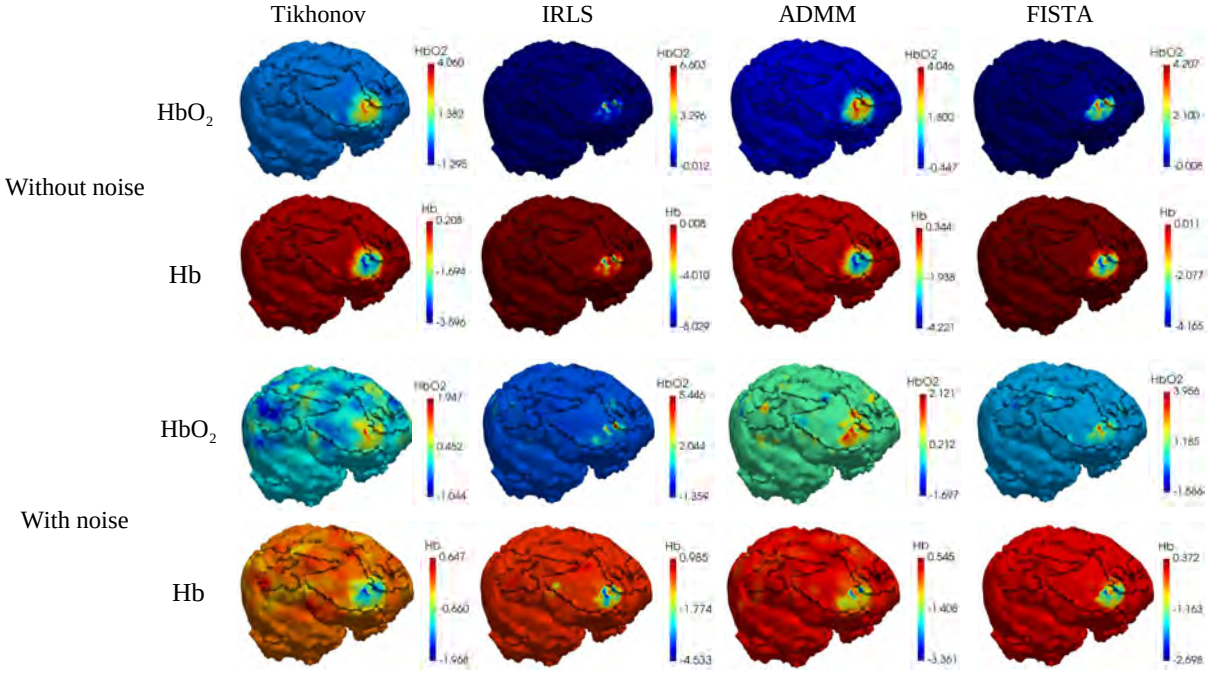


Figure 4.17: Reconstruction of HbO_2 and Hb using (L-R): Tikhonov for L_2-norm regularization; IRLS, ADMM, FISTA algorithms for L_1-norm regularization with different noise levels. First two rows: results with clean simulated data; Last two rows: those with noisy data.

Table 4.4: Three evaluation metrics for Hb on results by different methods

	Without noise				With noise			
	Tikhonov	IRLS	ADMM	FISTA	Tikhonov	IRLS	ADMM	FISTA
AC	0.86	0.95	0.96	**0.97**	0.72	0.91	0.91	**0.93**
PSNR	76.18	76.99	78.94	**79.07**	40.59	66.35	75.00	**76.26**

4.2.5.6 Experiments with phantom data

To evaluate the proposed algorithms on real experimental data, a multispectral, non-contact CW-DOT system designed for hand imaging [139, 140] was used to image a solid plastic cylindrical phantom (INO, Quebec, Canada) of radius 12.3mm and length 50mm. Boundary data was collected at five wavelengths (650nm, 710nm, 730nm 830nm and 930nm), in a transmission setup with a 7 x 5 grid of source positions on the underside of the phantom and a 11 x 9 grid of virtual detectors on top, displayed in Figure 4.18(b). The spatially constant, but spectrally varying optical properties of the phantom were measured previously in time resolved experiments [141]. The absorbing dye within the phantom was treated as a chromophore that has unit concentration in the bulk of the phantom, the extinction coefficient of which was modelled by the measured absorption coefficient. A heterogeneous version of the phantom was imaged which contained a cylindrical rod of radius 3mm and length 50mm, at a depth of 5mm (Figure 4.18(a)). The rod has twice the absorption coefficient of the bulk phantom which provides a 2:1 contrast in dye concentration compared to background (Figure 4.19 left). A homogeneous version was also imaged, enabling calibration of the model/data mismatch and any source or detector coupling variation. The mesh, as shown in Figure 4.18(a), consists of 85,205 nodes and 451,821 linear tetrahedral elements.

Ground truth data and images reconstructed with L_2 and L_1 methods are shown in Figure 4.19 respectively. The experiments described in the previous sections showed that the particular choice of L_1 method makes only a very small difference to the quality of the reconstruction, but there are very large differences in computational efficiency, with

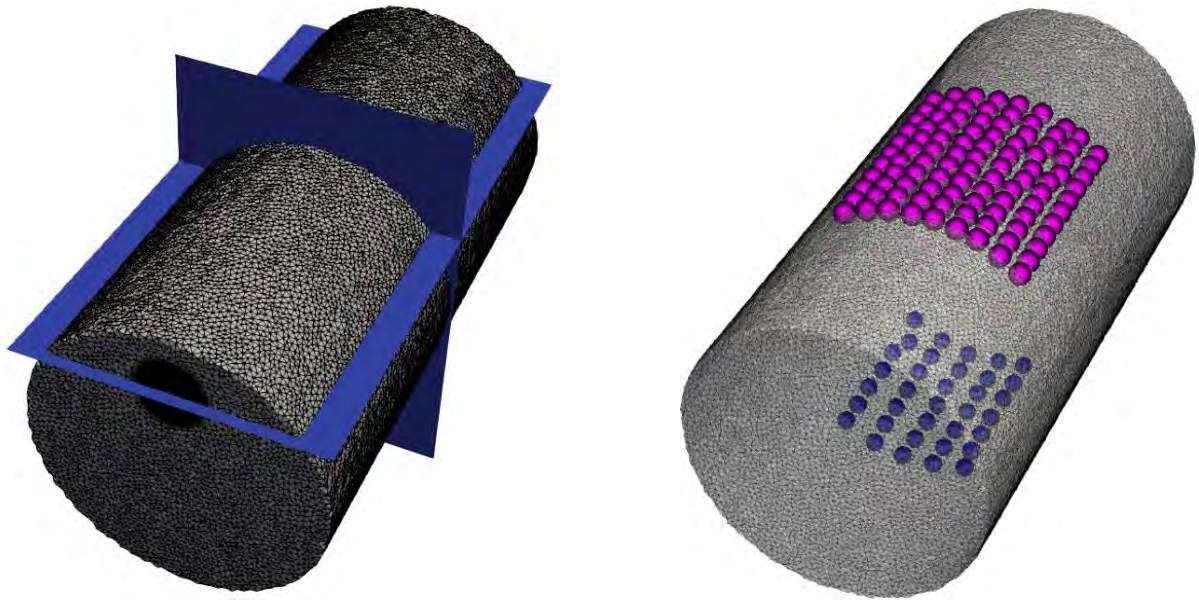


Figure 4.18: (a): Illustration of the overall distribution of slices. (b): Distribution of sources and detectors.

FISTA being far more efficient in this domain because of its superior ability to deal with large problems. Therefore in this experiment, we only use FISTA as the L_1 solver. It can be clearly observed from Figure 4.19 that L_2-norm regularization over-smooths the reconstructed images which have much lower image contrast than the ground truth. Some artefacts can be seen in the source and detector areas. We note that only the central region can be reconstructed in both cases because the sources and detectors are confined to this region, with very low sensitivity away from the centre. The image contrast reconstructed by L_1-norm regularization is much closer to the ground truth but with more compact results. We calculate the three evaluation metrics in the volume of illumination (Table 4.5) and these support the same conclusions.

4.2.5.7 Comparison of CPU time consumed in the inverse model

We now compare the computational efficiency of the proposed methods. All experiments are performed using Matlab 2013a (Mathworks, Natick, USA) on a Windows 7 (Microsoft, Redmond, USA) platform with an Intel Core CPU i7-6700 at 3.40GHz and 64.0GB memory. The simulated experiments described in Section 4.2.5.2, 4.2.5.3 and

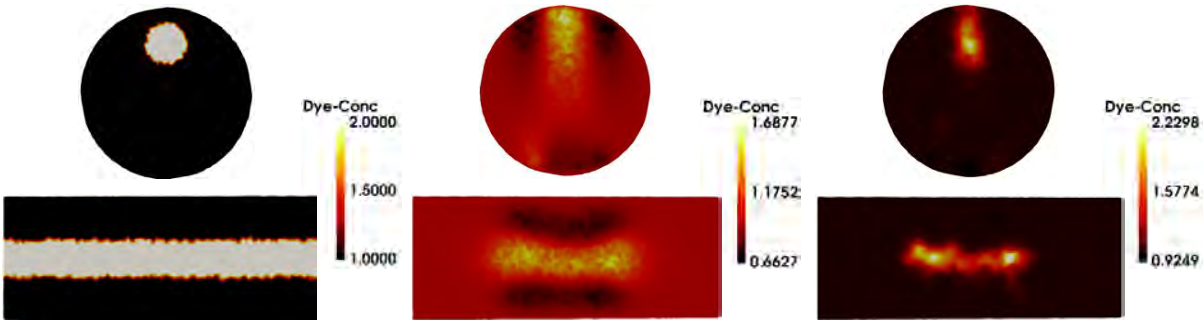


Figure 4.19: Ground truth and reconstruction results with different regularizations. From left to right: ground truth; results with L_2-norm regularization; results with L_1-norm regularization using FISTA algorithm.

Table 4.5: Evaluation of L_1-norm and L_2-norm regularization methods for reconstruction of a single rod inclusion in a tissue-simulating phantom.

	L_2-norm	L_1-norm with FISTA
AC	0.72	1.14
PC	0.66	0.73
PSNR	13.75	13.88

4.2.5.4 were used to perform this comparison. CPU times used in the inverse procedure only are measured. We run each method over ten realizations of noise at each of five noise levels to obtain reliable statistics. Figure 4.20 shows the CPU time consumed for the four different methods (Tikhonov, IRLS, ADMM, FISTA). In order to display the recorded times from ten repetitions clearly, CPU times for one iteration of the outer loop of the reconstruction algorithm are shown in Figure 4.20. Averaged total CPU times for iteration to convergence are given in Table 4.6, 4.7 and 4.8.

FISTA is clearly the fastest L_1 regularization method amongst those considered here, and it is faster even than Tikhonov regularization which does not use an inner iteration. FISTA only involves the computation of J^TJ which is much more computationally efficient than the computation of JJ^T. IRLS and ADMM are substantially slower because they require an inner iteration and inversion/multiplication of large matrices.

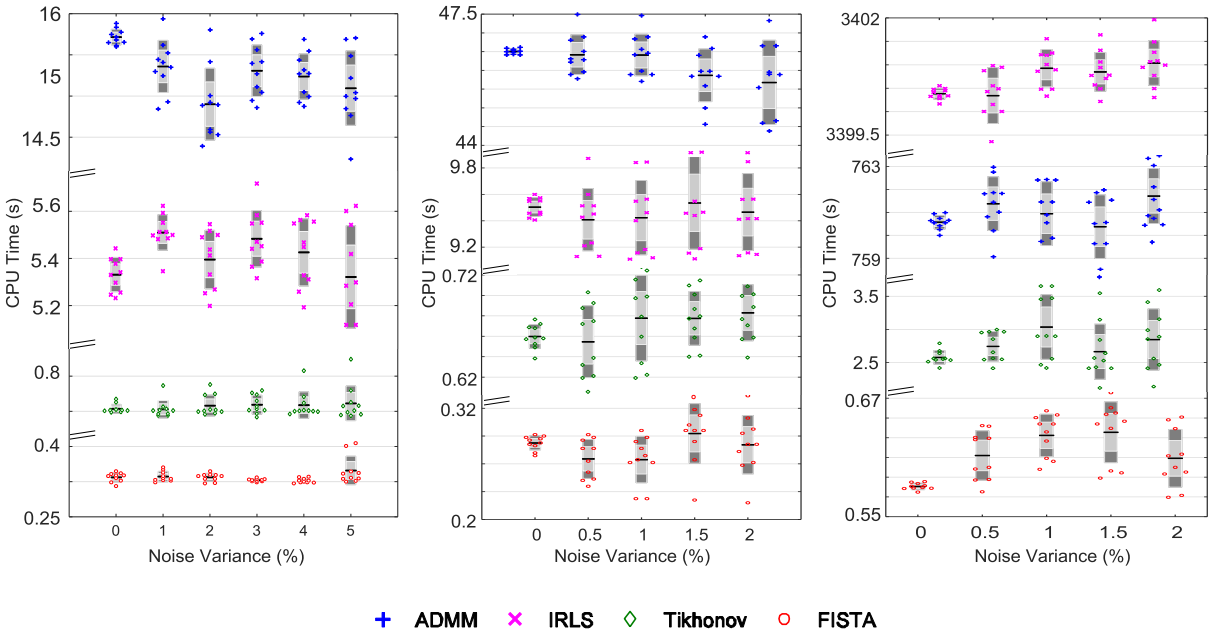


Figure 4.20: Total CPU time consumed in the experiments described in Section 4.2.5.2, 4.2.5.3 and 4.2.5.4

Table 4.6: Total CPU time(s) consumed in the inverse model for the experiments described in Section 4.2.5.2

Evaluation metric	CPU time (Mean)					
Chromophore	0%	1%	2%	3%	4%	5%
Tikhonov	34.52	5.43	5.30	5.20	5.63	5.32
IRLS	156.61	145.61	129.27	94.39	67.55	44.69
ADMM	204.59	140.48	132.19	107.54	85.19	50.07
FISTA	**1.04**	**1.05**	**1.13**	**0.84**	**0.84**	**0.83**

4.2.6 Conclusion and discussion

In this section, we introduce spectral-L_1 model for spectrally constrained DOT reconstruction. Numerical experiments showed that compared to the L_2-norm, L_1-norm regularization can reduce crosstalk and maintain image contrast by inducing sparsity. These findings were tested on real experimental data using a tissue-simulating phantom and similar results were found. Although all L_1-based methods perform similarly in terms of reconstruction quality, FISTA performs marginally better than ADMM and IRLS by the

Table 4.7: Total CPU time(s) consumed in the inverse model for the experiments described in Section 4.2.5.3

Evaluation metric	CPU time (Mean)				
Chromophore	0%	0.5%	1%	1.5%	2%
Tikhonov	17.15	3.65	3.74	2.61	2.63
IRLS	250.55	223.21	185.11	135.99	87.99
ADMM	853.02	751.08	350.99	252.59	154.60
FISTA	**1.13**	**1.20**	**1.18**	**1.21**	**1.22**

Table 4.8: Total CPU time(s) consumed in the inverse model for the experiments described in Section 4.2.5.4

Evaluation metric	CPU time (Mean)				
Chromophore	0%	0.5%	1%	1.5%	2%
Tikhonov	42.79	33.41	35.47	33.42	14.56
IRLS	10171.96	10171.95	6823.03	6821.99	6822.01
ADMM	9100.34	7444.31	6195.01	4193.57	2892.30
FISTA	**2.53**	**1.97**	**1.84**	**1.85**	**1.93**

measures of AC, PC, and PSNR, and is far more computationally efficient as it avoids direct matrix inversion and large matrix-matrix multiplications.

The contributions of this paper can be summarized as follows: 1) It is the first time that L_1-norm regularization methods and spectrally constrained DOT methods have been used together and it is their combination (i.e. spectral-L_1 model) that is original. We have given detailed descriptions of how this can be done, and performed systematic comparisons of the performance and efficiency of the different methods on both simulated and real data. We are not aware of any previously published work that proposes such a model and performs such a detailed analysis of its performance; 2) We have developed a method to automatically select the regularization parameters. This is based on the L-curve method but is modified for efficient application in this use-case; 3) We have con-

ducted extensive numerical experiments on simulated data and on real experimental data, and have performed comprehensive and robust qualitative and quantitative evaluations. To the best of our knowledge, this is the first systematic study in the area of spectral DOT reconstruction to perform such a comprehensive evaluation.

While promising results have been obtained, accurate reconstruction is still challenging. Regularizations involving the sparsity constraints are used under the assumption that the optical properties (representing the image) to be reconstructed are spatially sparse. These regularizations tend to oversparsify the distribution of the optical properties when such an assumption does not hold, as shown in Figure 4.19. In the case of multiple activations or complex injuries in the brain, the features of interest are not spatially localized and the optical properties relative to the background are therefore non-sparse [73]. A different approach is required to reconstruct images in which edges are preserved and features are not spatially sparse.

CHAPTER 5

GRAPH-BASED INVERSE MODEL FOR DIFFUSE OPTICAL TOMOGRAPHY

Tikhonov (L_2-norm) regularization is attractive for use in ill-posed problems because of its simplicity. However in functional neuroimaging studies, when the dynamic range of the measurement is large, incorporation of the maximum variance in the measurement will lead the minimization to bias the solution to specific data points, especially near the boundaries at source-detector locations in DOT [23]. To reduce the effect of bias, some researchers incorporate the spatial priors (RSP-I) from other medical modalities (for example MRI) which have relatively higher spatial resolution to improve quantitative accuracy and depth information of reconstructed images [22]. However, obtaining images from other modalities is time consuming and the pre-processing segmentation process is prone to errors.

In order to efficiently reduce above bias and alleviate the over-sparsifying effect existing in purely sparsity inducing regularization, we introduce spatial gradient differential operator to DOT reconstruction. Gradient differential operator is a highpass operator, which can impose smoothness to the solution and improve the conditioning of the minimization problem.

The contents of this chapter were the subject of the following publication.

- **Wenqi Lu, Jinming Duan, David Orive-Miguel, Lionel Herve, and Iain B. Styles,** ***Graph- and finite element-based total variation models for***

***the inverse problem in diffuse optical tomography*, Biomedical Optics Express 10, 2684-2707 (2019)**

5.1 Introduction

In image processing, the image domain is normally expressed as a Cartesian grid which is composed of squares (cubes in 3D) and is aligned with the Cartesian coordinate axes [68]. Cartesian grids can be generated easily with low computational effort. Under this expression, the differential operators, such as gradient, divergence, Laplacian and curvature, can be discretized straightforwardly using the finite difference method (FDM) [68]. However, the weakness of the Cartesian methods is the accuracy of the flow solution at boundaries, which are either curved or not oriented along the Cartesian coordinates. This problem becomes especially serious in the case of DOT, where the computational geometry is irregular (i.e. the multi-layer head) and accurate boundary flux are required. Hence, FDM is not practical in DOT. Normally, two representations can be employed to model irregular geometries: finite element and graph representations. In the former, the object geometry is represented by a polygon/polyhedron, over which a series of triangles (tetrahedra in 3D) are generated. In the latter, the object geometry is represented by an unstructured graph, defined by vertices, edges and weights. A vertex (plural vertices) or node is the fundamental unit of which graphs are formed and edge is each of the related pairs of vertices. In order to show the difference between the two representations clearly, in Figure 5.1, we use a circle as one example which is discretized with the two representations. For the FE representation (left), the circle is divided by a series of elements jointed at different vertices (nodes). In FE representation, we normally term a discrete geometry as a FE mesh. Within the mesh, a triangle (representing one element)

is highlighted comprising three disjoint vertices. For the graph representation (right), the circle is simply represented with a set of vertices and edges. In the graph representation

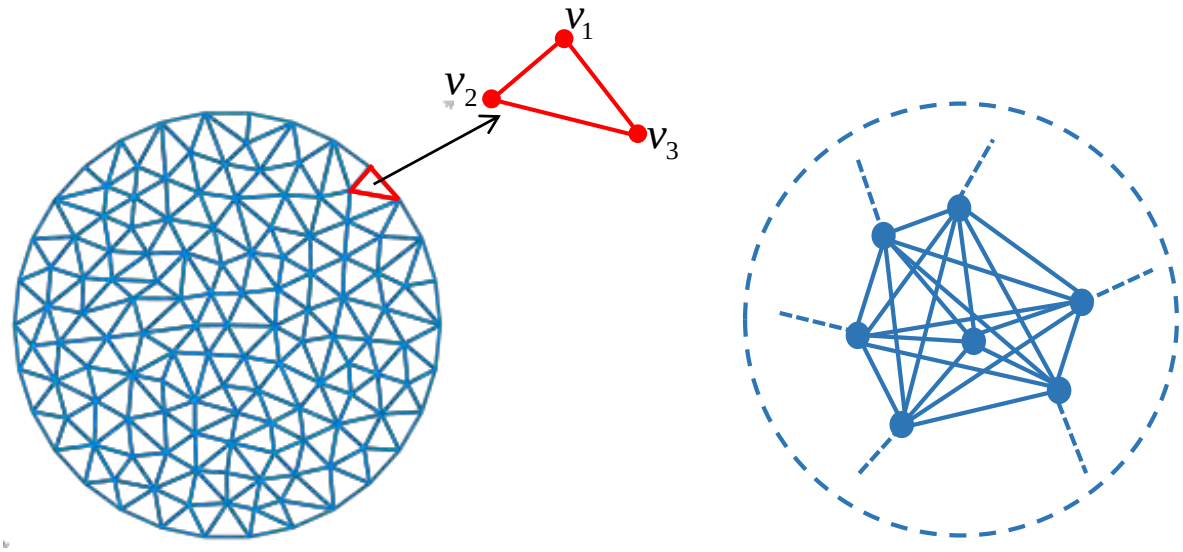


Figure 5.1: Modeling an unstructured geometry using finite element (left) and graph (right) representations.

For each representation, there is a systematic discretization scheme (finite element discretization or graph discretization) for the differential operators, which can be readily applied to the minimization problems. However, FEM implementations can be difficult and time-consuming, especially when higher-order polynomial basis (shape) functions are used for non-linear interpolation between vertices of high-order elements [24]. Finite element representation has been widely used in DOT reconstruction but we are not aware of any previous work that attempts to formulate the DOT reconstruction using a graph representation.

5.2 Graph representation

Under the graph representation [142, 143], the computational domain Ω is discretized using a weighted graph $G = (V, E, w)$, where $V = \{V_k\}_{k=1}^{N}$ denotes a finite set of N vertices, and $E \in V \times V$ represents a finite set of weighted edges. Here $V = V_{\Omega'} \cup V_\Gamma$ with $V_{\Omega'}$ representing vertices in the inner area Ω' and V_Γ vertices on the boundary Γ. In this study, we assume that G is an undirected simple graph (no multiple edges). Let $(i, j) \in E$ be an edge of E that connects the vertices i and j in V. The weight w_{ij} denotes the similarity between two vertices i and j. The computation of this quantity is discussed later in this section.

A typical graph in image processing is the 8-adjacency graph where vertices are associated to pixels and edges correspond to pixel adjacency relationships. This corresponds to a classical local processing. By changing the graph topology, we can naturally extend this idea to include nonlocal information. A vertex may have nonlocal edges with other vertices which are not spatially nearby. This idea is popularly used in image processing applications such as image denoising [142, 144] and image inpainting [121, 145]. Antoni et al. [144] pointed out that nonlocal methods can result in much greater post-filtering clarity, and less loss of detail in the image compared with local algorithms. Therefore, in order to fully leverage the power of graph-based discretization, we use the nonlocal vector calculus in this work. The corresponding nonlocal differential operators on the graph G can be defined as follows.

Definition (*Nonlocal gradient*). For a function $\mu_i : V \to \mathbb{R}$ and a nonnegative and symmetric weight function w_{ij}: $V \times V \to \mathbb{R}$, the nonlocal partial derivative can be written as

$$\partial_j \mu_i \triangleq (\mu_j - \mu_i) \sqrt{w_{ij}} : V \times V \to \mathbb{R}. \tag{5.1}$$

Therefore, the nonlocal gradient $\nabla_w \mu_i$ is defined as the vector of all partial derivatives:

$$\nabla_w \mu_i \triangleq (\mu_j - \mu_i) \sqrt{w_{ij}} \ : V \to \mathbb{R}. \tag{5.2}$$

Definition (*Nonlocal divergence*). Given a vector function $\boldsymbol{\nu}_i$: $V_{\Omega'} \to \mathbb{R}$ and a weight

function w_{ij}: $V \times V \to \mathbb{R}$, the nonlocal divergence operator div_w acting on $\boldsymbol{\nu}_i$ is

$$\mathrm{div}_w \boldsymbol{\nu}_i \triangleq \sum_{j=1}^{N} (\nu_{ij} - \nu_{ji}) \sqrt{w_{ij}} \ : V_{\Omega'} \to \mathbb{R}, \tag{5.3}$$

where ν_{ij} is the j'th element of $\boldsymbol{\nu}_i$.

Definition (*Nonlocal Laplacian*). Let $\mu_i : V \to \mathbb{R}$ and w_{ij}: $V \times V \to \mathbb{R}$. The linear nonlocal Laplace operator acting on Φ_i is defined based on equations (5.2) and (5.3):

$$\Delta_w \mu_i \triangleq \frac{1}{2} \mathrm{div}_w \ (\nabla_w \mu_i) = \sum_{j=1}^{N} (\mu_j - \mu_i) \, w_{ij} \ : V \to \mathbb{R}. \tag{5.4}$$

It should be noticed from the nonlocal differential operator definitions (Equations (5.2), (5.3), and (5.4)) that, in a full non-local scheme, each vertex has connections with all the vertices in V over Ω such that the constructed graph is fully connected. This can make the computational load extremely heavy and so approaches based on spectral graph theory [146, 147] or nearest neighbors [148], are typically employed to partition the vertices in the computational domain into groups according to their similarities. For example, Bertozzi [147] used spectral approaches along with the Nyström extension method to efficiently calculate the eigendecomposition of a dense graph Laplacian. The second eigenvector of the graph Laplacian was used to initialize the partitioning so that the weights between vertices in different groups are small and the weights between vertices within the same group are large. In this work, we build the graph by using the positions of the nodes and the connectivity between nodes in the finite element mesh as the vertices and edges in the graph to sparsify the graph for computational efficiency. Therefore, for each vertex i, we consider only those vertices that are directly connected to the vertex i for $\mathcal{N}_i$ (i.e. those vertices that share the edge with i). This approach constrains the nonlocal scheme to edges that exist between spatially connected vertices. Instead of using the differential operators defined under the FE representation, we use the ones defined under the nonlocal vector calculus to improve the discretization accuracy.

5.3 Graph based Tikhonov regularization (GTikhonov) for DOT reconstruction

5.3.1 Introduction

In this section, we combine the spatial gradient differential operator with the Tikhonov (L_2-norm) regularization to alleviate above bias in functional neuroimaging studies. Here we name the new proposed regularization as GTikhonov. We perform a systematic comparison between Tikhonov and GTikhonov regularizations on a digital head model where activations are simulated on the auditory cortex. The concentrations of HbO_2 and Hb were recovered from 3D continuous-wave DOT at two wavelengths (750 and 850 nm). The results from the two methods were compared with respect to their ability to localize the activation depth and to achieve higher accuracy. Localization error (LE), average contrast (AC) and peak signal-to-noise ratio (PSNR) in the recovered activation are adopted to quantitatively evaluate the performance of different regularizations. LE is defined as the Euclidian distance between the central node of the simulated activation region and the one of the recovered activation region while AC and PSNR were defined in the last section as equations (4.37) and (4.39).

5.3.2 Methodology

Functional imaging is based on the assumption that the change of absorption is very small. The image reconstruction problem is single-step and linear and remains ill-posed. The recovery of chromophore concentrations consists of two steps [125]. The first step is to recover a temporal change of absorption coefficient $\delta\mu_\lambda$ at each wavelength. This procedure needs to compare the boundary measurements before and after the activation at each wavelength. The objective function is given by

$$\delta\mu_\lambda^k = \arg\min_{\delta\mu_\lambda} \left\{ ||\delta\Psi_\lambda^{k-1} - \mathbf{J}_\lambda^{k-1}\delta\mu_\lambda||_2^2 + \lambda||\nabla_w(\delta\mu_\lambda)||_2^2 \right\}. \tag{5.5}$$

We compute an analytical solution by calculating the derivative of the objective function (Equation (5.5)) with respect to $\delta\mu_\lambda$ and setting the result to zero to find the minimum. We can get the solution

$$\delta\mu_\lambda^k = (\mathbf{J}_\lambda^{(k-1)\mathrm{T}}\mathbf{J}_\lambda^{(k-1)} - \lambda\Delta_w)^{-1}\mathbf{J}_\lambda^{(k-1)\mathrm{T}}\delta\Psi_\lambda^{k-1}. \tag{5.6}$$

where the laplacian operator Δ_w is defined in equation (5.4). w_{ij} is a nonnegative and symmetric weight function. There are many ways to define the weights, including contiguity weights, distance weights, and other weights. The simplest form of weights is binary weight where w_{ij}=1 for neighboring locations i and j, or when the distance between location i and j is less than a fixed distance, or for the fixed number of locations nearest to location i. Otherwise, w_{ij}=0. Another widely used weight is the spatial weight, defined by kernel, such that closer points are given higher weights. The Gaussian kernel is one of the most widely used kernel. In this work, in order to find a better way to guide the reconstruction in one-step functional imaging, we evaluate two weight functions which are defined using the Gaussian kernel: one is based on the spatial closeness (Equation (5.7)) and the other is based on the optical property difference (Equation (5.8)).

$$w_{ij} = \exp\left(-\frac{\|X_i - X_j\|_2^2}{2\sigma_d^2}\right), \tag{5.7}$$

$$w_{ij} = \exp\left(-\frac{\|\delta\mu_{\lambda,i} - \delta\mu_{\lambda,j}\|_2^2}{2\sigma_r^2}\right). \tag{5.8}$$

X_i and X_j are the coordinates of the two vertices i and j respectively while $\delta\mu_i$ and $\delta\mu_j$ are the corresponding difference of absorption coefficients. We name the regularization which uses equation (5.7) for the weight function definition as GTikhonov-SC while the one using equation (5.8) as GTikhonov-OD. In the second step, we decompose the change of chromophore concentrations from the calculated $\delta\mu_\lambda$ at different wavelength [125]

$$\begin{pmatrix} \delta\mathrm{HbO_2} \\ \delta\mathrm{Hb} \end{pmatrix} = \begin{pmatrix} \varepsilon_{\mathrm{HbO_2},\lambda_1} & \varepsilon_{\mathrm{Hb},\lambda_1} \\ \varepsilon_{\mathrm{HbO_2},\lambda_2} & \varepsilon_{\mathrm{Hb},\lambda_2} \end{pmatrix}^{-1} \begin{pmatrix} \delta\mu_{\lambda 1} \\ \delta\mu_{\lambda 2} \end{pmatrix}, \tag{5.9}$$

where $\varepsilon_{c,\lambda}$ is the molar extinction coefficient of chromophore c (c is either $\mathrm{HbO_2}$ or Hb) at wavelength λ.

5.3.3 Experiments

We compare the Tikhonov, GTikhonov-SC and GTikhonov-OD methods on a three dimensional(3D) head model derived from T1-weighted MPRAGE scans originally acquired by Eggebrecht et al [73]. The mesh for the image reconstruction has 200000 nodes with 1215434 elements. The imaging system is same to the one used in Section 4.2.5.4 which is a high density (HD) DOT system with 158 NIR light sources and 166 detectors. The source-detector (SD) separation distances range from 1.3cm to 4.8cm, resulting in 3478 associated SD pairs. 0.12%, 0.15%, 0.41% and 1.42% Gaussian random noise was added to first (13mm), second (30mm), third (40mm) and fourth (48mm) nearest neighbour measurements to provide realistic data [138]. Ten repeats were performed at different realizations of the noise. In this experiment, one activation is simulated in the auditory cortex with changes in HbO_2 of 5μM and Hb of -3μM, relative to the background concentrations as shown in Table 4.2. The distance from the activation to the head surface is between 7.5mm to 12mm.

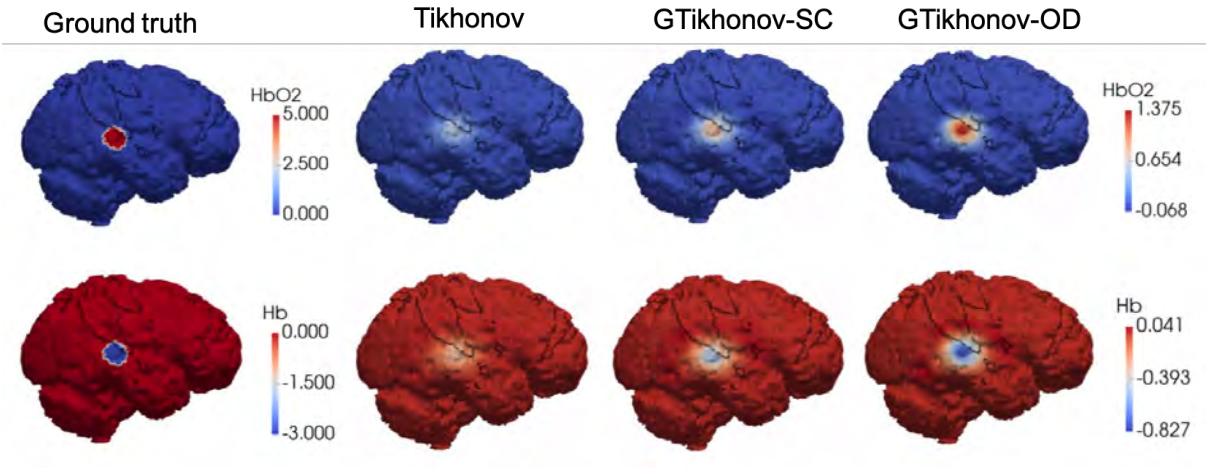


Figure 5.2: Reconstruction of HbO_2 (first row) and Hb (second row) using (L–R): Tikhonov; GTikhonov-SC and GTikhonov-OD regularization.

Reconstructed chromophore concentrations of the simulated activation using the Tikhonov model and the GTikhonov models are displayed in Figure 5.2. The results are scaled in the same range. It can be clearly seen that Tikhonov regularization results in the lowest image contrast. NLTikhonov-OD regularization can better localize the activation area

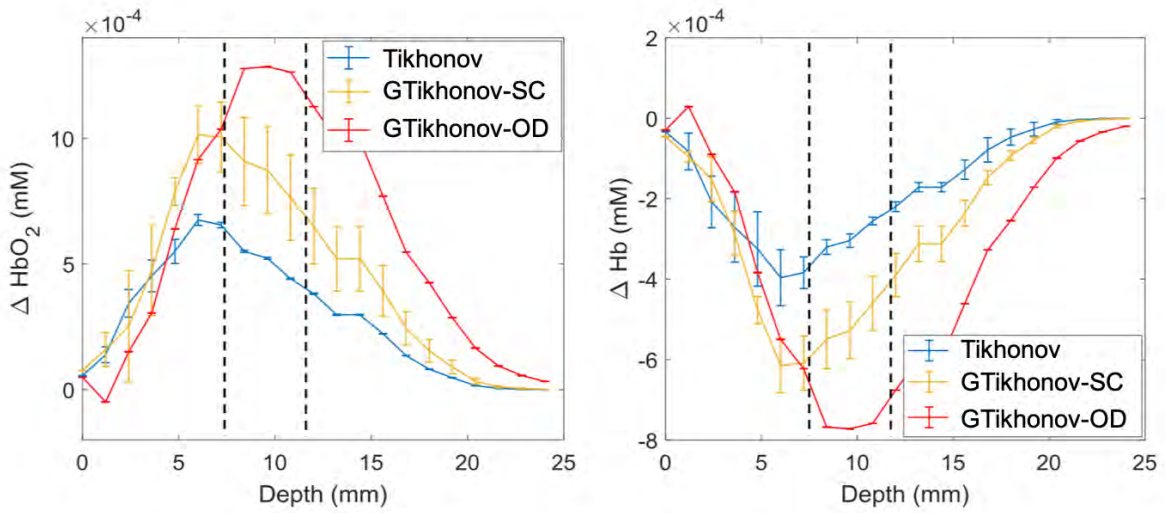


Figure 5.3: 1D cross section of the results in Figure 5.2.

Table 5.1: Evaluation on chromophore concentrations reconstructed by Tikhonov, GTikhonov regularization methods respectively

	HbO_2 (mean)			Hb (mean)		
	Tikhonov	GTikhonov-SC	GTikhonov-OD	Tikhonov	GTikhonov-SC	GTikhonov-OD
LE	4.03	3.83	**0.56**	4.27	3.74	**0.58**
AC	0.16	0.25	**0.26**	0.16	0.24	**0.22**
PSNR	61.34	80.70	**82.13**	61.31	86.12	**86.57**

and recover higher accuracy. In Figure 5.3, we plot the 1D cross section of results in Figure 5.2 along the line which crosses the nearest boundary node of the central simulated activation node and the central simulated activation node. The simulated activation is at a depth of from 7.5mm to 12mm, which is the area between the two black dash lines. We find that better depth localization can be achieved using the GTikhonov-OD regularization, while results by the GTikhonov-SC regularization and the standard Tikhonov model both deviate from the target area. We compare the three evaluation metrics in the volume of illumination (Table 5.1) and lower localization error and higher accuracy can be achieved by the GTikhonov-OD regularization.

5.3.4 Conclusion

GTikhonov regularization with a spatial gradient operator is proposed for image reconstruction in functional imaging. Simulated experiments have shown that GTikhonov regularization in which the similarity is mainly constrained by the optical properties can obtain higher accuracy and better depth localization than the standard Tikhonov regularization.

Even though graph-based gradient differential operators can help to locate the recovered activation, GTikhonov method still suffers from the over-smoothing effect because it is based on the Tikhonov regularization. In the next section, we introduce total variation regularization, which combines the L_1-norm regularization and graph-based gradient differential operator, to recover non-sparse images in which edges are preserved and features are not spatially sparse.

5.4 Graph based total variation regularization (GTV) for DOT reconstruction

5.4.1 Introduction

Total variation (TV) regularization, which uses the L_1-norm of the *gradient* of the solution as a regularizing term (details are given in Section 5.4.2 and 5.4.3), can be used to overcome the limitations associated with Tikhonov regularization $\|\delta\mu\|_2^2$, GTikhonov regularization $\|\nabla(\delta\mu)\|_2^2$, L_1-norm $\|\delta\mu\|_1$ or L_p-norm regularization $\|\delta\mu\|_p$. The gradient operator can transform the solution $\delta\mu^*$ to a sparse space where non-zero values only occur at sharp features. As such, TV can perform better than $\|\delta\mu\|_1$ or $\|\delta\mu\|_p$ at preserving edges of objects in images that are not sparse. Due to these advantages, TV has been adapted from applications in image processing [121, 149, 150] to various medical image reconstruction problems, including photoacoustic tomography (PAT) [151], bioluminescence tomography (BLT) [152], fluorescence tomography (FT) [153, 154], as well as

DOT [155, 156].

The minimization of a TV-associated problem can be non-trivial due to the non-linearity and non-differentiability of the TV regularization. In image processing, many efficient optimization algorithms have been developed for this task, including iteratively re-weighed least squares [117], primal dual [157], split Bregman [158], and fast iterative shrinkage-thresholding algorithm (FISTA) [109, 122]. Recently, alternating direction method of multipliers (ADMM) [68, 121, 159, 160] has become increasingly popular. The elegance of ADMM lies in decomposition of the original minimization problem into several simple subproblems, each of which either has a closed-form solution or can be iteratively solved with efficient numerical methods. However, since ADMM-based methods have been implemented mainly for Cartesian grids using a forward-backward FDM [161], it is not straightforward to generalize them to solve the inverse problem on an unstructured domain. Moreover, the non-linearity of the data fitting term further complicates the DOT reconstruction problem, making the overall minimization process difficult.

In this work, we address these limitations to develop TV regularization approaches to the inverse problem in DOT. More specifically, we make the following three distinct contributions: **(1)** We introduce finite element and graph representations to model unstructured geometries of DOT objects. The discrete differential operators resulting from each representation are then defined so that the minimization of the inverse problem associated with TV regularization can be carried out on unstructured domains. To the best of our knowledge, this is the first time that finite element-based discretizations have been provided in detail for DOT image reconstruction with TV regularization. Additionally, we are not aware of any previous work that attempts to formulate the TV-regularized inverse problem using a graph representation. **(2)** We propose an efficient algorithm based on ADMM to minimize the TV-regularized inverse problem. Our algorithm can handle unstructured geometries, and also reduced the computational difficulties arising from the non-differentiability and dual non-linearities in the inverse problem. **(3)** We further investigate the isotropic and anisotropic variants of TV regularization, and compare their

finite element- and graph-based implementations against the popular Tikhonov model, both qualitatively and quantitatively using extensive numerical experiments.

5.4.2 TV regularizations under finite element representation

Under the Galerkin FE method which we detailed in Section 3 and take a 2D geometry as one example, we first discretize the unstructured 2D domain Ω by M triangles jointed at N vertex nodes. $V = \{V_k\}_{k=1}^{N}$ denotes a finite number of N nodes. The continuous function piecewise-linear function $U(x, y) : \Omega \to \mathbb{R}$, approximating the optical properties on Ω, can be written in the form of

$$U = \sum_{i=1}^{N} \mu_i \varphi_i. \tag{5.10}$$

Here $\{\varphi_i\}_{i=1}^{N}$ are linear basis functions defined as $\varphi_j(V_i) = 1$ if $i = j$ and $\varphi_j(V_i) = 0$ if $i \neq j$. $\mu_i : V \to \mathbb{R}$ is the value of optical property on each node in the FE mesh, $i = 1, ..., N$.

Equation (5.10) means that the optical property value inside a triangle is associated with the optical property values on all nodes in the mesh. Given three nodes of a triangle T, i.e. $v_1 = (x_1, y_1)$, $v_2 = (x_2, y_2)$ and $v_3 = (x_3, y_3)$, there are three linear basis functions φ_i associated with the nodes, which are respectively expressed as

$$\begin{aligned} \varphi_1(x, y) &= a_1 x + b_1 y + c_1 \\ \varphi_2(x, y) &= a_2 x + b_2 y + c_2 : \quad \Omega \to \mathbb{R} \\ \varphi_3(x, y) &= a_3 x + b_3 y + c_3 \end{aligned} \tag{5.11}$$

where $a_1 = (y_2 - y_3)/(2A_T)$, $b_1 = (x_3 - x_2)/(2A_T)$, $c_1 = (x_2 y_3 - x_3 y_2)/(2A_T)$, $a_2 = (y_3 - y_1)/(2A_T)$, $b_2 = (x_1 - x_3)/(2A_T)$, $c_2 = (x_3 y_1 - x_1 y_3)/(2A_T)$, $a_3 = (y_1 - y_2)/(2A_T)$, $b_3 = (x_2 - x_1)/(2A_T)$ and $c_3 = (x_1 y_2 - x_2 y_1)/(2A_T)$. (x, y) represents any point inside of the triangle T. A_T denotes the triangular area of T, which is computed as $A_T = |x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2)|/2$.

In FE, one starts from a continuous problem and approximates the solution with a piecewise-polynomial function U. As such, we define the following anisotropic and

isotropic TV regularizations

$$\int_{\Omega} (|\partial_x U| + |\partial_y U|) dxdy = \|\mathbf{D}_x \mu\|_1 + \|\mathbf{D}_y \mu\|_1 \tag{5.12}$$

$$\int_{\Omega} \sqrt{(\partial_x U)^2 + (\partial_y U)^2} dxdy = \sum_{i=1}^{M} \sqrt{|(\mathbf{D}_x \mu)_i|^2 + |(\mathbf{D}_y \mu)_i|^2} \tag{5.13}$$

In equations (5.12) and (5.13), the continuous TV regularizations and their resulting discretized versions are shown on the left-hand side and right-hand side, respectively. The two discrete versions respectively are the anisotropic and isotropic definitions of TV regularization. ∂_x and ∂_y are continuous partial derivatives along the x and y directions, respectively. $\mathbf{D}_x$ is a matrix of size $M \times N$ which, when applied to μ, gives the discrete partial derivative of μ along the x direction. $\mathbf{D}_y$ is the derivative matrix along the y direction. $\mathbf{D}_x \mu$ and $\mathbf{D}_y \mu$ are therefore two vectors of size $M \times 1$, where M is the number of triangles in the mesh. We note that the main idea of FE is to break down the calculation domain Ω onto the local elements individually. Afterwards, the derived local matrices are assembled element by element to enable the final computation. Equations (5.12) and (5.13) can be proved by expressing the partial derivatives $\partial_x U$ and $\partial_y U$ in terms of a basis. To illustate this idea we prove the first term of equation (5.12):

$$\begin{aligned} \int_{\Omega} |\partial_x U| dxdy &= \sum_{i=1}^{M} \int_{T_i} |\partial_x U| \, dxdy = \sum_{i=1}^{M} \int_{T_i} |\sum_{j=1}^{N} \mu_j \partial_x \varphi_j| dxdy \\ &= \sum_{i=1}^{M} A_{T_i} \, |a_{i,1}\mu_{i,1} + a_{i,2}\mu_{i,2} + a_{i,3}\mu_{i,3}| \\ &= \sum_{i=1}^{M} |(\mathbf{D}_x \mu)_i| = \|\mathbf{D}_x \mu\|_1, \end{aligned} \tag{5.14}$$

where A_{T_i} denotes the area of triangle T_i and the subscripts $\{i,1\}$, $\{i,2\}$ and $\{i,3\}$ in $|a_{i,1}\mu_{i,1} + a_{i,2}\mu_{i,2} + a_{i,3}\mu_{i,3}|$ represent the indices of the nodes of the ith triangle. As $|a_{i,1}\mu_{i,1} + a_{i,2}\mu_{i,2} + a_{i,3}\mu_{i,3}|$ is a linear combination, we can thus construct the discrete derivative matrix $\mathbf{D}_x$ with the following steps:

- Initialize all-zeros matrix $\mathbf{D}_x$ of size $M \times N$.

- Loop over M triangles; for each triangle i, compute the coefficients a_1, a_2 and a_3 using the coordinates of the three nodes and fill in the three columns in the ith row of matrix $\mathbf{D}_x$ corresponding to the position of the three nodes in the node sequence.

The discrete derivative matrix $\mathbf{D}_y$ can be obtained in a similar way. Note that $\mathbf{D}_x$ and $\mathbf{D}_y$ are sparse matrices as most entries are zeros. With $\mathbf{D}_x$ and $\mathbf{D}_y$ defined, we can therefore minimize the TV regularization (either anisotropic (5.12) or isotropic (5.13) version) with the data fidelity term for DOT reconstruction over 2D unstructured geometries. The corresponding 3D counterparts were also implemented in this work, as shown in the experiments.

5.4.3 TV regularizations under graph representation

With these discrete differential operators defined on graph (Section 5.2), the anisotropic graph TV regularization is

$$\|\nabla_w \mu\|_1 = \sum_{i=1}^{N} \sum_{j \in \mathcal{N}_i} \left| (\mu_j - \mu_i) \sqrt{w_{ij}} \right|, \tag{5.15}$$

and the isotropic graph TV regularization is

$$\|\nabla_w \mu\|_1 = \sum_{i=1}^{N} \sqrt{\sum_{j \in \mathcal{N}_i} (\mu_j - \mu_i)^2 w_{ij}}, \tag{5.16}$$

where $\mathcal{N}_i = \{j \in V : (i,j) \in E\}$. We note that the 2D and 3D implementations of these differential operators are identical, making the resulting minimization processes of equations (5.15) and (5.16) more straightforward than the FE implementation.

5.4.4 Minimization of TV-associated DOT inverse problems

Due to the non-linearity of the data fitting term and the non-differentiability of the TV regularizations, it is non-trivial to minimize a TV-regularized inverse problem. It is harder than minimizing the standard L_1-regularized inverse problem [162] because of the existence of the gradient operator. In this section, we propose an efficient algorithm based on

ADMM to address this, the idea of which is to first linearize the non-linear inverse problem, and afterwards apply ADMM to the resulting linearized problem. The whole process is then iterated until convergence. We note that due to the use of the differential operators in Sections 5.4.3 and 5.4.2, the proposed algorithm can handle unstructured geometries, and also can ease the computational difficulties arising from the non-differentiability and non-linearities in the inverse problem. In equations (5.12), (5.13), (5.15) and (5.16), we have defined four types of TV regularizations using different representations. We then apply them to the regularization term, resulting in four linearized minimization problems in Table 5.2. These four TV-regularized minimization problems are A-FETV, I-FETV, A-GTV and I-GTV, representing anisotropic finite element total variation, isotropic finite element total variation, anisotropic graph total variation and isotropic graph total variation respectively.

Table 5.2: Four TV-regularized minimization problems obtained by applying different TV regularizations.

Name	Formulation
A-FETV	$\delta\mu^* = \arg\min_{\delta\mu} \left\{ \frac{1}{2}\|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \lambda\|\mathbf{D}_x(\delta\mu)\|_1 + \lambda\|\mathbf{D}_y(\delta\mu)\|_1 \right\}$
I-FETV	$\delta\mu^* = \arg\min_{\delta\mu} \left\{ \frac{1}{2}\|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \lambda \sum_{i=1}^{M} \sqrt{\left\|(\mathbf{D}_x(\delta\mu))_i\right\|^2 + \left\|(\mathbf{D}_y(\delta\mu))_i\right\|^2} \right\}$
A-GTV	$\delta\mu^* = \arg\min_{\delta\mu} \left\{ \frac{1}{2}\|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \lambda \sum_{i=1}^{N} \sum_{j\in\mathcal{N}_i} \left\| \left(\delta\mu_j - \delta\mu_i\right) \sqrt{w_{ij}} \right\| \right\}$
I-GTV	$\delta\mu^* = \arg\min_{\delta\mu} \left\{ \frac{1}{2}\|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \lambda \sum_{i=1}^{N} \sqrt{\sum_{j\in\mathcal{N}_i} \left(\delta\mu_j - \delta\mu_i\right)^2 w_{ij}} \right\}$

5.4.4.1 ADMM implementations

We begin with the ADMM implementation for A-FETV. Specifically, auxiliary splitting vectors $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$ are introduced to represent $\mathbf{D}_x(\delta\mu)$ and $\mathbf{D}_y(\delta\mu)$ respectively. Therefore, the A-FETV problem is transformed into the following unconstrained optimization

problem:

$$\delta\mu^n, \boldsymbol{\nu}_x^n, \boldsymbol{\nu}_y^n = \underset{\delta\mu,\boldsymbol{\nu}_x,\boldsymbol{\nu}_y}{\arg\min}\{\frac{1}{2}||\mathbf{J}\delta\mu - \delta\Psi||_2^2 + \lambda||\boldsymbol{\nu}_x||_1 + \lambda||\boldsymbol{\nu}_y||_1 + \frac{\theta}{2}||\boldsymbol{\nu}_x - \mathbf{D}_x(\delta\mu) - \boldsymbol{b}_x^{n-1}||_2^2 + \frac{\theta}{2}||\boldsymbol{\nu}_y - \mathbf{D}_y(\delta\mu) - \boldsymbol{b}_y^{n-1}||_2^2\}, \tag{5.17}$$

where superscript n denotes the n-th ADMM iteration and $\boldsymbol{b}_x$ and $\boldsymbol{b}_y$ are iterative parameters. In order to find the minimizer of equation (5.17), an alternating optimization method is used where equation (5.17) is split into several subproblems with respect to $\delta\mu$, $\boldsymbol{\nu}_x$, $\boldsymbol{\nu}_y$, $\boldsymbol{b}_x$ and $\boldsymbol{b}_y$, each of which can be solved separately.

First the iterative minimization approach requires us to solve the subproblem with respect to μ

$$\delta\mu^n = \underset{\delta\mu}{\arg\min}\{\frac{1}{2}||\mathbf{J}\delta\mu - \delta\Psi||_2^2 + \frac{\theta}{2}||\boldsymbol{\nu}_x^{n-1} - \mathbf{D}_x(\delta\mu) - \boldsymbol{b}_x^{n-1}||_2^2 + \frac{\theta}{2}||\boldsymbol{\nu}_y^{n-1} - \mathbf{D}_y(\delta\mu) - \boldsymbol{b}_y^{n-1}||_2^2\}, \tag{5.18}$$

which has the optimality condition

$$\left(\left(\mathbf{J}^T\mathbf{J} + \theta\left({\mathbf{D}_x}^T\mathbf{D}_x + {\mathbf{D}_y}^T\mathbf{D}_y\right)\right)\right)\delta\mu^n = \mathbf{J}^T\delta\Psi - \theta{\mathbf{D}_x}^T\left(\boldsymbol{b}_x^{n-1} - \boldsymbol{\nu}_x^{n-1}\right) - \theta{\mathbf{D}_y}^T\left(\boldsymbol{b}_y^{n-1} - \boldsymbol{\nu}_y^{n-1}\right). \tag{5.19}$$

As the inversion matrix of equation (5.19) has size $N \times N$, in order to achieve high efficiency, we use the gradient descent method with backtracking line search to optimize the functional iteratively, in which the step size controls how far the iterate moves along the gradient direction during the current iteration [122].

The next subproblem with respect to $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$ is given as

$$\boldsymbol{\nu}_x^n, \boldsymbol{\nu}_y^n = \underset{\boldsymbol{\nu}_x,\boldsymbol{\nu}_y}{\arg\min}\{\lambda||\boldsymbol{\nu}_x||_1 + \lambda||\boldsymbol{\nu}_y||_1 + \frac{\theta}{2}||\boldsymbol{\nu}_x - \mathbf{D}_x(\delta\mu^n) - \boldsymbol{b}_x^{n-1}||_2^2 + \frac{\theta}{2}||\boldsymbol{\nu}_y - \mathbf{D}_y(\delta\mu^n) - \boldsymbol{b}_y^{n-1}||_2^2\}. \tag{5.20}$$

It should be noticed that, in A-FETV, there is no coupling between $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$. We can explicitly compute the optimal value of $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$ using the generalized shrinkage operators

$$\begin{aligned}\boldsymbol{\nu}_x^n &= \max\left(\left|\mathbf{D}_x(\delta\mu^n) + \boldsymbol{b}_x^{n-1}\right| - \frac{\lambda}{\theta}, 0\right)\frac{\mathbf{D}_x(\delta\mu^n) + \boldsymbol{b}_x^{n-1}}{\left|\mathbf{D}_x(\delta\mu^n) + \boldsymbol{b}_x^{n-1}\right|}\\ \boldsymbol{\nu}_y^n &= \max\left(\left|\mathbf{D}_y(\delta\mu^n) + \boldsymbol{b}_y^{n-1}\right| - \frac{\lambda}{\theta}, 0\right)\frac{\mathbf{D}_y(\delta\mu^n) + \boldsymbol{b}_y^{n-1}}{\left|\mathbf{D}_y(\delta\mu^n) + \boldsymbol{b}_y^{n-1}\right|},\end{aligned} \tag{5.21}$$

with the convention that $0/0 = 0$. The last one is to update the iterative parameters $\boldsymbol{b}_x$ and $\boldsymbol{b}_y$, as

$$\begin{aligned} \boldsymbol{b}_x^n &= \boldsymbol{b}_x^{n-1} + \mathbf{D}_x(\delta\mu^n) - \boldsymbol{\nu}_x^n \\ \boldsymbol{b}_y^n &= \boldsymbol{b}_y^{n-1} + \mathbf{D}_y(\delta\mu^n) - \boldsymbol{\nu}_y^n. \end{aligned} \tag{5.22}$$

In I-FETV, using the same alternating optimization method, the original minimization problem can be transformed as

$$\begin{aligned} \delta\mu^n, \boldsymbol{\nu}_x^n, \boldsymbol{\nu}_y^n = \underset{\delta\mu, \boldsymbol{\nu}_x, \boldsymbol{\nu}_y}{\arg\min} \{ \frac{1}{2} ||\mathbf{J}\delta\mu - \delta\Psi||_2^2 + \lambda \left\| (\boldsymbol{\nu}_x, \boldsymbol{\nu}_y) \right\|_2 + \frac{\theta}{2} ||\boldsymbol{\nu}_x - \mathbf{D}_x(\delta\mu) - \boldsymbol{b}_x^{n-1}||_2^2 \\ + \frac{\theta}{2} ||\boldsymbol{\nu}_y - \mathbf{D}_y(\delta\mu) - \boldsymbol{b}_y^{n-1}||_2^2 \}, \end{aligned} \tag{5.23}$$

where

$$\left\| (\boldsymbol{\nu}_x, \boldsymbol{\nu}_y) \right\|_2 = \sum_{i=1}^{M} \sqrt{|(\boldsymbol{\nu}_x)_i|^2 + |(\boldsymbol{\nu}_y)_i|^2}, \tag{5.24}$$

and M is the number of finite elements. The first subproblem (L_2 component) with respect to $\delta\mu$ is the same as A-FETV. It should be noted that the $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$ variables cannot be decoupled as they were in A-FETV. In order to solve the subproblem with respect to $\boldsymbol{\nu}_x$ and $\boldsymbol{\nu}_y$, we can explicitly solve the minimization problem for $\left(\boldsymbol{\nu}_x^n, \boldsymbol{\nu}_y^n\right)$, using a generalized shrinkage formula

$$\begin{aligned} \boldsymbol{\nu}_x^n &= \max\left(s^n - \frac{\lambda}{\theta}, 0\right) \frac{\mathbf{D}_x(\delta\mu^n) + \boldsymbol{b}_x^{n-1}}{s^n} \\ \boldsymbol{\nu}_y^n &= \max\left(s^n - \frac{\lambda}{\theta}, 0\right) \frac{\mathbf{D}_y(\delta\mu^n) + \boldsymbol{b}_y^{n-1}}{s^n}, \end{aligned} \tag{5.25}$$

with the convention that $0/0 = 0$ and $s^n = \sqrt{\left|\mathbf{D}_x(\delta\mu^n) + \boldsymbol{b}_x^{n-1}\right|^2 + \left|\mathbf{D}_y(\delta\mu^n) + \boldsymbol{b}_y^{n-1}\right|^2}$. The iterative parameters $\boldsymbol{b}_x$ and $\boldsymbol{b}_y$ are then updated as shown in A-FETV.

The ADMM-based algorithm for A-FETV and I-FETV is given in Algorithm 1, where *inner_loop* is the maximum number of iterations for the ADMM-based algorithm.

We then propose an ADMM-based algorithm to address the minimizations of A-GTV and I-GTV. For A-GTV, we first introduce an auxiliary splitting vector variable $\boldsymbol{\nu}$, an iterative parameter $\boldsymbol{b}$, and a positive penalty parameter θ. The sizes of $\boldsymbol{\nu}$ and $\boldsymbol{b}$ are both

Algorithm 1: ADMM-based algorithm for A-FETV and I-FETV.

INPUT: $\delta\Psi$, $\mathbf{J}$, y, *inner_loop*, ϵ_1, regularization parameter $\theta > 0$, $\lambda > 0$

Initialization: $\boldsymbol{\nu}_x^0 = \boldsymbol{\nu}_y^0 = \boldsymbol{b}_x^0 = \boldsymbol{b}_y^0 = 0$

for $n = 1 : inner\_loop$

1: Update $\delta\mu^n$ using equation (5.18)

2: Update $\boldsymbol{\nu}_x^n$ and $\boldsymbol{\nu}_y^n$ using equation (5.21) for A-FETV or (5.25) for I-FETV

3: Update $\boldsymbol{b}_x^n$ and $\boldsymbol{b}_y^n$ using equation (5.22)

4: Stop if $n = inner\_loop$ or $\frac{||\delta\mu^n - \delta\mu^{n-1}||_1}{||\delta\mu^{n-1}||_1} \leq \epsilon_1$.

end for

RETURN $\delta\mu^k = \delta\mu^n$

of $N \times N$. The A-GTV problem can be reformulated as the following unconstrained optimization problem

$$\delta\mu^n, \boldsymbol{\nu}^n = \arg\min_{\delta\mu,\boldsymbol{\nu}} \left\{ \frac{1}{2}\|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \lambda \sum_{i=1}^{N} \|\boldsymbol{\nu}_i\|_1 + \frac{\theta}{2} \sum_{i=1}^{N} \|\boldsymbol{\nu}_i - \nabla_w(\delta\mu_i) - \boldsymbol{b}_i^{n-1}\|_2^2 \right\}. \tag{5.26}$$

Since equation (5.26) is a multivariate minimization problem, we first solve the subproblem with respect to $\delta\mu$

$$\delta\mu^n = \arg\min_{\delta\mu} \left\{ \frac{1}{2}\|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \frac{\theta}{2} \sum_{i=1}^{N} \|\boldsymbol{\nu}_i^{n-1} - \nabla_w(\delta\mu_i) - \boldsymbol{b}_i^{n-1}\|_2^2 \right\}, \tag{5.27}$$

which gives the the optimality condition

$$\left(\mathbf{J}^T\mathbf{J}\delta\mu - \mathbf{J}^T\delta\Psi\right)_i + \theta \mathrm{div}_w\left(\boldsymbol{\nu}_i^{n-1} - \nabla_w(\delta\mu_i) - \boldsymbol{b}_i^{n-1}\right) = 0, \quad i = 1, ..., N. \tag{5.28}$$

With the definition of the *nonlocal divergence operator* (Equation (5.3)) and the *nonlocal Laplace operator* (Equation (5.4)), the point-wise equation system (Equation (5.28)) can be equivalently converted to the following matrix-based equation system

$$(\mathbf{J}^T\mathbf{J} - \theta\mathbf{L})\delta\mu = \mathbf{J}^T\delta\Phi - \theta g^{n-1}. \tag{5.29}$$

$\mathbf{L}$ above is the graph Laplacian in matrix form, whose entries are

$$L_{i,j} = \begin{cases} -\sum\limits_{j \in \mathcal{N}_i} w_{ij} & \text{if } i = j \\ w_{ij} & \text{otherwise} \end{cases}.$$

In equation (5.29), the vector $g^{n-1} = \sum\limits_{j \in \mathcal{N}_i} \sqrt{w_{ij}} \left(\nu_{ji}^{n-1} - \nu_{ij}^{n-1} \right) + \sum\limits_{j \in \mathcal{N}_i} \sqrt{w_{ij}} \left(b_{ji}^{n-1} - b_{ij}^{n-1} \right)$. Equation (5.29) is a system of linear equations. The solution $\delta\mu^n$ can be acquired iteratively using the same method in A-FETV. Then we minimize the following subproblem with respect to ν

$$\boldsymbol{\nu}^n = \arg\min_{\boldsymbol{\nu}} \left\{ \lambda \sum_{i=1}^{N} \|\boldsymbol{\nu}_i\|_1 + \frac{\theta}{2} \sum_{i=1}^{N} \|\boldsymbol{\nu}_i - \nabla_w(\delta\mu_i^n) - \boldsymbol{b}_i^{n-1}\|_2^2 \right\}, \tag{5.30}$$

which has an analytical solution, calculated from the generalized shrinkage formula

$$\nu_{ij}^n = \max\left(\left| \sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) + b_{ij}^{n-1} \right| - \frac{\lambda}{\theta}, 0 \right) \frac{\sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) + b_{ij}^{n-1}}{\left| \sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) + b_{ij}^{n-1} \right|}, \tag{5.31}$$

with the convention that $0/0 = 0$. Lastly, we update the iterative parameter $\boldsymbol{b}$ with

$$b_{ij}^n = b_{ij}^{n-1} + \sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) - \nu_{ij}^n. \tag{5.32}$$

We can similarly apply ADMM to the minimization of I-GTV, which can be transformed into the following unconstrained problem with the auxiliary splitting vector variable $\boldsymbol{\nu}$, an augmented Lagrangian multiplier $\boldsymbol{b}$, and a positive penalty parameter θ.

$$\delta\mu^n, \boldsymbol{\nu}^n = \arg\min_{\delta\mu, \boldsymbol{\nu}} \left\{ \frac{1}{2} \|\mathbf{J}\delta\mu - \delta\Psi\|_2^2 + \lambda \sum_{i=1}^{N} \|\boldsymbol{\nu}_i\|_2 + \frac{\theta}{2} \sum_{i=1}^{N} \|\boldsymbol{\nu}_i - \nabla_w(\delta\mu_i) - \boldsymbol{b}_i^{n-1}\|_2^2 \right\}. \tag{5.33}$$

The L_2 subproblem with respect to $\delta\mu$ is the same as the one in A-GTV and can be computed with equation (5.29). We then fix $\delta\mu$ to minimize the second subproblem with respect to $\boldsymbol{\nu}$:

$$\boldsymbol{\nu}^n = \arg\min_{\boldsymbol{\nu}} \left\{ \lambda \sum_{i=1}^{N} \|\boldsymbol{\nu}_i\|_2 + \frac{\theta}{2} \sum_{i=1}^{N} \|\boldsymbol{\nu}_i - \nabla_w(\delta\mu_i^n) - \boldsymbol{b}_i^{n-1}\|_2^2 \right\}, \tag{5.34}$$

which can be solved with the following soft thresholding equation

$$\nu_{ij}^n = \max\left(\sqrt{\sum_{j \in \mathcal{N}_i} \left(\sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) + b_{ij}^{n-1} \right)^2} - \frac{\lambda}{\theta}, 0 \right) \frac{\sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) + b_{ij}^{n-1}}{\sqrt{\sum_{j \in \mathcal{N}_i} \left(\sqrt{w_{ij}} \left(\delta\mu_j^n - \delta\mu_i^n \right) + b_{ij}^{n-1} \right)^2}}, \tag{5.35}$$

with the convention that $0/0 = 0$. The update of iterative parameter $\boldsymbol{b}$ is the same as for A-GTV, as shown in equation (5.32). The ADMM-based algorithm for A-GTV and I-GTV is given in Algorithm 2.

Algorithm 2: ADMM-based algorithm for I-GTV and A-GTV.

INPUT: $\delta\Psi$, $\mathbf{J}$, y, *inner_loop*, ϵ_1, regularization parameter $\theta > 0$, $\lambda > 0$

Initialization: $\boldsymbol{\nu}^0 = \boldsymbol{b}^0 = 0$

for $n = 1 : inner\_loop$

1: Update $\delta\mu^n$ using equation (5.29)

2: Update $\boldsymbol{\nu}^n$ using equation (5.31) for A-GTV or equation (5.35) for I-GTV

3: Update $\boldsymbol{b}^n$ using equation (5.32)

4: Stop if $n = inner\_loop$ or $\frac{||\delta\mu^n - \delta\mu^{n-1}||_1}{||\delta\mu^{n-1}||_1} \leq \epsilon_1$.

RETURN $\delta\mu^k = \delta\mu^n$

Therefore, the whole procedure for minimizing the TV-regularized inverse problem is given in Algorithm 3, in which *outer_loop* represents the maximum number of iterations required for the DOT reconstruction.

Algorithm 3: Algorithm for minimizing the TV-associated inverse problem.

INPUT: y, $\Psi(\cdot)$, μ^0, *outer_loop*, ϵ_2

for $k = 1 : outer\_loop$

1: Compute $\Psi\left(\mu^{k-1}\right)$ and $\mathbf{J}^{k-1}$

2: Set $\delta\Psi^{k-1} = \Psi^{\mathrm{M}} - \Psi\left(\mu^{k-1}\right)$

3: Compute $\delta\mu^k$ by introducing $\delta\Psi^{k-1}$ and $\mathbf{J}^{k-1}$ to one of **Algorithm 1-2**

4: Update $\mu^k = \delta\mu^k + \mu^{k-1}$

5: Stop if $k = outer\_loop$ or $\frac{||\Psi(\mu^k) - y||_2^2 - ||\Psi(\mu^{k-1}) - y||_2^2}{||\Psi(\mu^{k-1}) - y||_2^2} \leq \epsilon_2$.

end for

RETURN μ^k

5.4.5 Experiments

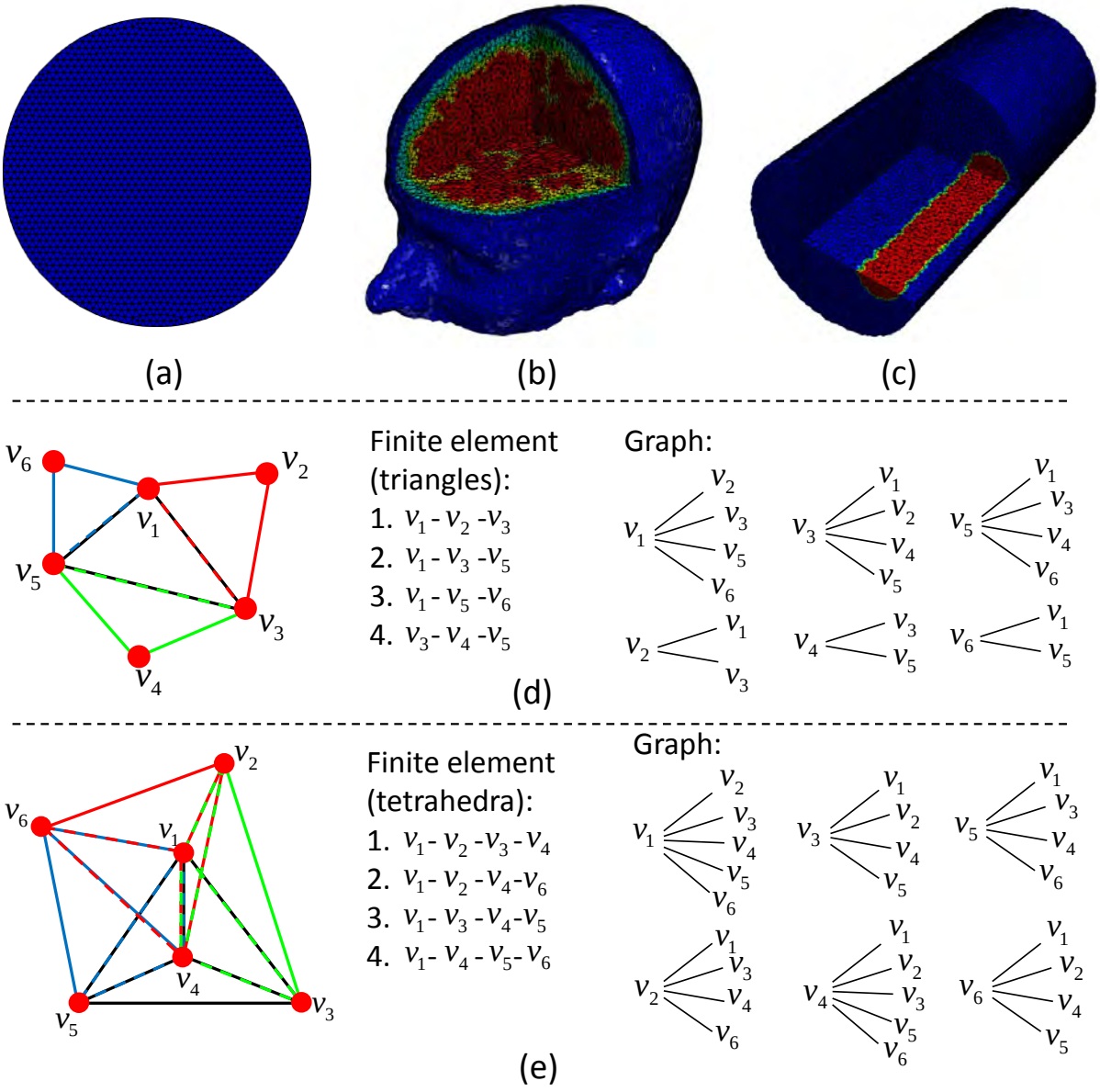


Figure 5.4: (a)-(c): Discretized computational domain of the three experimental samples; (d): Detailed mesh composition of 2D geometry in finite element and graph representation respectively; (e): Detailed mesh composition of 3D geometry in finite element and graph representation respectively.

In this section, we describe extensive experiments to qualitatively and quantitatively evaluate the performance of finite element and graph representations on the two variants of TV regularization in DOT. We use the finite element representation for the forward modelling in all the experiments and use both the finite element and graph representa-

tions to discretize the TV regularization term during the solution of the inverse problem. We first define four evaluation metrics to quantify the quality of the reconstructed images. Then we describe simulated numerical experiments on 2D circle and 3D head samples, and real experiments performed on phantom samples. Figure. 5.4 shows the unstructured grids of the three computational domains. Red dots represent the vertices in the computational domain. In 2D, using the finite element representation, the computational domain is discretized with a finite number of triangles (Figure. 5.4 (d)) while in 3D, tetrahedra are taken as the basic element (Figure. 5.4 (e)). However the graph representation is the same in both 2D and 3D because the graph method requires only vertices and edges of the mesh. For simulated experiments in which measurement noise was added, ten repeats were performed. For all experiments in this section, the forward model was implemented using the NIRFAST package [110] in Matlab R2017a (Mathworks, Natick, USA). The simulated experiments conducted are all based on single wavelength continuous-wave (CW) measurements where the optical property to be recovered is the tissue absorption coefficient μ_a at that wavelength. We set *inner_loop* to 100, *outer_loop* to 40, threshold value ϵ_1 to 0.001 and ϵ_2 to 0.02 for all experiments in this work.

5.4.5.1 Quantitative evaluation metrics

Four evaluation metrics are used to perform the quantitative evaluation: the localization error (as defined in section 5.3), average contrast (as defined in equation (4.37)), peak signal-to-noise ratio (PSNR) (as defined in equation (4.39)) and relative recovered volume. The relative recovered volume is formulated as $\mathrm{V}_{RRV} = \mathrm{V}_r/\mathrm{V}_s \times 100\%$ where V_r and V_s denote the volume of the recovered activation region and simulated activation region, respectively. If the reconstructed image is identical to the ground truth image, the localization error is 0, average contrast and relative recovered volume are both equal to 1. PSNR is higher if the reconstructed image is closer to the ground truth image.

5.4.5.2 Experiments on anisotropic TV regularization

Anisotropic TV regularization is easy to implement because the partial derivatives along different directions can be decoupled as explained in Section 5.4.4.1. It is based on the assumption that the shape of the inhomogeneity that matters is aligned with the coordinate axes. Its minimization favors horizontal and vertical structures, because oblique structures cause the TV regularization to increase [163]. In DOT, this assumption does not necessarily hold as the region of interest is normally random and structures are not normally aligned with the coordinate system. Therefore, anisotropic TV regularization seems to be a poor choice for discrete TV in DOT, as it yields 'blocky' artefacts. However no research has been carried out about the relationship between the 'blocky' artefacts and the representation employed to discretize over the unstructured computational domain. In this section we investigate the anisotropic TV regularization in DOT reconstruction and compare their FE- and graph-based implementations. The effect of the representation method adopted on anisotropic TV regularization will be evaluated.

A 2D circular geometry is simulated with one anomaly centered at (-10mm,10mm). The 2D model has a radius of 43mm while the radius of the anomaly is 10mm. Sixteen source-detector fibres are placed equidistant around the external boundary for CW boundary data acquisition. When one fibre as a source is turned on, the rest are used as detectors, leading to 240 total boundary data points per wavelength. All sources were positioned one scattering distance within the outer boundary because the source is assumed to be spherically isotropic. In order to evaluate the effect of mesh resolution on the representation method, two reconstruction meshes are created with different spatial resolutions. The coarser mesh has 1785 nodes and 3418 linear triangle elements with the average element size 1.6977mm^2 (Figure. 5.5 (a)) while the finer one has 5133 nodes and 10013 elements with the average element size 0.5801mm^2 (Figure. 5.5 (d)). The background absorption coefficient μ_a is set as 0.01mm^{-1} and μ_a for the anomaly is set as 0.03mm^{-1} (Figure. 5.5 (b) and (e)). μ_s remains constant as 1mm^{-1}. To represent various realistic cases, normally distributed randomly generated Gaussian noise ranging from 0%

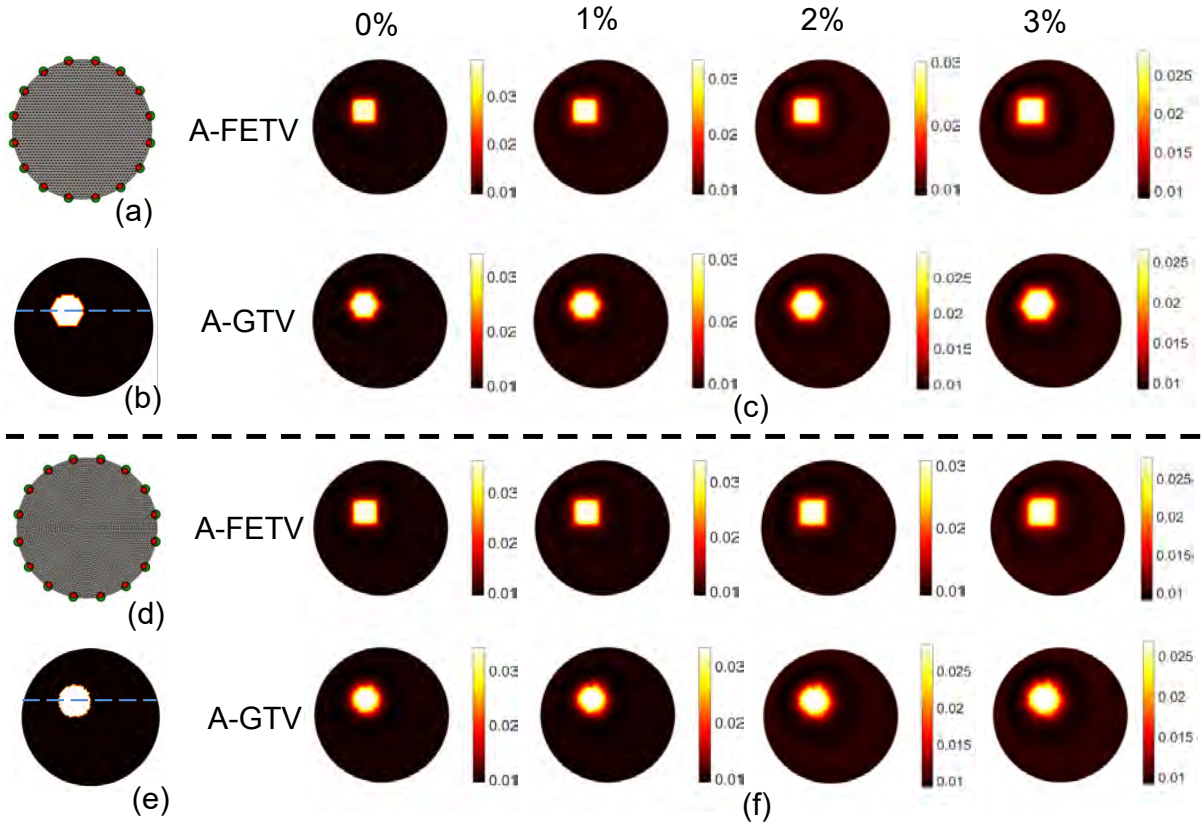


Figure 5.5: (a)-(c): Reconstruction on the 2D mesh with low spatial resolution. (d)-(f): Reconstruction on the 2D mesh with high spatial resolution. (a) and (d): 2D reconstruction mesh with sixteen co-located sources and detectors. (b) and (e) give the original target distributions. First row in (c) and (f) represents the results using A-FETV on 0% , 1% , 2% and 3% noisy data while the second row shows the results using A-GTV.

to 3% at 1% intervals was added to the boundary measurements. Reconstructed images of the absorption coefficient are shown in Figure. 5.5 (c) and (f).

5.4.5.3 Experiments on isotropic TV regularization

5.4.5.3.1 Two dimensional circular experiments Using the same reconstruction meshes described in Section 5.4.5.2, we compare I-FETV, I-GTV against a baseline Tikhonov model. To represent various realistic cases, normally distributed randomly generated noise ranging from 0% to 3% at 1% intervals was added to the amplitude of the boundary data. Reconstructed images of absorption coefficient are shown in Figure. 5.6 (c) and (f). The 1D cross sections and evaluation metrics comparisons are displayed in

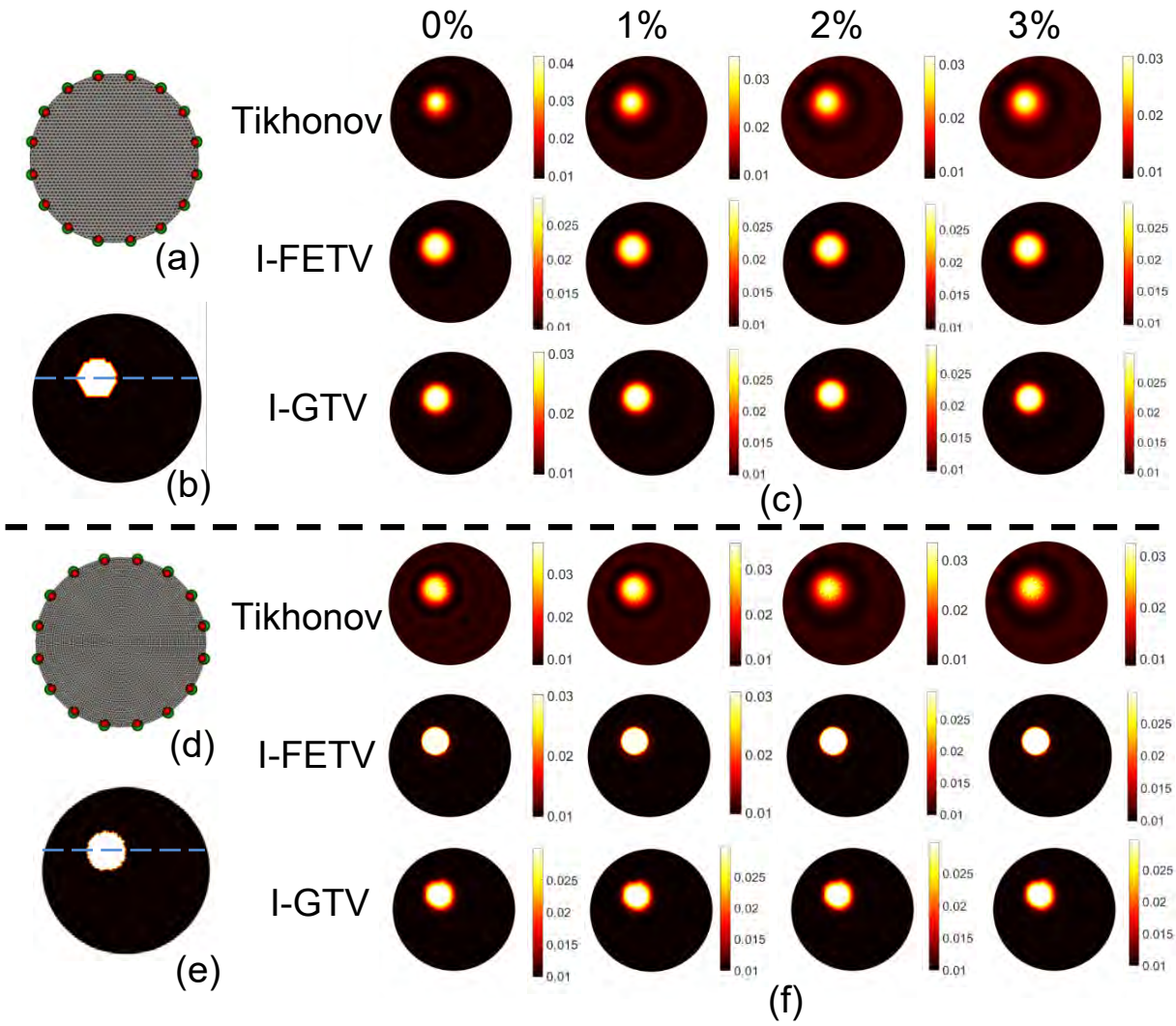


Figure 5.6: (a)-(c): Reconstruction on the 2D mesh with low spatial resolution; (d)-(f): Reconstruction on the 2D mesh with high spatial resolution. (a) and (d): 2D reconstruction mesh with sixteen co-located sources and detectors. (b) and (e) give the original target distributions. First row in (c) and (f) represents the results using I-FETV on 0% , 1% , 2% and 3% noisy data while the second row shows the results by I-GTV.

Figure. 5.7 and Figure. 5.8.

5.4.5.3.2 Three dimensional head numerical experiments We now evaluate the isotropic TV model with two discrete differential operator definitions on the physically realistic 3D head model which is described in Section 4.2.5.4. The reconstruction mesh consists of 50721 nodes associated with 287547 tetrahedral elements, with the average element size 9.2676mm^3. Each node is labeled by one of the five segmented head tissue

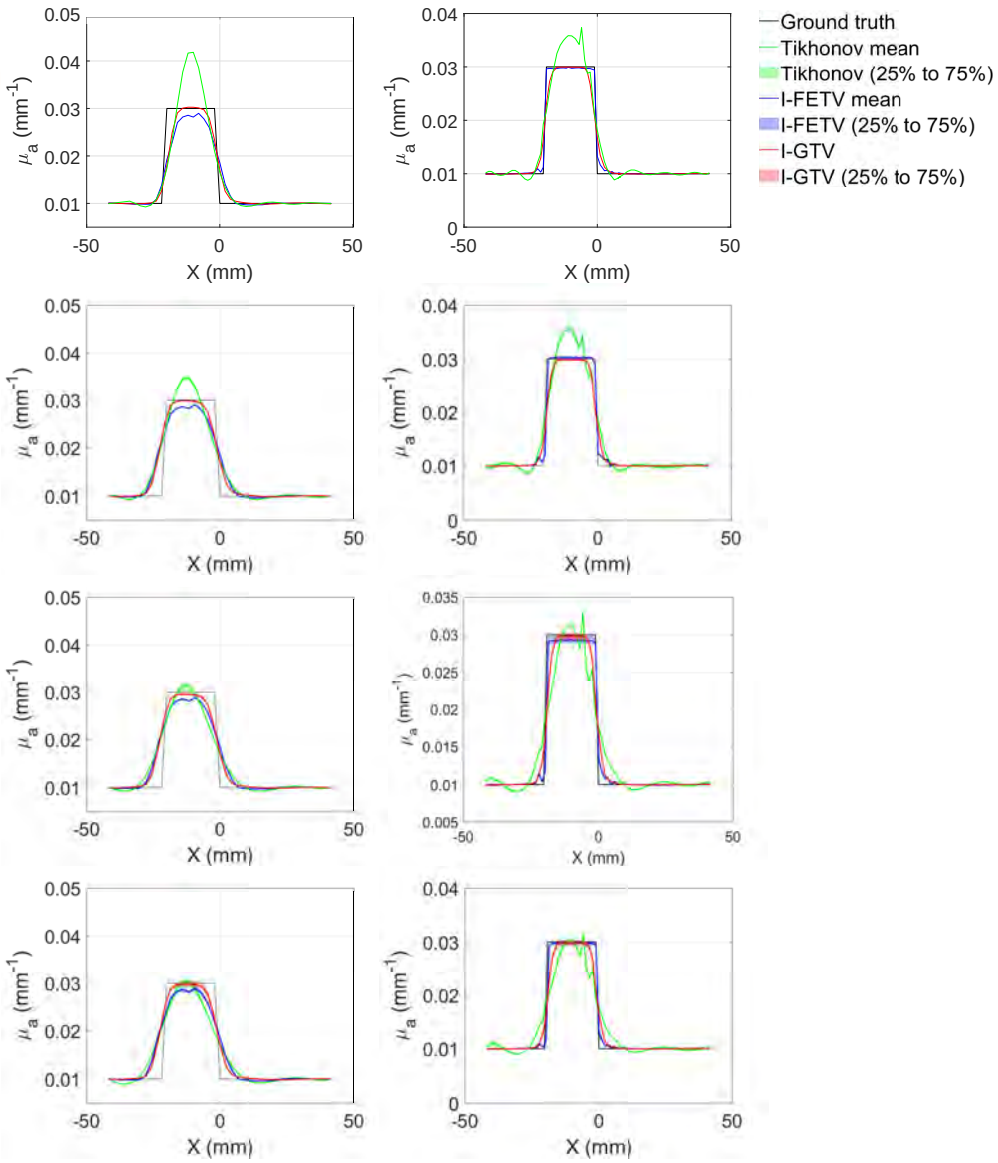


Figure 5.7: 1D cross section of images recovered in Figure. 5.6. First column corresponds to Figure. 5.6 (c) where the spatial resolution of the reconstruction mesh is lower. Second column corresponds to Figure. 5.6 (f) where the spatial resolution of the reconstruction mesh is higher. Top to bottom row: 0%, 1%, 2% and 3% added Gaussian noise.

types. Absorption coefficients assigned to each layer are from an in vivo study [6] at 750nm (Table 4.2).

A high-density (HD) imaging array with 158 sources and 166 detectors (Figure. 5.9 first column) [73] was placed over the whole head, with source-detector (SD) separation distances ranging from 1.3 to 4.8cm. 3478 differential measurements per wavelength were used to image hemodynamic changes in the brain. Two distinct anomalies were simulated

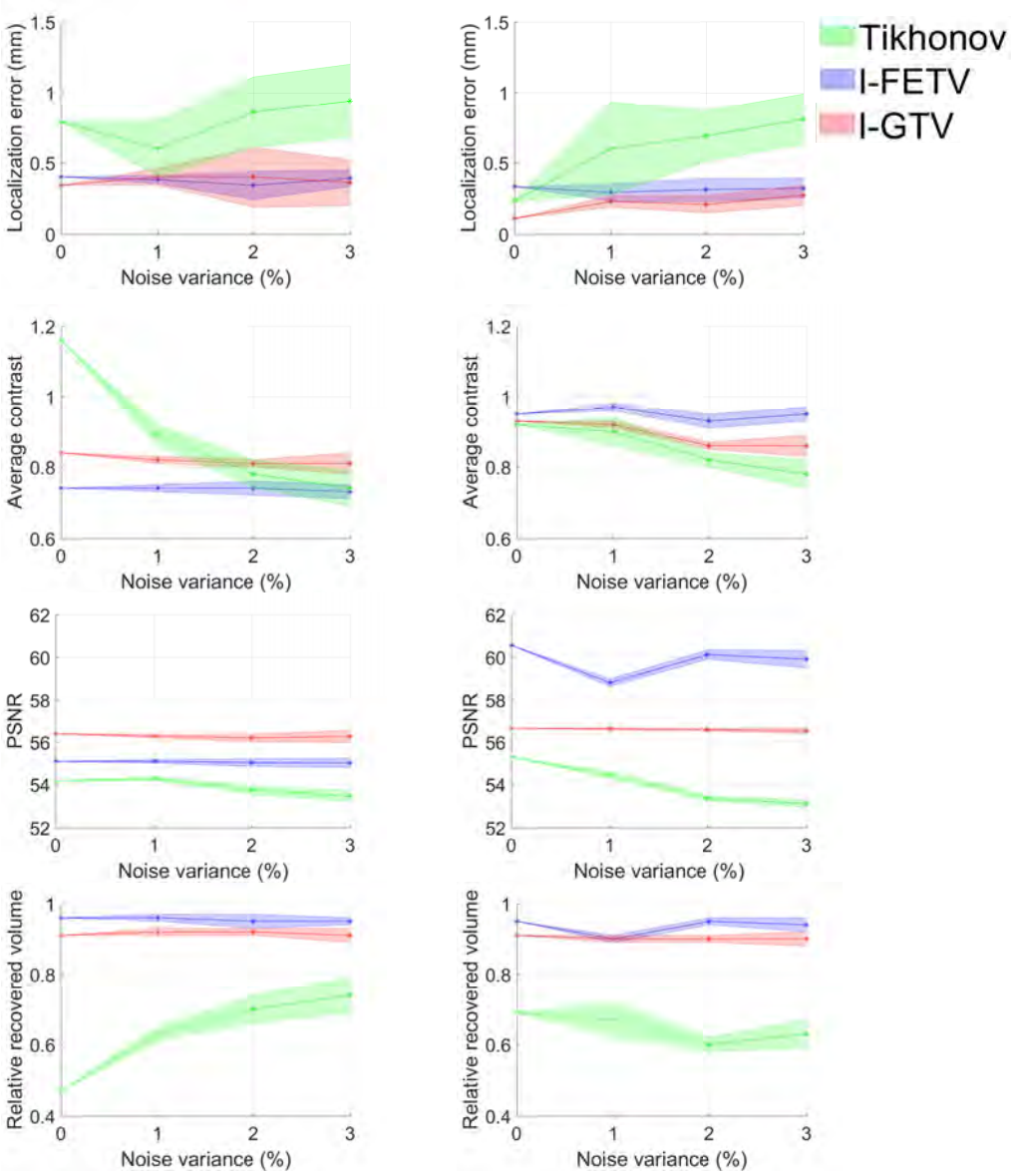


Figure 5.8: Evaluation metrics comparing the performance of different methods at four different noise levels. Top to bottom row: localization error index; average contrast index; PSNR index and relative recovered volume. Left column corresponds to the reconstructions in Figure. 5.6 (c) where the reconsturction mesh resolution is low. Right column corresponds to Figure. 5.6 (f) where the reconsturction mesh resolution is relatively high.

simultaneously in the brain, with each 15mm radius. In order to simulate a traumatic brain injury (TBI) case where tissue oxygen saturation (StO_2) is normally between 50% and 75% [164, 165], the absorption coefficient in the two anomalies are calculated using Beer's law [110] with 55% StO_2 (Figure. 5.9 second column). In line with the expected in vivo performance of the imaging system, 0.12%, 0.15%, 0.41% and 1.42% Gaussian

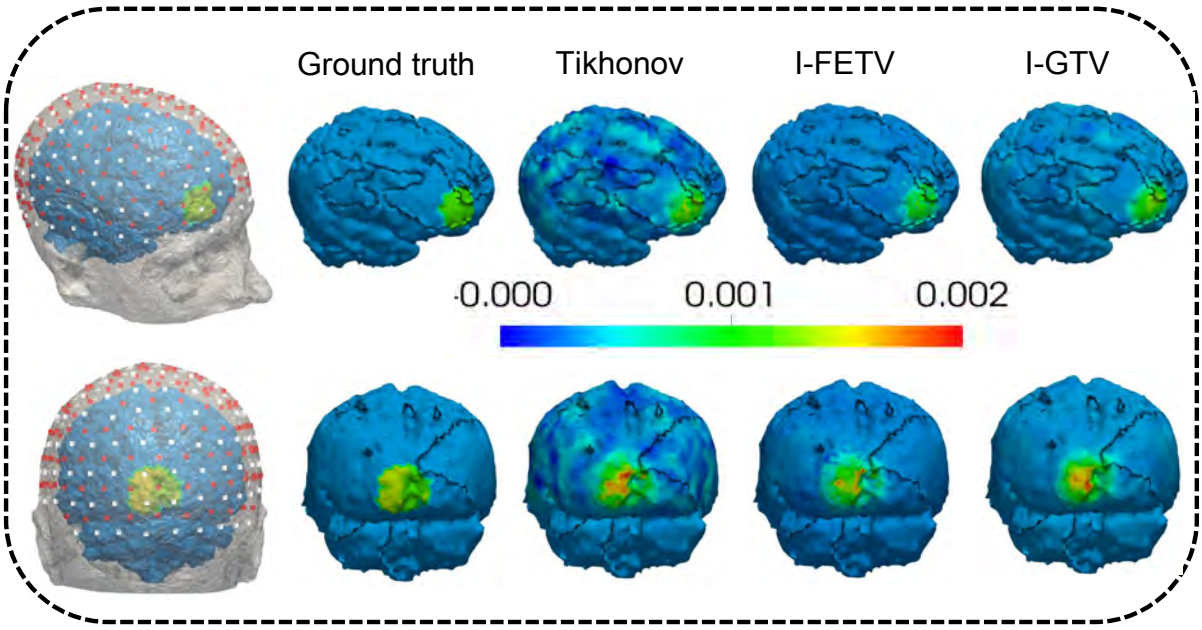


Figure 5.9: First column: distribution of the imaging array with 158 sources (red dots) and 166 detectors (white dots) and the positions of the two simultaneous simulated anomalies. Second to final column: Ground truth and reconstructions by Tikhonov, I-FETV and I-GTV

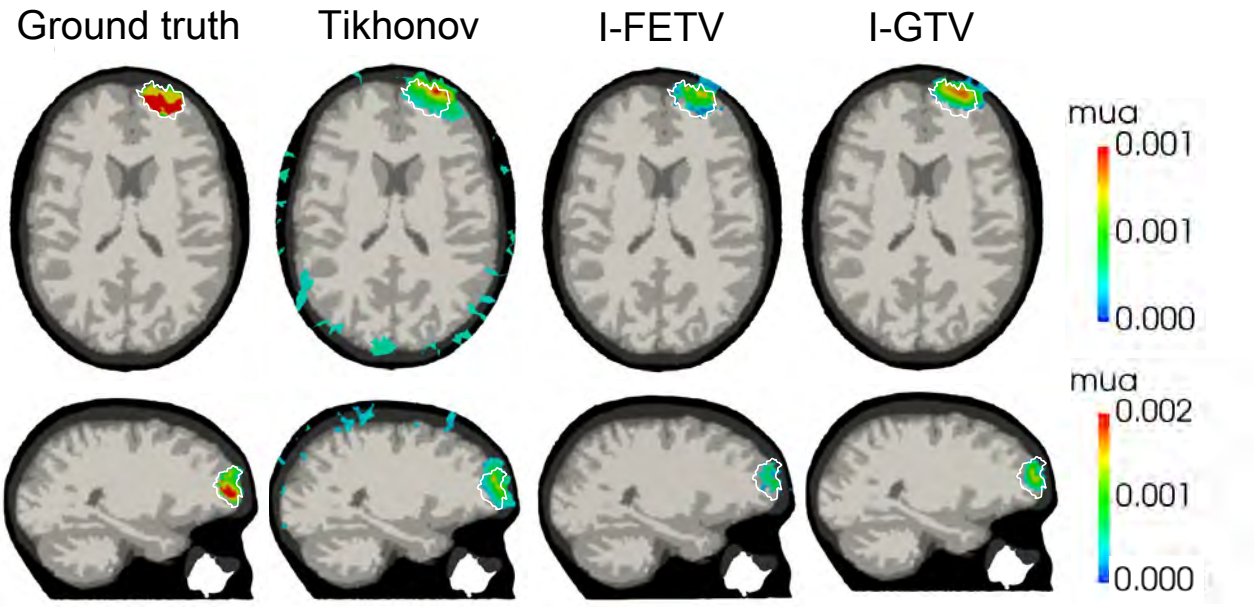


Figure 5.10: 2D slices of the reconstructions of the absorption coefficient changes on the forehead anomaly (first row in Figure. 5.9). The ground truth areas are highlighted in white boundary.

random noise was added to first (13mm), second (30mm), third (40mm) and fourth (48mm) nearest neighbor measurements to provide realistic data [138]. Reconstructed

absorption coefficients using different model are displayed in the third to fifth column of Figure. 5.9. Corresponding 2D slices are displayed in Figures. 5.10 and 5.11 and the

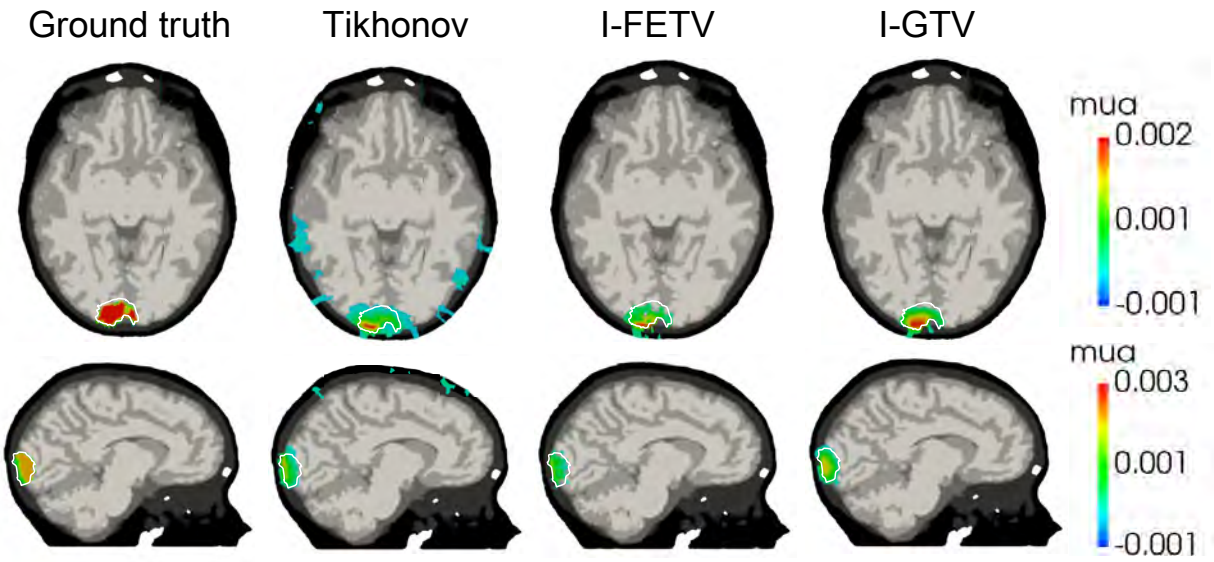


Figure 5.11: 2D slices of the reconstructions of the absorption coefficient changes on the back-head anomaly (second row in Figure. 5.9). The ground truth areas are highlighted in white boundary.

5.4.5.3.3 Experiments with phantom data In the final experiment we evaluate different methods on real experimental data which is collected from the solid plastic cylindrical phantom described in Section 4.2.5.6. The reconstruction mesh consists of 9082 nodes and 48099 linear tetrahedral elements with the average tetrahedral elements size 0.4218mm^3. Ground truth data and images reconstructed with Tikhonov, I-FETV and I-GTV are shown in Figure. 5.13(c). The four evaluation metrics in the volume of illumination are given in Table 5.3. For all the experiments above, the regularization parameter λ is determined using an L-curve method [162].

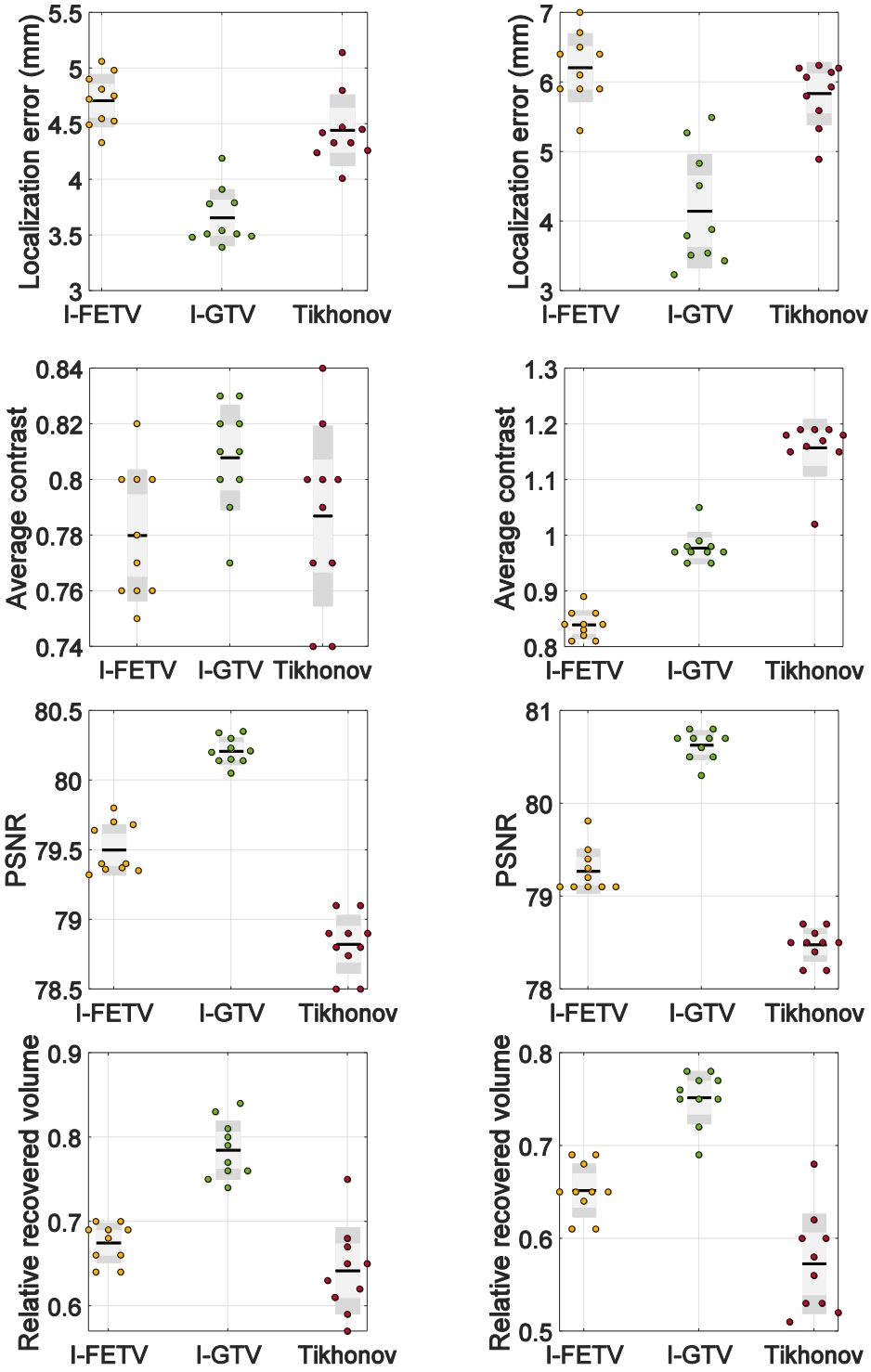


Figure 5.12: Evaluation metrics comparing the performance of different methods on a 3D head model. The left column represents the reconstruction of the forehead anomaly (first row in Figure. 5.9), while the right column gives the reconstruction of the back-head anomaly (second row in Figure. 5.9).

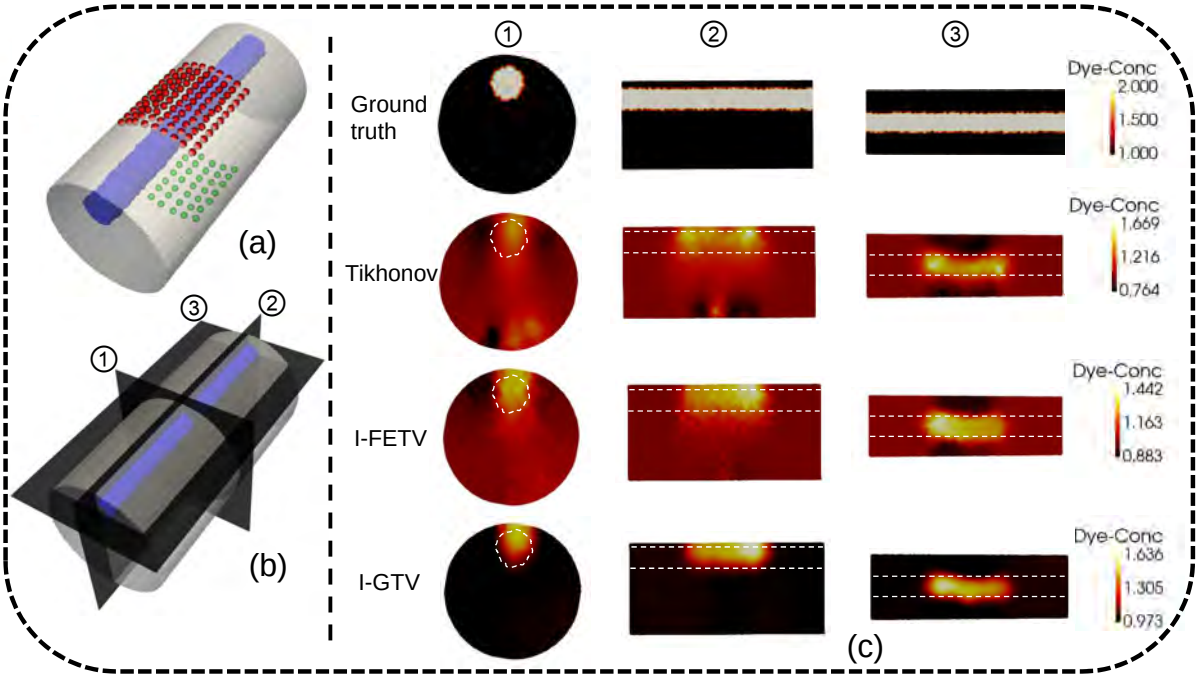


Figure 5.13: (a): Distribution of sources and detectors. (b): Illustration of the overall distribution of three slices. (c): Ground truth and reconstruction results with different methods. From top to bottom: ground truth; results with Tikhonov regularization; results with I-FETV regularization and results with I-GTV regularization.

Table 5.3: Evaluation of different methods for reconstruction on a tissue-simulating phantom. LE presents localization error. AC presents average contrast. RRV represents relative recovered volume.

	LE / mm	AC / -	PSNR / -	RRV / %
Tikhonov	2.90	0.74	13.74	40
I-FETV	2.81	0.69	14.77	48
I-GTV	3.16	0.79	16.71	46

5.4.6 Discussion

The 2D images reconstructed using A-FETV and A-GTV are shown in Figure. 5.5, together with the original target distributions. The results show that A-FETV reconstructs the target with boundaries that align with the coordinate axes which is consistent with the assumption of the anisotropic TV regularization. In addition, some artefacts are

observed inside the recovered region of interest when the mesh resolution is low. The reconstructions using A-GTV do not feature these artefacts, and the recovered shape is more accurate, with no bias towards the coordinate axes. The experiments reveal that the blocky artefacts of anisotropic TV regularization are associated with the discretization method used. The blocky artefacts are clearly visible in reconstructions based on the finite element representation, but not in the ones based on the graph representation. This is because in the graph representation, the region is discretized along all edge-based directions, leading to nearly isotropic solutions. Therefore, in DOT, A-GTV can adapt to the ground truth solution. However, it is not a good method to preserve anisotropy if anisotropy is a desired property of the solution.

For the 2D case which uses isotropic TV regularization, Figure. 5.6, Tikhonov reconstruction over-smooths the results and smears the edges. The results become smoother with increases in measurement noise. Little difference can be visually observed between the reconstruction by I-FETV and I-GTV when the reconstruction mesh resolution is low. However when the reconstruction mesh has higher resolution, Figure. 5.6 (f), the results by I-FETV is visually closer to the ground truth than the ones by I-GTV. Similar findings are observed from the corresponding 1D cross sections (Figure. 5.7). Tikhonov reconstruction produces a single peaked distribution in the piecewise constant target area, and edges of the objects are over-smoothed. Both TV methods are able to reconstruct a piecewise constant distribution. However when the mesh resolution is lower (first column in Figure. 5.7), fluctuations in the target regions are observed in the results by I-FETV. When the mesh resolution is higher (second column in Figure. 5.7), the cross-section from I-FETV reconstruction is almost identical to the ground truth. In Figure. 5.8, red and blue areas represent 25% to 75% value among the ten repeats' experiments. We see that the performance of I-FETV improves with an increase of mesh resolution: by 25% in localization error, 26% in average contrast and 11% in PSNR, while the performance of I-GTV improves 20% in localization error, 11% in average contrast and 2% in PSNR. These 2D experiments confirm that comparing with the performance of I-FETV, the one

of I-GTV has relatively fewer difference with different mesh resolution and the graph representation is more stable and robust to changes in mesh resolution.

The 3D images reconstructed from the head geometry represent a physically realistic case, in which two anomalies are simulated simultaneously in the brain. From Figure. 5.9, Tikhonov reconstruction lead to many visible artefacts near the source and detector area. Due to smoothing induced by Tikhonov regularization, sharp features are not present in the image recovered. I-FETV and I-GTV both can eliminate the surface artefacts resulting from Tikhonov regularization and reconstruct tightly localized results. These findings can be clearly observed in the 2D slice images shown in Figure. 5.10 and Figure. 5.11. It should be noticed that, in Figure. 5.10, the colorbar values corresponding to the green and red parts remain 0.001. It is because only three digits are selected after the radix point and in this study we use rounding off to constrain the three digits. From the visualization of the results, there is no obvious difference between the reconstruction performance of I-FETV and I-GTV because both are based on TV regularization. However it can be observed from the evaluation metrics comparison in Figure. 5.12 that I-GTV achieves lower localization error (4mm), higher peak signal-to-noise ratio (79.5) and average contrast (0.8) much closer to 1. The average relative recovered volume achieved by I-GTV is 77%, compared with I-FETV (66%) and Tikhonov (64%). The underestimation of the magnitude of localized changes result from the partial volume effect, wherein the tissue inhomogeneity occupies only a small portion of the optically interrogated volume.

In the experiments with phantom data, only the central region is reconstructed in all the cases because the positions of sources and detectors lead to very low sensitivity away from the centre. It can be seen from the second row of Figure. 5.13(c), that Tikhonov regularization over-smooths the reconstructed images which have much lower image contrast than the ground truth, especially in the first slice image. Artefacts are clearly observed near the source and detector areas. Even though total variation regularization can alleviate the over-smoothing effect caused by Tikhonov regularization, discretization methods still play an important role in the reconstruction performance. It should be

noticed that I-FETV can alleviate the artefacts near to the sources and detectors but introduce some artefacts (staircase effect) in the non-anomaly area and does not preserve edges. I-GTV is seen to recover the anomaly with clear edges and high image contrast. It is interesting to compare these results to those of our previous work [162] where L_1-norm regularization was applied to the phantom data. Reconstructions by L_1-norm regularization were found to be over-sparsified and over-compact. In this work, TV regularization, which induces sparsity to the gradient of the solution, is seen to effectively alleviate the over-sparsifying effect of L_1-norm regularization and is therefore suitable for non-sparse coefficient distributions. We calculate the four evaluation metrics in the volume of illumination (Table 5.3) and these support the same conclusions. Similar localization errors are obtained by the different methods with only 1mm difference. Comparing to I-FETV, I-GTV can obtain the highest average contrast and PSNR values with similar relative recovered volume.

5.4.7 Conclusion

In this work, we introduce finite element and graph representations to discretize the TV regularization term in DOT reconstruction. Isotropic and anisotropic variants of the TV regularization are also investigated and compared between their FE- and graph-based implementations. The ADMM-based algorithms are proposed for each TV-regularized inverse problem. Experiments on the anisotropic TV regularization reveal that finite element representation yields the 'blocky' artefacts which is the designed in feature in the anisotropic TV regularization. However the graph representation favors the underlying shape of the region of interest so that the 'blocky' artefacts are not realized. Graph discretization on anisotropic TV regularization can adapt to the ground truth solution, but is not a good way to preserve anisotropy.

Numerical experiments on isotropic TV regularization illustrate that, comparing to Tikhonov regularization, TV regularization can alleviate the over-smoothing effect of Tikhonov regularization and localize the anomaly by inducing sparsity of the gradient

of the solution. These findings were tested on real experimental data using a tissue-simulating phantom. I-FETV does not perform well on low resolution reconstruction meshes because of the discrete nature of the finite element representation. Because the finite element representation operates on each element, the discretization becomes more accurate when the mesh resolution increases. I-GTV is shown to be more stable and robust to changes in mesh resolution because I-GTV is discretized on the graph directly, having no information of elements. Hence I-GTV can give more accurate discretization when the reconstruction mesh is a coarse mesh which is the usual case in DOT. However, I-FETV will outperform I-GTV when an reconstruction mesh with high resolution is available.

CHAPTER 6

GRAPH-BASED FORWARD MODEL FOR CONTINUOUS-WAVE DIFFUSE OPTICAL TOMOGRAPHY

In DOT, the spatial distribution of light remitted from the object's surface is measured for each source fibre, and this information is used to estimate the object's internal optical properties by iteratively refining the optical properties of a forward model of light propagation in the object until the model predictions match the measured surface remittance. As such, the forward model of light propagation must be able to accurately model the main interactions (i.e. absorption and scattering) between light and the object so as to recover internal properties faithfully.

In this chapter, the graph-based methods previously applied to the inverse model are developed to model the forward problem of modelling light transport for continuous-wave(CW) DOT.

The contents of this chapter were the subject of the following publication.

- **Wenqi Lu, Jinming Duan, Joshua Deepak Veesa, and Iain B Styles, *New nonlocal forward model for diffuse optical tomography*, Biomedical Optics Express 10, 6227-6241 (2019)**

6.1 Introduction

As we mentioned in Section 2.2, when its wave nature is neglected and light is interpreted as a stream of particles (photons), the main interactions between light and biological tissue are characterized as absorption and scattering and are modelled by the radiative transfer equation (RTE) which is generally accepted to accurately describe how light propagates in biological tissues [31, 34]. Analytical solutions exist for the RTE only for simple geometries with nearly homogeneous interior structure [41, 43]. Although a number of algorithms exist to find numerical solutions for more complex inhomogeneous domains [44, 47], they are extremely computationally expensive, especially for large 3D volumes. Therefore, a range of stochastic or deterministic approximation schemes are frequently adopted to simplify the RTE. The Monte Carlo method is the most commonly used stochastic model [48, 54]. It is however costly in computational time, because millions of photons need to be tracked to acquire meaningful statistics. Light propagation can also be modelled by a deterministic diffusion equation (DE) using the P_1 approximation of the RTE. It is based on the assumption that the radiance in an optical medium in which multiple scattering occurs is almost isotropic, and scattering in that medium is dominant over absorption. The modified Beer–Lambert law (MBLL) is also sometimes used to model thick tissues [166, 167]. However Boas et al. [168] observed that standard MBLL analysis cannot accurately quantify relative changes in the concentration of chromophores. Bhatt et al. [169] proposed a generalized Beer–Lambert model to overcome this limitation and have applied this method widely to near-infrared spectroscopy (NIRS) studies. In DOT, technically, such interactions can be accurately described by a diffusion equation (DE) which is derived from the radiative transfer equation (RTE) [31] under the assumption that the radiance in an optical medium is almost isotropic, and that the scattering interactions dominate over absorption [170]. The DE for a continuous wave (CW) imaging system is given in equation (2.3) with $w = 0$. The corresponding Robin boundary condition is described in equation (2.5).

Mathematically, Equation (2.3) is an elliptic partial differential equations, the differential operators (i.e. gradient or divergence) in which are defined using the classical vector calculus. A general approach to analytically solve the DE (with its RBC) is to apply the Green function, but analytical solutions are only known for homogeneous objects [65, 171, 172]. Numerical techniques are required if more complex geometries are modelled. When Equation (2.3) with its RBC (2.5) are discretized for a numerical solution, a computational mesh must be established, which should essentially describe the geometry of the object modelled. Such generated mesh can be structured (regular) or unstructured (irregular). For a structured mesh where each element represents one pixel in 2D or voxel in 3D, finite difference method (FDM) can be applied [68, 134] to discretize the computational domain. For an unstructured mesh, finite element method (FEM) is a more natural choice [15, 65, 67, 70, 173, 174]. In this discretization, the computational domain Ω is divided into a series of elements (triangles in 2D, tetrahedra in 3D). However, FEM implementations can be difficult and time-consuming, especially when higher-order polynomial basis (shape) functions are used for non-linear interpolation between vertices of high-order elements [24]. Other numerical algorithms such as finite volume method (FVM) [175] and boundary element method (BEM) [176, 177] have also been employed in DOT, where only some specialized applications are observed. In Chapter 5, we introduced a graph representation to discretize a total-variation regularization term for the inverse problem in DOT where the object geometry is represented by an unstructured graph, defined by vertices, edges and weights. The graph was constructed by exploring neighborhood relationships between vertices.

In order to fully leverage the power of graph-based discretization, one must use the nonlocal vector calculus. In the classical local vector calculus, the differential operators are numerically evaluated using purely local information. In the nonlocal calculus, the operators include more pixel information in the domain. For example, in image processing, some well-known PDEs and variational techniques such as nonlocal image denoising [142, 144] and inpainting [121, 145] have explored the advantages of nonlocal vector

calculus [142, 143]. When applied to these problems, local operators include information from only neighbouring pixels whilst nonlocal methods include information from a wider area and are naturally formulated in a graph-based representation instead of in terms of the classical local difference operators.

In image processing, nonlocal methods are shown to have several advantages over local methods, including preservation of important image features such as texture and ability to handle unstructured geometries. It has also been observed that many PDE-based physical processes, minimizations and computational methods, such as CT image processing and reconstruction [178], can be generalized to be nonlocal. Therefore we expect that such a framework may be useful for the physical modelling in DOT.

As such, we propose a nonlocal diffusion equation (NDE) as a new forward model for DOT. The concept of differential operators under the nonlocal vector calculus [121, 142, 143, 145] is used to formulate a new forward model that can accurately simulate light propagation in turbid media. The discretization for the NDE is performed using a graph-based numerical method (GNM). As a result, the proposed method naturally applies without modification to complex, unstructured DOT geometries in both two and three dimensions. The accuracy of the proposed model is compared against the conventional diffusion equation implemented by FEM and to the existing analytical solution on both homogeneous and heterogeneous slabs. We also compare the image reconstruction accuracy of different forward models on a 2D circular model and a 3D human head model. It should be noted that the diffusion equation is also used to model light propagation in imaging techniques such as diffuse correlation spectroscopy and near infrared spectroscopy, and our results can be applied to any technique that uses a diffusion-based model of light propagation.

6.2 Methodology

Our approach is based on reformulating the diffusion equation (Equation (2.3)) in terms of nonlocal differential operators. We denote $\nabla_w(\cdot)$, $\mathrm{div}_w(\cdot)$ and $\mathcal{N}_w(\cdot)$ as the nonlocal

gradient, the nonlocal divergence and the nonlocal normal derivative, respectively. Their definitions are given in equations (5.2), (5.3) and (6.3). In the computational domain Ω with boundary surface Γ and internal domain Ω' (i.e. $\Omega{=}\Omega'{\cup}\Gamma$ and $\Omega'{\cap}\Gamma = \emptyset$), we simply replace the differential operators in equation (2.3) with their nonlocal counterparts and solve the new NDE under the framework of nonlocal vector calculus:

$$-\operatorname{div}_w\left(\kappa\left(r\right)\nabla_w\Phi\left(r\right)\right)+\mu_a\left(x\right)\Phi\left(r\right)=q_0\left(r\right)\quad\text{for } r\in\Omega'. \tag{6.1}$$

Similarly, we reformulate the RBC with the nonlocal normal derivative and the nonlocal gradient to give a nonlocal boundary condition (NBC):

$$2A\mathcal{N}_w\left(\kappa\left(r\right)\nabla_w\Phi\left(r\right)\right)+\Phi\left(r\right)=0\ \text{ for } r\in\Gamma. \tag{6.2}$$

We now formulate a graph-based numerical method to discretize the NDE with its NBC. Following the graph method introduced in Section 5.2, the computational domain Ω is discretized using a weighted graph $G=(V,E,w)$, including finite number of vertices V, edges E and similarity matrix w. The nonlocal differential operators required by the new forward model on the graph G are nonlocal gradient, nonlocal divergence, nonlocal normal derivative and nonlocal Laplacian. The definitions of nonlocal gradient, nonlocal divergence and nonlocal Laplacian are already given in equations (5.2), (5.3) and (5.4). The nonlocal normal derivative which is a nonlocal analogue of the normal derivative operator at the boundary encountered in the classical differential vector calculus (i.e. $\widehat{\mathrm{n}}$ in equation (2.5)) is defined as follows:

Definition (*Nonlocal normal derivative*). Given a function $\boldsymbol{\nu}_i$: $V_\Gamma\to\mathbb{R}$ and a weight function w_{ij}: $V\times V\to\mathbb{R}$, the nonlocal normal operator acting on $\boldsymbol{\nu}_i$ is

$$\mathcal{N}_w\boldsymbol{\nu}_i\triangleq-\sum_{j=1}^{N}\left(\nu_{ij}-\nu_{ji}\right)\sqrt{w_{ij}}\ :V_\Gamma\to\mathbb{R}. \tag{6.3}$$

Note that div_w in equation (5.3) and $\mathcal{N}_w$ in equation (6.3) have similar definitions but differ in their signs and the regions over which $\operatorname{div}_w\boldsymbol{\nu}_i$ and $\mathcal{N}_w\boldsymbol{\nu}_i$ are calculated. Also note that the mapping $\boldsymbol{\nu}_i\mapsto\mathcal{N}_w\boldsymbol{\nu}_i$ is scalar-valued which is analogous to the local differential divergence of a vector function in equation (2.3). Finally, with the definitions of div_w

and $\mathcal{N}_w$, the nonlocal divergence theorem is $\int_{\Omega'} \mathrm{div}_w \boldsymbol{\nu} dx = \int_\Gamma \mathcal{N}_w \boldsymbol{\nu} dx$, which essentially relates the flow (i.e. flux) of a nonlocal vector field through a boundary/surface to the behaviour of the nonlocal vector field inside the boundary/surface.

It should be noticed from the nonlocal differential operator definitions (Equations (5.2), (5.3), (6.3) and (5.4)) that, in a full non-local scheme, each vertex has connections with all the vertices in V over Ω such that the constructed graph is fully connected. This can make the computational load extremely heavy. Therefore we build the graph by using the positions of the nodes and the connectivity between nodes in the finite element mesh as the vertices and edges in the graph to sparsify the graph for computational efficiency. We have learned from Chapter 5 that the graph-based nonlocal method is more stable and robust to changes in mesh resolution. Therefore for each vertex i, we consider only those vertices that are directly connected to the vertex i for $\mathcal{N}_i$ (i.e. those vertices that share the same edge with i). With this structure and the nonlocal discrete differential operators, we can derive the following discretized versions of equation (6.1) and (6.2):

$$\begin{aligned} &\sum_{j\in\mathcal{N}_i} (\kappa_i + \kappa_j)(\Phi_i - \Phi_j) w_{ij} + \mu_{ai}\Phi_i = q_{0i} \quad \text{for } i \in \Omega' \\ &2A \sum_{j\in\mathcal{N}_i} (\kappa_i + \kappa_j)(\Phi_i - \Phi_j) w_{ij} + \Phi_i = 0 \quad \text{for } i \in \Gamma \end{aligned} \tag{6.4}$$

The nonnegative and symmetric weight function w_{ij} between two connected vertices i and j has many possible choices. In order to reduce the number of hyper-parameters, in this work, we use the Euclidean distance directly to define the weight function. Specifically, we first obtain the similarity w_{ij} by simply using the inverse of the square of the Euclidean distance d_{ij} between two nodes. Then we normalize the similarity using $w_{ij}/\sum_{j\in\mathcal{N}_i} w_{ij}$ to convert the similarities into probabilities and ensure that the probabilities sum to one.

We note that due to the nature of the graph representation, the implementation of equation (6.4) is identical for a 2D or 3D geometry. It should also be noted that increasing the number of vertices and edges will decrease the sparsity of the graph and increase the computational burden with no change in the implementation. Under these assumptions,

equation (6.4) can be rewritten in matrix form as

$$\boldsymbol{\mathcal{M}}\Phi = Q. \tag{6.5}$$

$\boldsymbol{\mathcal{M}}$ is a $N \times N$ sparse matrix and a diagonally dominant and positive definite real-value matrix, whose entries are

$$\mathcal{M}_{i,j} = \begin{cases} \sum\limits_{j \in \mathcal{N}_i} (\kappa_i + \kappa_j)\, w_{ij} + \mu_{ai} & \text{if } i = j \in \Omega' \\ \sum\limits_{j \in \mathcal{N}_i} (\kappa_i + \kappa_j)\, w_{ij} + \frac{1}{2A} & \text{if } i = j \in \Gamma \\ -(\kappa_i + \kappa_j)\, w_{ij} & \text{if } i \neq j \text{ and } j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases}.$$

Q is a $N \times N_s$ sparse matrix where N_s is the number of sources and each column represents one distributed Gaussian source.

6.3 Experimental results

Numerical experiments are conducted to quantitatively evaluate the performance of the proposed NDE method. The NDE method with the GNM implementation will be compared against the original DE with the FEM implementation. We evaluate the light propagation performance of the proposed method in a 3D homogeneous rectangular slab and then a heterogeneous two-layer rectangular slab where the analytical solutions are known, followed by two dimensional (2D) and three dimensional (3D) image reconstruction examples. All the experiments are performed using Matlab 2018b on a Windows 7 platform with an Intel Xeon CPU i7-6700 (3.40 GHz) and 64 GB memory.

6.3.1 Forward modelling on a 3D rectangular-slab model

To quantitatively compare our GNM method with classical FEM approaches, we model a rectangular-slab of size: 200×100×100 mm^3, as shown in Figure 6.1. The mesh is composed of 442381 nodes corresponding to 2620541 tetrahedral elements, with the average nodal distance of 1.5 mm. For the forward model based on FEM, such a discrete structure can be directly employed for the finite element method. However, the forward model

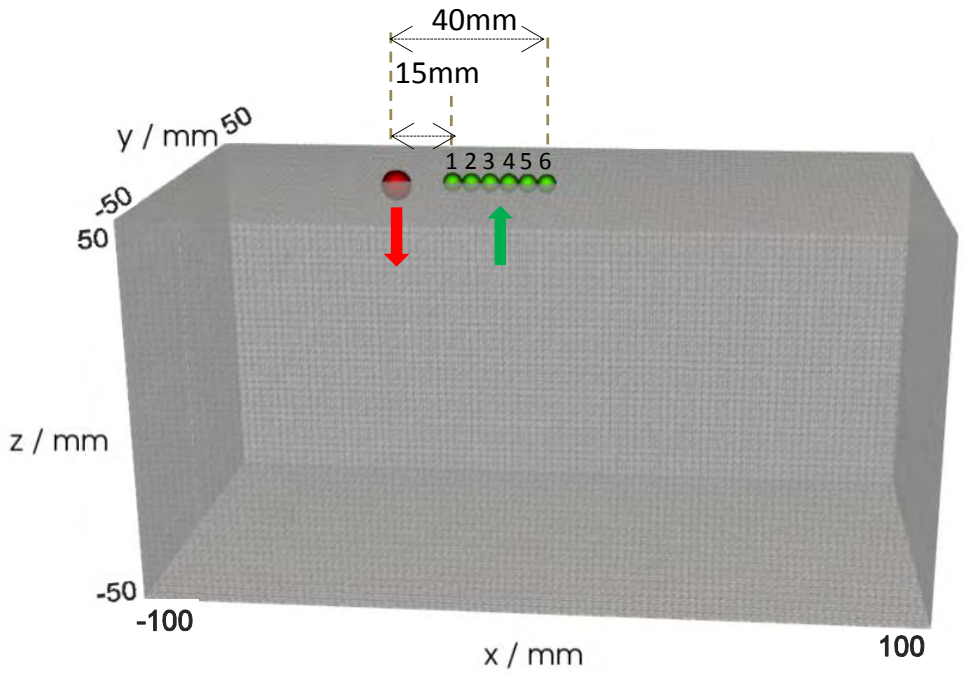


Figure 6.1: Rectangular-slab mesh with one source (red dot) and six detectors (green dots). The distance between the source and the six detectors varies from 15 mm to 40 mm, in 5 mm increments.

6.3.1.1 3D homogeneous rectangular-slab model

For a 3D homogeneous rectangular-slab model, the optical parameters μ_a and μ_s' in the slab were set to 0.01 mm^{-1} and 1 mm^{-1}, respectively. The analytical solution of the BF has the form [171]:

$$I(\rho) = \frac{1}{4\pi}\left[\frac{1}{\mu_a+\mu_s'}\left(\mu_{\text{eff}}+\frac{1}{r_1}\right)\frac{e^{-\mu_{\text{eff}}r_1}}{{r_1}^2}+\frac{3+4A}{3(\mu_a+\mu_s')}\left(\mu_{\text{eff}}+\frac{1}{r_2}\right)\frac{e^{-\mu_{\text{eff}}r_2}}{{r_2}^2}\right], \quad (6.6)$$

where ρ represents the distance from the source, A is the internal reflection parameter for the air-tissue interface, μ_{eff} is the effective attenuation coefficient which is $\sqrt{3\mu_a(\mu_a+\mu_s')}$, $r_1 = \sqrt{1/(\mu_a+\mu_s')^2+\rho^2}$ and $r_2 = \sqrt{(3+4A)^2/(3(\mu_a+\mu_s'))^2+\rho^2}$.

In Figure 6.2(a), we plot the normalized photon flux at the boundary (NBF). We normalize the BF to remove any constant offset resulting from the use of different propagation models. It can be seen that the NBF from both forward models match the analytical solution. In order to observe the difference clearly, in Figure 6.2(b), we plot the percentage of error between the analytical solution and the other two methods with regards to NBF. The percentage of error is calculated by, for each source-detector channel, dividing the absolute difference between each forward model and the analytical solution by the analytical solution. We average the percentage errors along the six source-detector pairs. The forward models based on FEM and GNM are both shown to reproduce the analytical solution to within 7% on average.

We then compare the FR calculated at the vertices inside of the medium. The analytical solution of the FR is [172]:

$$\Phi(r_d,z) = \frac{P\mu_{\text{eff}}^2}{4\pi\mu_a}\left[\left(\frac{\exp\left\{-\mu_{\text{eff}}\left[(z-z_0)^2+r_d^2\right]^{1/2}\right\}}{-\mu_{\text{eff}}\left[(z-z_0)^2+r_d^2\right]^{1/2}}\right)-\left(\frac{\exp\left\{-\mu_{\text{eff}}\left[(z+z_0)^2+r_d^2\right]^{1/2}\right\}}{-\mu_{\text{eff}}\left[(z+z_0)^2+r_d^2\right]^{1/2}}\right)\right], \quad (6.7)$$

where P is the source power. z_0 is the depth of the source which is $1/\mu_s'$. z represents the depth under the surface which is $z = 50\text{mm}$ in our case. r_d is the distance between a given vertex and the source on the X-Y plane. Note that $\sqrt{(z-z_0)^2+r_d^2}$ represents the distance between a given vertex and the source.

In Figure 6.3, we compare the FR calculated using equation (6.7) and the FEM and

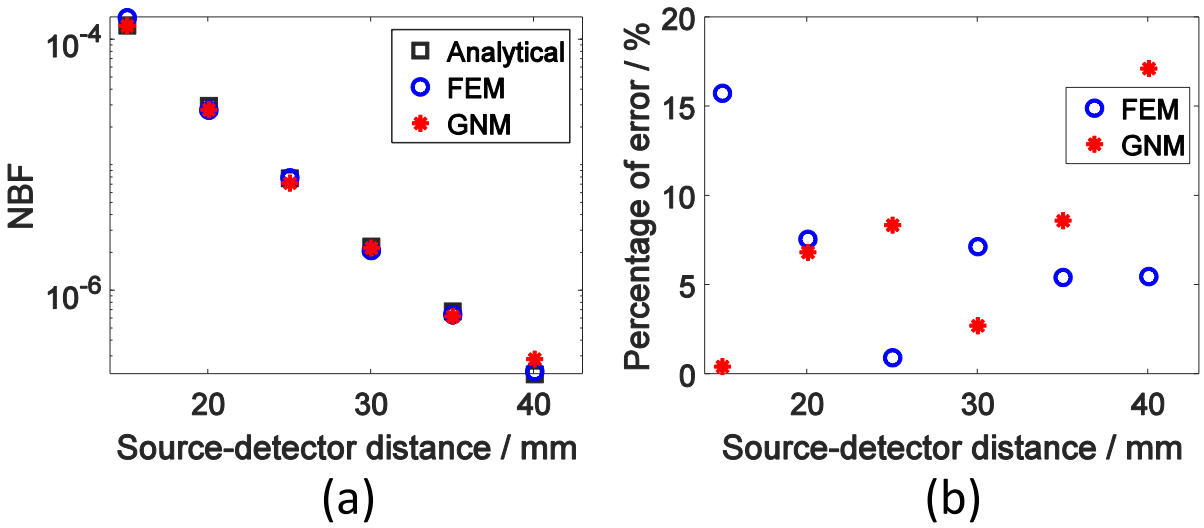


Figure 6.2: The flux measurements on the boundary versus the source-detector distance. (a): NBF; (b): Percentage of error based on NBF.

GNM methods. We choose the vertical plane across the source-detector positions as the region of interest (ROI). For each method, in order to remove any constant offset resulting from the use of different propagation models, we rescaled FR onto the range $[0, 1]$ by dividing the FR with the highest FR value in the ROI and name the rescaled FR as NFR. This is necessary because in FEM, point sources are distributed across the nodes belonging to the element in which the source is placed, whereas in GNM, the source is fully attached to the nearest node. The two methods can therefore have different initialization states for the same source. In Figure 6.3(a)-(c), we plot the NFR at each vertex in the ROI calculated using the analytical method, and the FEM and GNM models, respectively. We also plot its logarithm in (d)-(f), corresponding to the NFR in (a)-(c) respectively. It can be observed that the light propagation in the medium modelled by the proposed forward model is comparable to the one modelled by the forward model based on FEM. In order to see the difference clearly, in Figure 6.4, we plot the descending tendency of the NFR calculated by different propagation methods. Specifically, we plot the logarithm of NFR along the z axis starting from the source position. As can be seen, for all methods the fluence rate gradually drops as the light penetrates deeper. The descending tendency of the curves derived from the both forward methods are almost parallel to the one from the analytical solution. Therefore we can see that all the three models can generate the

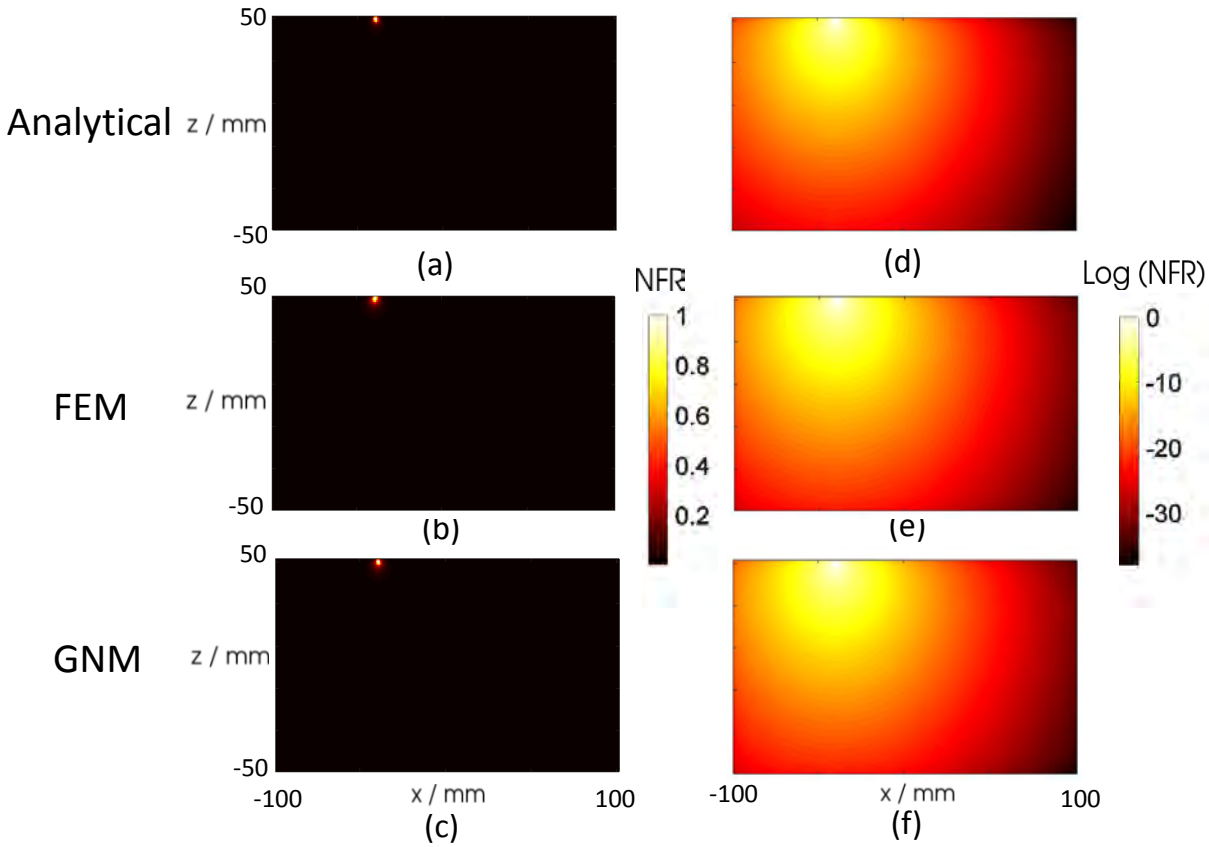


Figure 6.3: (a)-(c): NFR at each vertex in the ROI calculated using the analytical solution, forward models based on FEM and the one based on GNM, respectively; (d)-(f): logarithm of the NFRs, corresponding to (a)-(c).

6.3.1.2 3D heterogeneous rectangular-slab model

We further evaluate the light propagation performance of FEM and GNM on a heterogeneous model which has the same geometry as the previous example, but contains two optically different layers. We choose the thickness of the first layer to be relevant to one potential application of the layered model: DOT for measurements of cerebral oxygenation, where the thickness of the tissues above the brain is around 10 mm. Figure 6.5 gives the 2D section of the two-layered model in which the thickness of the first layer is 7.5mm. The optical properties (μ_a, μ_s') in the two layers were set to (0.015 mm^{-1}, 1.5 mm^{-1}) and (0.01 mm^{-1}, 1 mm^{-1}), respectively. The positions of sources and detectors are the same as that in Section 6.3.1.1. The analytical solution of the BF for two-layered

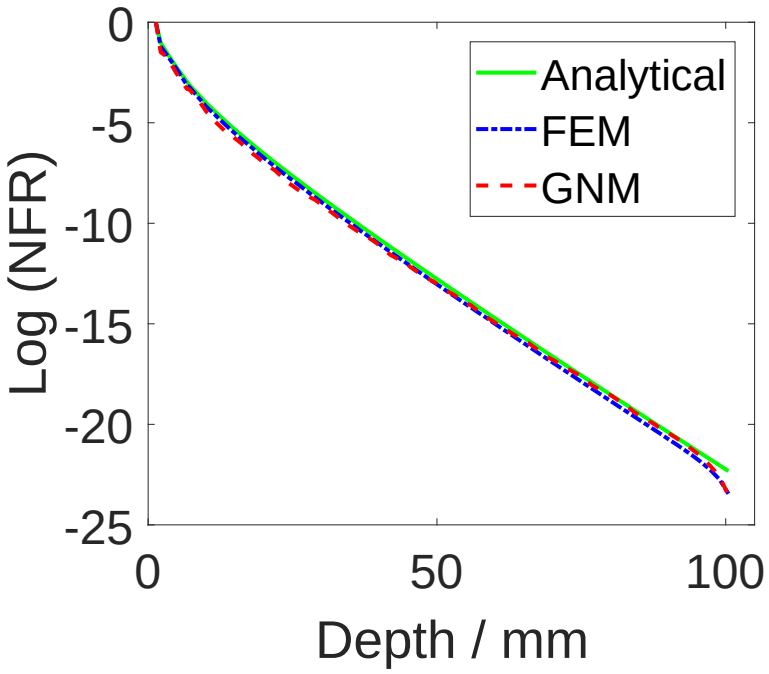


Figure 6.4: Descending tendency of the NFR from the source to the medium along the z axis.

medium is given in [179] and we use it as the ground truth to evaluate the accuracy of our proposed forward model. In Figure 6.6, we first plot the normalized photon flux (NBF) at the boundary versus the source-detector separation and then give the percentage of error between the analytical solution and the other two methods. The forward models based on FEM and GNM are both shown to reproduce the analytical solution to within 8.5% on average. Similar to what we observed in the previous section, GNM has less error than FEM when the source-detector distance is short (15mm) while the error becomes greater when the source-detector distance is as long as 40mm. There is no significant difference in the percentage errors when the source-detector distance is between 20mm to 35mm. Therefore similar conclusions can be achieved from the experiments in Section 6.3.1.1 and 6.3.1.2.

6.3.1.3 Computational time

After evaluating the accuracy of the fluence rates and boundary measurements modelled by different forward models, in Figure 6.7, we compare the computational efficiency of

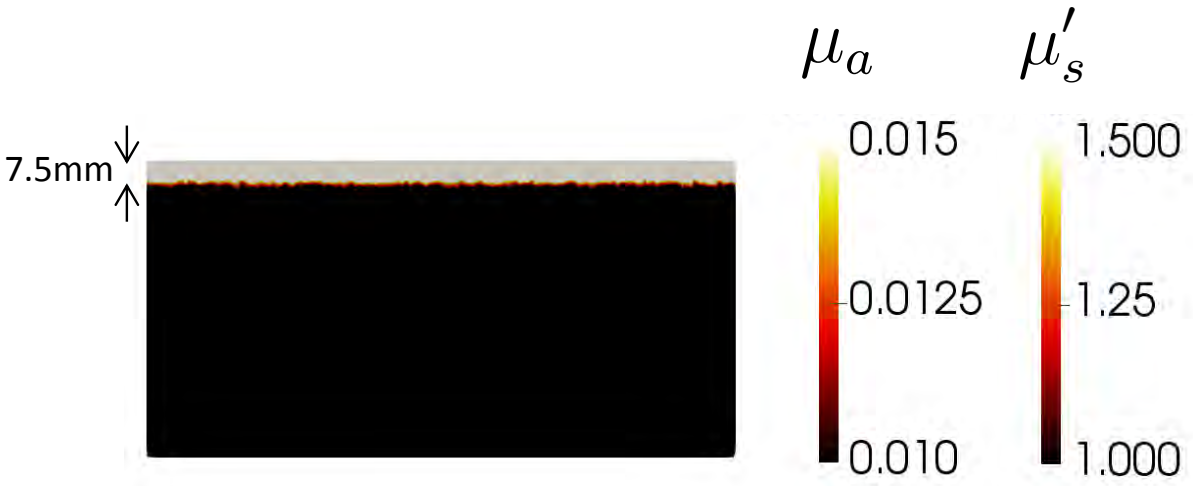


Figure 6.5: Scheme of the heterogeneous rectangular-slab model.

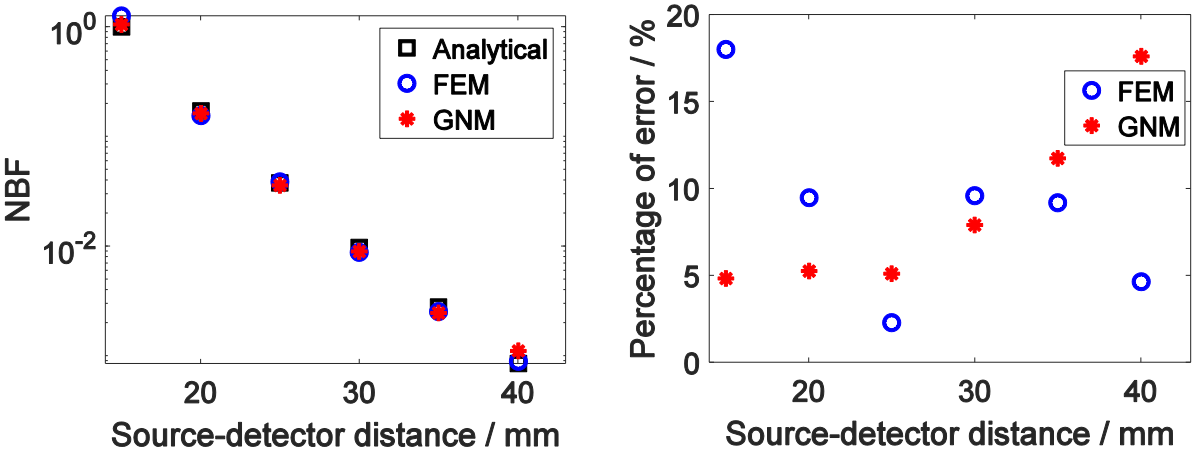


Figure 6.6: The flux measurements on the boundary versus the source-detector distance. (a): NBF; (b): Percentage of error based on NBF.

FEM and GNM forward models on the 3D homogeneous rectangular-slab model. We run each model on six mesh structures with different average nodal distance of 1.5, 2, 2.5, 3, 3.5 and 4 mm respectively. The mesh spatial resolution becomes lower when the nodal distance is larger. We run each forward modelling process ten times and record the mean and standard deviation of the CPU time consumed for computing one source-detector channel. For a fair comparison, we use a direct solver with Cholesky decomposition to solve the linear equation resulting from each forward model.

For all mesh resolutions, based on each source-detector channel, the CPU times required by the FEM model are larger than for GNM. When the mesh resolution is low (for

example the case where the average nodal distance is 4 mm) the CPU time consumed by the FEM approach (0.11s) is 175% larger than the time required by the GNM approach (0.04s). When the mesh resolution is high (average nodal distance is 1.5 mm), the CPU time consumed by the FEM approach (14.6s) is only 14% longer than the GNM approach (12.7s). It should be noticed that the CPU time plotted in Figure 6.7 is only for one source-detector channel. GNM is 14% faster than the FEM approach when average nodal distance is low (mesh resolution is high). For DOT which normally has thousands of source-detector channels in some clinical scenarios, this 14% improvement at each channel would make a significant difference in the total computational time. This result demonstrates the computational efficiency of the proposed forward model.

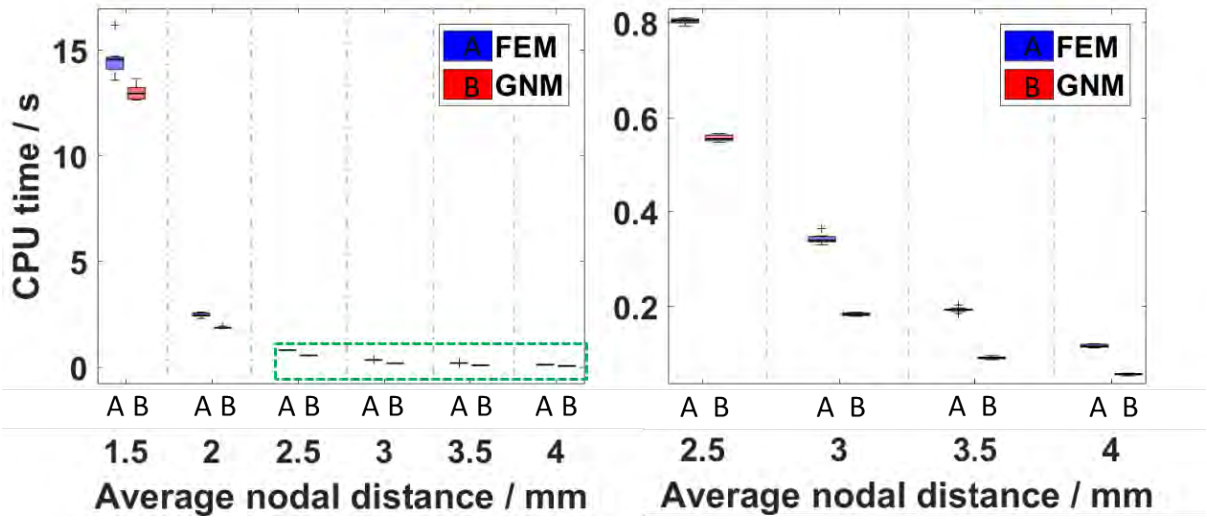


Figure 6.7: CPU time (s) consumed at one source-detector channel using different forward models. 'A' represents the FEM approach while 'B' represents the GNM approach. Right figure is the zoomed-in plot of the area in the green dash line of the left figure.

6.3.2 Image reconstruction using different forward models

We now consider the recovery of the optical properties at each vertex within the medium using both forward models. The image reconstruction process is implemented by iteratively refining the optical properties of the forward model until the forward model prediction matches the boundary measurements [110]. It can be implemented by solving

the following minimization problem:

$$\mu^* = \arg\min_{\mu} \left\{ \| y - \mathcal{F}(\mu) \|_2^2 + \lambda \mathcal{R}(\mu) \right\}, \tag{6.8}$$

where $\mathcal{F}$ is the non-linear operator induced from the forward model, $\mathcal{R}$ is a general regularization term, and λ is a weight that determines the extent to which regularization will be imposed on the solution μ^*. In this work, we adopt the popular quadratic Tikhonov-type regularization for fair comparison [110]. Four quantitative evaluation metrics are considered to evaluate the reconstruction results: the average contrast (AC) [162], peak signal-to-noise ratio (PSNR) [162], structural similarity index (SSIM) [68] and root mean square error (RMSE) [68]. If the reconstruction is identical to the ground truth image, AC is equal to 1. For PSNR and SSIM, the recovered image has higher quality if higher PSNR or SSIM values are obtained. Lower RMSE represents better reconstruction results. Randomly generated Gaussian noise is added to the amplitude of the measurement vector to simulate real noise in a CW system. In order to reduce the randomness resulting from the randomly distributed Gaussian noise, we run each experiment ten times and record the average (mean) and standard deviation (SD) of the four evaluation metrics.

6.3.2.1 Image reconstruction on a homogeneous circle model

We consider a 2D homogeneous circular geometry containing one target activation region (Figure 6.8(b)). The model has a radius of 43mm and is composed of 1785 nodes and 3418 linear triangle elements. Sixteen source-detector fibres are placed equidistant around the external boundary for data acquisition (Figure 6.8(a)). When one fibre as a source is turned on, the rest are used as detectors, leading to 240 total boundary measurements. All sources were positioned one scattering distance within the outer boundary because the source is assumed to be spherically isotropic. The background absorption coefficient is set to 0.01 mm^{-1}. One 10mm radius target region is centred at (20mm, 0mm) with 0.03 mm^{-1} absorption coefficient. The reduced scattering coefficient is set to be homogeneous throughout the whole computational domain with the value of 1 mm^{-1}. 1% normally distributed Gaussian noise was added to the amplitude of the measurement vector.

Figure 6.8(c) shows the reconstruction results using the forward model based on FEM (Equation (2.3)) and GNM (Equation (6.1)) on 0% and 1% noisy data respectively. By visual inspection, it is evident that for the same level of Gaussian noise, the image recovered using the GNM approach is similar to the one recovered using the FEM approach. Figure 6.9 gives the 1D cross section of the results recovered in Figure 6.8 along the horizontal line across the centre of the target (20mm, 0mm). It can be seen that the curves resulting from different forward models have similar edge smoothing resulting from the Tikhonov regularization and slightly different peak values. This is consistent with our visual observation from the reconstructed images in Figure 6.8.

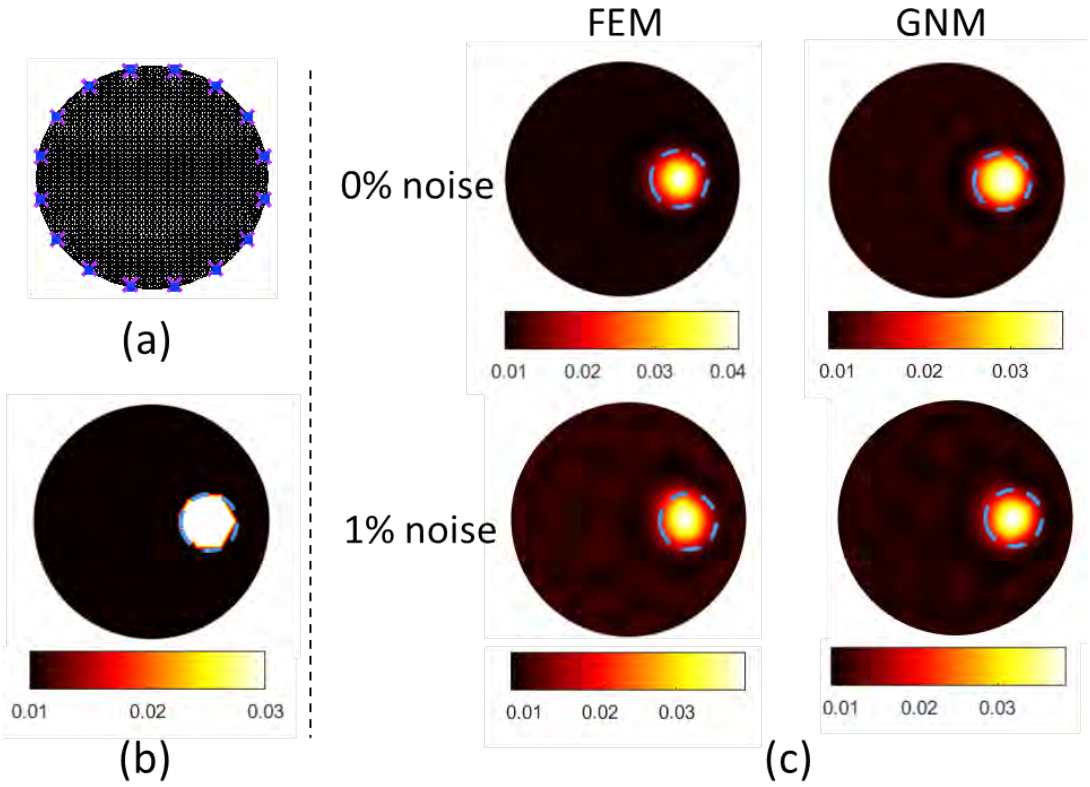


Figure 6.8: (a): A typical circle mesh with sixteen co-located sources and detectors; (b): True distribution of μ_a; (c): Images reconstruction of μ_a using the forward model based on FEM and GNM (from left to right column) on 0% (top part) and 1% (bottom part) noisy data.

In Table 6.1, the values of the metrics AC, PSNR, SSIM and RMSE are shown to qualitatively evaluate the results in Figure 6.8. It can be observed that when the data is

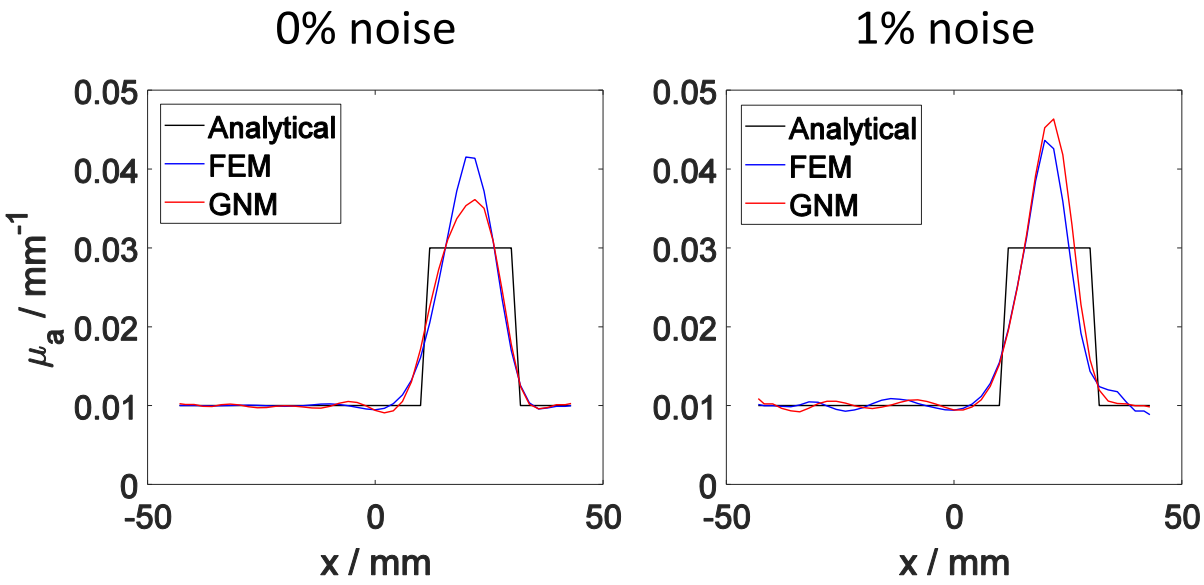


Figure 6.9: 1D cross sections of images recovered in Figure 6.8 along the horizontal line across the centre of the target. Left to right column: 0% and 1% added Gaussian noise.

clean, GNM gives AC closer to 1, slightly higher PSNR and SSIM, and lower RMSE than FEM. For the noisy data, GNM achieves similar AC, PSNR, SSIM and RMSE values with the FEM approach. This experiment quantitatively validates the forward modelling capacity of our proposed model and the consistency between these two forward models.

Table 6.1: Evaluation metrics for the recovered results using FEM and GNM on data with 0% and 1% added noise.

	0% noise		1% noise (Mean)	
	FEM	GNM	FEM	GNM
AC	1.1	1.0	1.1	1.1
PSNR	54.5	55.8	54.3	54.3
SSIM	0.996	0.997	0.996	0.996
RMSE	0.0019	0.0016	0.0019	0.0019

6.3.2.2 Image reconstruction on a heterogeneous head model

We now evaluate both forward models on a physically realistic three dimensional heterogeneous head model. This head model is composed of three tissue layers which are scalp,

skull and brain. The reconstruction mesh consists of 50721 nodes associated with 287547 tetrahedral elements, with the average element size 9.3mm^3. Each node is assigned to one of the three layers. Absorption coefficients assigned to each layer refer to an in vivo study [6] at 750nm.

A large rectangular imaging array with 36 sources and 37 detectors was placed over the back-head area (Figure 6.10(a)), allowing use of multiple sets of overlapping measurements which can improve both the spatial resolution and quantitative accuracy [180]. The source-detector (SD) separation distances ranges from 1.3 to 4.8cm, leading to 590 overlapping, multi-distance measurements. One anomaly with 15mm radius is simulated in the brain (Figure 6.10(b)). In order to simulate traumatic brain injury (TBI) cases where

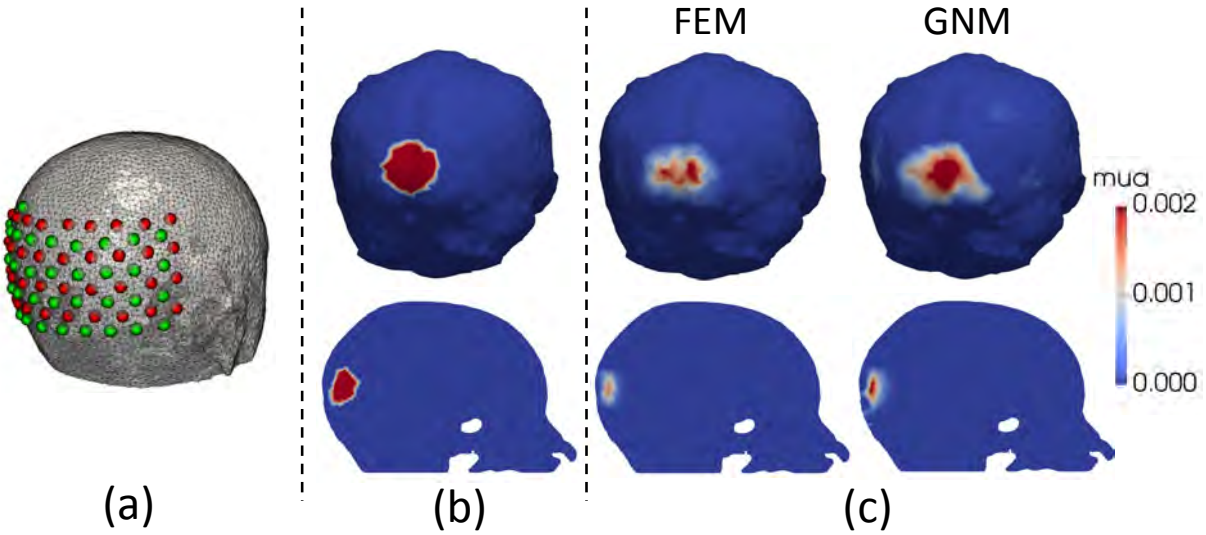


Figure 6.10: (a): Three-dimensional head mesh and distribution of the rectangular imaging array with 36 sources (red dots) and 37 detectors (green dots); (b): Ground truth; (c) Reconstruction with the forward model based on FEM and GNM, respectively.

using Beer's law [110] with 55% StO_2. In line with the current in vivo performance of the imaging system, 0.12%, 0.15%, 0.41% and 1.42% Gaussian random noise was added to first (13mm), second (30mm), third (40mm) and fourth (48mm) nearest neighbor measurements to provide realistic data [138]. Reconstructed absorption coefficients of the

simulated anomaly using different models are displayed in the third to fourth column of Figure 6.10. Corresponding 2D cross section is given in the second row. The visualization suggests that GNM can achieve better reconstruction performance with optical property values closer to the ground truth. This may be because the simulated anomaly is close to the outer surface and GNM gives lower errors when the source detector distance is low. We can also see that the results by both methods are smoothed and the volume sizes of the recovered anomaly are smaller than the ground truth. Evaluation metrics are given in Table 6.2. Even though there is slight visual difference between the two methods , no obvious difference between these two reconstruction models can be observed from the four evaluation metrics. These findings further quantitatively validate the consistency between these two forward models.

Table 6.2: Evaluation metrics for μ_a on the recovered results shown in Figure 6.10.

	FEM	GNM
AC	0.8	0.9
PSNR	78.9	79.0
SSIM	0.99	0.99
RMSE	0.0001	0.0001

6.4 Conclusion

In this chapter, we proposed a new formulation of the forward model for DOT that is based on the concepts of differential operators under a nonlocal vector calculus. The discretization of the new forward model is performed using an efficient graph-based numerical method. Our proposed model is shown to be able to accurately model the light propagation in the medium and is quantitatively comparable with both analytical and FEM forward models. Compared with the conventional forward model based on FEM, our proposed model has the following two advantages: 1) according to the experiments in Section 6.3.1, our proposed model is shown to be more computationally efficient with

an average speed improvement of 30% compare to the FEM forward model due to the simple graph-based discretization; 2) it allows identical implementation for geometries in different dimensions thanks to the nature of the graph representation. In addition, the proposed graph-based discretization method can also be applied to other imaging techniques which are modelled using a diffusion equation, where has potential to improve the computational efficiency and simplicity.

CHAPTER 7

GRAPH-BASED FORWARD MODEL FOR FREQUENCY-DOMAIN DIFFUSE OPTICAL TOMOGRAPHY

7.1 Introduction

In chapters 5 and 6, we introduced graph discretization to DOT reconstruction and forward modelling. Specifically we proposed a graph-based forward model and inverse model with a total-variation regularization term for a continuous wave (CW) DOT imaging system. In this chapter, we extend the graph-based forward model to frequency domain DOT where intensity modulated near-infrared light is used as the excitation and the boundary data includes both the amplitude (or intensity) and phase shift of the measured signal. The traditional DE in frequency domain (FD) was described in equation (2.3) with the Robin boundary condition (Equation (2.5)).

The concept of differential operators under the nonlocal vector calculus [121, 142, 143, 145] is used to formulate a new forward model that can accurately simulate light propagation in turbid media. The discretization for the NDE is performed using a graph-based numerical method (GNM). The nonlocal diffusion equation (NDE) in FD system is a linear PDE. After its discretization with the GNM, the resulting equations are a linear system, which must be solved for the numerical solution of photon fluence Φ. Existing solvers for the linear equations are either direct or iterative. Direct solvers use

a finite number of operations to find an exact solution to a finite set of linear equations. Unfortunately, they are not feasible when the linear system becomes increasingly large and complex. Instead of exactly solving linear equations, iterative solvers aim to find an approximate solution iteratively up to a tolerance. They are expected to be more computationally efficienct than direct solvers for large sized linear systems. In this work, we study several representative linear solvers for the linear systems that result from the discretization of the NDE with its nonlocal RBC for FD system. We will validate the effectiveness of the new FD NDE model and the efficiency of different solvers through numerical experiments.

The work on the FD system can be easily extended to the time domain (TD) cases by replacing $i\omega$ with $\partial/\partial t$. In the TD system, tissue is irradiated by ultrashort (picosecond order) leaser pulses, and the number of the emerging photons is recorded over time to generate a temporal point spread function (TPSF). Using the TPSF, it is possible to extract several important quantities from the boundary measurement, including total intensity, transient time, variance, skew and other higher order moments. In this study, we limit ourselves to the case of FD measurement systems. However, the proposed approach can be generalized to TD measurements without difficulty.

7.2 Nonlocal vector calculus

7.2.1 Nonlocal diffusion equation

With the nonlocal differential operators at hand (Equations (5.2), (5.3)), one readily rewrites the DE (Equation 2.3) in the sense of the nonlocal vector calculus. The NDE is given by

$$-\operatorname{div}_w \left[\kappa\left(r\right)\nabla_w\left[\Phi\right]\left(r\right)\right]\left(r,\omega\right)+f\left(r,\omega\right)\Phi\left(r,\omega\right)=q_0\left(r,\omega\right)\quad\text{for } r\in\Omega'. \tag{7.1}$$

The left-hand side of the NDE can be derived by using the definitions of the nonlocal gradient (Equation 5.2) and the nonlocal divergence (Equation 5.3). $f(r,\omega)$ in equation

7.1 is defined in equation 2.4. The equivalent formulation of the NDE is thereby given by

$$\int_{\Omega} \left(\kappa\left(r\right)+\kappa\left(y\right)\right)\left(\Phi\left(r,\omega\right)-\Phi\left(y,\omega\right)\right)w\left(r,y\right)dy+f\left(r\right)\Phi\left(r,\omega\right)=q_0\left(r,\omega\right) \quad \text{for } r\in\Omega'. \tag{7.2}$$

The corresponding nonlocal RBC of the NDE can be written as

$$2A\mathcal{N}_w\left[\kappa\left(r\right)\nabla_w\left[\Phi\right]\left(r\right)\right]\left(r,\omega\right)+\Phi\left(r,\omega\right)=0 \quad \text{for } r\in\Gamma. \tag{7.3}$$

In accordance with the definitions of the nonlocal gradient (Equation 5.2) and the nonlocal normal derivative (Equation 6.3), the nonlocal RBC can be derived in a similar manner as equation 7.2

$$2A\int_{\Omega}\left(\kappa\left(r\right)+\kappa\left(y\right)\right)\left(\Phi\left(r,\omega\right)-\Phi\left(y,\omega\right)\right)w\left(x,y\right)dy+\Phi\left(r,\omega\right)=0 \quad \text{for } r\in\Gamma. \tag{7.4}$$

Note that the first term on the left-hand side of equation (7.4) has the same form as that of equation (7.2), except that there is a weighted factor $2A$ imposed on equation (7.4). Both equation (7.2) and equation (7.4) are linear equations.

With the graph discretization notations explained in Chapter 6, we have the following two simultaneous discrete equivalents of (7.2) and (7.4)

$$\begin{cases} \sum\limits_{j\in\mathcal{N}_i}\left(\kappa_i+\kappa_j\right)\left(\Phi_i-\Phi_j\right)w_{i,j}+f_i\Phi_i=q_{0i} & \text{for } i\in\Omega' \\ 2A\sum\limits_{j\in\mathcal{N}_i}\left(\kappa_i+\kappa_j\right)\left(\Phi_i-\Phi_j\right)w_{i,j}+\Phi_i=0 & \text{for } i\in\Gamma \end{cases}. \tag{7.5}$$

7.2.2 Linear system

If we rewrite equations (7.5) in the matrix form, we can have the following system of linear equations

$$\boldsymbol{\mathcal{M}}\Phi=Q, \tag{7.6}$$

where $\boldsymbol{\mathcal{M}}$ is a $nN \times nN$ sparse matrix, whose entries are

$$\mathcal{M}_{i,j} = \begin{cases} \sum_{j \in \mathcal{N}_i} (\kappa_i + \kappa_j)\, w_{ij} + f_i & \text{if } i = j \in \Omega' \\ \sum_{j \in \mathcal{N}_i} (\kappa_i + \kappa_j)\, w_{ij} + \frac{1}{2A} & \text{if } i = j \in \Gamma \\ -(\kappa_i + \kappa_j)\, w_{ij} & \text{if } i \neq j \text{ and } j \in \mathcal{N}_i \\ 0 & \text{otherwise} \end{cases}.$$

Q is a $N \times N_s$ sparse matrix where N_s is the number of sources and each column represents one distributed Gaussian source. Figure 7.1 visualizes the sparsity patterns of $\boldsymbol{\mathcal{M}}$ resulting from the 3D slab and 3D head models, as shown in Section 7.4.2 and 7.4.3, respectively. Although the structures of $\boldsymbol{\mathcal{M}}$ are all symmetric and sparse, $\boldsymbol{\mathcal{M}}$ is a non-Hermitian matrix. This is because some entries on the main diagonal of $\boldsymbol{\mathcal{M}}$ are complex values, which makes $\boldsymbol{\mathcal{M}}$ non-Hermitian. The linear system (Equation 7.6) for the FD system can be solved either exactly or approximately by the methods introduced in section 7.3.

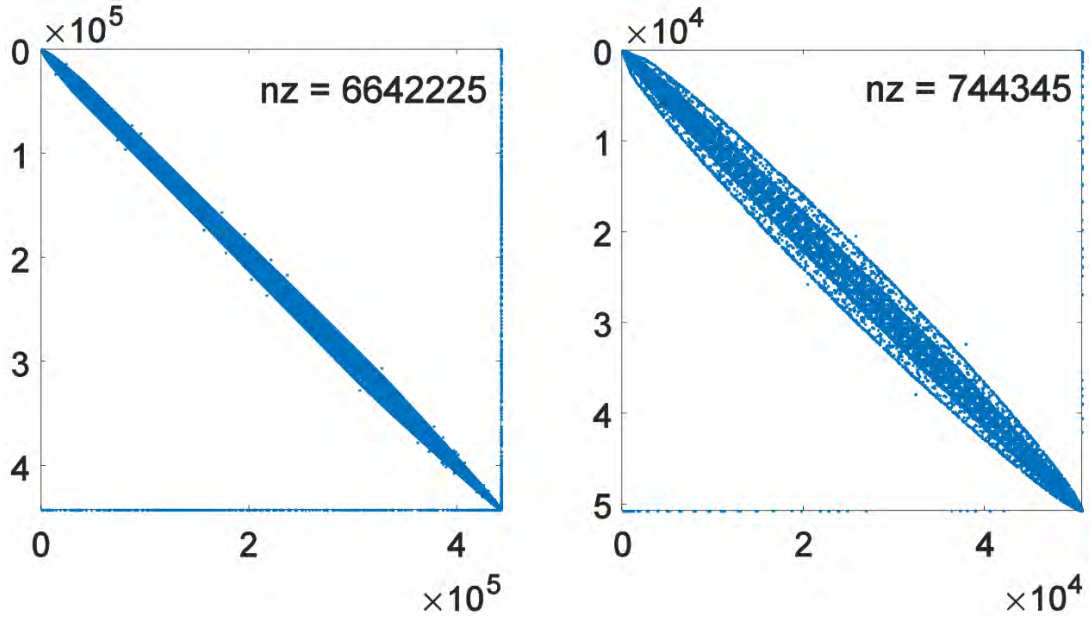


Figure 7.1: Visualization of the sparsity pattern of $\boldsymbol{\mathcal{M}}$. From left to right are the results from the slab and head models shown in Section 7.4.2 and 7.4.3. As evident, the structures of $\boldsymbol{\mathcal{M}}$ are all symmetric and sparse. However, $\boldsymbol{\mathcal{M}}$ is the non-Hermitian matrix and nz denotes the number of non-zeros.

7.3 Solvers for linear system

There are a large number of solvers that are able to handle equation (7.6), which can be direct, iterative, multigrid, and domain decomposition approaches. In the following, we shall focus on some representative direct and iterative solvers in detail.

7.3.1 Direct solvers

When $\boldsymbol{\mathcal{M}}$ is small, direct methods may be the best choice. Among them is the popular Gaussian elimination. It has two steps: forward elimination and backward substitution. The former reduces the matrix to an upper triangular matrix (i.e. echelon form), while the latter reduces the matrix to a reduced row echelon form. Both steps use elementary row operations. Another popular approach is factorization techniques that rewrite the matrix as a product of two or more matrices, each of which has a meaningful structure that can be exploited to make more efficient algorithms. The LU factorization is suitable for the linear system with non-symmetric matrix from the FD system.

7.3.2 Iterative solvers

When $\boldsymbol{\mathcal{M}}$ is large, direct methods would not be a plausible solution. To overcome the computational bottleneck of direct solvers, various approximated solvers have been proposed in the literature. They can be mainly classified into three categories: iterative, multigrid, and domain decomposition approaches. Multigrid and domain decomposition can be very efficient regardless of the linear system size. However, they require additional input parameters which might be hard to choose for different problems. Instead, iterative solvers [181] are very generic and require little or no additional inputs from users. Therefore, they are suitable choices to solve the linear system (Equation 7.6).

Iterative methods for the linear system with non-symmetric matrix rely on the Krylov subspace. Krylov subspace methods include the conjugate gradient (CG) method [182], generalized minimal residual (GMRES) method [183], biconjugate gradient stabilised

(BiCGSTAB) method [184], etc. Among them, GMRES and BiCGSTAB suit non-symmetric/-Hermitian system (i.e. FD system).

7.3.3 Preconditioners

The computational speed of the iterative methods in Krylov subspace can be dramatically increased by using the preconditioning technique for the linear system (Equation 7.6). The main idea of the technique is to apply a transformation, called the preconditioner, to reduce the condition number of the problem and make it more suitable for iterative solvers. By choosing a suitable preconditioner, the condition number of a linear system can be reduced such that convergence can be achieved with fewer iterations. Common preconditioners include the incomplete Cholesky (IC) factorization [185], incomplete LU (ILU) factorization [185], factorized sparse approximate inverse (FSAI) [186]. In practice, choosing the best preconditioner is usually a trial and error procedure. For simplicity's sake, we investigate only the ILU preconditioner in this paper, which is applied to our non-Hermitian linear system. In summary, we present the solvers and preconditioners for the FD system in Table 7.1.

Table 7.1: Solvers and preconditioners for the FD system.

Direct Solver	Iterative Solver+Preconditioner	
LU Factorization	GMRES+ILU	BiCGSTAB+ILU

7.4 Experimental results

Extensive numerical experiments are conducted to qualitatively and quantitatively evaluate the performance of the proposed NDE method and the solvers introduced in Section 7.3. The NDE method with the GNM implementation will be compared against the original DE approach with the FEM implementation. In all the simulated experiments (Section 7.4.2 and 7.4.3), the modulation frequency ω in $f(x)$ in equation (2.4) is chosen as 100MHz and the source term q_0 is set to $e^{i\epsilon}$, where ϵ is a very small positive value.

With such q_0, we assume that the injected light has no phase shift and its amplitude is unity. In this section, we shall first define the relative error metric used to quantify the accuracy of the solution obtained by the iterative solvers in Table 7.1, followed by the experiments to the proposed NDE method. We finish with an image reconstruction example on real human brain data. All the experiments are performed using Matlab 2018b on a Windows 7 platform with an Intel Xeon CPU i7-6700 with 3.40 GHz and 64 GB memory.

7.4.1 Relative error metric

In order to compare the convergence rate of different iterative solvers and guarantee that these solvers do not lead large errors to the final solution, the following relative error R_Φ^k is defined to monitor the accuracy of an iterative solver over iterations

$$R_\Phi^k = \frac{\|\Phi_{itr}^k - \Phi_{dir}\|_2}{\|\Phi_{dir}^k\|_2}, \tag{7.7}$$

where $\|\cdot\|_2$ denotes the L^2-norm. Φ_{dir} is the photon fluence calculated using a direct solver, which is taken as the ground truth here. Φ_{itr}^k is the fluence computed using an iterative solver at the kth iteration. The iterative algorithm shall converge if $R_\Phi^k = 0$. In practice, the iteration normally terminates when R_Φ^k drops below a preset small positive value instead of waiting for $R_\Phi^k = 0$. This value is set to 1×10^{-12} for all the iterative solvers used in the following experiments.

7.4.2 3D slab model

In this section, we continue to use the homogeneous rectangular-slab mesh which is described in section 6.3.1 to compare the light propagation modelling by different forward models. We conduct simulations using a FD source for which we can analytically calculate the photon flux measurement on the boundary as well as the fluence rate at each vertex. The analytical solutions are then compared with the solutions from the forward model based on FEM and GNM. In FD system, both the amplitude and phase data types of

the diffusing wave are measured. The additional information from the phase permits simultaneous determination of μ_a and κ.

In Figure 7.2, we plot the normalized amplitude and phase of boundary measure-

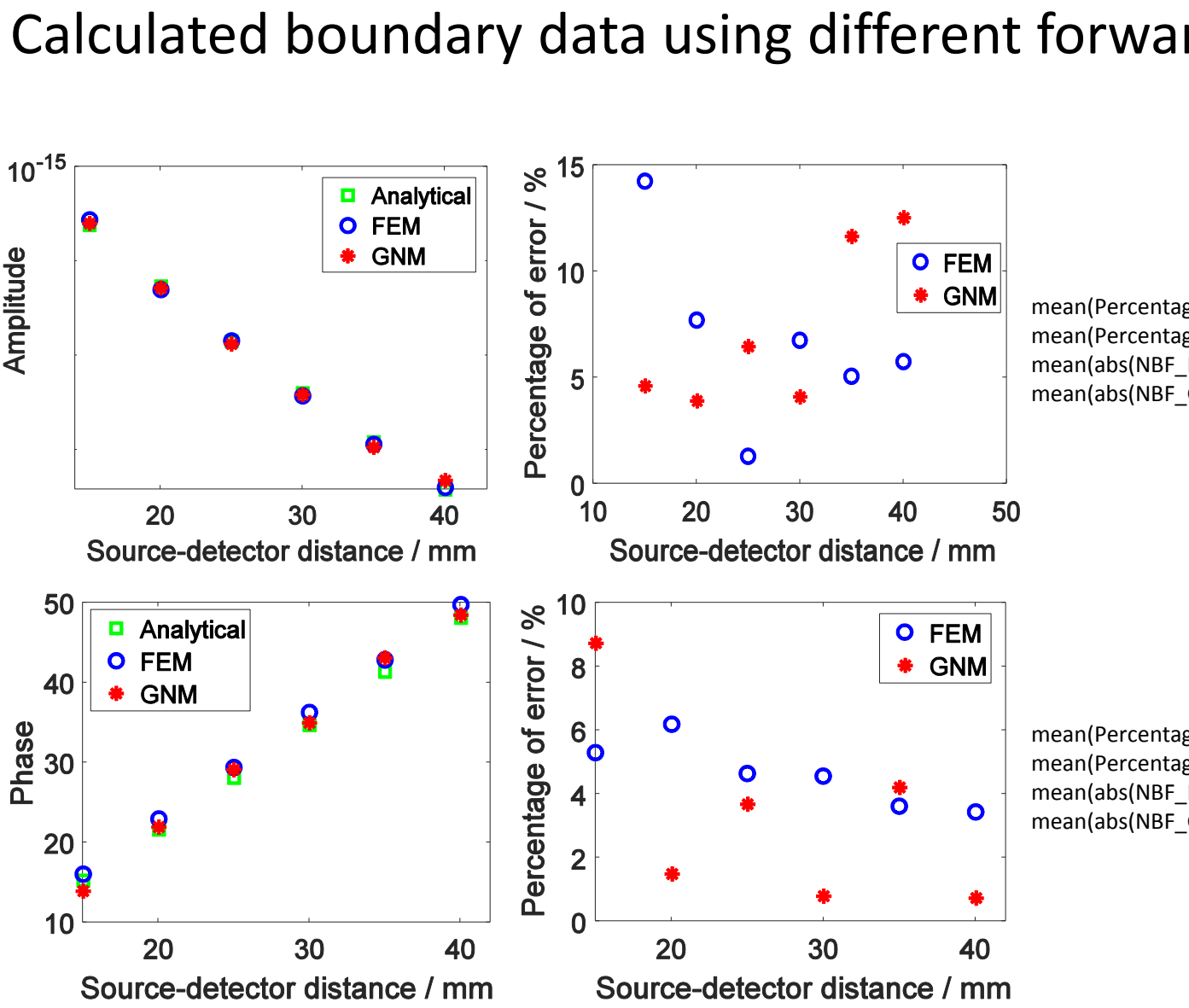


Figure 7.2: The boundary measurements versus the source-detector distance. First row: normalized amplitude and the corresponding percentage of error; second row: phase and the corresponding percentage of error.

the other two forward models with regards to amplitude and phase, respectively. The percentage of error is calculated by, for each source-detector channel, dividing the absolute difference between each forward model and the analytical solution by the analytical solution. We average the error percentages across all the source-detector pairs. We can

see that on average, the percentage of error with regards to amplitude by both forward models are around 7%. For the phase data type, the percentage of error by the GNM is around 3% which is lower than the one by the FEM which is around 4.5%. For the amplitude, the error by the GNM increases when the source-detector distance becomes larger, leading to 17.5% error when the source-detector distance is 40mm. We then evaluate the error by using the alternative definition of the effective attenuation coefficient $\mu_{\text{eff}} = \sqrt{3\mu_a\mu_s'}$. We can see from Figure 7.3 that the error of GNM reduces at all the source-detector separations. Under the new effective attenuation coefficient definition, the percentage of error by GNM is around 5% on average, while the error was 7% when $\mu_{\text{eff}} = \sqrt{3\mu_a\left(\mu_a + \mu_s'\right)}$. The error at 40mm source-detector separation appears the largest reduction, reducing from 17.5% to 13.5%.

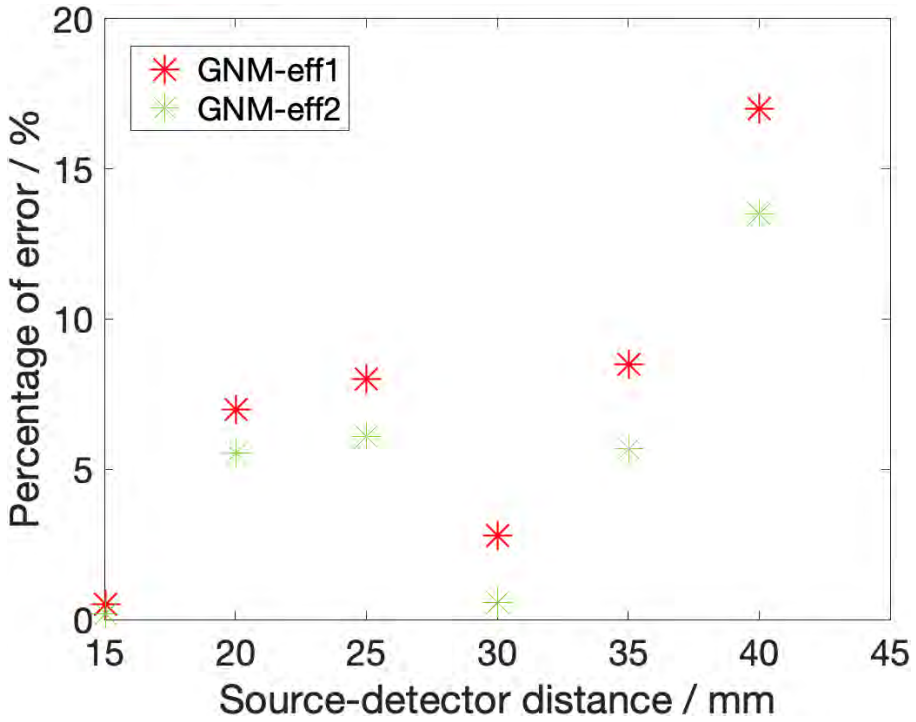


Figure 7.3: Percentage of error of amplitude by GNM using two different effective attenuation coefficient definitions. GNM-eff1: $\mu_{\text{eff}} = \sqrt{3\mu_a\left(\mu_a + \mu_s'\right)}$; GNM-eff2: $\mu_{\text{eff}} = \sqrt{3\mu_a\mu_s'}$.

We then compare the fluence rate calculated at each vertex inside of the medium by using the analytical, FEM and GNM approaches, respectively. In Figure 7.4, we plot the values on a 2D cross section which is positioned though the source-detector and parallel to the X-Z plane. For each method, we linearly rescaled the amplitude onto the range

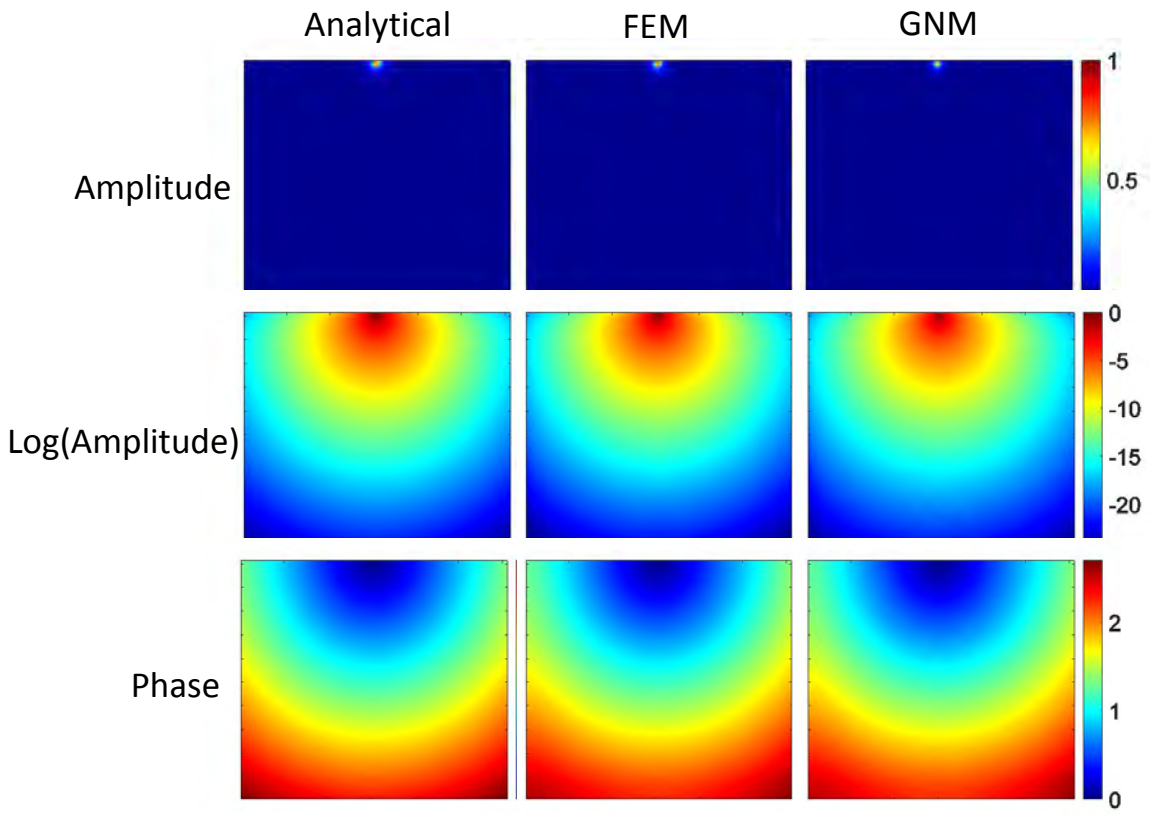


Figure 7.4: 2D cross sections of light propagation images by different methods. First to third row: the amplitude, logarithm of the amplitude and the phase. First to third column: results by the analytical, FEM and GNM approaches respectively.

$[0, 1]$. We can see from this figure that the overall visual results by the forward models based on FEM and GNM are compatible.

Table 7.2: Absolute difference on the results plotted in Figure 7.4. FEM-ANA: absolute difference between results by the FEM approach and the analytical solution; GNM-ANA: absolute difference between results by the GNM approach and the analytical solution.

	Amplitude	Log(Amplitude)	Phase
FEM-ANA	2.9	4300	160.1
GNM-ANA	7.2	3900	32.5

In Figure 7.5, we compare the 1D signal profiles describing the light propagation by the FEM and GNM approaches, as shown in Figure 7.4, from the source position and along the Z axis. For all methods the amplitude gradually drops and the phase gradually increases as the light penetrates deeper. It can be seen from the zoomed-in images (right

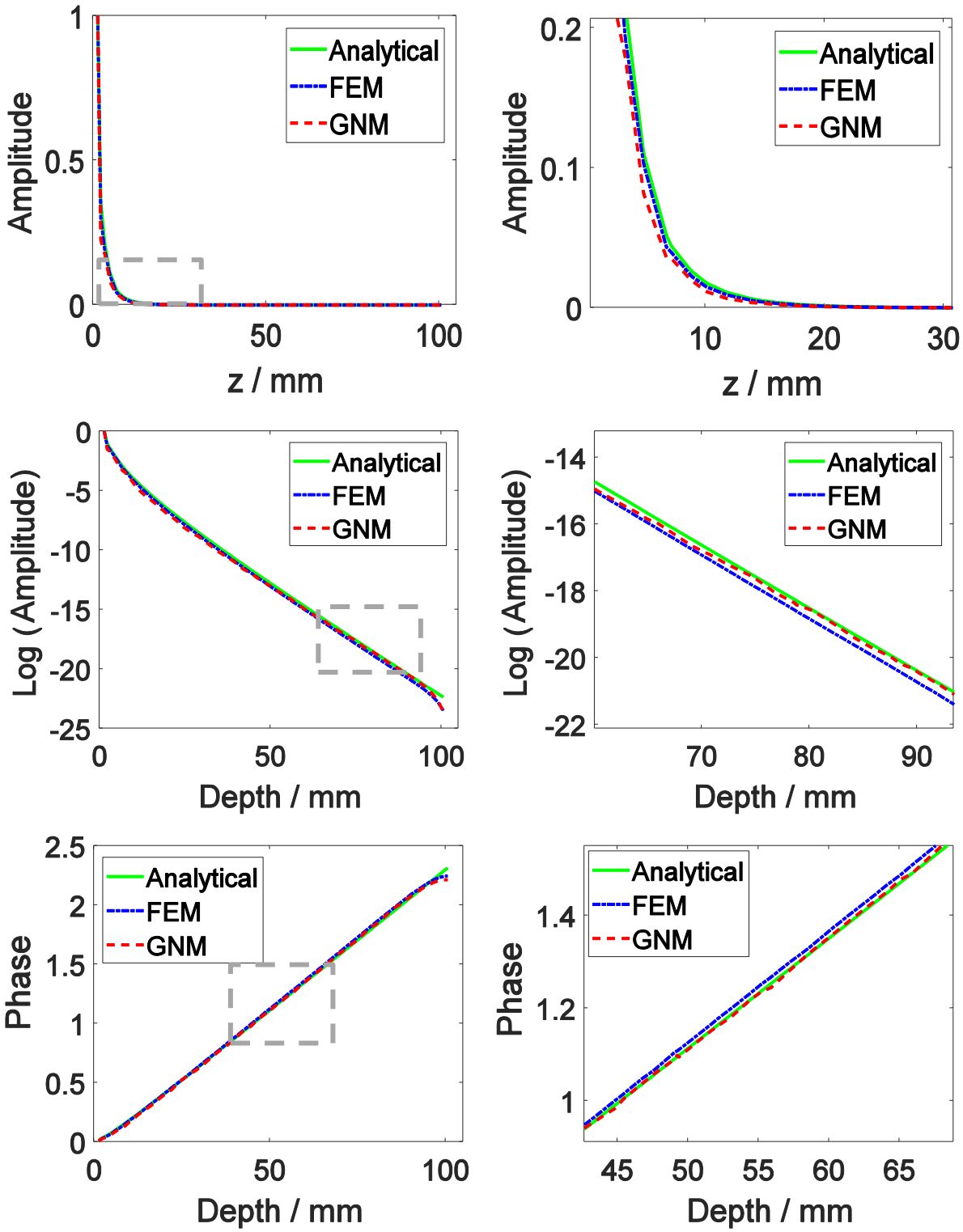


Figure 7.5: Dependency of the fluence rate as a function of distance from the source along the z axis. Right column gives the zoomed-in detail of the grey rectangle area on the left corresponding image.

Calcul
using
mode

- D:\MATLAB\Non local\3DNLMultipleWV\forward\results\comparision with different mesh resolution\NEW DATA\mesh_440k\results_440k\FD

column in Figure 7.5) that the amplitude calculated using the forward model based on GNM is closer to the analytical solution when the light penetrates deeper. The difference of the amplitudes become more evident after taking their logarithm. For the phase data,

the curves derived from both forward methods are almost parallel to the one from the analytical solution and the one by the GNM approach is much closer to the analytical solution. This finding can also be validated from table 7.1 in which the absolute difference between the two forward models and the analytical solution on the results plotted in Figure 7.4 are given.

7.4.3 3D head model

In order to illustrate the light propagation performance of different methods on a more realistic model, in this section, we use the heterogeneous 3D head model which we described in section 4.2.5.4. Five tissue types are included: scalp, scull, cerebrospinal fluid (CSF), gray matter and white matter. The optical properties for each segmented tissue region at 750nm wavelength are taken from Eggebrecht et al. [6] and listed in Table 4.2 .

First, in Table 7.3, we compare the CPU time consumed in building the sparse matrix $\mathcal{M}$ for the linear equation (7.6) using both FEM and GNM approaches. Five head meshes are created with different mesh resolutions where the number of mesh nodes is 500k, 400k, 270k, 140k and 58k respectively. In order to reduce any instability, we run each process at each setting ten times and record the mean consumed time and standard deviation (Mean $\pm$ Standard deviation). We can see that the time consumed by the GNM approach is ten times lower than the one consumed by the FEM approach. The reason is that the FEM approach needs to go through each element inside of the computational domain where the number of elements is much larger than the number of nodes and then to calculate the basis function at the corresponding nodes in each element. However GNM just needs to go though each vertex and find its connectivity with the surrounding vertices.

Next we take the head mesh with 58886 nodes as an example, to evaluate the light propagation performance by different forward models, as well as the computational efficiency of each linear solver. As shown in Figure 7.6, we place one light source on the surface of the right side of the forehead. Both the GNM and FEM approaches are implemented on the mesh, which results in two 58886×58886 sparse linear systems. Due

Table 7.3: Time consumption (sec) of building the sparse matrix $\mathcal{M}$ for the linear equation (7.6).

	500k	400k	270k	140k	58k
FEM	14.9 ± 0.2	11.7 ± 0.1	7.3 ± 0.1	3.6 ± 0.1	1.2 ± 0.1
GNM	1.6 ± 0.0	1.3 ± 0.0	0.8 ± 0.0	0.4 ± 0.0	0.1 ± 0.0

to the large system size, the direct solver becomes less efficient than the iterative solvers. Note that because different solvers lead to almost identical results, we only present the light propagation results by the BiCGSTAB+ILU solver in Figure 7.6. One can observe that the qualitative results by the FEM and GNM approaches are similar. When the light propagates away from the source, the amplitude is decreasing while the phase is increasing. These results demonstrate the consistency between FEM and GNM. In Figure 7.7, we present the plots of convergence rate (1st column) and the boxplots of computational time (2nd column) for the solvers given in Table 7.1. We define a solver converges if its relative residual defined in equation (7.7) is smaller than 1×10^{-12}. It is clear from the results in the first column that preconditioned solvers converge faster than their non-preconditioned counterparts, and that BiCGSTAB+ILU is the solver with the fastest convergence rate (less than 20 iterations). In addition, the iterative solvers for the FEM approach converge faster than those for the GNM approach, but they are more computationally expensive as confirmed from the boxplots in the second column and Table 7.3. Comparing to the FEM approach, GNM approach leads to 12%, 47%, 50% and 68% computational improvement with the LU factorization, GMRES+ILU, BiCGSTAB and BiCGSTAB+ILU linear solvers respectively.

7.4.4 Image reconstruction

In the last section, we have evaluated the graph-based forward model on the continuous wave simulations and promising results have been achieved. In this section, we consider the reconstruction performance by different strategies on real experimental data which is

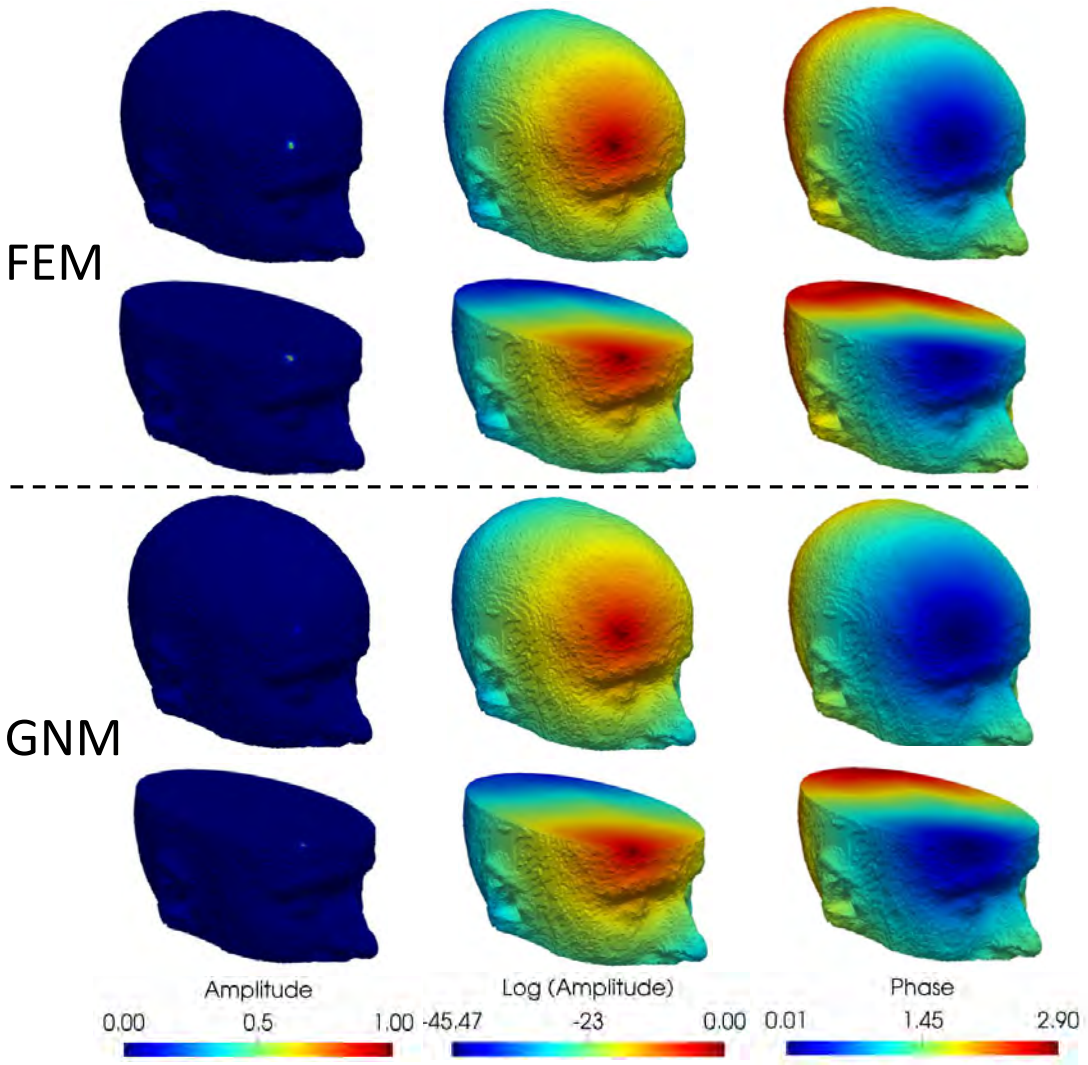


Figure 7.6: Comparison of the light propagation results obtained by FEM and GNM approaches through the FD system, as shown in the first and last two rows, respectively; Row 2 and 4 represent the corresponding clipped results in Row 1 and 3; Column 1-3 show the amplitude, logarithm of amplitude and phase of photon fluence Φ, respectively.

measured over the visual cortex of a human subject. The frequency domain measuring device is a functional brain imaging system using infrared photons (IMAGENT^{TM}, ISS Inc., Illinois). This system is composed of 32 sources and 30 detectors. The sources are modulated at 140 MHz and coupled to laser diodes emitting at 690 nm and 830 nm. The subject was recruited from the University of Birmingham, contacted via word of mouth and required to sign informed consent forms prior to participating. The subject was seated facing an adjustable screen while the imaging pad was attached over the occipital cortex with hook and loop strapping. Visual stimuli consisted of rotating logarithmic

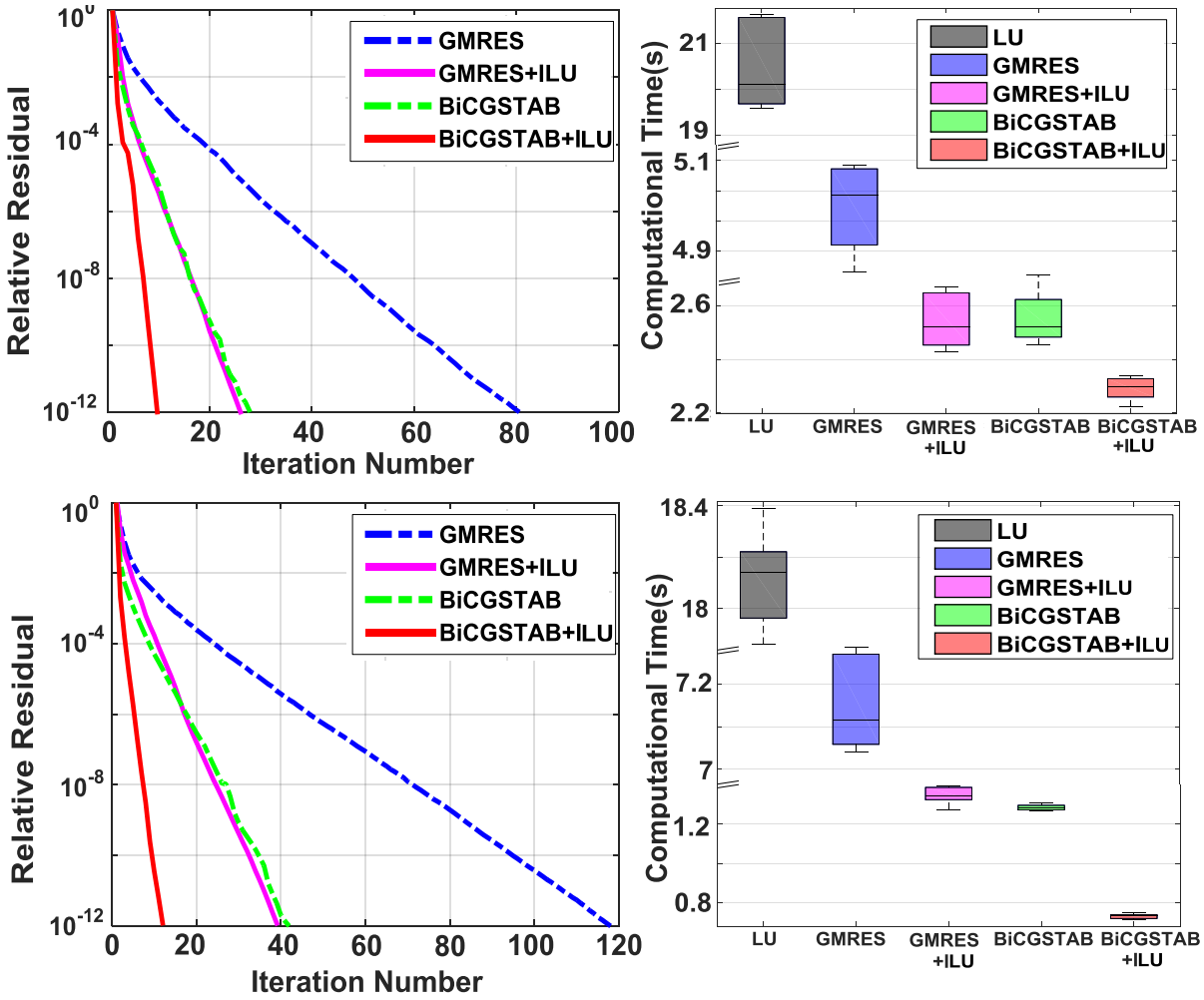


Figure 7.7: Comparison of convergence rate and computational time of different solvers using the FD system. First column: the plots of relative residual versus number of iterations; Second column: the boxplots of computational time versus different solvers; First row gives the plots by the FEM approach while the second row by the GNM approach.

checkerboards, reversing intensity at 10 Hz, on a 50% intensity gray background. The grids rotate around a white cross located at the center of the visual field. Each rotation completed in 36 seconds and 10 rotations occurred in the run. Gray screens are presented for 30 seconds before and after the total scanning session as the baseline periods, resulting in 7 minutes total scanning session time.

The imaging cap consisted of 24 source positions (two NIR wavelengths 690 nm and 830 nm at each position) and 28 detectors interlaced in a high-density array (Figure 7.8). First to fourth nearest neighbour (13, 30, 39, 48 mm respectively) optode pairs were used, giving a total of 348 total possible measurements, at a frame rate of 40 Hz.

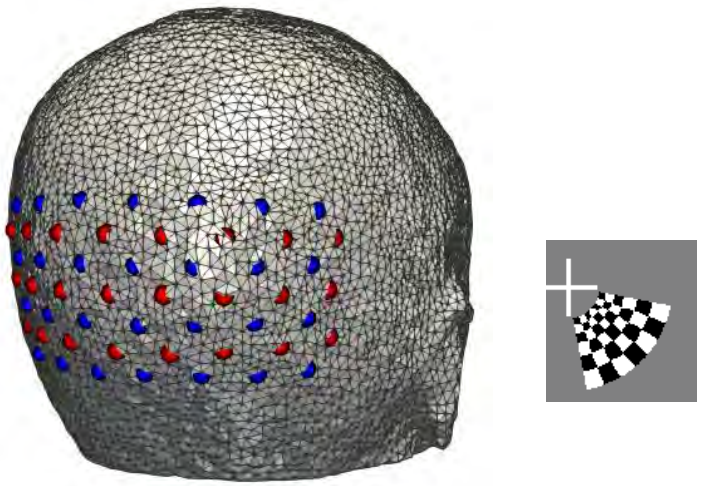
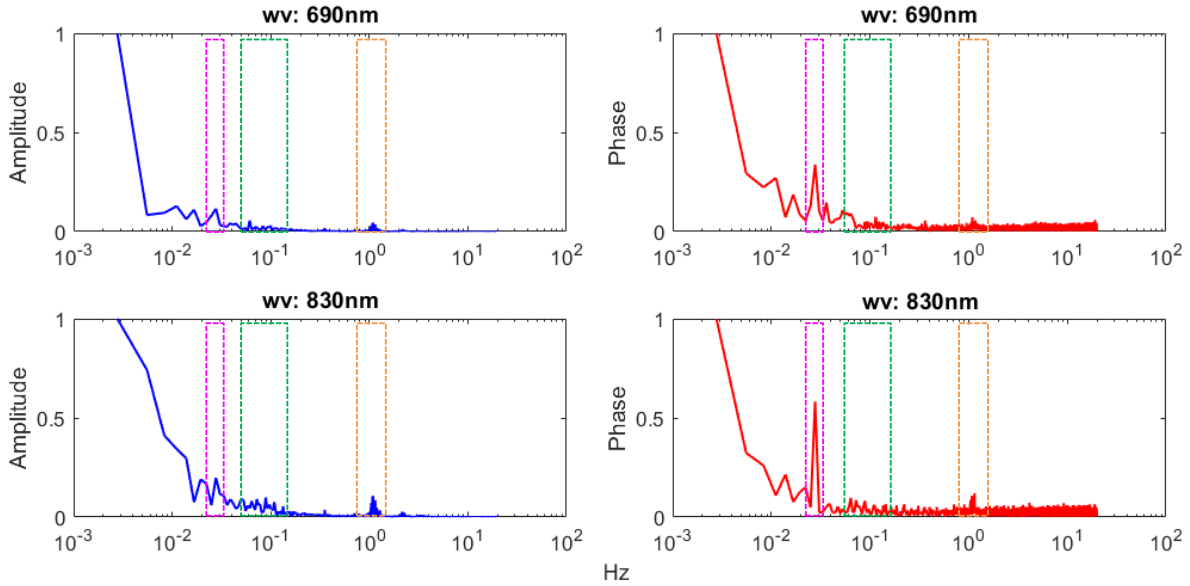


Figure 7.8: High-density imaging grid with the distribution of 24 sources (red) and 28 detectors (blue).

The measurement amplitude was converted to log-ratio to allow for the recovery of change of optical properties. High-pass filter (0.01Hz cutoff) and low-pass filter (0.09Hz cutoff) are used to remove the unwanted pulse signals. The channels which standard deviation of the log-ratio of the data is larger than 7.5% were removed to alleviate the effect of channels with high noise. In Figure 7.9, we plot the fourier transforms of the time trace for one first nearest source-detector pair. From the amplitude data (left column), the peak due to the stimuli (∼0.028Hz) is not obvious because the low separation only samples the superficial tissues. However for the phase data, the peak due to the stimuli is evident even in the nearest separation (13mm), which illustrates that the phase data can sample the deeper brain tissue. The peaks due to breathing and cardiac pulse are much smaller than the one due to the visual stimuli. Therefore the phase data is less sensitive to signal contamination from the superficial tissues. This finding is consistent with the findings of Doulgerakis et al. [187] who showed that the FD measurements can be used to resolve activations in deeper regions. We remove a global signal derived from first-nearest-neighbor measurements from all measurement amplitudes by linear regression. After down-sampling to 1 Hz, and block averaged in 10 blocks of 72 s each, the time traces were used for image reconstruction.

NN1



- So only do regression on amplitude

Figure 7.9: Fourier transforms of the time traces obtained using the FD system on the first nearest neighbour separation. The y-axis shows power (arbitrary units) and x-axis depicts frequency (Hz). First row: 690 nm; second row: 830 nm. Left column (red plots) correspond to amplitude data, whereas the right column (blue plots) correspond to Phase data. Areas denoted by pink, green and yellow dashed lines represent the rotation frequency of the stimulus wedge (∼0.028Hz), Breathing (∼0.1 Hz) and Cardiac pulse (∼1Hz).

For reconstruction, a five-layer, hemispheric head model was used with around 70k discretized nodes (Figure 7.8). Differential measurements which are measured at two near-infrared wavelengths can be used to recover the chromophore concentration changes (δ HbO_2 and δHb). Therefore in this work, we calculated the absorption coefficient at each wavelength and then unmix them spectrally to recover δ HbO_2 and δHb respectively. The image recovery process for functional imaging follows Doulgerakis et al. [187] but under two different discretization strategies: FEM and GNM. Comparing with the whole medium, the regions of activation are sparse so the vector corresponding to the difference in the chromophore concentrations relative to the background is sparse with only a few non-zero elements. Therefore in this work, instead of using the total variation regularization, we use the L_1-norm based regularization which is a pure sparsity preserv-

ing regularization [162] to recover the activation. We also compare it with the results by Tikhonov regularization [187] which is popular used in functional imaging reconstruction.

Figure 7.10 and Figure 7.11 give the reconstruction results under FEM and GNM discretization strategy respectively. Color maps in the first, third rows were normalized within each frame to positive maximum contrast and the negative values were set as zero. No further thresholding is applied on the color map. It can be seen that more localized activation can be recovered by the L_1-norm regularization under GNM in response to each presentation of the stimulus. It further verifies the superior performance of the proposed discretization and regularization methods on real experimental data.

7.5 Conclusion

In this work, by using the concepts of differential operators in the framework of the nonlocal vector calculus, we proposed the nonlocal diffusion equation (NDE) in the frequency domain as a new forward model to accurately describe light propagation in biological tissues. The discretization of the NDE was performed using an effective and efficient algorithm based on the graph scheme, and the resulting linear systems were solved by various linear solvers. Extensive numerical experiments on simulated and realistic meshes validated the performance of the proposed approaches for modelling light transmission through different complex geometries in either 2D or 3D space. The contributions can be summarized as follows:

- We proposed a new light migration model by applying the nonlocal differential operators to the original diffusion equation (DE) and its Robin boundary condition (RBC). To the best of our knowledge, this is the first time that the nonlocal vector calculus is used for the frequency domain forward modelling in the area of diffuse optical tomography (DOT) imaging.

- We developed the graph-based numerical method (GNM) to discretize the proposed NDE with its nonlocal RBC. Such numerical algorithm has several advantages over

the finite element method (FEM) for the DE with its local RBC, including its capability of easily processing complex geometric shapes of objects, of being more computationally efficient, and of allowing identical implementation for both 2D and 3D geometries.

- We studied various linear solvers so as to find efficient ones for the linear systems resulted from applying the GNM to the NDE for the FD DOT measurement systems. We have found that the iterative solvers from Krylov subspace are generally more efficient than other solvers. Furthermore, the computational speed of the Krylov subspace methods can be drastically increased via the preconditioning technique. For the frequency domain (FD) measurement system, the experimental results show that the biconjugate gradient stabilised method with incomplete LU preconditioner (BiCGSTAB+ILU) is the fastest solver in terms of both computational time and convergence rate.

- We compared our NDE model implemented with the GNM with the original DE model implemented with the FEM, and validated via extensive experiments that the proposed NDE method is compatible to the DE approach in terms of performance of both light propagation and image reconstruction. However, our model is more efficient.

Ongoing research includes further boosting the computational efficiency of the proposed NDE method by investigating more advanced preconditioners and using graphics processing unit (GPU) programming and applying the NDE and its GNM to the time domain measurement system.

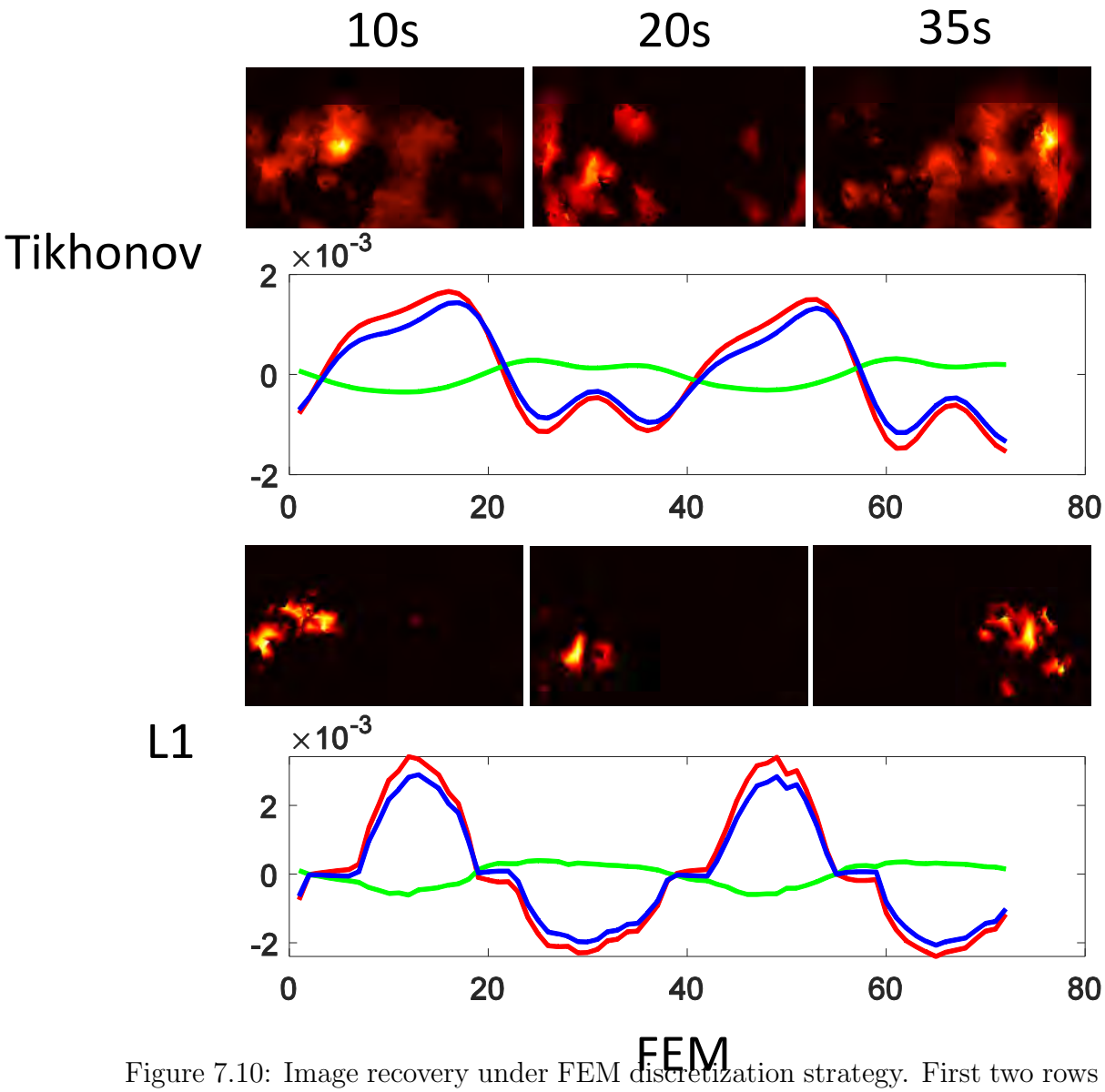


Figure 7.10: Image recovery under FEM discretization strategy. First two rows represent the results using Tikhonov regularization while the last two rows represent the results using L_1-norm based regularization. In each group, first row: response of δHbO_2 at three time points (10s, 20s and 35s); second row: time trace of hemoglobin concentrations in a single voxel (δHbO_2, red; δHb, green; δHbT, blue, where $\delta HbT = \delta HbO_2 + \delta Hb$) over 72s.

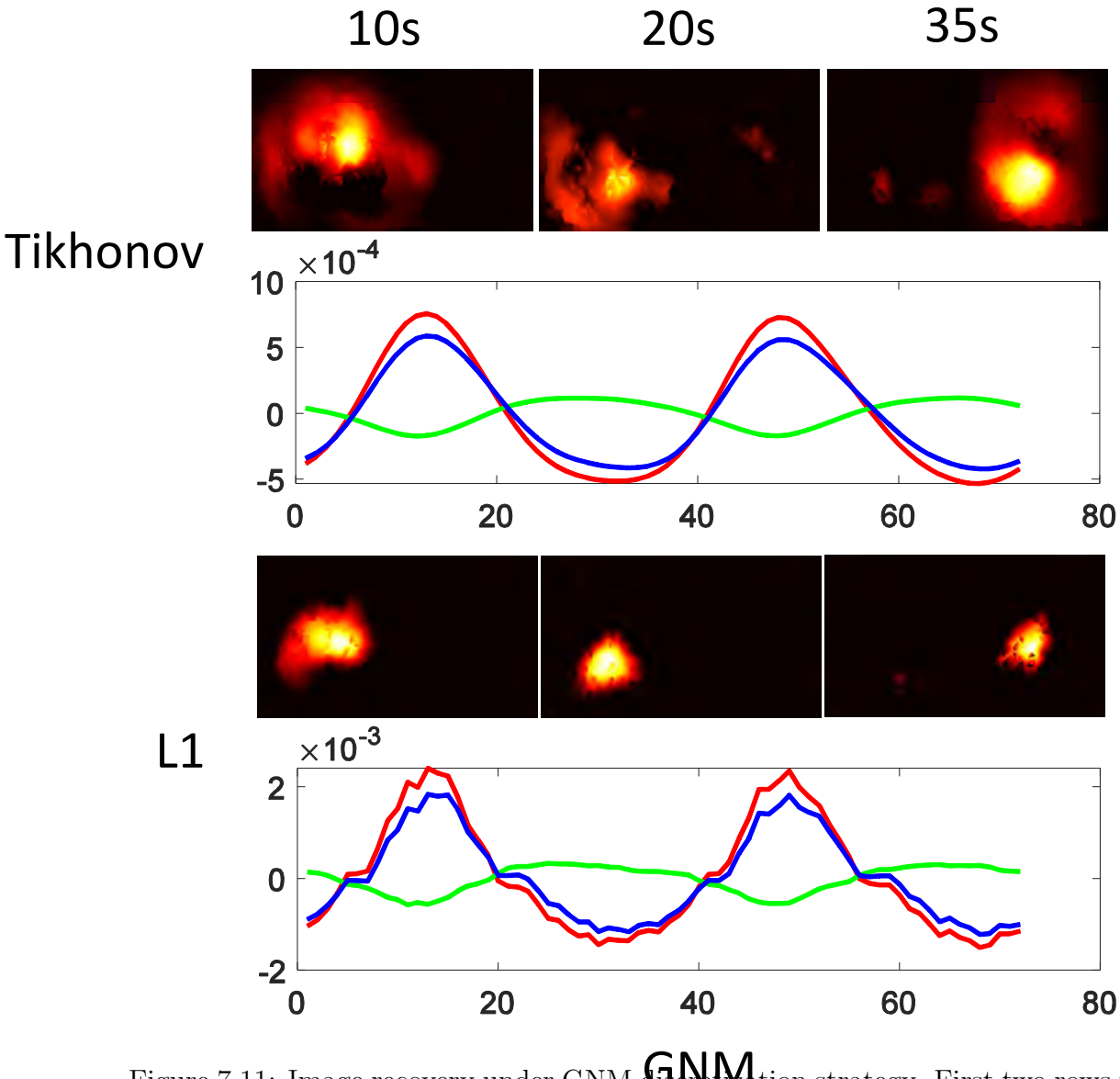


Figure 7.11: Image recovery under GNM discretization strategy. First two rows represent the results using Tikhonov regularization while the last two rows represent the results using L_1-norm based regularization. In each group, first row: response of δHbO_2 at three time points (10s, 20s and 35s); second row: time trace of hemoglobin concentrations in a single voxel (δHbO_2, red; δHb, green; δHbT, blue, where $\delta HbT = \delta HbO_2 + \delta Hb$) along 72s.

CHAPTER 8

SUMMARY AND FUTURE WORK

8.1 Summary

In this thesis, we proposed advanced regularization and discretization methods for modeling light propagation and image reconstruction in diffuse optical tomography (DOT). DOT is a non-invasive and non-ionising imaging technology, in which the near-infrared light with wavelength from 650−900nm is injected into the object to be imaged (e.g. breast, head and mouse). The light, which propagates through the internal tissue and emerges at the surface of the object, is then measured and utilized to reconstruct the internal distribution of tissue's optical properties. The image recovery process in DOT is divided into two steps: forward modeling and inverse reconstruction. The forward model is a fundamental pillar for such reconstruction, because it must be able to accurately model the main interactions (i.e. absorption and scattering) between light and the object so as to recover internal properties faithfully. Due to limited number of measurements, the inverse reconstruction process is ill-posed and regularization methods are needed to yield the physiologically and anatomically plausible reconstructed solutions.

This thesis focused on developing both forward modeling methods and advanced regularization methods to improve the accuracy DOT image reconstruction. For forward and inverse models, we also proposed new discretization methods for accurate implementations. Of note, the methods proposed are not limited to DOT only. Our proposed discretization methods can be further applied to any technique that uses a diffusion-based

model of light propagation, such as diffuse correlation spectroscopy and near infrared spectroscopy. The proposed regularization methods can be applied to other imaging techniques to constrain the reconstruction process to get plausible solutions.

This thesis consists of four main distinguishable parts related to the publications that have been produced by the author as a part of the PhD studies [162, 188–190]. In the following, we summarize the thesis by highlighting our contributions which have been introduced in more details in Chapters 4, 5, 6 and 7.

- Chapter 2 introduces diffuse optical tomography, including detailed descriptions of modelling of light propagation (forward model) and image reconstruction (inverse model). Stochastic or deterministic approximate light propagation models for simplifying the radiative transfer equation are discussed in detail. This chapter also discussed the linear approach and nonlinear iterative approach for the inverse model.

- Chapter 3 describes how the 2D and 3D partial differential equations in diffuse optical tomography imaging reconstruction can be numerically solved by the Galerkin finite element method.

- Chapter 4 reviews the regularization methods proposed in the DOT community and proposes L_1-norm regularization based nonlinear DOT reconstruction for spectrally constrained diffuse optical tomography. It is the first time that L_1-norm regularization methods and spectrally constrained DOT methods have been used together. It gives detailed descriptions of how this can be done, and performed systematic comparisons of the performance and efficiency of the different methods on both simulated and real data. Compared to the L_2-norm, L_1-norm regularization is proved to reduce crosstalk and maintain image contrast by inducing sparsity. Among all the algorithms to solve the proposed model, FISTA performs marginally better than ADMM and IRLS by the measures of AC, PC, and PSNR, and is more than 10 times faster than ADMM and IRLS as it avoids direct matrix inversion and large

matrix-matrix multiplications.

- Chapter 5 introduces the gradient differential operator to alleviate the over-sparsifying effect induced from the pure sparsity preserving regularizations on the non-sparse images. Two regularizations are proposed: Tikhonov regularizationwith a spatial gradient operator and Total variation (TV) regularization, which uses the L_1-norm of the gradient of the solution. Graph-based numerical methods are proposed in this chapter to model unstructured geometries of DOT objects. Results show that TV regularization can alleviate the over-smoothing effect of Tikhonov regularization and localize the anomaly by inducing sparsity of the gradient of the solution. Comparing with the popular used finite element-based numerical method, the inverse model implemented under the graph-based numerical method is shown to be more stable and robust to changes in mesh resolution. From the work in section 5.4.5.3.1, the performance of FETV improves with an increase of mesh resolution: by 25% in localization error, 26% in average contrast and 11% in PSNR, while the performance of GTV has relatively fewer difference with different mesh resolutions.

- Chapter 6 applies the proposed graph-based numerical method to the forward model in the continuous-wave(CW) DOT. A new formulation of the forward model for CW DOT is proposed and it is based on the concepts of differential operators under a nonlocal vector calculus. The proposed model is shown to be able to accurately model the light propagation in the medium and is quantitatively comparable with both analytical and FEM forward models. It is more computational efficient and allows identical implementation for geometries in any dimension. This computational improvement is validated in section 6.3.1. Our proposed model is shown to be more computationally efficient with an average speed improvement of 30% compare to the FEM forward model.

- Chapter 7 proposed the nonlocal diffusion equation (NDE) in frequency domain as a new forward model to accurately describe light propagation in biological tis-

sues. The discretization of the NDE was performed using an effective and efficient algorithm based on the graph scheme, and the resulting linear systems were tackled by various linear solvers. Extensive numerical experiments on simulated and realistic meshes validated the performance of the proposed approaches for modelling light transmission through different complex geometries in either 2D or 3D space. The biconjugate gradient stabilised method with incomplete LU preconditioner (BiCGSTAB+ILU) is proved to be the fastest solver in terms of both computational time and convergence rate. Comparing to the FEM approach, GNM approach leads to 12%, 47%, 50% and 68% computational improvement with the LU factorization, GMRES+ILU, BiCGSTAB and BiCGSTAB+ILU linear solvers respectively.

8.2 Future work

Although the methods proposed in the thesis have achieved significant success, there is still room for improvement of these methods. For example, (1) for a large-scale forward model (e.g. from a 3D head mesh) discretized using either the FEM and the graph method, we end up with a large linear system, for which it is very time consuming to solve it numerically. (2) In the inverse model, for each subject the smoothness hyper-parameter λ needs to be tuned carefully to deliver a precise reconstruction. While a too small λ leads to an irregular and non-smooth reconstruction, setting it too large reduces the reconstruction magnitude and therefore loses the ability to modeling anatomical structure accurately. (3) The hand-crafted regularization term itself is another hyper-parameter, which is usually selected based on the assumption about the solution, such as piecewise constant, piecewise smooth, etc. However, existing assumptions may be too simple to capture complex changes of image content associated with biological tissues. (4) The inverse model is nonlinear and therefore needs to be optimized iteratively, which is a very time-consuming process especially for high-dimensional data inputs.

Recently, many deep learning approaches have been very popular for image classifi-

cation and recognition. Good feature extractors designed by hand normally (such as the regularizations proposed in the thesis) require a considerable amount of engineering skill and domain expertise. This however can be avoided if good features can be learned automatically using a general-purpose learning procedure. This is why deep learning starts to play an important role in medical image analysis by automatically learning morphological and/or patterns from images without using man craft features. Deep learning methods have achieved state-of-the-art performance across various medical image processing and reconstruction applications. Breakthrough improvements were achieved by using a large data set of medical images from which deep models can find more generalised features and thus lead to improved performance. In addition, due to the powerful GPU, deep learning approaches are much faster. Our next future work therefore will focus on image reconstruction using deep learning techniques, with the aim to overcome the existing limitations mentioned above.

Bibliography

[1] Robert Percuoco. Plain radiographic imaging. In *Clinical Imaging*, pages 1–43. Elsevier, 2014.

[2] Turgut Durduran, Regine Choe, Wesley B Baker, and Arjun G Yodh. Diffuse optics for tissue monitoring and tomography. *Reports on Progress in Physics*, 73(7):076701, 2010.

[3] David J Davies, Zhangjie Su, Michael T Clancy, Samuel JE Lucas, Hamid Dehghani, Ann Logan, and Antonio Belli. Near-infrared spectroscopy in the monitoring of adult traumatic brain injury: a review. *Journal of Neurotrauma*, 32(13):933–941, 2015.

[4] Diffuse optical tomography and imaging. http://www.atomic.physics.lu.se/biophotonics/research/diffuse-optical-tomography-and-imaging/.

[5] Yoko Hoshi and Yukio Yamada. Overview of diffuse optical tomography and its clinical applications. *Journal of Biomedical Optics*, 21(9):091312, 2016.

[6] Adam T Eggebrecht, Brian R White, Silvina L Ferradal, Chunxiao Chen, Yuxuan Zhan, Abraham Z Snyder, Hamid Dehghani, and Joseph P Culver. A quantitative spatial comparison of high-density diffuse optical tomography and fmri cortical mapping. *Neuroimage*, 61(4):1120–1128, 2012.

[7] W Calvo, JW Hopewell, HS Reinhold, and TK Yeung. Time-and dose-related changes in the white matter of the rat brain after single doses of x rays. *The British journal of radiology*, 61(731):1043–1052, 1988.

[8] Mrinalini Balki, Yung Lee, Stephen Halpern, and Jose CA Carvalho. Ultrasound imaging of the lumbar spine in the transverse plane: the correlation between estimated and actual depth to the epidural space in obese parturients. *Anesthesia & Analgesia*, 108(6):1876–1881, 2009.

[9] Joel J Niederhauser, Michael Jaeger, Robert Lemor, Peter Weber, and Martin Frenz. Combined ultrasound and optoacoustic system for real-time high-contrast vascular imaging in vivo. *IEEE transactions on medical imaging*, 24(4):436–440, 2005.

[10] Thilaka S Sumanaweera, John R Adler Jr, Sandy Napel, and Gary H Glover. Characterization of spatial distortion in magnetic resonance imaging and its implications for stereotactic surgery. *Neurosurgery*, 35(4):696–704, 1994.

[11] Ian L Pykett, Jeffrey H Newhouse, Ferdinando S Buonanno, Thomas J Brady, Mark R Goldman, J Philip Kistler, and Gerald M Pohost. Principles of nuclear magnetic resonance imaging. *Radiology*, 143(1):157–168, 1982.

[12] Theodore J Huppert, Richard D Hoge, Solomon Gilbert Diamond, Maria Angela Franceschini, and David A Boas. A temporal comparison of bold, asl, and nirs hemodynamic responses to motor stimuli in adult humans. *Neuroimage*, 29(2):368–382, 2006.

[13] AP Gibson, JC Hebden, and Simon R Arridge. Recent advances in diffuse optical imaging. *Physics in Medicine and Biology*, 50(4):R1–R43, 2005.

[14] Subhadra Srinivasan, Brian W Pogue, Shudong Jiang, Hamid Dehghani, Christine Kogel, Sandra Soho, Jennifer J Gibson, Tor D Tosteson, Steven P Poplack, and Keith D Paulsen. Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured in vivo by near-infrared breast tomography. *Proceedings of the National Academy of Sciences*, 100(21):12349–12354, 2003.

[15] Hamid Dehghani, Brian W Pogue, Steven P Poplack, and Keith D Paulsen. Multi-wavelength three-dimensional near-infrared tomography of the breast: initial simulation, phantom, and clinical results. *Applied Optics*, 42(1):135–145, 2003.

[16] DA Boas, K Chen, D Grebert, and MA Franceschini. Improving the diffuse optical imaging spatial resolution of the cerebral hemodynamic response to brain activation in humans. *Optics letters*, 29(13):1506–1508, 2004.

[17] Benjamin W Zeff, Brian R White, Hamid Dehghani, Bradley L Schlaggar, and Joseph P Culver. Retinotopic mapping of adult human visual cortex with high-density diffuse optical tomography. *Proceedings of the National Academy of Sciences*, 104(29):12169–12174, 2007.

[18] Anna Custo, David A Boas, Daisuke Tsuzuki, Ippeita Dan, Rickson Mesquita, Bruce Fischl, W Eric L Grimson, and Williams Wells. Anatomical atlas-guided diffuse optical tomography of brain activation. *Neuroimage*, 49(1):561–567, 2010.

[19] Haijing Niu, Sabin Khadka, Fenghua Tian, Zi-Jing Lin, Chunming Lu, Chaozhe Zhu, and Hanli Liu. Resting-state functional connectivity assessed with two diffuse optical tomographic systems. *Journal of biomedical optics*, 16(4):046006–046006, 2011.

[20] Hamid Dehghani, Brian W Pogue, Jiang Shudong, Ben Brooksby, and Keith D Paulsen. Three-dimensional optical tomography: resolution in small-object imaging. *Applied Optics*, 42(16):3117–3128, 2003.

[21] Xu Heng, Roger Springett, Hamid Dehghani, Brian W Pogue, Keith D Paulsen, and Jeff F Dunn. Magnetic-resonance-imaging–coupled broadband near-infrared tomography system for small animal brain studies. *Applied optics*, 44(11):2177–2188, 2005.

[22] Phaneendra K Yalavarthy, Brian W Pogue, Hamid Dehghani, Colin M Carpenter, Shudong Jiang, and Keith D Paulsen. Structural information within regulariza-

tion matrices improves near infrared diffuse optical tomography. *Optics Express*, 15(13):8043–8058, 2007.

[23] Phaneendra K Yalavarthy, Brian W Pogue, Hamid Dehghani, and Keith D Paulsen. Weight-matrix structured regularization provides optimal generalized least-squares estimate in diffuse optical tomography. *Medical Physics*, 34(6Part1):2085–2098, 2007.

[24] Daniel R Lynch. *Numerical partial differential equations for environmental scientists and engineers: a first practical course*. Springer Science & Business Media, 2004.

[25] Judith R Mourant, Tamika Fuselier, James Boyer, Tamara M Johnson, and Irving J Bigio. Predictions and measurements of scattering and absorption over broad wavelength ranges in tissue phantoms. *Applied Optics*, 36(4):949–957, 1997.

[26] Hugo J Van Staveren, Christian JM Moes, Jan van Marie, Scott A Prahl, and Martin JC Van Gemert. Light scattering in lntralipid-10% in the wavelength range of 400–1100 nm. *Applied Optics*, 30(31):4507–4514, 1991.

[27] Brian W Pogue and Michael S Patterson. Frequency-domain optical absorption spectroscopy of finite tissue volumes using diffusion theory. *Physics in Medicine & Biology*, 39(7):1157, 1994.

[28] AH Hielscher, AY Bluestone, GS Abdoulaev, AD Klose, J Lasker, M Stewart, U Netz, and J Beuthan. Near-infrared diffuse optical tomography. *Disease Markers*, 18(5-6):313–337, 2002.

[29] Florian EW Schmidt, Martin E Fry, Elizabeth MC Hillman, Jeremy C Hebden, and David T Delpy. A 32-channel time-resolved instrument for medical optical tomography. *Review of Scientific Instruments*, 71(1):256–265, 2000.

[30] Davide Contini, Lucia Zucchelli, Lorenzo Spinelli, Matteo Caffini, Rebecca Re, Antonio Pifferi, Rinaldo Cubeddu, and Alessandro Torricelli. Brain and muscle near

infrared spectroscopy/imaging techniques. *Journal of Near Infrared Spectroscopy*, 20(1):15–27, 2012.

[31] James J Duderstadt and William Russell Martin. Transport theory. *Transport Theory., by Duderstadt, JJ; Martin, WR. Chichester (UK): John Wiley & Sons, 10+ 613 p.*, 1979.

[32] Akira Ishimaru. *Wave propagation and scattering in random media*, volume 2. Academic press New York, 1978.

[33] Pei Wang, Simon R Arridge, and Ming Jiang. Radiative transfer equation for media with spatially varying refractive index. *Physical Review A*, 90(2):023803, 2014.

[34] Alexander D Klose. The forward and inverse problem in tissue optics based on the radiative transfer equation: a brief review. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 111(11):1852–1853, 2010.

[35] Ashley J Welch, Martin JC Van Gemert, et al. *Optical-thermal response of laser-irradiated tissue*, volume 2. Springer, 2011.

[36] Alexander D Klose, Uwe Netz, Jürgen Beuthan, and Andreas H Hielscher. Optical tomography using the time-independent equation of radiative transferpart 1: forward model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 72(5):691–713, 2002.

[37] Tanja Tarvainen, Marko Vauhkonen, Ville Kolehmainen, and Jari P Kaipio. Hybrid radiative-transfer–diffusion model for optical tomography. *Applied Optics*, 44(6):876–886, 2005.

[38] Andreas H Hielscher, Raymond E Alcouffe, and Randall L Barbour. Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues. *Physics in Medicine and Biology*, 43(5):1285, 1998.

[39] Andreas H Hielscher, Alexander D Klose, Alexander K Scheel, Bryte Moa-Anderson, Marina Backhaus, Uwe Netz, and Jürgen Beuthan. Sagittal laser optical tomography for imaging of rheumatoid finger joints. *Physics in Medicine and Biology*, 49(7):1147, 2004.

[40] Alexander D Klose and Andreas H Hielscher. Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer. *Medical Physics*, 26(8):1698–1707, 1999.

[41] G Eason, AR Veitch, RM Nisbet, and FW Turnbull. The theory of the back-scattering of light by blood. *Journal of Physics D: Applied Physics*, 11(10):1463, 1978.

[42] Michael S Patterson, Britton Chance, and Brian C Wilson. Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties. *Applied Optics*, 28(12):2331–2336, 1989.

[43] Simon R Arridge, M Cope, and DT Delpy. The theoretical basis for the determination of optical pathlengths in tissue: temporal and frequency analysis. *Physics in Medicine and Biology*, 37(7):1531, 1992.

[44] William H Reed. New difference schemes for the neutron transport equation. *Nuclear Science and Engineering*, 46(2):309–314, 1971.

[45] KD Lathrop. Discrete-ordinates methods for the numerical solution of the transport equation. *Reactor Technol*, 15(2):107–135, 1972.

[46] T Tarvainen, M Vauhkonen, and SR Arridge. Gauss–newton reconstruction method for optical tomography using the finite element solution of the radiative transfer equation. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 109(17):2767–2778, 2008.

[47] Pedro González-Rodríguez and Arnold D Kim. Comparison of light scattering models for diffuse optical tomography. *Optics Express*, 17(11):8756–8774, 2009.

[48] BC Wilson and G Adam. A monte carlo model for the absorption and flux distributions of light in tissue. *Medical Physics*, 10(6):824–830, 1983.

[49] Lihong Wang and Steven L Jacques. Monte carlo modeling of light transport in multi-layered tissues in standard c. *The University of Texas, MD Anderson Cancer Center, Houston*, pages 4–11, 1992.

[50] David A Boas, JP Culver, JJ Stott, and AK Dunn. Three dimensional monte carlo code for photon migration through complex heterogeneous media including the adult human head. *Optics Express*, 10(3):159–170, 2002.

[51] Eiji Okada and David T Delpy. Near-infrared light propagation in an adult head model. i. modeling of low-level scattering in the cerebrospinal fluid layer. *Applied Optics*, 42(16):2906–2914, 2003.

[52] Toshiyuki Hayashi, Yoshihiko Kashio, and Eiji Okada. Hybrid monte carlo-diffusion method for light propagation in tissue with a low-scattering region. *Applied Optics*, 42(16):2888–2896, 2003.

[53] Y Phaneendra Kumar and Ram M Vasu. Reconstruction of optical properties of low-scattering tissue using derivative estimated through perturbation monte-carlo method. *Journal of Biomedical Optics*, 9(5):1002–1012, 2004.

[54] Juha Heiskala, Mika Pollari, Marjo Metsäranta, P Ellen Grant, and Ilkka Nissilä. Probabilistic atlas can improve reconstruction from optical imaging of the neonatal brain. *Optics Express*, 17(17):14977–14992, 2009.

[55] Amir H Gandjbakhche and George H Weiss. V: Random walk and diffusion-like models of photon migration in turbid media. *Progress in Optics*, 34:333–402, 1995.

[56] George H Weiss, Josep M Porrà, and Jaume Masoliver. The continuous-time random walk description of photon motion in an isotropic medium. *Optics Communications*, 146(1):268–276, 1998.

[57] Victor Chernomordik, David Hattery, Amir H Gandjbakhche, Antonio Pifferi, Paola Taroni, Alessandro Torricelli, Gianluca Valentini, and Rinaldo Cubeddu. Quantification by random walk of the optical parameters of nonlocalized abnormalities embedded within tissuelike phantoms. *Optics Letters*, 25(13):951–953, 2000.

[58] Victor Chernomordik, David W Hattery, Israel Gannot, Giovanni Zaccanti, and Amir Gandjbakhche. Analytical calculation of the mean time spent by photons inside an absorptive inclusion embedded in a highly scattering medium. *Journal of Biomedical Optics*, 7(3):486–492, 2002.

[59] Victor Chernomordik, David W Hattery, Dirk Grosenick, Heidrun Wabnitz, Herbert Rinneberg, K Thomas Moesta, Peter M Schlag, and Amir Gandjbakhche. Quantification of optical properties of a breast tumor using random walk theory. *Journal of Biomedical Optics*, 7(1):80–87, 2002.

[60] Songwan Jin, Zsolt Zador, and AS Verkman. Random-walk model of diffusion in three dimensions in brain extracellular space: comparison with microfiberoptic photobleaching measurements. *Biophysical Journal*, 95(4):1785–1794, 2008.

[61] Sarah Kathryn Patch. Recursive recovery of markov transition probabilities from boundary value data. 1994.

[62] FA Grunbaum. Diffuse tomography: the isotropic case. *Inverse Problems*, 8(3):409, 1992.

[63] FA Grunbaum and Jorge P Zubelli. Diffuse tomography: computational aspects of the isotropic case. *Inverse Problems*, 8(3):421, 1992.

[64] Simon R Arridge and Jeremy C Hebden. Optical imaging in medicine: Ii. modelling and reconstruction. *Physics in Medicine and Biology*, 42(5):841, 1997.

[65] Simon R Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41, 1999.

[66] Martin Schweiger, SR Arridge, M Hiraoka, and DT Delpy. The finite element method for the propagation of light in scattering media: boundary and source conditions. *Medical Physics*, 22(11):1779–1792, 1995.

[67] Hamid Dehghani, Ben Brooksby, Karthik Vishwanath, Brian W Pogue, and Keith D Paulsen. The effects of internal refractive index variation in near-infrared optical tomography: a finite element modelling approach. *Physics in Medicine & Biology*, 48(16):2713, 2003.

[68] Wenqi Lu, Jinming Duan, Zhaowen Qiu, Zhenkuan Pan, Ryan Wen Liu, and Li Bai. Implementation of high-order variational models made easy for image processing. *Mathematical Methods in the Applied Sciences*, 39(14):4208–4233, 2016.

[69] Simon R Arridge and John C Schotland. Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12):123010, 2009.

[70] Simon R Arridge. Photon-measurement density functions. part i: Analytical forms. *Applied Optics*, 34(31):7395–7409, 1995.

[71] Simon Robert Arridge, Martin Schweiger, and David T Delpy. Iterative reconstruction of near-infrared absorption images. In *Inverse Problems in Scattering and Imaging*, volume 1767, pages 372–383. International Society for Optics and Photonics, 1992.

[72] MA OLeary, DA Boas, B Chance, and AG Yodh. Experimental images of heterogeneous turbid media by frequency-domain diffusing-photon tomography. *Optics Letters*, 20(5):426–428, 1995.

[73] Adam T Eggebrecht, Silvina L Ferradal, Amy Robichaux-Viehoever, Mahlega S Hassanpour, Hamid Dehghani, Abraham Z Snyder, Tamara Hershey, and Joseph P Culver. Mapping distributed brain function and networks with diffuse optical tomography. *Nature Photonics*, 8(6):448, 2014.

[74] Nicholas M Gregg, Brian R White, Benjamin W Zeff, Andrew J Berger, and Joseph P Culver. Brain specificity of diffuse optical imaging: improvements from superficial signal regression and tomography. *Frontiers in Neuroenergetics*, 2:14, 2010.

[75] Vadim A Markel and John C Schotland. Inverse problem in optical diffusion tomography. i. fourier–laplace inversion formulas. *JOSA A*, 18(6):1336–1347, 2001.

[76] Vadim A Markel, Vivek Mital, and John C Schotland. Inverse problem in optical diffusion tomography. iii. inversion formulas and singular-value decomposition. *JOSA A*, 20(5):890–902, 2003.

[77] Xingde Li, Deva N Pattanayak, Turgut Durduran, Joseph P Culver, Britton Chance, and Arjun G Yodh. Near-field diffraction tomography with diffuse photon density waves. *Physical Review E*, 61(4):4295, 2000.

[78] DA Dunavant. High degree efficient symmetrical gaussian quadrature rules for the triangle. *International Journal for Numerical Methods in Engineering*, 21(6):1129–1148, 1985.

[79] Patrick Keast. Moderate-degree tetrahedral quadrature formulas. *Computer Methods in Applied Mechanics and Engineering*, 55(3):339–348, 1986.

[80] William F Mitchell. How high a degree is high enough for high order finite elements? In *ICCS*, pages 246–255, 2015.

[81] David A Boas, Dana H Brooks, Eric L Miller, Charles A DiMarzio, Misha Kilmer, Richard J Gaudette, and Quan Zhang. Imaging the body with diffuse optical tomography. *IEEE Signal Processing Magazine*, 18(6):57–75, 2001.

[82] Brian W Pogue, Troy O McBride, Judith Prewitt, Ulf L Österberg, and Keith D Paulsen. Spatially variant regularization improves diffuse optical tomography. *Applied Optics*, 38(13):2950–2961, 1999.

[83] Brian W Pogue, Keith D Paulsen, Chris Abele, and Howard Kaufman. Calibration of near-infrared frequency-domain tissue spectroscopy for absolute absorption coefficient quantitation in neonatal head-simulating phantoms. *Journal of Biomedical Optics*, 5(2):185–194, 2000.

[84] Jong Chul Ye, Su Yeon Lee, and Yoram Bresler. Exact reconstruction formula for diffuse optical tomography using simultaneous sparse representation. In *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1621–1624. IEEE, 2008.

[85] Hector RA Basevi, Kenneth M Tichauer, Frederic Leblond, Hamid Dehghani, James A Guggenheim, Robert W Holt, and Iain B Styles. Compressive sensing based reconstruction in bioluminescence tomography improves image resolution and robustness to noise. *Biomedical Optics Express*, 3(9):2131–2141, 2012.

[86] Jaya Prakash, Calvin B Shaw, Rakesh Manjappa, Rajan Kanhirodan, and Phaneendra K Yalavarthy. Sparse recovery methods hold promise for diffuse optical tomographic image reconstruction. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(2):74–82, 2013.

[87] Venkaiah C Kavuri, Zi-Jing Lin, Fenghua Tian, and Hanli Liu. Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography. *Biomedical Optics Express*, 3(5):943–957, 2012.

[88] Nannan Cao, Arye Nehorai, and Mathews Jacob. Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm. *Optics Express*, 15(21):13695–13708, 2007.

[89] Calvin B Shaw and Phaneedra K Yalavarthy. Effective contrast recovery in rapid dynamic near-infrared diffuse optical tomography using 1-norm-based linear image reconstruction method. *Journal of Biomedical Optics*, 17(8):086009, 2012.

[90] Shinpei Okawa, Yoko Hoshi, and Yukio Yamada. Improvement of image quality of time-domain diffuse optical tomography with lp sparsity regularization. *Biomedical Optics Express*, 2(12):3334–3348, 2011.

[91] Qin Lyu, Zhouchen Lin, Yiyuan She, and Chao Zhang. A comparison of typical p minimization algorithms. *Neurocomputing*, 119:413–424, 2013.

[92] Simon R Arridge and Martin Schweiger. A gradient-based optimisation scheme for optical tomography. *Optics Express*, 2(6):213–226, 1998.

[93] Deguang Kong and Chris Ding. Non-convex feature learning via lp, inf operator. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[94] Emmanuel Candes, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *arXiv preprint math/0409186*, 2004.

[95] David L Donoho. For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 59(6):797–829, 2006.

[96] Chen Chen, Fenghua Tian, Hanli Liu, and Junzhou Huang. Diffuse optical tomography enhanced by clustered sparsity for functional brain imaging. *IEEE Transactions on Medical Imaging*, 33(12):2323–2331, 2014.

[97] Maureen A O'Leary. *Imaging with diffuse photon density waves*. PhD thesis, University of Pennsylvania Philadelphia, 1996.

[98] Ben A Brooksby, Hamid Dehghani, Brian W Pogue, and Keith D Paulsen. Near-infrared (nir) tomography breast image reconstruction with a priori structural information from mri: algorithm development for reconstructing heterogeneities. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(2):199–209, 2003.

[99] Ben A Brooksby, Shudong Jiang, Hamid Dehghani, Brian W Pogue, Keith D Paulsen, John B Weaver, Christine Kogel, and Steven P Poplack. Combining near-infrared tomography and magnetic resonance imaging to study in vivo breast tissue: implementation of a laplacian-type regularization to incorporate magnetic resonance structure. *Journal of Biomedical Optics*, 10(5):051504, 2005.

[100] Ben Brooksby, Subhadra Srinivasan, Shudong Jiang, Hamid Dehghani, Brian W Pogue, Keith D Paulsen, John Weaver, Christine Kogel, and Steven P Poplack. Spectral priors improve near-infrared diffuse tomography more than spatial priors. *Optics Letters*, 30(15):1968–1970, 2005.

[101] Ang Li, Eric L Miller, Misha E Kilmer, Thomas J Brukilacchio, Tina Chaves, Jonathan Stott, Quan Zhang, Tao Wu, MaryAnn Chorlton, Richard H Moore, et al. Tomographic optical breast imaging guided by three-dimensional mammography. *Applied Optics*, 42(25):5181–5190, 2003.

[102] Ang Li, Greg Boverman, Yiheng Zhang, Dana Brooks, Eric L Miller, Misha E Kilmer, Quan Zhang, Elizabeth MC Hillman, and David A Boas. Optimal linear inverse solution with multiple priors in diffuse optical tomography. *Applied Optics*, 44(10):1948–1956, 2005.

[103] Haijing Niu, Fenghua Tian, Zi-Jing Lin, and Hanli Liu. Development of a compensation algorithm for accurate depth localization in diffuse optical tomography. *Optics Letters*, 35(3):429–431, 2010.

[104] Limin Zhang, Yan Zhao, Shudong Jiang, Brian W Pogue, and Keith D Paulsen. Direct regularization from co-registered anatomical images for mri-guided near-infrared spectral tomographic image reconstruction. *Biomedical Optics Express*, 6(9):3618–3630, 2015.

[105] Murad Althobaiti, Hamed Vavadi, and Quing Zhu. Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix. *Journal of Biomedical Optics*, 22(2):026002, 2017.

[106] Junzhou Huang, Tong Zhang, and Dimitris Metaxas. Learning with structured sparsity. *Journal of Machine Learning Research*, 12(Nov):3371–3412, 2011.

[107] Richard G Baraniuk, Volkan Cevher, Marco F Duarte, and Chinmay Hegde. Model-based compressive sensing. *arXiv preprint arXiv:0808.3572*, 2008.

[108] Volkan Cevher, Piotr Indyk, Chinmay Hegde, and Richard G Baraniuk. Recovery of clustered sparse signals from compressive measurements. Technical report, RICE UNIV HOUSTON TX DEPT OF ELECTRICAL AND COMPUTER ENGINEERING, 2009.

[109] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[110] Hamid Dehghani, Matthew E Eames, Phaneendra K Yalavarthy, Scott C Davis, Subhadra Srinivasan, Colin M Carpenter, Brian W Pogue, and Keith D Paulsen. Near infrared optical tomography using nirfast: Algorithm for numerical model and image reconstruction. *Communications in Numerical Methods in Engineering*, 25(6):711–732, 2009.

[111] Alper Corlu, Turgut Durduran, Regine Choe, Martin Schweiger, Elizabeth MC Hillman, Simon R Arridge, and Arjun G Yodh. Uniqueness and wavelength optimization in continuous-wave multispectral diffuse optical tomography. *Optics Letters*, 28(23):2339–2341, 2003.

[112] Hamid Dehghani, Subhadra Srinivasan, Brian W Pogue, and Adam Gibson. Numerical modelling and image reconstruction in diffuse optical tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1900):3073–3093, 2009.

[113] Subhadra Srinivasan, Brian W Pogue, Shudong Jiang, Hamid Dehghani, and Keith D Paulsen. Spectrally constrained chromophore and scattering near-infrared tomography provides quantitative and robust reconstruction. *Applied Optics*, 44(10):1858–1869, 2005.

[114] Subhadra Srinivasan, Brian W Pogue, Ben Brooksby, Shudong Jiang, Hamid Dehghani, Christine Kogel, Wendy A Wells, Steven P Poplack, and Keith D Paulsen. Near-infrared characterization of breast tumors in vivo using spectrally-constrained reconstruction. *Technology in Cancer Research & Treatment*, 4(5):513–526, 2005.

[115] Ang Li, Quan Zhang, Joseph P Culver, Eric L Miller, and David A Boas. Reconstructing chromosphere concentration images directly by continuous-wave diffuse optical tomography. *Optics Letters*, 29(3):256–258, 2004.

[116] Xavier Intes, Clemence Maloux, Murat Guven, Birzen Yazici, and Britton Chance. Diffuse optical tomography with physiological and spatial a priori constraints. *Physics in Medicine & Biology*, 49(12):N155, 2004.

[117] Ingrid Daubechies, Ronald DeVore, Massimo Fornasier, and C Sinan Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 63(1):1–38, 2010.

[118] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. Fast image recovery using variable splitting and constrained optimization. *IEEE Transactions on Image Processing*, 19(9):2345–2356, 2010.

[119] Manya V Afonso, José M Bioucas-Dias, and Mário AT Figueiredo. An augmented lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Transactions on Image Processing*, 20(3):681–695, 2010.

[120] Jinming Duan, Zhenkuan Pan, Xiangfeng Yin, Weibo Wei, and Guodong Wang. Some fast projection methods based on chan-vese model for image segmentation. *EURASIP Journal on Image and Video Processing*, 2014(1):7, 2014.

[121] Jinming Duan, Zhenkuan Pan, Baochang Zhang, Wanquan Liu, and Xue-Cheng Tai. Fast algorithm for color texture image inpainting using the non-local ctv model. *Journal of Global Optimization*, 62(4):853–876, 2015.

[122] Tom Goldstein, Christoph Studer, and Richard Baraniuk. A field guide to forward-backward splitting with a fasta implementation. *arXiv preprint arXiv:1411.3406*, 2014.

[123] Martin Schweiger and Simon R Arridge. The toast++ software suite for forward and inverse modeling in optical tomography. *Journal of Biomedical Optics*, 19(4):040801, 2014.

[124] Brian W Pogue and Michael S Patterson. Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry. *Journal of Biomedical Optics*, 11(4):041102, 2006.

[125] Yuxuan Zhan, Adam T Eggebrecht, Joseph P Culver, and Hamid Dehghani. Singular value decomposition based regularization prior to spectral mixing improves crosstalk in dynamic imaging using spectral diffuse optical tomography. *Biomedical Optics Express*, 3(9):2036–2049, 2012.

[126] Calvin B Shaw and Phaneendra K Yalavarthy. Performance evaluation of typical approximation algorithms for nonconvex p-minimization in diffuse optical tomography. *JOSA A*, 31(4):852–862, 2014.

[127] Bernard Martinet. Regularisation, d'inéquations variationelles par approximations succesives. *Revue Francaise d'informatique et de Recherche operationelle*, 1970.

[128] Antonin Chambolle, Ronald A De Vore, Nam-Yong Lee, and Bradley J Lucier. Nonlinear wavelet image processing: variational problems, compression, and noise

removal through wavelet shrinkage. *IEEE Transactions on Image Processing*, 7(3):319–335, 1998.

[129] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences*, 57(11):1413–1457, 2004.

[130] Mário AT Figueiredo and Robert D Nowak. An em algorithm for wavelet-based image restoration. *IEEE Transactions on Image Processing*, 12(8):906–916, 2003.

[131] Elaine T Hale, Wotao Yin, and Yin Zhang. A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing. *CAAM TR07-07, Rice University*, 43:44, 2007.

[132] Cédric Vonesch and Michael Unser. A fast iterative thresholding algorithm for wavelet-regularized deconvolution. In *Wavelets XII*, volume 6701, page 67010D. International Society for Optics and Photonics, 2007.

[133] Stephen J Wright, Robert D Nowak, and Mário AT Figueiredo. Sparse reconstruction by separable approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493, 2009.

[134] JP Culver, R Choe, MJ Holboke, L Zubkov, T Durduran, A Slemp, V Ntziachristos, B Chance, and AG Yodh. Three-dimensional diffuse optical tomography in the parallel plane transmission geometry: Evaluation of a hybrid frequency domain/continuous wave clinical system for breast imaging. *Medical Physics*, 30(2):235–247, 2003.

[135] Jaya Prakash and Phaneendra K Yalavarthy. A lsqr-type method provides a computationally efficient automated optimal choice of regularization parameter in diffuse optical tomography. *Medical Physics*, 40(3):033101, 2013.

[136] Xue Wu, Adam T Eggebrecht, Silvina L Ferradal, Joseph P Culver, and Hamid Dehghani. Quantitative evaluation of atlas-based high-density diffuse optical tomography for imaging of the human visual cortex. *Biomedical Optics Express*, 5(11):3882–3900, 2014.

[137] John Ashburner and Karl J Friston. Image segmentation. *Human Brain Function*, 2, 2003.

[138] Hamid Dehghani, Brian R White, Benjamin W Zeff, Andrew Tizzard, and Joseph P Culver. Depth sensitivity and image reconstruction analysis of dense imaging arrays for mapping brain function with diffuse optical tomography. *Applied Optics*, 48(10):D137–D143, 2009.

[139] Hao Yang Wu, Andrew Filer, Iain Styles, and Hamid Dehghani. Development of a multi-wavelength diffuse optical tomography system for early diagnosis of rheumatoid arthritis: simulation, phantoms and healthy human studies. *Biomedical Optics Express*, 7(11):4769–4786, 2016.

[140] Daniel Lighter, Andrew Filer, and Hamid Dehghani. Multispectral diffuse optical tomography of finger joints. In *European Conference on Biomedical Optics*, page 104120N. Optical Society of America, 2017.

[141] James A Guggenheim, Ilaria Bargigia, Andrea Farina, Antonio Pifferi, and Hamid Dehghani. Time resolved diffuse optical spectroscopy with geometrically accurate models for bulk parameter recovery. *Biomedical Optics Express*, 7(9):3784–3794, 2016.

[142] Guy Gilboa and Stanley Osher. Nonlocal operators with applications to image processing. *Multiscale Modeling & Simulation*, 7(3):1005–1028, 2008.

[143] Max Gunzburger and Richard B Lehoucq. A nonlocal vector calculus with application to nonlocal boundary value problems. *Multiscale Modeling & Simulation*, 8(5):1581–1598, 2010.

[144] Antoni Buades, Bartomeu Coll, and J-M Morel. A non-local algorithm for image denoising. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 60–65. IEEE, 2005.

[145] Jinming Duan, Zhenkuan Pan, Wangquan Liu, and Xue-Cheng Tai. Color texture image inpainting using the non local ctv model. *Journal of Signal and Information Processing*, 4(03):43, 2013.

[146] Andrea L Bertozzi and Arjuna Flenner. Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Modeling & Simulation*, 10(3):1090–1118, 2012.

[147] Ekaterina Merkurjev, Tijana Kostic, and Andrea L Bertozzi. An mbo scheme on graphs for classification and image processing. *SIAM Journal on Imaging Sciences*, 6(4):1903–1930, 2013.

[148] Xavier Bresson, Xue-Cheng Tai, Tony F Chan, and Arthur Szlam. Multi-class transductive learning based on 1 relaxations of cheeger cut and mumford-shah-potts model. *Journal of Mathematical Imaging and Vision*, 49(1):191–201, 2014.

[149] Stanley Osher, Martin Burger, Donald Goldfarb, Jinjun Xu, and Wotao Yin. An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489, 2005.

[150] Jinming Duan, Wenqi Lu, Christopher Tench, Irene Gottlob, Frank Proudlock, Niraj Nilesh Samani, and Li Bai. Denoising optical coherence tomography using second order total generalized variation decomposition. *Biomedical Signal Processing and Control*, 24:120–127, 2016.

[151] Lei Yao and Huabei Jiang. Enhancing finite element-based photoacoustic tomography using total variation minimization. *Applied Optics*, 50(25):5031–5041, 2011.

[152] Hao Gao and Hongkai Zhao. Multilevel bioluminescence tomography based on radiative transfer equation part 1: l1 regularization. *Optics Express*, 18(3):1854–1871, 2010.

[153] Alexander B Konovalov and Vitaly V Vlasov. Total variation based reconstruction of scattering inhomogeneities in tissue from time-resolved optical projections. In *Saratov Fall Meeting 2015: Third International Symposium on Optics and Biophotonics and Seventh Finnish-Russian Photonics and Laser Symposium (PALS)*, volume 9917, page 99170S. International Society for Optics and Photonics, 2016.

[154] Manuel Freiberger, Christian Clason, and Hermann Scharfetter. Total variation regularization for nonlinear fluorescence tomography with an augmented lagrangian splitting approach. *Applied Optics*, 49(19):3741–3747, 2010.

[155] Keith D Paulsen and Huabei Jiang. Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization. *Applied Optics*, 35(19):3447–3458, 1996.

[156] Jinping Tang, Bo Han, Weimin Han, Bo Bi, and Li Li. Mixed total variation and regularization method for optical tomography based on radiative transfer equation. *Computational and Mathematical Methods in Medicine*, 2017, 2017.

[157] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. An algorithm for minimizing the mumford-shah functional. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1133–1140. IEEE, 2009.

[158] Tom Goldstein and Stanley Osher. The split bregman method for l1-regularized problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009.

[159] Konstantinos Papafitsoros and Carola-Bibiane Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of Mathematical Imaging and Vision*, 48(2):308–338, 2014.

[160] Jinming Duan, Zhaowen Qiu, Wenqi Lu, Guodong Wang, Zhenkuan Pan, and Li Bai. An edge-weighted second order variational model for image decomposition. *Digital Signal Processing*, 49:162–181, 2016.

[161] Jinming Duan, Wil OC Ward, Luke Sibbett, Zhenkuan Pan, and Li Bai. Introducing diffusion tensor to high order variational model for image reconstruction. *Digital Signal Processing*, 69:323–336, 2017.

[162] Wenqi Lu, Daniel Lighter, and Iain B Styles. L 1-norm based nonlinear reconstruction improves quantitative accuracy of spectral diffuse optical tomography. *Biomedical Optics Express*, 9(4):1423–1444, 2018.

[163] Gerardo González, Ville Kolehmainen, and Aku Seppänen. Isotropic and anisotropic total variation regularization in electrical impedance tomography. *Computers & Mathematics with Applications*, 74(3):564–576, 2017.

[164] Michael Clancy, Antonio Belli, David Davies, Samuel J. E. Lucas, Zhangjie Su, and Hamid Dehghani. Comparison of neurological nirs signals during standing valsalva maneuvers, pre and post vasopressor injection. In *European Conference on Biomedical Optics*, page 953817. Optical Society of America, 2015.

[165] Carole Ichai, Hervé Quintard, and Jean-Christophe Orban. *Metabolic Disorders and Critically Ill Patients: From Pathophysiology to Treatment*. Springer, 2017.

[166] László Kocsis, Peter Herman, and Andras Eke. The modified beer–lambert law revisited. *Physics in Medicine & Biology*, 51(5):N91, 2006.

[167] Wesley B Baker, Ashwin B Parthasarathy, David R Busch, Rickson C Mesquita, Joel H Greenberg, and AG Yodh. Modified beer-lambert law for blood flow. *Biomedical optics express*, 5(11):4053–4075, 2014.

[168] David A Boas, Tom Gaudette, Gary Strangman, Xuefeng Cheng, John JA Marota, and Joseph B Mandeville. The accuracy of near infrared spectroscopy and imaging during focal changes in cerebral hemodynamics. *Neuroimage*, 13(1):76–90, 2001.

[169] Manish Bhatt, Kaylan R Ayyalasomayajula, and Phaneendra K Yalavarthy. Generalized beer–lambert model for near-infrared light propagation in thick biological tissues. *Journal of Biomedical Optics*, 21(7):076012, 2016.

[170] David A Boas et al. *Diffuse photon probes of structural and dynamical properties of turbid media: theory and biomedical applications*. PhD thesis, Citeseer, 1996.

[171] Thomas J Farrell, Michael S Patterson, and Brian Wilson. A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical Physics*, 19(4):879–888, 1992.

[172] V Allen and AL McKenzie. The modified diffusion dipole model. *Physics in Medicine & Biology*, 36(12):1621, 1991.

[173] SR Arridge, M Schweiger, M Hiraoka, and DT Delpy. A finite element approach for modeling photon transport in tissue. *Medical Physics*, 20(2):299–309, 1993.

[174] Simon R Arridge, Hamid Dehghani, Martin Schweiger, and Eiji Okada. The finite element model for the propagation of light in scattering media: a direct method for domains with nonscattering regions. *Medical physics*, 27(1):252–264, 2000.

[175] Kui Ren, Gassan S Abdoulaev, Guillaume Bal, and Andreas H Hielscher. Algorithm for solving the equation of radiative transfer in the frequency domain. *Optics letters*, 29(6):578–580, 2004.

[176] Jorge Ripoll and Vasilis Ntziachristos. Iterative boundary method for diffuse optical tomography. *JOSA A*, 20(6):1103–1110, 2003.

[177] Jenni Heino, Simon Arridge, Jan Sikora, and Erkki Somersalo. Anisotropic effects in highly scattering media. *Physical Review E*, 68(3):031908, 2003.

[178] Hao Zhang, Dong Zeng, Hua Zhang, Jing Wang, Zhengrong Liang, and Jianhua Ma. Applications of nonlocal means algorithm in low-dose x-ray ct image processing and reconstruction: A review. *Medical physics*, 44(3):1168–1185, 2017.

[179] Alwin Kienle, Michael S Patterson, Nora Dögnitz, Roland Bays, Georges Wagnières, and Hubert van Den Bergh. Noninvasive determination of the optical properties of two-layered turbid media. *Applied optics*, 37(4):779–791, 1998.

[180] David A Boas, Anders M Dale, and Maria Angela Franceschini. Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *Neuroimage*, 23:S275–S288, 2004.

[181] Yousef Saad. *Iterative methods for sparse linear systems*, volume 82. siam, 2003.

[182] Boris Teodorovich Polyak. The conjugate gradient method in extremal problems. *USSR Computational Mathematics and Mathematical Physics*, 9(4):94–112, 1969.

[183] Youcef Saad and Martin H Schultz. Gmres: A generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM Journal on scientific and statistical computing*, 7(3):856–869, 1986.

[184] Henk A Van der Vorst. Bi-cgstab: A fast and smoothly converging variant of bi-cg for the solution of nonsymmetric linear systems. *SIAM Journal on scientific and Statistical Computing*, 13(2):631–644, 1992.

[185] Thomas A Manteuffel. An incomplete factorization technique for positive definite linear systems. *Mathematics of computation*, 34(150):473–497, 1980.

[186] L Yu Kolotilina and A Yu Yeremin. Factorized sparse approximate inverse preconditionings i. theory. *SIAM Journal on Matrix Analysis and Applications*, 14(1):45–58, 1993.

[187] Matthaios Doulgerakis, Adam T Eggebrecht, and Hamid Dehghani. High-density functional diffuse optical tomography based on frequency-domain measurements improves image quality and spatial resolution. *Neurophotonics*, 6(3):035007, 2019.

[188] Wenqi Lu, Jinming Duan, Joshua Deepak Veesa, and Iain B Styles. New non-local forward model for diffuse optical tomography. *Biomedical Optics Express*, 10(12):6227–6241, 2019.

[189] Wenqi Lu, Jinming Duan, David Orive-Miguel, Lionel Herve, and Iain B Styles. Graph-and finite element-based total variation models for the inverse problem in diffuse optical tomography. *Biomedical optics express*, 10(6):2684–2707, 2019.

[190] Wenqi Lu and Iain B Styles. Nonlocal differential operators improve image reconstruction in diffuse optical tomography. In *Optical Tomography and Spectroscopy*, pages JTu3A–32. Optical Society of America, 2018.