Novel mathematical techniques for structural inversion and image reconstruction in medical imaging governed by a transport equation

A thesis submitted to the University of Manchester for the degree of Doctor of Philosophy in the Faculty of Engineering and Physical Sciences

2015

Kernel Enrique Prieto Moreno
School of Mathematics

Contents

1 Introduction **15**

1.1 Motivation . . . 15

1.2 State-of-the-art . . . 16

1.3 Diffuse Optical Tomography . . . 19

1.4 Aims and structure of the thesis . . . 24

2 Mathematical background **28**

2.1 The Radiative Transport Equation . . . 28

2.1.1 Diffusive Approximation . . . 34

2.1.2 Purely absorbing medium . . . 35

2.1.3 Well-posedness of the forward problem in DOT . . . 37

2.2 Regularization . . . 43

2.2.1 Truncated Singular Value Decomposition . . . 43

2.2.2 Filters . . . 50

2.2.3 Tikhonov regularization of linear operator equations . . . 52

2.2.4 Tikhonov regularization of nonlinear operator equations . . . 53

2.3 Iterative reconstruction techniques . . . 58

2.3.1 Kaczmarz method . . . 58

2.3.2 Landweber, Landweber-Kaczmarz and Loping-LK methods . . . 61

3 Reconstructions using the L,LK and a LS methods **69**

3.1 Inverse problem in DOT . . . 70

3.2 Level set method: calculus of variations approach . . . 79

3.3 Shape reconstruction problem in DOT . . . 82

3.4 Smooth gradient . . . 90

3.5 Shape and contrast value reconstructions in DOT . . . 93
3.6 Discretization of the forward and inverse problems . . . 96
3.7 Numerical Results . . . 101

4 Level set evolution by shape derivative for RTE **111**
4.1 Shape Derivative . . . 111
4.2 Level set method: Front tracking by a Hamilton-Jacobi type equation . . . 122
4.2.1 Eulerian and Lagrange formulations . . . 122
4.2.2 Motion under an external velocity field . . . 125
4.2.3 Motion Involving Mean Curvature . . . 127
4.2.4 General Motion . . . 128
4.3 Incorporating the shape derivative into the level set method . . . 131
4.4 Numerical Results . . . 135

5 Sparsity regularization **140**
5.1 Basic concepts of sparsity and compressive sensing . . . 140
5.2 Linearised Bregman iteration . . . 147
5.3 Linearised Bregman Iteration for the ℓ_1-minimisation: linear case . . . 150
5.4 Sparsity method . . . 154
5.5 Numerical Results . . . 161

6 Total variation regularization **175**
6.1 Introduction . . . 175
6.2 Split Bregman method . . . 177
6.3 Limited-memory BFGS algorithm . . . 181
6.4 Lagged Diffusivity . . . 186
6.5 Discrete Derivatives . . . 194
6.6 Numerical Results . . . 197

7 Conclusions and Future Work **202**

Word Count: 54062

List of Figures

1.1 Resolution with intrinsic contrast characteristics of biomedical imaging techniques. 19
1.2 Visualization of possible paths of photons through the medium. 20
1.3 Absorption spectra of oxy-haemoglobin, deoxy-haemoglobin and water. . 21
1.4 DOT applications. Top row: DOT apparatus for screening the breast. Bottom row: DOT instrumentation and arrangement for brain imaging [115]. 22

2.1 Kaczmarz method in two dimensions. 60

3.1 Configuration for the imaging problem 81
3.2 Contrast value reconstruction of μ. 105
3.3 Contrast value reconstruction of b. 106
3.4 Simultaneous contrast value reconstruction of μ. 107
3.5 Simultaneous contrast value reconstruction of b. 107
3.6 Simultaneous contrast value reconstruction of μ and b. 108
3.7 Simultaneous shape reconstruction of μ. 108
3.8 Simultaneous shape value reconstruction of b. 109
3.9 Simultaneous shape reconstruction with μ and b. 109
3.10 Norm of residuals of the misfit functional of Simultaneous reconstruction of μ and b using the LK and LS method versus using the LK method, then LK with LS method 110

4.1 Configuration for the imaging problem 112
4.2 Level set function. 123
4.3 Lagrange formulation of the level set equation 125

4.4 Shape reconstruction for μ using algorithm 7 for one obstacle 135
4.5 Shape reconstruction for μ using algorithm 7 for two obstacles. 136
4.6 Shape reconstruction for μ using algorithm 8 with V_{ext} for one obstacle . 137
4.7 Shape reconstruction for μ using algorithm 8 with V_{ext} for two obstacles . 138
4.8 Norm of residuals of the cost of μ using the shape derivative and the LK method. 139

5.1 Spheres in L^p. 143
5.2 The Bregman distance $D^p_{\mathcal{K}}(\tilde{\sigma}, \upsilon)$. 148
5.3 The approximation of ϑ_ϵ to the absolute value function. 151
5.4 The approximation of ϑ'_ϵ to the absolute value function. 152
5.5 The approximation of $s_{\lambda,\epsilon}$ to the soft-thresholding function s_λ. 153
5.6 Comparison between the LK method and the LK method with monotonicity condition 163
5.7 Comparison between the LK method and L-LK methods without monotonicity condition 165
5.8 Comparison between the LK method and L-LK methods with monotonicity condition 167
5.9 Reconstructions using sparsity regularization 168
5.10 Profiles, residuals and errors using sparsity regularization 169
5.11 Reconstruction of two closely-spaced inclusions 170
5.12 Robustness of the sparsity reconstruction with respect of the regularizer . 171
5.13 Reconstruction of scattering coefficient using sparsity regularization . . . 172
5.14 Simultaneous reconstruction of μ and b coefficients using sparsity regularization 173
5.15 Simultaneous reconstruction of μ and b coefficients using only the LK method 174

6.1 Reconstructions using Split-Bregman method of μ 197
6.2 Norm of residuals of the cost of μ using the SB and the SB with continuation scheme. 198
6.3 Reconstructions of absorption coefficient μ using total variation regularization. 199

6.4 Profiles, residuals and errors using total variation regularization 200
6.5 Shape reconstruction for σ using the L-BFGS method 201

List of Tables

2.1 Kaczmarz method. 60

3.1 Experimental set-up of the absorption coefficient reconstruction. 102
3.2 Experimental set-up of the scattering coefficient reconstruction. 102
3.3 Experimental set-up of the simultaneous reconstruction of μ and b. 102

5.1 Experimental set-up for simultaneous reconstruction using sparsity method 162

List of Algorithms

1 Landweber . . . 78
2 LK . . . 78
3 l-LK . . . 79
4 LK and LS methods . . . 90
5 Smooth Gradient . . . 93
6 LK and LS methods with contrast value. . . . 96
7 Shape reconstruction using the LS method with shape derivative . . . 134
8 Shape reconstruction using the LS method with shape derivative and extended V. . . . 134
9 Landweber using sparsity regularization . . . 159
10 LK using sparsity regularization . . . 159
11 LK using our sparsity method . . . 160
12 Split Bregman iteration for total variation regularization . . . 180
13 Split Bregman iteration with continuation scheme for TV . . . 181
14 Backtracking Line Search . . . 183
15 L-BFGS . . . 185
16 L-BFGS Method with inexact line search . . . 185
17 Lagged diffusivity method for total variation regularization . . . 194

The University of Manchester

Kernel Enrique Prieto Moreno
Doctor of Philosophy
Novel mathematical techniques for structural inversion and image reconstruction in medical imaging governed by a transport equation
August 2, 2015

Since the inverse problem in Diffusive Optical Tomography (DOT) is nonlinear and severely ill-posed, only low resolution reconstructions are feasible when noise is added to the data nowadays. The purpose of this thesis is to improve image reconstruction in DOT of the main optical properties of tissues with some novel mathematical methods. We have used the Landweber (L) method, the Landweber-Kaczmarz (LK) method and its improved Loping-Landweber-Kaczmarz (L-LK) method combined with sparsity or with total variation regularizations for single and simultaneous image reconstructions of the absorption and scattering coefficients. The sparsity method assumes the existence of a sparse solution which has a simple description and is superposed onto a known background. The sparsity method is solved using a smooth gradient and a soft thresholding operator. Moreover, we have proposed an improved sparsity method. For the total variation reconstruction imaging, we have used the split Bregman method and the lagged diffusivity method. For the total variation method, we also have implemented a memory-efficient method to minimise the storage of large Hessian matrices. In addition, an individual and simultaneous contrast value reconstructions are presented using the level set (LS) method. Besides, the shape derivative of DOT based on the RTE is derived using shape sensitivity analysis, and some reconstructions for the absorption coefficient are presented using this shape derivative via the LS method.

Whereas most of the approaches for solving the nonlinear problem of DOT make use of the diffusion approximation (DA) to the radiative transfer equation (RTE) to model the propagation of the light in tissue, the accuracy of the DA is not satisfactory in situations where the medium is not scattering dominant, in particular close to the light sources and to the boundary, as well as inside low-scattering or non-scattering regions. Therefore, we have solved the inverse problem in DOT by the more accurate time-dependant RTE in two dimensions.

Declaration

No portion of the work referred to in this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

Copyright Statement

i the author of this thesis (including any appendices and/or schedules to this thesis) owns certain copyright or related rights in it (the ”Copyright”) and s/he has given The University of Manchester certain rights to use such Copyright, including for administrative purposes.

ii Copies of this thesis, either in full or in extracts and whether in hard or electronic copy, may be made **only** in accordance with Copyright, Designs and Patents Act 1988 (as amended) and regulations issued under it or, where appropriate, in accordance with licensing agreements which the University has from tome to time. This page must form part of any such copies made.

iii The ownership of certain Copyright, patents, designs, trade marks and other intellectual property (the ”Intellectual Property”) and any reproductions of copyright works in the thesis, for example graphs and tables (”Reproductions”), which may be described in this thesis, may not be owned by the author and may be owned by third parties. Such Intellectual Property and Reproductions cannot and must not be made available for use without the prior written permission of the owner(s) of the relevant Intellectual Property and/or Reproductions.

iv Further information on the conditions under which disclosure, publication and commercialisation of this thesis, the Copyright and any Intellectual Property and/or Reproductions described in it may take place is available in the University IP Policy (see http://documents.manchester.ac.uk/DocuInfo.aspx?DocID=487, in any relevant Thesis restriction declarations deposited in the University Library, The University Library’s regulations (see http://www.library.manchester.ac.uk/aboutus/regulations and in The University’s Policy on Presentation of Theses.

Acknowledgements

I thank my supervisor Oliver Dorn for his support, motivation and guidance throughout this work.

I thank deeply to my examiners Bill Lionheart and Ville Kolehmainen for their recommendations. I am sure that this work has improved vastly thanks to them. I am very grateful to the Mexican taxpayers who fully sponsored me through CONACyT. I also deeply and sincerely thank my wife Irma Gonzalez Gonzalez for all her support, patience and encouragement through the entire course of my PhD. I am very grateful to my beloved mother for all her love, patience and financial support. Also, I thank sincerely to Rafael Alcaraz, Michael Crabb, Javier Hernandez, Jack Williams and Julio Pérez for their technical assistance, advices and for making the Alan Turing building an enjoyable place to work.

Abbreviations

DOT	Diffuse Optical Tomography
RTE	Radiative Transfer Equation
LS	Level Set
L	Landweber
LK	Landweber-Kaczmarz
L-LK	Loping-Landweber-Kaczmarz
TV	Total variation
LD	Lagged diffusivity
L-BFGS	Limited-memory Broyden-Fletcher-Goldfarb-Shanno
SB	Split Bregman
BB	Barzilai-Borwein
NIR	Near-infrared
H-G	Henyey-Greenstein
$2D$	Two-dimensional
$3D$	Three-dimensional

Notations

S^{n-1}	Angular space
Ω	Domain
Γ	Boundary of the domain
x	Position vector
θ	Direction velocity vector
t	Time
u	Density of photons
$\bar{\nu}$	Outward unit normal
μ	Absorption coefficient
b	Scattering coefficient
$\eta(\cdot)$	Scattering phase function
U	Space of states of u
P	Space of parameters μ , b
Z	Space of data recorded
$\mathcal{G}[\cdot]$	Measurement operator
$\mathcal{R}[\cdot]$	Residual operator
Υ	Domain of the inclusion
$\partial\Upsilon$	Boundary of the inclusion
$\mathcal{V}$	Velocity vector field
V	Normal component of the vector field $\mathcal{V}$
$s(\cdot)_\lambda$	Soft-thresholding function
$\|\cdot\|$	Euclidean norm
$\mathrm{Re}(\cdot)$	Real part of a complex valued function
τ	step size

Chapter 1

Introduction

1.1 Motivation

In many imaging scenarios it is important to noninvasively obtain some relevant information of the interior of an object, without cutting or taking sections of an object or organism, which represents a potentially dangerous process or comes with an increased cost, we want to obtain important information of interior anatomical or physiological values. There are many applications in image reconstruction, such as detection of anti-personal land mines in post-war remediation areas, the monitoring of industrial processes, the localization and characterisation of oil reservoirs, satellite recognition of objects in the space, the detection and reconstruction of images of objects in luggage at airports, or in noninvasive biomedical imaging techniques [122]. In biomedical imaging, in particular breast cancer and brain imaging studies have seen increased growth in the last decades.

Breast cancer has one of the highest mortality rate at present, and is the most frequent womens cancer. More than one million people in the world were diagnosed with breast cancer in 2000. Estimates show that every 2.5 minutes a woman is diagnosed with breast cancer and every 7.5 minutes a woman dies from this illness within the European Union [96, 173]. Breast cancer had the second highest mortality rate in United States, and every year more than 200,000 American women are diagnosed with breast cancer [140]. Benign lesions (cysts and fibroadenomas) and mastalgia are non-mortal breast diseases. Male breast cancer is much less common but is more mortal than female cases [131]. Therefore, improved methods for early detection for preventing the metastasis of the tumour and for obtaining accurate diagnosis of cancer treatment, visualization of small breast tumours,

classification of breast illness and distinction between benign and malign lesions at low cost are essential to decrease unnecessary biopsies and mortality rates [104, 122].

Brain diseases or disorders are of high interest in Health institutions worldwide. In Europe alone it has been estimated that 165 million people were affected by mental disorders in 2011. In fact, 37 percent of the European population suffer from some form of brain disorder. In Europe those disorders cost 800 billion Euros each year [123, 124, 170]. The story in United States is quite similar, one in out of 76 of American adults were qualified for Security Disability insurance or Supplemental security Income in 2007, while in 1987 the rate was two and a half times less. The rate increased 30 times for children in the same period. For children, mental illness is the leading cause of disability. The National Institute of Mental Health (NIMH) conducted a survey for adult people between 2001-2003, which indicated that 46% had at least one mental illness within four different types of them at some point in their lives [4].

Mental disorders include anxiety disorder, obsessive-compulsive disorder, panic disorder, phobias, addictions, post-traumatic stress disorder, schizophrenia,"mood disorders" like bipolar disorder and depression, deficit/hyperactivity disorder, dementia, insomnia and new disorders. Neurological disorders include dementia, epilepsy and multiple sclerosis. Anxiety disorder and dementia are the most common disorders in Europe. In fact, mental disorders are the main contributor to the morbidities burden in Europe, it has been announced that disorders of the brain as the core health challenge of the 21st century [123]. In order to contribute to work on this challenge, health research must be carried out in many levels. One of these levels correspond to obtain contrast images of high resolution of the interior of the head and the technique should be portable, relatively nonexpensive and safe. Other brain problems include head injuries (traumatic injuries), strokes, multiple sclerosis, epilepsy, Alzheimer's disease, disorders of consciousness after a brain injury and brain tumours [80].

1.2 State-of-the-art

X-ray mammography, magnetic resonance imaging (MRI) and ultrasound (US) are the main imaging techniques presently used for breast cancer detection and for brain imaging respectively. Also, MRI, electrical impedance tomography (EIT), positron emission

tomography (PET) techniques and more recently microwave imaging are used for this propose nowadays [20]. We now introduce rather briefly the main traditional imaging techniques used nowadays.

X-Ray Mammography

Roentgen discovered X-rays in 1895, and since the early 1970's, they were used in biomedical imaging [14]. High energy X-rays are used for probing mediums in X-ray Mammography. X-rays are directed to biological tissue and the X-rays are absorbed or scattered in the tissue. The energy of X-rays used to examine breast for early tumour detection varies from 15 Kev-20 Kev [22, 65, 118]. The accuracy of biomedical imaging techniques is measured by the sensitivity. Sensitivity in biomedical imaging is calculated by dividing the number of positive breast cancer cases detected by the total sum of those that were detected plus those that were missed and then multiplied by 100 [134]. Another important measure in biomedical imaging techniques is the specificity. The specificity is given by dividing the number of negative breast cancer cases detected by the total sum of those that were truly negative and were detected as negative breast cancer cases plus those that were incorrectly detected to be positive and then multiplied by 100 [134]. Whereas X-ray mammography is the current biomedical imaging technique preferred for detecting breast cancer tumours, its use has been questioned since specificity in mammography varies from 90-98% [22, 65, 118]; 70-90 % of biopsies performed, based on suspicious mammograms were negative [41, 96, 105, 117, 132], which is an expensive procedure. Moreover, 30% or 15% of tumours may not be detected by the X-ray scan [66, 104, 105, 117, 132]; the sensitivity of mammography has been reported in the range of 80-95% [22, 65, 118]. Hence, its use results in false negative and positive diagnoses. The reason for this incorrect diagnoses is the relative poor contrast between diseased and healthy tissues in mammographies. Another drawback of mammography is its use of ionising radiation as well as being uncomfortable and painful for patients. X-ray computed tomography (CT) is another imaging technique used to detect breast tumours and although the accuracy of X-ray CT is better than mammography, it leads to incorrect early-stage breast tumour detections and high radiation dose has to be used [105, 117, 132]. In addition, mammography is not suitable for male breast cancer detection at present [173].

Magnetic Resonance Imaging (MRI)

In MRI, permanent magnets in the range of .25-3.9 Tesla are used in MR scans. Hy-

drogen atoms H^+, which constitute the majority of atoms in livings tissue, absorb the wave energy at resonance frequency of 60 MHz when they are placed in a 1.5 Tesla magnetic field [28]. These Hydrogen atoms re-emit the energy absorbed after a period of time depending on the magnetic properties of the tissue. This signal is captured and converted point by point in a k-space data of a plane, making this procedure slow. Finally, a Fourier transform is applied to the k-space date to obtain the reconstruction of the magnetic properties of tissue [28]. Deficiencies of MRI are that it is a slow procedure, has high cost and importable [145]. In addition, patients with metallic pieces inserted in past surgical interventions are unable to be imaged by using MRI. Another disadvantage of MRI is its low specificity (64 %). Some advantages of MRI are that it is nonionising and its accurate sensitivity (98%) [89, 95]. Another benefit is its high predictability of residual disease (88%) and multicentricity (90%) [94].

Ultrasound (US)

Sound waves of frequency in the range of 1 to 10 MHz are used in ultrasound modality imaging techniques. Sound waves are guided through the tissue, and the reflected sound waves, called echoes, are collected using ultrasound transducers. There exists a trade-off between resolution and the imaging depth, which depends on the frequency of the sound waves used [28]. US, which is nonionising, is used to scan pregnant women. The sensitivity and specificity values of US are 97% and 85% respectively [89, 134], some studies report that US provides relatively poor information to classify certain solid breast masses. US and mammography are relatively poor at identifying the presence or extent of residual disease [94]. One of the main drawbacks of US is its relatively poor resolution for imaging deep-tissues and other disadvantage of US is that it is operator dependent. In addition, US is not suitable for male breast cancer detection because of the close proximity of the ribs and lungs [173]. There has been an increase in hybrid imaging modalities in the last decades such as photoacoustic tomography (PAT) or ultrasound modulated optical tomography (UMOT).

Finally, we show a comparison of resolution performance with respect to intrinsic contrast characteristics of traditional biomedical imaging techniques with biomedical optical imaging with respect to spatial resolution in figure 1.1, where NIR stands for Near-infrared light.

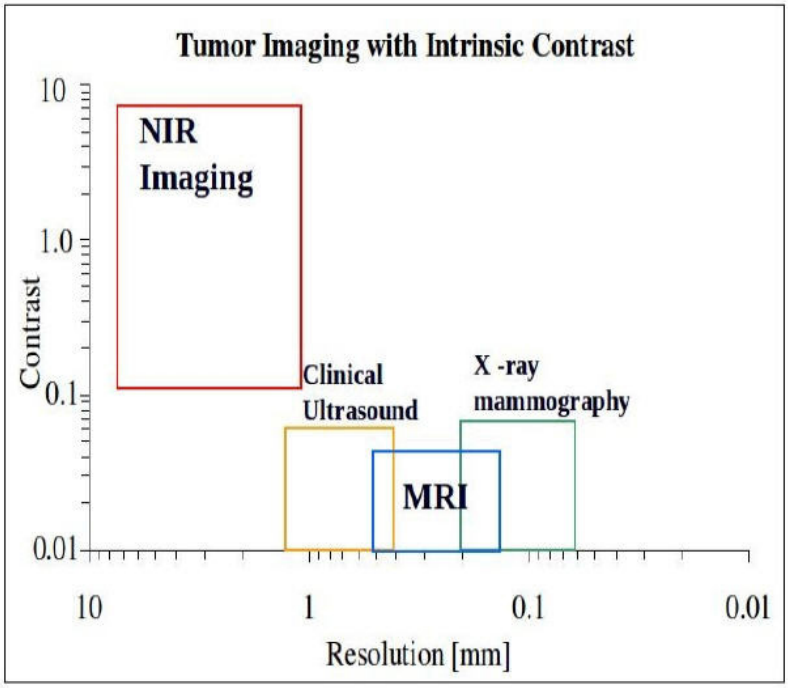


Figure 1.1: Comparison of resolution with respect to intrinsic contrast characteristics of traditional biomedical imaging techniques with biomedical optical imaging with respect to spatial resolution [133].

1.3 Diffuse Optical Tomography

DOT is a noninvasive imaging modality that uses near-infrared light to interrogate optical properties of biological tissue. DOT is an imaging technique which provides information about the spatial distribution of the optical properties within a tissue based on noninvasive exitance or the outward directed photon flux (or transmitted/or reflected light) measurements on the periphery. Light from a near-infrared laser source located on the surface of the object is guided into the medium and the amount of transmitted/reflected light is measured using light sensitive detectors (fibre bundles) located on the boundary of the object and this measurement process is repeated for all laser sources. Figure 1.2 illustrates the possibilities of the propagation of a photon through the medium.

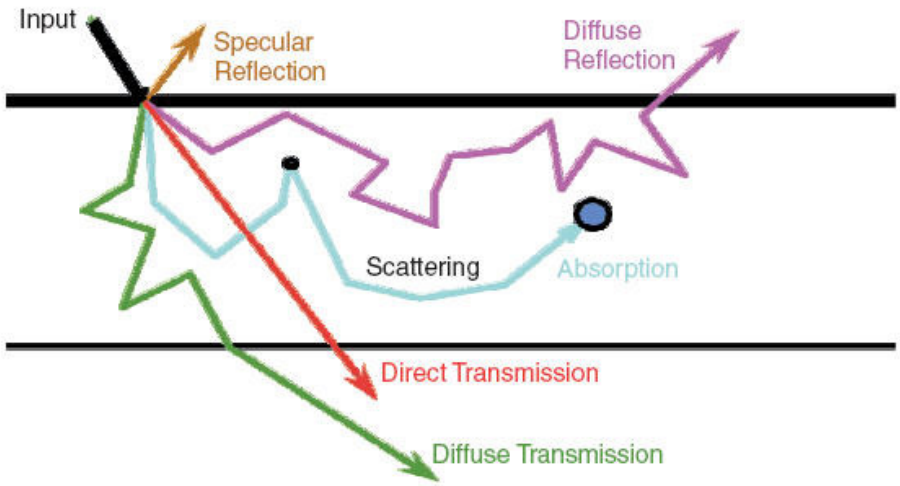


Figure 1.2: Visualization of possible paths of photons through the medium: purple, green, blue and brown colours, as opposed to X-ray tomography in which there is direct transmission: red colour.

Water, oxy-haemoglobin and desoxy-haemoglobin are the main absorbers of the interior of tissues such as brain and breast [14, 21, 115, 159]. Collagen, which is found in the skin and bone, is a relevant absorber in this near-infrared frequency range [159]. Information of the oxygen and blood volume content of the absorbers and scatterers have application in diagnosis of brain and breast cancer and cerebral activity monitoring. The simultaneous relative weak absorption of the oxy-haemoglobin, deoxy-haemoglobin and water in the near-infrared region spectrum 600-1000 nm allows imaging tissues at centimetre depths in DOT, also the absorption of water increases fast at frequencies higher than 1000 nm, interfering with the identification of the haemoglobin [14, 20, 21, 115], as it can be seen in figure 1.3. The range of frequency of the infrared light is called the "NIR window". Also, note in figure 1.3 that the spectrum of the oxy-haemoglobin and deoxy-haemoglobin differ in the near-infrared region spectrum (600-1000 nm), then DOT can be used to distinguish between deoxy-haemoglobin from oxy-haemoglobin, this is used to discriminate among tumours with different degrees of malignancy [20]. Note in figure 1.3 that the absorption spectrum of water is not higher than absorption spectrum of HbO_2 or Hb in the interval of 600-1000 nm, allowing to measure the oxygen content in patient's blood. There are applications of this in functional brain monitoring and neonatal research [81, 161]. Similar experimental results of the absorbers of these quantities (water, haemoglobin and oxy-haemoglobin) are found in [14, 20, 173, 177]. A review of the optical properties of different kind of tissues such as liver, aorta and muscle and for some living organisms can be found in [37, 177] and the references therein.

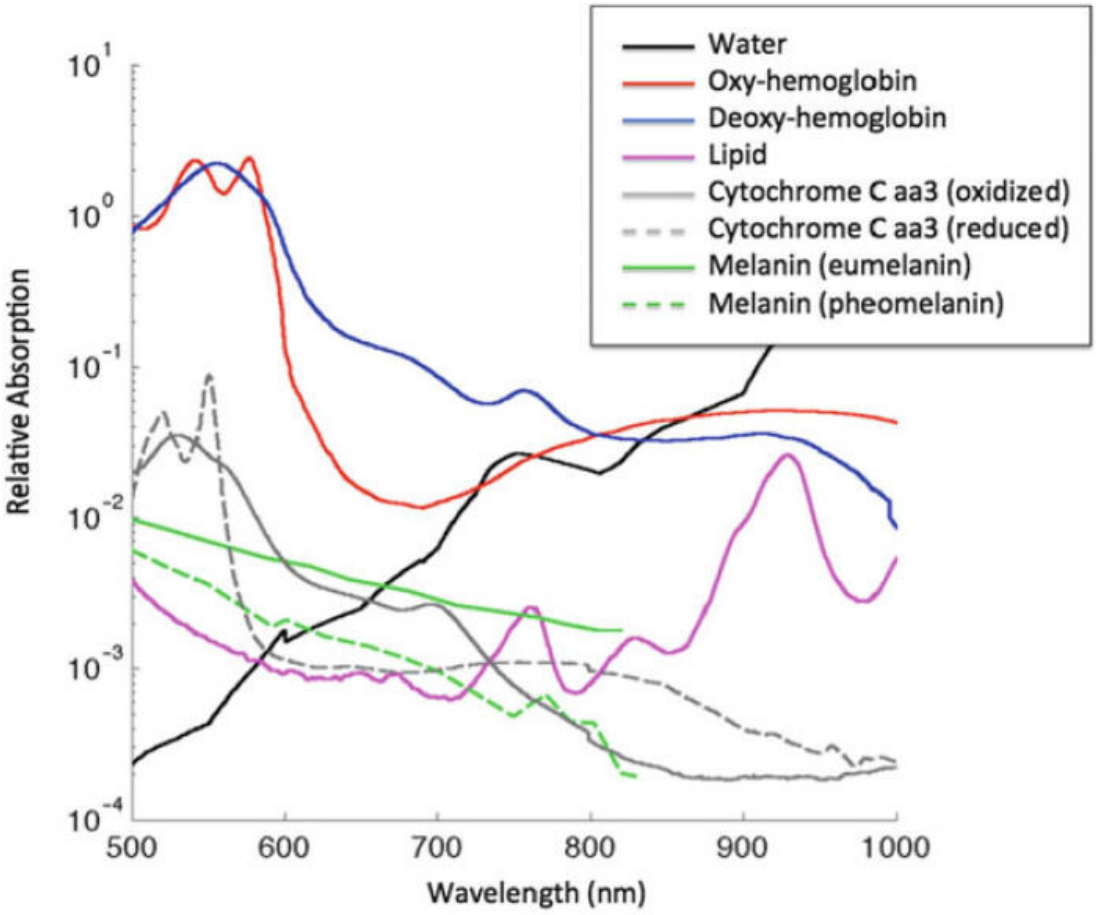
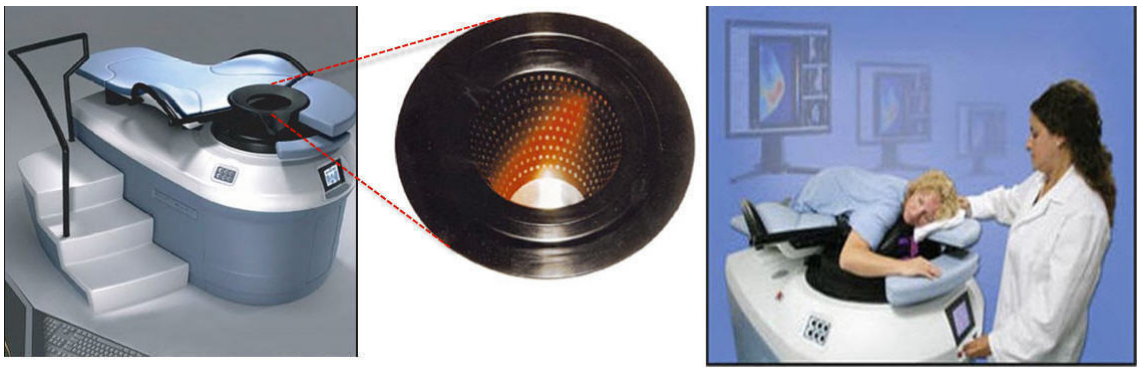


Figure 1.3: Absorption spectrum of oxy-haemoglobin HbO_2, deoxy-haemoglobin Hb, Water, Lipid, Cytochrome C aa3 oxidized and reduced and Melanin. Oxi/deoxy-haemoglobin, water and lipids from Data stored by Oregon Medical Laser Centre and [14, 20, 159, 177]. Melanin and Cytochrome C aa3 oxidized/reduced absorption spectras from [115, 173].

The most relevant applications of DOT are the detection of breast cancer, and the study of the brain, including stroke, haemorrhage, and brain function. DOT may enable one to recognise ischemic from haemorrhagic stroke, which is very important since neuroprotective drugs for treating ischematic stroke given to patients with haemorrhagic stroke yields to expeditious death [18, 42, 81]. It is more complicated to image the brain than the breast in DOT because of the head structure, that is, the photons must pass through the skin and skull and the cerebrospinal fluid [145].

The interest for using DOT for imaging the living brain has increased in recent decades [18] since brain activity and haematoma are associated with increased volume of blood called vascularization and tumours are strongly linked with haemoglobin concentration via angiogenesis. The localization of those places which have an increased concentration of haemoglobin is possible because they have different absorption optical values from the

surrounding healthy tissue [20, 82]. For instance, DOT has been used to localize blood flow increase during epileptic episodes and imaging arterial pulsation [18].



(a)

(b)

Figure 1.4: DOT applications. Top row: DOT apparatus for screening the breast. Bottom row: DOT instrumentation and arrangement for brain imaging [115].

Advantages and disadvantages of DOT

Some benefits of using DOT include:

- DOT is noninvasive, nonionising, relatively inexpensive, unobtrusive, low-power and portable making possible use for ambulatory and continuous bedside monitoring, and produce valuable information about detection and classification of breast cancer, functional brain studies and imaging newborn infant head [20, 82, 111, 150, 153].

- DOT provides very high intrinsic contrast for discriminating benign and malign tumour in comparison with some traditional imaging modalities as can be seen in fig. 1.1.

- DOT can provide functional information, whereas functional information is not provided by some traditional modalities like mammography and US.

- DOT is a relatively cheap imaging technique with faster data collection than MRI.
- DOT does not pose a radiation risk since it is nonionising and it is not uncomfortable nor painful, and has potential to be used as a bedside monitoring [93].
- DOT has been widely applied to scan functional brain activities with adults and neonates [100, 161].
- Imaging reconstruction of functional haemodynamics is based on the fact that DOT provides information of variations of blood and oxygen content of a tissue due to cerebral activity of heads of living organisms [21, 42, 160]. Successful $3D$ methods have been used in bedside monitor systems for reconstructing the functional haemodynamics of the human brain and the activation of the motor and sensory cortex [18, 82, 93].
- DOT has been applied to scan pre-term and term infants with brain injuries. A brain injury is caused by the interruption of blood or oxygen which may lead to permanent disability or death. The scattering coefficient has also shown to be linked with brain activity [85].
- Compared with alternative imaging modalities such as fMRI, DOT has several advantages, including simplicity, portability, reduced cost, and faster acquisition speeds. It is also possible to simultaneously obtain both oxy- and deoxy-haemoglobin and blood volume with satisfactory temporal resolution [20].
- DOT may detect male breast cancer [173].

Tissue is a turbid media with strong scattering coefficient, thus light follows an extremely complicated path 1.2 and the signal strength attenuates rapidly according to the Beer-Lambert law [115].

Some disadvantages of using DOT include:

- Due to the physics of light propagation, the inverse problem is ill-posed; this means that small errors in the measured data may cause arbitrarily large error in the quantity of interest of the medium, causing inverse solutions to be wildly erratic and nonphysical unless additional *a priori* knowledge is assumed [107].

- The inverse problem in general is underdetermined; the number of voxels of the mesh of unknown of absorption and scattering coefficients may be one or two orders of magnitude greater than the number of measurements [20].

- The spatial resolution of DOT is relatively poor (around 4mm to 7mm) compared with traditional biomedical imaging techniques [134].

Finally, we mention that interest of DOT has been raising in the industry process because of the radiation emission of light is safe, nonintrusive and has fast response time [169]. Some nonbiomedical applications include nonintrusive measurements flow concentration profiles and velocity profiles, in multiphase flows or complex mixtures of diffusive fluids such as fibre suspensions or the monitoring of pulp flows in pipelines in paper industry [103] or more recently the measurements of mass flow rate, velocity of beads in flow rig, flame imaging for combustion rate estimation and bubbles flow [106].

1.4 Aims and structure of the thesis

Aims

It was around the 1990s that the reconstruction of optical properties of biological tissues with data from the external boundary was first formulated as an inverse problem [14]. Although the development of theory, application of regularization methods have made rapid progress since then [8], including the development of hybrid imaging methods in DOT [142, 155], a huge improvement of existent methods and the development of new methods still need to be done in order to address significant practical difficulties of this modality. This is, amongst others, due to the nonlinearity of the inverse problem in DOT, its ill-posedness and the fact that it is highly undetermined. This dissertation aims at combining some modern regularization techniques (such as level set shape evolution, sparsity and total variation) with advanced and accurate forward modelling techniques (here the RTE) in order to verify the inherent potential of DOT as a future imaging modality in several of the above listed applications. It mainly aims at performing proof-of-concept (and therefore is kept at a manageable two-dimensional setup) of a variety of advanced techniques rather than providing off-the-shelf industrial 3D imaging codes. However, all the algorithms and ideas provided in these techniques are kept at a very general level using formulations involving PDE's rather than specific discretizations, such that we are very

confident that all our techniques and numerical simulations generalize with the appropriate modifications to real-world 3D situations [6, 8, 20]. Such generalizations have been beyond the scope of this thesis.

Tikhonov regularization so far is the method of choice for performing the image reconstruction in DOT. Inverting a large ill-conditioned Hessian matrix for solving the corresponding inverse problem is usually achieved by adding a suitable (tuning) value to the diagonal of the Hessian matrix, which makes it a regularization parameter. Large regularization parameters result in over-smoothing (blurring, or a lack of sharp features) of the image, while a small regularization parameters results in an image with increased noise (artifacts).

Recent work has shown that applying TV or sparsity regularization in DOT based on the diffusion equation shows a promising improvement on denoising and helps to address the lack of sharp features (increasing the spatial resolution). Therefore, we have implemented TV and Sparsity regularization methods with different modalities to overcome this problem. In contrast to most approaches taken so far, our work is directly based on the RTE, which is the most complete model for the propagation of light in tissue.

Notice that the radiative transfer equation (RTE) provides a reasonable model in medical imaging not only for Diffuse optical tomography (DOT), but also for fluorescence diffuse tomography (FDOT), photoacoustic imaging and bioluminiscence. Therefore, the results of this thesis should also be of interest to those communities and applications. For more information about FDOT we refer to [14, 17, 59], photoacoustics see [134, 142, 153, 154, 164], bioluminiscence tomography see [76, 77] and the references therein. Some applications of inverse transport theory outside of medical imaging include astrophysics, neutron physics, atmospheric physics and geophysics.

In biomedical imaging, whereas most of the approaches for solving the nonlinear problem of DOT (and so far many of the above mentioned related modalities) make use of the diffusion approximation (DA) to the RTE to model the propagation of the light in tissue [7, 70, 91, 99, 144], it is argued that the accuracy of the DA is not satisfactory in situations where the medium is not scattering dominant. This is in particular a problem close to the light sources and to the boundary, as well as inside low-scattering or non-scattering regions such as the cerebrospinal fluid which wraps the brain and fills the ventricles [9, 139, 145, 154]. Therefore, we have solved the inverse problem in DOT

by the more accurate RTE (as mentioned, so far in 2D) in time domain. Since there are only few analytical results for the solution of the RTE assuming very specific geometrical configurations, numerical methods need to be applied for solving the RTE. Statistical approaches such as the Monte Carlo method are as well used for solving the RTE in practice. However, Monte Carlo methods are commonly computationally intensive, such that they are usually only employed for checking the accuracy of faster numerical methods [115].

Level set methods provide another tool which has been applied successfully in many fields including DOT, usually based on the diffusive approximation. In this thesis we have implemented two types of level set approaches, combined with contrast value reconstruction, for DOT based on the RTE. In connection to implementing novel level set techniques, we have calculated the shape derivative of DOT based on the RTE, which provides us with an alternative model for shape evolutions in the level set approach.

As already mentioned above, we want to emphasise once again that all techniques proposed and tested in this dissertation can be applied with only few changes to other imaging modalities such as photoacoustic tomography (PAT), fluorescence tomography (FDOT) and bioluminiscence.

Structure of thesis

The thesis contains seven chapters whose content is outlined in the following.

Chapter 1 outlines general approaches to the imaging of breast and brain diseases and gives an introduction into the current state of art of the traditional biomedical imaging tools highlighting some of their benefits and drawbacks. Then a more in-depth introduction into DOT is given with its advantages and limitations.

Chapter 2 presents and summarizes some mathematical background which is needed in later chapters of this dissertation. It covers the basics of the Radiative Transfer Equation, followed by a brief proof of the well-posedness of the forward problem in DOT. A summary of the Tikhonov regularization scheme, and the more recent Landweber-Kaczmarz regularization methods together with some of its variants is given.

In chapter 3, the inverse and shape reconstruction problems in DOT are formulated. A regularized gradient of the residual operator, needed later, is defined, and an algorithm for computing it is provided. An individual/simultaneous shape and contrast value reconstruction technique for the optical properties of tissue in DOT by using the Landweber-Kaczmarz (LK) method and a Level set approach introduced by Santosa [141] and its

variants are presented. Also, the Finite Differences method implemented in this thesis is described. Finally, numerical results are presented using these methods.

In chapter 4, the concept of a shape derivative is introduced and expressions for a shape derivative in DOT based on the RTE are derived. These are then tested via the level set method. For this purpose, the level set evolution scheme using a Hamilton-Jacobi-type equation is formulated. Two different variants of the numerical shape evolution in DOT are presented, which correspond to using a narrowband on the one hand and an extended velocity field in the whole domain on the other hand. In addition, one of these variants is compared with the shape reconstruction method presented in chapter 3.

In chapter 5, we again consider situations where inclusions have sharp boundaries or are highly localized. Here, sparsity regularization with ℓ_1 penalty terms instead of the standard ℓ_2 terms present an interesting alternative to the standard tools. The basics of sparsity regularization, compressed sensing and Bregman distances are reviewed. In addition, a linearised Bregman iteration is presented for the ℓ_1-minimisation for a linear problem. Moreover, a sparsity regularization scheme for DOT and some variants are described in detail. Some numerical results using sparsity regularization are compared with the LK method.

In chapter 6, the concept of total variation regularization is reviewed and the split Bregman and lagged diffusivity methods are presented in details. For the case of the lagged diffusivity method, some novel discretization schemes are presented and tested. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is introduced, which is then integrated into the TV regularization technique for DOT.

Finally, in chapter 7, some conclusion and directions for future work are given.

Chapter 2

Mathematical background

In this chapter, we collect some mathematical background which is relevant to this thesis and which is not easy to find currently in a single reference. This chapter also intents to fix notation and to introduce some concepts and abbreviations used further on in the thesis. To start with, some basic mathematical properties of the Radiative Transfer Equation (RTE) are briefly reviewed since we have modelled the forward problem in DOT using this model. In section 2.1, the general RTE is described in details. In particular, subsection 2.1.2 discusses the purely absorbing RTE demonstrating an exponential behaviour of the general RTE. The derivation of the diffusion equation is then sketched in subsection 2.1.1. Then we move on to discussing some relevant regularization theory background. Here, in subsections 2.2.1 and 2.2.3, the basics of the conventional truncated Singular Value Decomposition regularization approach and the Tikhonov regularization approach are reviewed, respectively, which are important to understand before discussing other types of regularization techniques. Then, in subsection 2.3.1 the Kaczmarz method and related techniques are briefly discussed. Finally, in subsection 2.3.2 the Landweber (L), Landweber-Kaczmarz (LK) and the loping-Landweber-Kaczmarz (L-LK) methods are presented together with some of their properties.

2.1 The Radiative Transport Equation

The mathematical description of the propagation of light travelling in random media is classified according to the spatial length scales. Usually, Maxwell Equations or geometric optics are used at the microscale, the Radiative Transfer Equation (RTE) or transport

equation at the mesoscale, and the Diffusion theory at the macroscale. We will focus on this thesis mainly on the mesoscale model which deals with turbid media such as tissue, a cloud or milk, that is, light propagation bahaves like photons following meandering trajectories through the media, rather than a propagating ray or a wave. A derivation of the transport equation from scalar wave models is presented in [9, 20]. The transition from the transport equation to the diffusive regime is presented below in subsection 2.1.1. One can roughly distinguish three regimes of the propagation of particles in the Radiative Transfer model according to the strength and frequency of interaction between the particles and the material in which they propagate. When there is no interaction between particles and background material, one talks about the ballistic regime. When there exists moderate interaction, it is called transport regime, and finally when there exists strong interaction (i.e. a large number of scattering events on a typical particle trajectory) it is called the diffusive regime [13]. The RTE can be applied to all of these regimes, and we will be interested in particular in this thesis to situations where mixed environments are present which contain diffusive areas as well as ballistic or transport dominated areas.

There exist three kind of systems of data acquisition in typical optical and fluorescence tomography setups, the first one is the continuous wave system, the second one is the frequency domain system, and the third one is the time domain system [107]. The continuous wave system measures the transmitted intensity at the detectors, while the frequency domain system measures the modulation amplitude and the phase shift of the light sources [107]. In this thesis we will focus in particular on the time domain system for acquiring the data. Here, the time-dependent RTE is the model we use.

The RTE is a linearised form of the more general Boltzmann equation [33], which can be derived from kinetic theory [153]. Diffuse Optical Tomography (DOT) treats light propagation in tissue as a transport of photons. DOT is a non-invasive technique for probing heterogeneous turbid media. The aim is to obtain images of their optical properties in the near-infrared (NIR) regime $600 - 1000$nm [107]. NIR light from laser sources is guided to the body and the amount of transmitted light is measured at detector locations along the boundary of the body. It is assumed that the photons are travelling with constant speed c through the medium, that is, one assumes a constant refractive index. To deal with the RTE with nonconstant refractive index, or variable speed of the photons, see for example [9].

Next we consider the mathematical formulation of the RTE in order to fix notation. In this thesis, the one-speed time-dependent RTE is used with c being normalized to 1 (in appropriate physical units). Let the space X be defined by

$$X := \Omega \times S^{n-1} \times [0, T],$$

where Ω is the domain of interest which is assumed to be a compact, convex subset of $\mathbb{R}^n$ with boundary $\partial X := \partial\Omega \times S^{n-1} \times [0, T]$. The boundary ∂X can be decomposed as $\partial X = \partial X_+ \bigcup \partial X_-$, where

$$\partial X_\pm := \{(x, \theta, t) \in \partial X,\ \pm\theta \cdot \bar{\nu} > 0\},$$

is assumed to be Lipschitz, which implies that the outward normal vector $\bar{\nu}$ exists almost everywhere on ∂X. Thus, the boundary of ∂X is decomposed into the inflow part ∂X_+ and the outflow part ∂X_-. The function $u(x, \theta, t) \in X$ is physically considered to describe the density of the photons travelling in the region Ω at position x with velocity direction θ at time t. The density of the photons u and the source term q are scalar quantities and are assumed nonnegative functions in $L^1(X)$. The following equation is called the *one-speed time-dependent RTE*

$$\frac{\partial u}{\partial t} + \theta \cdot \nabla u + (\mu(x) + b(x))\, u - b(x) \int_{S^{n-1}} \eta(\theta \cdot \theta') u(x, \theta', t) d\theta' = q(x, \theta, t) \quad \text{in} \quad X, \tag{2.1a}$$

with initial condition

$$u(x, \theta, 0) = 0 \qquad \text{in} \quad \Omega \times S^{n-1}, \tag{2.1b}$$

and boundary condition

$$u(x, \theta, t) = 0 \qquad \text{on} \quad \partial X_-. \tag{2.1c}$$

Here, q is the source term given by

$$q(x,\theta,t) = \begin{cases} q_j(x,\theta,t) & x \in \bigcup_j^M p_j, \\ 0 & x \in \partial\Omega \backslash \bigcup_j p_j, \end{cases} \tag{2.2}$$

with p_j being source positions along $\partial\Omega$. In time domain techniques, these sources q_j are often picosecond laser pulses which can be described mathematically by the expressions

$$q_j(x,\theta,t) = \delta_x(p_j)\delta_\theta(\theta_j)\delta_t(0),$$

where δ_x, δ_θ, and δ_t are Dirac delta distributions corresponding to the position, angular and time variables, respectively; the coefficients μ and b in (2.1a) are the absorption and scattering coefficients of the medium, respectively, and θ is a unit direction vector pointing into the direction of velocity of the photons.

In this thesis, where we are aiming mainly at proof of concept, we restrict ourselves to the RTE in $2D$, Since numerical solutions to the RTE are computationally intensive due to the dependence on space, angle, and time [6, 8, 20], we restrict ourselves to the RTE in $2D$ in this thesis, that is, we have tested our regularization, level set methods and the shape derivative (simulations) in a two-dimensional setting, where the domain $\Omega \subset \mathbb{R}^2$ and the direction vectors θ are in S^1. Moreover, we are only aiming at 'proof of concept', and did not use any real data. Physically the $2D$ and $3D$ versions are not that different from each other. All observed phenomena in $2D$ should scale up to $3D$. Certainly, it would be more desirable to work in $3D$, but using the time-dependent RTE in $3D$ simply takes quite long to run in $3D$ and we would have less time for investigating interesting mathematical concepts this way.

The scattering function $\eta(\theta \cdot \theta')$, also known as the scattering kernel [33], is assumed to be independent of the position of the scattering event and to depend on the cosine of the angle denoted by υ, which is the angle between the initial direction θ and the deviated direction θ' (after the scattering event has occurred), that is, $\cos\upsilon = \theta \cdot \theta'$. Particle conservation in pure scattering events is expressed by the additional requirement

$$\int_{S^{n-1}} \eta(\theta \cdot \theta')d\theta' = 1. \tag{2.3}$$

Often the Henyey-Greenstein phase function is used in DOT as a convenient model for the scattering function. In $2D$ it has the form

$$\eta(\cos\upsilon) = \frac{1-g^2}{2\pi(1+g^2-2g\cos\upsilon)}, \tag{2.4}$$

whereas in $3D$ it reads

$$\eta(\cos\upsilon) = \frac{1-g^2}{4\pi(1+g^2-2g\cos\upsilon)^{\frac{3}{2}}}, \tag{2.5}$$

where g is the mean cosine of the scattering function η with values between $-1 < g < 1$. In this model, values of g close to 1 indicate that the scattering is primarily forward directed, values close to zero indicate that the scattering is almost isotropic, whereas values near to -1 indicate that the scattering is primarily backward directed. Typical values for g based on animal tissue in Diffuse Optical Tomography are $0.9 \leq g \leq 0.99$ [11]. The mean free path is the mean distance travelled by a photon before changing its velocity direction, and is given in the Henyey-Greenstein model by

$$l = \frac{1}{(1-g)(\mu+b)}.$$

The total attenuation coefficient a is $a(x) := \mu(x) + b(x)$. Using the total attenuation coefficient, the mean free path is then $l = \big((1-g)a\big)^{-1}$.

The RTE is a conservation law for the number of photons. The following brief derivation can be found in [14, 20, 33]. Firstly, we define the vector quantity $\Phi(x,\theta,t) := u(x,\theta,t)\theta$ called the angular photon current density. Using this quantity, the number of photons flowing out of $\partial\Omega$ with velocity direction θ per time unit [second] and for a differential surface element dS is given by $\Phi(x,\theta,t)\cdot\nu dS$. The RTE describes a balance of the variation of the photon density inside a small but otherwise arbitrary region $\Upsilon \subseteq \mathbb{R}^n$ around position x and time t which originally travel with velocity direction θ. In more details, the time variation of the flux, i.e. the quantity

$$\frac{\partial}{\partial t}\int_{\Upsilon} u(x,\theta,t)\,dx, \tag{2.6}$$

is balanced by two loss terms, one being due to absorption and scattering events,

$$-\int_{\Upsilon} (\mu(x) + b(x))\, u\, dx, \tag{2.7a}$$

and one due to the photons flowing out the region Υ through the boundary

$$\int_{\partial\Upsilon} \Phi(x, \theta, t) \cdot \nu\, dS, \tag{2.7b}$$

and two gain terms, namely one describing those photons which were travelling with velocity direction θ' and now being scattered in direction θ

$$b(x) \int_{\Upsilon} \int_{S^1} \eta(\theta \cdot \theta') u(x, \theta', t)\, d\theta' dx, \tag{2.7c}$$

and the other one being due to an applied source term

$$\int_{\Upsilon} q(x, \theta, t)\, dx. \tag{2.7d}$$

Applying the Stokes' theorem to the above boundary flow term results in

$$\int_{\partial\Upsilon} \Phi(x, \theta, t) \cdot \nu\, dS = \int_{\Upsilon} \nabla \cdot (u\theta)\, dx = \int_{\Upsilon} \theta \cdot \nabla u\, dx.$$

The last identity $\nabla \cdot (u\theta) = \theta \cdot \nabla u$ holds because θ is considered as an independent variable in $u(x, \theta, t)$. Since the region Υ is arbitrary and small, the terms (2.6) and (2.7) produce in the limit of infinitesimal domains the RTE (2.1a).

Notice that an alternative way of describing laser sources at the boundary is to use an inhomogeneous boundary condition instead of (2.1c) and a zero source $q = 0$ term on the right side of equation (2.1a). Both expressions are equivalent [45, 53]. We will use the homogeneous RTE in chapter 3 with laser sources at the boundary. For more detailed reviews of the RTE in various settings, see e.g. [8, 11, 33, 45, 60], and for more details of the time-dependent RTE and its stability in particular, see [12, 13]. A deterministic method has been chosen in this thesis to solve the RTE (2.1). For statistical methods like Monte Carlo methods for solving the time-dependent forward and backward propagation problems in DOT, see [8, 9, 20, 107] and for statistical methods for inverse problems in general [101, 152]. In the following two subsections we will demonstrate ideas by

considering in more details two specific cases of the radiative transport equation [107], namely the situations of diffusive and ballistic behaviour of light propagation.

We conclude this section by formulating our *direct or forward problem* in the time domain technique for DOT. Given a pair of absorption and scattering coefficients, μ and b, a scattering function η with the property (2.3), and sources q (2.2), the goal of the forward or direct problem is to find the density of photons $u(x, \theta, t)$ such that equation (2.1) holds.

2.1.1 Diffusive Approximation

We present a summary of a derivation of the diffusive approximation of the RTE given in [6, 8, 19, 41, 87, 97, 115]. Firstly, we define the photon radiance $\Phi(x, t)$ as

$$\Phi(x,t) = \int_{S^{n-1}} u(x,\theta,t)\, d\theta,$$

and the energy density current $J(x, t)$ by

$$J(x,t) = \int_{S^{n-1}} \theta u(x,\theta,t)\, d\theta.$$

One way to derive the diffusive approximation of the RTE is to expand the angular part of the density of photons, phase, and source functions in spherical harmonics [6] as follows

$$u(x,\theta,t) = \sum_{l=N}^{\infty} \sum_{n=-l}^{l} \sqrt{\frac{2l+1}{4\pi}} \psi_{lm}(x,t) Y_{lm}(\theta), \tag{2.8}$$

$$\eta(\theta \cdot \theta') = \sum_{l=N}^{\infty} \sum_{n=-l}^{l} \sqrt{\frac{2l+1}{4\pi}} \eta_{lm}(\theta \cdot \theta') Y^*_{lm}(\theta') Y_{lm}(\theta), \tag{2.9}$$

$$q(x,\theta,t) = \sum_{l=N}^{\infty} \sum_{n=-l}^{l} \sqrt{\frac{2l+1}{4\pi}} q_{lm}(x,t) Y_{lm}(\theta). \tag{2.10}$$

A P_N approximation of the RTE is obtained if the first N terms are taken into account of the above expressions, in particular the P_1 approximation is obtained after substituting

(2.8) into (2.1a) and truncating the series of (2.8) after the $l = 1$ terms, resulting in

$$\frac{\partial \Phi(x,t)}{\partial t} + \mu\Phi(x,t) + \nabla J(x,t) = q_0(x,t), \tag{2.11a}$$

$$\frac{\partial J(x,t)}{\partial t} + \frac{J(x,t)}{3D(x)} + \frac{1}{3}\nabla\Phi(x,t) = q_1(x,t). \tag{2.11b}$$

Here $D(x)$ is known as the diffusion coefficient given by

$$D(x) = \frac{1}{3(\mu + b')}, \quad b' := (1-\eta)b,$$

where b' is known as the reduced scattering coefficient. The diffusive approximation follows now by assuming that

$$\frac{\partial J}{\partial t} = 0, \quad q_1 = 0.$$

The first assumption is equivalent to requesting that the medium is scattering dominant, that is, $\mu << b'$. After these assumptions, equation (2.11b) converts into Fick's law

$$J(x,t) = -\kappa_0 \nabla\Phi(x,t). \tag{2.12}$$

Finally, substituting Fick's law (2.12) into equation (2.11a), one obtains the diffusion approximation

$$\frac{\partial \Phi(x,t)}{\partial t} - \nabla \cdot \kappa_0 \nabla\Phi(r,t) + \mu\Phi(x,t) = q_0(x,t).$$

Therefore, the diffusion approximation is a special case of the P_1 approximation of the RTE where the radiance function $\Phi(x,t)$ does not depend on the direction velocity vector θ as opposed to the density of photons $u(x,\theta,t)$ in the RTE. Another derivation of the diffusive approximation of the RTE is found in [14, 41].

2.1.2 Purely absorbing medium

Now we consider the Radiative Transfer Equation with constant refractive index (2.1a) in a purely absorbing medium, i.e., $b = 0$. It is written as

$$\frac{\partial u(x,\theta,t)}{\partial t} + \theta \cdot \nabla u(x,\theta,t) + \mu(x)u(x,\theta,t) = q(x,\theta,t). \tag{2.13}$$

Let the variables x' and t' be given by

$$x' = x - \theta t, \quad t' = t. \tag{2.14}$$

Since, (using the variables (2.14))

$$\begin{aligned}\frac{\partial u(x,\theta,t)}{\partial t'} &= \frac{\partial u}{\partial x}\frac{\partial x}{\partial t'} + \frac{\partial u}{\partial t}\frac{\partial t}{\partial t'},\\ &= \nabla u \cdot \theta + \frac{\partial u}{\partial t},\end{aligned}$$

Equation (2.13) can be written as

$$\frac{\partial u(x',\theta,t')}{\partial t'} + \mu(x'+\theta t')u(x',\theta,t') = q(x'+\theta t',\theta,t'). \tag{2.15}$$

The integrating factor of equation (2.15) is

$$\exp\left[\int_{t_0}^{t'} \mu(x'+\theta t)\,dt\right].$$

Using the above integrating factor, equation (2.15) may now be written as

$$\frac{\partial}{\partial t'}\left\{\exp\left[\int_{t_0}^{t'} \mu(x'+\theta t)\,dt\right]u(x',\theta,t')\right\} = \exp\left[\int_{t_0}^{t'} \mu(x'+\theta t)\,dt\right]q(x'+\theta t',\theta,t'). \tag{2.16}$$

Integrating equation (2.16) over t' from 0 to t yields

$$\begin{aligned}\int_0^t \left\{\frac{\partial}{\partial t'}\exp\left[\int_0^{t'} \mu(x'+\theta t)\,dt\right]u(x',\theta,t')\right\}dt' = \int_0^t \exp\left[\int_0^{t'} \mu(x'+\theta t)\,dt\right]\\ \times\, q(x'+\theta t',\theta,t')\,dt',\end{aligned}$$

and we obtain

$$\begin{aligned}u(x,\theta,t) = \exp\left[-\int_0^t \mu(x+\theta(\tau-t))d\tau\right]\Bigg\{u(x-\theta t,\theta,0)\\ + \int_0^t \exp\left[\int_0^{\tau'} \mu(x+\theta(\tau-t))d\tau\right] q(x+\theta(\tau'-t),\theta,\tau')\,d\tau'\Bigg\}\end{aligned}$$

or equivalently

$$u(x,\theta,t) = \exp\left[-\int_0^t \mu(x+\theta(\tau-t))d\tau\right] u(x-\theta t,\theta,0) + \int_0^t \exp\left[-\int_{\tau'}^t \mu(x+\theta(\tau-t))d\tau\right] q(x+\theta(\tau'-t),\theta,\tau')\,d\tau'. \tag{2.17}$$

The first term on the right hand side of (2.17) physically corresponds to those photons which were at the position $x-\theta t$ with velocity direction θ at time $t=0$ and have escaped absorption (the exponential attenuation term) while propagating along straight lines (or 'rays') $I = p + l\theta$ for a time t in the direction θ, with $l \in \mathbb{R}$ and p is position of the source q. The second term represents the density of photos created by the source q [33].
Using the initial condition (2.1b), the solution $u(x,\theta,t)$ (2.17) reduces to

$$u(x,\theta,t) = \int_0^t \exp\left[-\int_{\tau'}^t \mu(x+\theta(\tau-t))d\tau\right] q(x+\theta(\tau'-t),\theta,\tau')\,d\tau'.$$

2.1.3 Well-posedness of the forward problem in DOT

Similar to the derivation of the solution of the RTE in pure absorbing media, a description of the solution of the general RTE can be found. Thus, formally replacing the source term $q(x,\theta,t)$ and the absorption coefficient μ in (2.13) by

$$q(x,\theta,t) \to q(x,\theta,t) + b(x)\int_{S^{n-1}} \eta(\theta\cdot\theta')u(x,\theta,t)\,d\theta',$$
$$\mu(x) \to a(x) := \mu(x) + b(x),$$

the implicit solution of the general RTE (2.1a) with initial condition (2.1b) is

$$u(x,\theta,t) = Q(x,\theta,t) + \int_0^t \exp\left[-\int_{\tau'}^t a(x+\theta(\tau-t))d\tau\right] b(x+\theta(\tau-t)) \times \int_{S^{n-1}} \eta(\theta\cdot\theta')u(x+\theta(\tau'-t),\theta',\tau')\,d\theta' d\tau', \tag{2.18}$$

with

$$Q(x,\theta,t) = \int_0^t \exp\left[-\int_{\tau'}^t a(x+\theta(\tau-t))d\tau\right] q(x+\theta(\tau'-t),\theta,\tau')\,d\tau'. \tag{2.19}$$

This is the integral formulation of the transport equation [45] which is a Volterra equation of the second kind [33]. The form (2.18) can be used to prove the existence and uniqueness of solution of the RTE [33, 50]. Similar to the proof of the existence and uniqueness for the solution of the RTE, the existence and uniqueness of the solution of an adjoint of the RTE can be shown, which will be presented in chapter 3.

The following definition will be used in the next theorem addressing well-posedness of the RTE.

Definition 1 *The coefficients μ and b are called admissible if μ, b are positive and $\mu, b \in L^{\infty}(\Omega)$.*

The next theorem is proved in [33, 45, 50, 60]. It states that for admissible coefficients μ, b, and $q \in L^1(X)$, the solution u of (2.1) exists and is unique in $L^1(X)$. In other words, the theorem states that the forward problem in DOT is well-posed. We will give a brief proof of this important Theorem.

Theorem 2 *Let η be the positive scattering function, and let the pair of coefficients μ, b be admissible. Furthermore, assume that $q \in L^1(X)$, with*

$$\hat{Q} = \int_X q(x, \theta, t)\, dx d\theta dt < \infty. \tag{2.20}$$

Then, the solution u of (2.1) exists and is unique in $L^1(X)$ and satisfies the inequality

$$\int_{S^{n-1}} \int_{\Omega} u(x, \theta, t) dx d\theta \leq \hat{Q} \exp(\beta t) \tag{2.21}$$

for any $t > 0$ and $\beta := \sup_{x \in \Omega} b(x)$.

Proof. Equation (2.18) may be rewritten as

$$u(x, \theta, t) = Q(x, \theta, t) + Ku(x, \theta, t), \tag{2.22}$$

where

$$Ku(x, \theta, t) = \int_0^t \int_{\Omega} \int_{S^{n-1}} \delta(x' - x + \theta(t - \tau')) \tag{2.23}$$

$$\times \exp\left[- \int_{\tau'}^{t} a(x + \theta(\tau - t) d\tau \right] b(x') \eta(\theta \cdot \theta') u(x', \theta', \tau')\, d\theta' dx' d\tau'. \tag{2.24}$$

Here $\delta(\cdot)$ is the Dirac delta distribution, $a = \mu + b$ is the attenuation as defined in the previous section, and $Q(x, \theta, t)$ is as in (2.19). The functions μ, b, q are considered to be equal to zero outside of X. Equation (2.22) can also be written as

$$u(x, \theta, t) = (I - K)^{-1} Q(x, \theta, t),$$

where I is the identity operator. Making use of the power series expansion of the function $(I - K)^{-1}$, we have

$$(I - K)^{-1} = I + K + K^2 + \ldots = \sum_{n=0}^{\infty} K^n, \quad \|K\|_{L^1(X)} < 1.$$

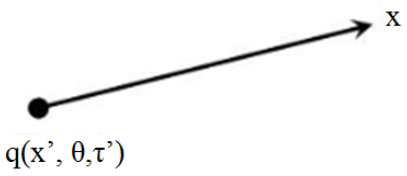
Therefore, we obtain the following Neumann series representation of the solution

$$u(x, \theta, t) = \sum_{n=0}^{\infty} K^n Q(x, \theta, t). \tag{2.25}$$

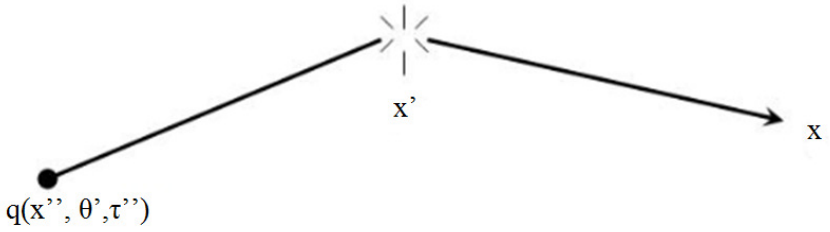
We need to investigate convergence of this series. To see if $\|K\|_{L^1(X)} < 1$ holds, we can

examine the first few terms of the Neumann series (2.25)

$$u(x,\theta,t) = \int_0^t \exp\left[-\int_{\tau'}^t a(x+\theta(\tau-t))d\tau\right] q(x+\theta(\tau'-t),\theta,\tau')\,d\tau',$$



$$+\int_0^t \int_\Omega \int_{S^{n-1}} \delta(x'-x+\theta(t-\tau'))\exp\left[-\int_{\tau'}^t a(x+\theta(\tau-t))d\tau\right]$$
$$\times\, b(x')\eta(\theta\cdot\theta')\left\{\int_0^\tau \exp\left[-\int_{\tau''}^\tau a(x'+\theta'(\tau'-\tau))d\tau'\right]\right.$$
$$\left.\times\, q(x'+\theta'(\tau'-\tau),\theta',\tau')\,d\tau''\right\} d\theta' dx' d\tau' + ...,$$



(2.26)

The first term of (2.25) physically represents photons arriving at x directly (without scattering) from the source term $q(x',\theta,\tau')$, i.e., those photos are virgin source photons; the second term corresponds to the second generation photons that have scattered once at x' before arriving at x. Therefore, the Neumann series is an expansion in the number of times that the photons have been scattered before arriving at a given point at a given time with a certain direction. Then, from physical reasoning we would expect that in an absorbing medium the Neumann series (2.25) will converge.

Now let us check mathematically that the Neumann series (2.25) will converge. To do this, we will use the following equivalent formulation of the Neumann series

$$u(x,\theta,t) = \sum_{n=0}^{\infty} u_n(x,\theta,t), \tag{2.27}$$

with

$$u_0(x,\theta,t) = Q(x,\theta,t) \tag{2.28a}$$

$$u_n(x,\theta,t) = Ku_{n-1}(x,\theta,t), \tag{2.28b}$$

and Q as in (2.19). Integrating equation (2.28a) over the spatial and angular variables x and θ and taking into account (2.20), we have

$$\begin{aligned}\int_{S^{n-1}}\int_{\Omega} u_0(x,\theta,t)\,dxd\theta &= \int_{S^{n-1}}\int_{\Omega} Q(x,\theta,t)\,dxd\theta \\ &= \int_{S^{n-1}}\int_{\Omega}\int_0^t \exp\left[-\int_{\tau'}^t a(x+\theta(\tau-t))d\tau\right] \\ &\quad \times q(x+\theta(\tau'-t),\theta,\tau')\,d\tau'\,dxd\theta \\ &\leq \int_{S^{n-1}}\int_{\Omega}\int_0^t q(x+\theta(\tau'-t),\theta,\tau')\,d\tau'\,dxd\theta \\ &= \hat{Q}.\end{aligned} \tag{2.29}$$

Using the iteration formula (2.28), we have

$$\begin{aligned}\int_{S^{n-1}}\int_{\Omega} u_1(x,\theta,t)\,dxd\theta &= \int_{S^{n-1}}\int_{\Omega} Ku_0(x,\theta,t)\,dxd\theta \\ &= \int_{S^{n-1}}\int_{\Omega} KQ(x,\theta,t)\,dxd\theta,\end{aligned}$$

and using equations (2.23), (2.3) and the inequality (2.29), the right hand side of the above

expression is equal to

$$
\begin{aligned}
&= \int_{S^{n-1}} \int_{\Omega} \Bigg\{ \int_0^t \int_{\Omega} \int_{S^{n-1}} \delta(x' - x - \theta(\tau' - t)) \exp\left[- \int_{\tau'}^t a(x + \theta(\tau - t)) d\tau \right] \\
&\times b(x') \eta(\theta \cdot \theta') Q(x', \theta', \tau') d\theta' dx' d\tau' \Bigg\} \, dx d\theta \\
&\leq \int_{S^{n-1}} \int_{\Omega} \Bigg\{ \int_0^t \int_{\Omega} \int_{S^{n-1}} \delta(x' - x - \theta(\tau' - t)) \beta \eta(\theta \cdot \theta') \\
&\times Q(x', \theta', \tau') d\theta' dx' d\tau' \Bigg\} \, dx d\theta \\
&= \int_0^t \int_{\Omega} \int_{S^{n-1}} \beta Q(x', \theta', \tau') \Bigg\{ \int_{S^{n-1}} \int_{\Omega} \delta(x' - x - \theta(\tau' - t)) \eta(\theta \cdot \theta') \\
&\times \; dx d\theta \Bigg\} \, d\theta' dx' d\tau' \\
&= \int_0^t \int_{\Omega} \int_{S^{n-1}} \beta Q(x', \theta', \tau') \Bigg\{ \int_{S^{n-1}} \eta(\theta \cdot \theta') \int_{\Omega} \delta(x' - x - \theta(\tau' - t)) \\
&\times \; dx d\theta \Bigg\} \, d\theta' dx' d\tau' \\
&= \int_0^t \int_{\Omega} \int_{S^{n-1}} \beta Q(x', \theta', \tau') \Bigg\{ \int_{S^{n-1}} \eta(\theta \cdot \theta') d\theta \Bigg\} \, d\theta' dx' d\tau' \\
&= \int_0^t \int_{\Omega} \int_{S^{n-1}} \beta Q(x', \theta', \tau') \; d\theta' dx' d\tau' \\
&= \beta \int_0^t \left(\int_{\Omega} \int_{S^{n-1}} Q(x', \theta', \tau') \; d\theta' dx' \right) d\tau' \\
&\leq \beta \int_0^t \hat{Q} \; d\tau' = \beta \hat{Q} \int_0^t \; d\tau' \\
&= \hat{Q} \beta t.
\end{aligned}
$$

Continuing the iterations (by induction), one obtains

$$
\int_{S^{n-1}} \int_{\Omega} u_j(x, \theta, t) \; dx d\theta \leq \hat{Q} \frac{\beta^j t^j}{j!}, \quad \forall \; j \geq 0,
$$

which implies

$$
\begin{aligned}
\int_{S^{n-1}} \int_{\Omega} u(x, \theta, t) \; dx d\theta = \int_{S^{n-1}} \int_{\Omega} \sum_{j=0}^{\infty} u_j(x, \theta, t) \; dx d\theta &= \sum_{j=0}^{\infty} \int_{S^{n-1}} \int_{\Omega} u_j(x, \theta, t) \; dx d\theta \\
&\leq \sum_{j=0}^{\infty} \hat{Q} \frac{\beta^j t^j}{j!} = \hat{Q} \exp(\beta t).
\end{aligned}
$$

The above inequality proves (2.21), which means that the Neumann series converges. ■

For more details on the proof of the existence and uniqueness of the solution of the RTE we refer to [33, 50].

2.2 Regularization

In this section we summarize a few results from regularization theory which are of importance to later sections.

2.2.1 Truncated Singular Value Decomposition

Let H_1 and H_2 be separable Hilbert spaces of finite or infinite dimensions, $\mathcal{A} : H_1 \to H_2$ a compact operator. An operator $\mathcal{A} : H_1 \to H_2$ is called *linear compact* if $\mathcal{A}$ is linear and if for every bounded subset $D \subset X$, the closure $\overline{\mathcal{A}(D)}$ is compact in H_2, in other words if the image $\mathcal{A}(D)$ is relatively compact. Any compact operator $\mathcal{A} : H_1 \to H_2$ can be written in the form

$$\mathcal{A} : x \mapsto \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \langle x, u_j \rangle v_j,$$

where $\lambda_j \to 0$ as $j \to \infty$, then $\|\mathcal{A} - \mathcal{A}_N\| \to 0$ as $N \to \infty$, where

$$\mathcal{A}_N : x \mapsto \sum_{j=1}^{N} \frac{1}{\lambda_j} \langle x, u_j \rangle v_j.$$

Let H_1 and H_2 be separable Hilbert spaces of finite or infinite dimension and $\mathcal{A} : H_1 \to H_2$ a linear operator. Given $b \in H_2$, one can consider the problem of finding $x \in H_1$ such that

$$\mathcal{A}x = b. \tag{2.30}$$

Equation (2.30) is called a *Fredholm equation of the first kind* when the operator $\mathcal{A}$ in (2.30) is compact. A solution x of (2.30) for a linear or a compact operator $\mathcal{A}$ exists if and only if $b \in \mathrm{Ran}(\mathcal{A})$, and this solution x is unique if and only if $\ker(\mathcal{A}) = 0$ [101, 162].

The definition of *well-posedness in the sense of Hadamard* is given in [14, 51, 63, 96, 167] as follows

Definition 3 *A problem (2.30) is called* well-posed *if all the following properties are satisfied*

i A solution of (2.30) exists for all admissible data b.

ii The solution is unique for all admissible data b.

iii The solution of (2.30) depends continuously on the data b.

A problem which does not satisfy one of the properties of definition 3 is called *ill-posed problem*. Inverse problems typically do not satisfy all of the Hadamard's conditions. Sometimes, condition (i) can be re-established by relaxing the notion of an exact solution, while if condition (ii) is not satisfied, one must determine which solution is of interest [63].

Hadamard's conditions in Definition 3 will fail for example in the following situations

1. If $\mathrm{Ran}(\mathcal{A}) \neq H_2$, that is, if $\mathcal{A}$ is not surjective, then the problem (2.30) might not have a solution.

2. If $\ker(\mathcal{A}) \neq \emptyset$, that is, if $\mathcal{A}$ is not injective, then the problem (2.30) does not have an unique solution.

3. If the inverse of $\mathcal{A}$ exists but is not continuous, then the solution of (2.30) does not depend continuously on the data b.

The failure of condition (iii) will produce instability in the inversion process. This condition states that small perturbations in the data b correspond to small perturbations on the solution x, in other words the solution is stable with respect to small changes in data b:

- If $b \to b'$, $\mathcal{A}x' = b'$ and $\mathcal{A}x = b$, then $x \to x'$, that is, the solution of (2.30) is stable with respect small changes in the data b [162].

For a general operator $\mathcal{A}$ and infinite-dimensional H_1, H_2, the inverse of $\mathcal{A}$ can exist but might not be continuous. Even if H_1, H_2 are finite-dimensional, the solution of (2.30) might not depend continuously on the data b in practical situations. In these situations, when the problem (2.30) is discretized, the operator A is represented by a matrix with a large condition number, such that condition (iii) is not satisfied [96, 101]. In practice the data b are always corrupted by some noise δ, so instead of equation (2.30) we have to consider

$$\mathcal{A}x = b + \delta.$$

Regularization methods are applied to overcome the difficulties that ill-posed problems present. Regularization methods solve uniquely a nearby problem instead of (2.30) and ensure stability of the inversion scheme. Therefore, we present a generalized notion of the solution of the problem (2.30) as follows:

Definition 4 *Let H_1, H_2 be Hilbert spaces and $\mathcal{A} : H_1 \to H_2$ a bounded linear operator. $x_{ls} \in H_1$ is called a* least-squares solution of 2.30 *if*

$$\|\mathcal{A}x_{ls} - b\| \leq \|\mathcal{A}x - b\| \quad \textit{for each } x \in H_1.$$

Also, $x \in H_1$ is called best-approximate solution of (2.30) *if x is a least-squares solution of (2.30) with minimal norm, i.e.,*

$$\|x\| = \arg\min_{z \in H_1} \{\|z\| \,\big|\, z \textit{ solves } \mathcal{A}x = b\},$$

which is also known as the least-squares minimum norm solution $x_{lsmn} \in H_1$ of (2.30) *defined by*

$$x_{lsmn} = \arg\min_{z} \|z\|,$$

where $z \in x_{ls} + \ker(\mathcal{A})$ and $\|\cdot\|$ denotes the standard Euclidean norm.

Definition 5 *Let $\mathcal{A} : H_1 \to H_2$ be a compact operator with adjoint $\mathcal{A}^* : H_2 \to H_1$. The square roots $\lambda_j = \sqrt{\sigma_j}$ of the eigenvalues σ_j of the self-adjoint operator $\mathcal{A}^*\mathcal{A} : H_1 \to H_1$ are called singular values of $\mathcal{A}$.*

All the eigenvalues σ of $\mathcal{A}^*\mathcal{A}$ are nonnegative. The following theorem addresses the Singular Value Decomposition (SVD) of a compact operator A [14, 101, 162].

Theorem 6 *Let $\mathcal{A} : H_1 \to H_2$ be a compact operator and $\mathcal{A}^*$ its adjoint operator. Then*

1. *The spaces H_1, H_2 allow orthogonal decompositions.*

$$H_1 = \ker(\mathcal{A}) \oplus (\ker(\mathcal{A}))^{\perp} = \ker(\mathcal{A}) \oplus \overline{\mathrm{Ran}(\mathcal{A}^*)},$$
$$H_2 = \overline{\mathrm{Ran}(\mathcal{A})} \oplus (\mathrm{Ran}(\mathcal{A}))^{\perp} = \overline{\mathrm{Ran}(\mathcal{A})} \oplus \ker(\mathcal{A}^*).$$

2. *There exist orthonormal bases $(v_j) \in H_1$ and $(u_j) \in H_2$ for $\ker(\mathcal{A})^{\perp}$ and $\overline{\mathrm{Ran}(\mathcal{A})}$, respectively, and a strictly positive non-increasing sequence of numbers λ_j, $(\lambda_1 \geq$*

$\lambda_2 \geq \ldots > 0$), *and if* $\mathrm{Ran}(\mathcal{A})$ *is infinite-dimensional then one has* $\lambda_j \to 0$ *as* $j \to \infty$, *such that the operator* $\mathcal{A}$ *acting on* x *can be represented as*

$$\mathcal{A}x = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle u_j. \tag{2.31}$$

Moreover, the following equations hold

$$\mathcal{A}v_j = \lambda_j u_j \quad \textit{and} \quad \mathcal{A}^* u_j = \lambda_j v_j.$$

3. *Furthermore, the problem (2.30) has a solution represented by*

$$x = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \langle b, u_j \rangle v_j, \tag{2.32}$$

if and only if

$$b = \sum_{j=1}^{\infty} \langle b, u_j \rangle u_j,$$

and

$$\sum_{j=1}^{\infty} \frac{1}{\lambda_j^2} |\langle b, u_j \rangle|^2 < \infty. \tag{2.33}$$

4. *If* $b \in \mathrm{Ran}(\mathcal{A}) \oplus (\mathrm{Ran}(\mathcal{A}))^{\perp}$, *we have*

$$\mathcal{A}^{\dagger} b = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} \langle b, u_j \rangle v_j, \tag{2.34}$$

where $X^{\perp}$ denotes the ortho-complement of the subspace $X \subset H$. The triplet (λ_j, v_j, u_j) is called the *singular system of the operator* $\mathcal{A}$ and the representation of $\mathcal{A}$ in terms of this singular system is called the *singular value decomposition (SVD) of* $\mathcal{A}$. The linear operator $\mathcal{A}^{\dagger}$ defined in (2.34) is called the *Moore-Penrose generalized inverse or the pseudo-inverse of* $\mathcal{A}$.

Note that $\mathrm{Ran}(\mathcal{A})$ is not necessarily equal to $\overline{\mathrm{Ran}(\mathcal{A})}$. Also, we have the following theorem

Theorem 7 *If* $b \in \mathrm{Ran}(\mathcal{A}) \oplus (\mathrm{Ran}(\mathcal{A}))^{\perp}$, *then the least squares minimum norm solution*

(or the best-approximate solution) of (2.30) satisfies

$$x_{lsmn} = \mathcal{A}^{\dagger} b,$$

and $\mathcal{A}^{\dagger}$ is bounded if and only if $\mathrm{Ran}(\mathcal{A})$ *is closed.*

In the finite-dimensional setting, we have the following SVD of a matrix A.

Theorem 8 *Every matrix $A \in \mathbb{R}^{M\times N}$ allows a decomposition*

$$A = U\Lambda V^T, \tag{2.35}$$

where the matrices U and V are constructed by the vectors u_j and v_j respectively, as follows

$$U = [u_1, ..., u_M] \qquad\qquad V = [v_1, ..., v_N],$$

and $\Lambda \in \mathbb{R}^{M\times N}$ is diagonal with nonnegative elements such that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{\min(M,N)} \geq 0$. The matrix λ is described as follows

- *if $M < N$,*

  $$\Lambda = [\mathrm{diag}(\lambda_1, ..., \lambda_M), 0],$$

  where 0 of the last expression denotes a zero matrix of size $M \times (N - M)$.

- *if $M > N$*

  $$\Lambda = \begin{bmatrix} \mathrm{diag}(\lambda_1, ..., \lambda_N) \\ 0 \end{bmatrix}.$$

  where 0 of the last expression denotes a zero matrix of size $(M - N) \times N$.

- *if $M = N$*

  $$\Lambda = [\mathrm{diag}(\lambda_1, ..., \lambda_M)].$$

Note that the vectors $(u_j)_{j=1}^M$ and $(v_j)_{j=1}^N$ are eigenvectors of AA^* and A^*A, respectively.

Another difficulty when attempting to solve (2.30) arises when b has a nonzero components in the subspace orthogonal to $\mathrm{Ran}(A)$ as it can be seen as follows.

Let $\mathcal{A}$ be a compact operator (where H_1, H_2 can be infinite or finite dimensional) and $P : H_2 \to \overline{\mathrm{Ran}(\mathcal{A})}$ be the orthogonal projection on $\overline{\mathrm{Ran}(\mathcal{A})}$ which is defined by

$$b \mapsto \sum_j \langle b, u_j \rangle u_j. \tag{2.36}$$

Then, for any $x \in H_1$, it holds

$$\|\mathcal{A}x - b\|^2 = \|\mathcal{A}x - Pb\|^2 + \|(1-P)b\|^2 \geq \|(1-P)b\|^2.$$

Therefore, when b has a nonzero component in the subspace orthogonal to $\mathrm{Ran}(\mathcal{A})$, then the equation (2.30) does not have an exact solution [101]. In such cases, instead of solving the equation (2.30), we can try to solve the projected equation

$$\mathcal{A}x = P\mathcal{A}x = Pb. \tag{2.37}$$

However, for corrupted data b, the convergence condition in (2.33) may not hold since the components $\langle b, u_j \rangle$ may not decrease towards zero sufficently fast. In this case, the solution of (2.30) would not exist.

To avoid the above mentioned difficulty, we can make use of the following projection operators. Let $P_k : H_2 \to span\{u_1, ..., u_k\}$ denote the finite-dimensional orthogonal projection, i.e.,

$$b \mapsto \sum_{j=1}^{k} \langle b, u_j \rangle u_j.$$

We have $P_k b \in \mathrm{Ran}(\mathcal{A})$ for all $k \in \mathbb{N}$ since P_k is finite dimensional, and $P_k b \to Pb$ in H_2 when $k \to \infty$. Now one can consider to solving instead of (2.37) the following finite-dimensional projected equation

$$\mathcal{A}x = P_k b, \tag{2.38}$$

which is solvable with solution given below.

Definition 9 *Let $\mathcal{A} : H_1 \to H_2$ be a compact operator with the singular triplet (λ_j, v_j, u_j). The problem of finding $x \in H_1$ and $x \perp \ker(\mathcal{A})$ such that equation (2.38) holds for some $k \in \mathbb{N}$ is called the truncated SVD solution of the problem (2.30).*

If one takes the inner product with u_j on both sides of eq. (2.38), we obtain

$$\lambda_j \langle x, u_j \rangle = \begin{cases} \langle b, u_j \rangle & 1 \leq j \leq k, \\ 0 & j > k. \end{cases} \tag{2.39}$$

Therefore, the solution of the problem given in Definition 9 has a solution x^k given by

$$x^k = x^0 + \sum_{j=1}^{k} \frac{1}{\lambda_j} \langle b, u_j \rangle, \tag{2.40}$$

where $x^0 \in \ker(\mathcal{A})$. In the finite-dimensional setting [51, 101, 162], the solution of the projected equation $Ax = Pb$ is given by

$$x = x^0 + A^\dagger b,$$

where x^0 is an arbitrary vector in $\ker(A)$ and $\mathcal{A}^\dagger$ is the pseudo-inverse of A defined in (2.34), which in the finite-dimensional case has the following explicit expression

$$A^\dagger = V \Lambda^\dagger U, \tag{2.41}$$

where the matrix $\Lambda^\dagger \in \mathbb{R}^{N \times M}$ is given by

$$\begin{pmatrix} 1/\lambda_1 & 0 & \cdots & 0 \\ 0 & 1/\lambda_2 & & \\ & & \ddots & \\ \vdots & & 1/\lambda_p & \vdots \\ & & & 0 \\ & & & & \ddots \\ 0 & & \cdots & 0 \end{pmatrix}.$$

with $p := \min(M, N)$. The condition number of a matrix A is defined by

$$\mathrm{cond}(A) = \|A\| \|A^\dagger\|.$$

It provides an approximate measure to classify well and ill-posed (finite-dimensional)

problems. Ill-posed problems are characterized by a large condition number $\mathrm{cond}(A)$.

2.2.2 Filters

When singular values λ_j in (2.32) become too small, this will lead to an unstable solution x. In order to avoid this numerical instability, a filter function $w(\lambda_j^2)$ can be incorporated such that the product $w_\alpha(\lambda_j^2)(\lambda_j)^{-1} \to 0$ as $\lambda_j \to 0$ and $w_\alpha(\lambda_j^2) \approx 1$ for large values of λ_j^2. Here α refers to the regularized parameter. Regularizing an ill-posed problem in this way can be interpreted as 'filtering' the SVD components of the solution x [51, 107]. When incorporating this filter into (2.32) we obtain

$$x = \sum_{j=1}^{\infty} w_\alpha(\lambda_j^2)\frac{1}{\lambda_j}\langle b, u_j\rangle v_j, \tag{2.42}$$

In the infinite case, the filter of the TSVD for a compact operator $\mathcal{A}$ is given by

$$w_\alpha(\lambda_j^2) = \begin{cases} 1 & \text{if } j \leq k, \\ 0 & \text{if } j > k. \end{cases} \tag{2.43}$$

Equation (2.43) resembles (and is analogous to) equation (2.39) which describes the infinite-dimensional projected property presented in the TSVD discussion. Usually, filter (2.43) removes 'high frequency' components which often correspond to small singular values λ_j. In the finite-dimensional case, where $A \in \mathbb{R}^{M\times N}$, the following formula for the truncated SVD can be used [107, 162]

$$w_\alpha(\lambda_j^2) = \begin{cases} 1 & \text{if } \lambda_j^2 > \alpha, \\ 0 & \text{if } \lambda_j^2 \leq \alpha. \end{cases} \tag{2.44}$$

By incorporating the filter function (2.44) into the approximate solution representation (2.32) we are removing those singular values λ_j which produce instabilities. Thus, the approximate solution is given by

$$x = \sum_{j=1}^{M} w_\alpha(\lambda_j^2)\langle b, u_j\rangle v_j = \sum_{\substack{j=1 \\ \lambda_j^2>\alpha}}^{M} \lambda_j^{-1}\langle b, u_j\rangle v_j. \tag{2.45}$$

The expression (2.45) is called *truncated singular value decomposition* [63, 162]. As mentioned, the parameter α plays the role of a 'regularization parameter', which corresponds here to the threshold of the filtering process. The target in regularization techniques is to find an optimal value for α in an appropriate sense. Clearly, when α is chosen too small, the noise still may produce an unstable solution x, while if α is chosen too large, most of the main components $\langle b, u_j \rangle v_j$ are removed, giving an inadequate or incorrect solution [162].

Another filter is the *Landweber filter* given by

$$w_\alpha(s^2) = 1 - (1 - \beta s^2)^\alpha, \tag{2.46}$$

where $\beta \in \mathbb{R}^+$. If one substitutes the Landweber filter (2.46) into eq. (2.42) for a linear bounded operator $\mathcal{A}$, and truncates the sum until k terms, we obtain the following *Landweber regularized solution*

$$\begin{aligned} x^k &= \sum_{j=1}^{\infty} w_k(\lambda_j^2) \frac{1}{\lambda_j} \langle b, u_j \rangle v_j = \sum_{j=1}^{\infty} \frac{1}{\lambda_j} (1 - (1 - \beta \lambda_j^2)^k) \langle b, u_j \rangle v_j, \\ &= \sum_{i=0}^{k-1} \sum_{j=1}^{\infty} \beta \lambda_j (1 - \beta \lambda_j^2)^i \langle b, u_j \rangle v_j = \sum_{i=0}^{k-1} \sum_{j=1}^{\infty} \beta \lambda_j (1 - \beta \lambda_j^2)^i \langle b, u_j \rangle v_j, \\ &= \sum_{i=0}^{k-1} R^i \beta \mathcal{A}^* b, \end{aligned}$$

where the equality in the second row holds due the geometric sum formula. Here

$$R = I - \beta \mathcal{A}^* \mathcal{A}.$$

Inductively, the last expression yields

$$x^k = \sum_{i=0}^{k-1} R^i \beta \mathcal{A}^* b, \tag{2.47}$$

which, for $x^0 = 0$, is equivalent to the following iteration

$$x^{k+1} = x^k - \beta \mathcal{A}^* (\mathcal{A} x^k - b), \quad k = 0, 1, \ldots . \tag{2.48}$$

In effect, for $k = 1$, eq. (2.47) is given by

$$x_1 = \beta \mathcal{A}^* b.$$

Therefore, eq. (2.48) for $k = 1$ is equal to $x_1 = \beta \mathcal{A}^* b$. Assuming that eq. (2.47) holds, one obtains

$$x^{k+1} = x^k - \beta \mathcal{A}^*(\mathcal{A}x^k - b) = Rx^k + \beta \mathcal{A}^* b = R \sum_{i=0}^{k-1} R^i \beta \mathcal{A}^* b = \sum_{i=0}^{k} R^i \beta \mathcal{A}^* b.$$

Equation (2.48) is known as the *Landweber iteration* which converges for $0 < \beta < \dfrac{2}{\|\mathcal{A}\|^2}$. Therefore, the regularization parameter α for the Landweber scheme is an index $\alpha = k$ which indicates the iteration count where the scheme (2.48) is terminated.

Another typical and stable choice for a regularization filter is the *Tikhonov regularization filter*. It is given by

$$w_\alpha(s^2) = \frac{s^2}{s^2 + \alpha}.$$

In the finite-dimensional case, the Tikhonov regularized solution is expressed by

$$x = \sum_{j=1}^{N} w_\alpha(\lambda_j^2) \frac{1}{\lambda_j} \langle b, u_j \rangle v_j = (A^T A + \alpha)^{-1} A^T b. \tag{2.49}$$

We will discuss Tikhonov regularization in more details in the following section.

2.2.3 Tikhonov regularization of linear operator equations

In the linear case, i.e., when $\mathcal{A}$ is a linear operator, Tikhonov regularization is best known as the minimization of the following "Tikhonov functional"

$$\mathcal{F}_\alpha(x) = \|\mathcal{A}x - b\|^2 + \alpha \|x\|^2, \tag{2.50}$$

where $\alpha > 0$ is the regularizer parameter. In the framework of Variational Calculus the parameter α is related to a Lagrange multiplier. Here one is looking for the minimum of $\|\mathcal{A}x - b\|^2$ subject to the restriction $\|x\|^2 = R$ for some $R > 0$. Let $\mathcal{A} : H_1 \to H_2$ be a compact operator. Then, the Tikhonov regularized solution x_α exists, is unique and

furthermore can be expressed by the formula

$$x_\alpha = (\mathcal{A}^*\mathcal{A} + \alpha I)^{-1}\mathcal{A}^* b = \sum_{j=1}^{\infty} \frac{\lambda_j}{\lambda_j^2 + \alpha} \langle b, u_j \rangle v_j.$$

As *generalized Tikhonov regularization* we understand the minimization of the more general functional

$$\mathcal{F}_\alpha(x) = \|\mathcal{A}x - b\|^2 + \alpha \mathcal{D}(x), \tag{2.51}$$

where $\mathcal{D}(x)$ is a nonnegative functional and $\alpha > 0$. The standard version of this penalty functional is

$$\mathcal{D}(x) = \|\mathcal{L}(x - x_0)\|^2, \tag{2.52}$$

where $\mathcal{L}$ is a linear operator and $x^0 \in H_1$ is the available a priori information on x. The operator $\mathcal{L}$ may for example be the identity operator or a differential operator. In discrete or finite-dimensional cases, typically the operator $\mathcal{L}$ is chosen as either the identity matrix, or an approximation for a differential operator, or a diagonal weighting matrix [51, 107].

In the finite-dimensional setting, the minimiser of the Tikhonov functional (2.51), (2.52) is given by

$$x_\alpha = (A^T A + \alpha L^T L)^{-1}(A^T b + \alpha L^T L x^0).$$

This minimiser is unique if $\ker(A) \cap \ker(L) = \emptyset$ or, which is equivalent to ask, if $A^T A + \alpha L^T L$ is positive definite. The minimizer of the Tikhonov functional (2.51), (2.52) can be obtained by minimizing

$$\mathcal{F}_\alpha(x) = \left\| \begin{bmatrix} A \\ \sqrt{\alpha} L \end{bmatrix} x - \begin{bmatrix} b \\ \sqrt{\alpha} L x^0 \end{bmatrix} \right\|^2,$$

for which the minimizer is

$$x_\alpha = K^\dagger \begin{bmatrix} b \\ \sqrt{\alpha} L x^0 \end{bmatrix} \quad \text{with} \quad K = \begin{bmatrix} A \\ \alpha L \end{bmatrix}.$$

Wang et. al. [167] have suggested some extrapolated methods to accelerate the standard Tikhonov regularization method if the true solution x is assumed to be smooth and the 'extrapolated' parameters are selected properly.

2.2.4 Tikhonov regularization of nonlinear operator equations

The imaging modalities discussed in the thesis involve nonlinear forward operators which require some slight generalizations of the above discussed theory on Tikhonov regularization. Some of these are outlined in the following for future reference.

Let now $\mathcal{G} : H_1 \to H_2$ be a *nonlinear* operator. One can consider the problem of finding $x \in H_1$ such that

$$\mathcal{G}(x) = g, \tag{2.53}$$

where $g \in H_2$ represents the exact data. In practice, certainly, we measure perturbed (noisy) data g_δ, and assume that the following inequality holds

$$\|g_\delta - g\| \leq \delta, \tag{2.54}$$

with δ indicating the level of noise.

Typically the ill-posed problem (2.53)is written as the following least-square functional

$$\mathcal{F}(x) = \|\mathcal{G}(x) - g\|^2. \tag{2.55}$$

The problems (2.53) we are interested are *ill-posed* in the sense that even small changes in the measured data g_δ may cause large errors in the solution vector x. A standard Tikhonov regularization can be formulated for such ill-posed nonlinear problems, which amounts to finding the minimiser of the functional

$$\mathcal{F}_\alpha(x) = \|\mathcal{G}(x) - g\|^2 + \alpha\|x\|^2. \tag{2.56}$$

For practically finding minimizers of Tikhonov functional involving nonlinear operators, the following concept of a Fréchet derivative is of importance.

Definition 10 *Let $z, x^0 \in H_1$. The operator $\mathcal{G} : H_1 \to H_2$ is Fréchet differentiable at a point x^0 if it allows an expansion*

$$\mathcal{G}(x^0 + z) = \mathcal{G}(x^0) + \mathcal{G}'_{x^0}(z) + \mathcal{W}(x^0, z),$$

here $\mathcal{G}'_{x^0} : H_1 \to H_2$ *is a continuous linear operator and*

$$\|\mathcal{W}(x^0, z)\| \leq \|z\|\epsilon(x^0, z),$$

where the functional $\epsilon(x^0, z)$ *tends to zero when* $z \to 0$.

Assuming that $\mathcal{G}$ is Fréchet differentiable, the functional (2.56) can be approximated as

$$\begin{aligned} \mathcal{F}_\alpha(x) \approx \tilde{\mathcal{F}}_\alpha(x; x^0) &= \|\mathcal{G}(x^0) + \mathcal{G}'_{x^0}(x - x^0) - g\|^2 + \alpha\|x\|^2 \\ &= \|\mathcal{G}'_{x^0}x - \mathfrak{g}(y, x^0)\|^2 \end{aligned} \tag{2.57}$$

where

$$\mathfrak{g}(y, z) = g - \mathcal{G}(z) + \mathcal{G}'_z z. \tag{2.58}$$

Applying now Tikhonov regularization for the linearised operator $\mathcal{G}'_{x^0}$, the minimizer of the functional (2.57) is given by

$$x = \left((\mathcal{G}'_{x^0})^*\mathcal{G}'_{x^0} + \alpha I\right)^{-1} (\mathcal{G}_{x^0})^*\mathfrak{g}(g, x^0). \tag{2.59}$$

A discrete version of the Tikhonov regularization method (2.59) is given next in form of a pseudo-code [101]:

1. Select an initial guess x^0 and set $k = 0$.
2. Calculate

$$x = \left((\mathcal{G}'_{x^k})^*\mathcal{G}'_{x^k} + \alpha I\right)^{-1} (\mathcal{G}_{x^k})^*\mathfrak{g}(g, x^k),$$

   with $\mathfrak{g}(g, x^k)$ as defined in (2.58), and define $\delta x = x - x^k$.
3. Calculate the minimiser $s > 0$ of the function

$$f(s) = \|\mathcal{G}(x^k + s\delta x) - g\|^2 + \|x^k + s\delta x\|^2.$$

4. Set $x^{k+1} = x^k + s\delta x$ and $k \leftarrow k + 1$.
5. Repeat steps (2-4) until convergence.

A different derivation of the expression (2.59) for finite-dimensional nonlinear operator $\mathcal{G}$ is as follows. Let $\mathcal{G} : \mathbb{R}^N \to \mathbb{R}^M$ be a nonlinear operator, $g = (g_1, ..., g_M) \in \mathbb{R}^M$ the data vector and $x \in \mathbb{R}^N$ a vector variable. The nonlinear problem is to find $x \in \mathbb{R}^N$ such that

$$\mathcal{G}(x) = g. \tag{2.60}$$

In order to solve the ill-posed problem (2.60), one can attempt to solve the following least-square functional

$$\mathcal{F}(x) = \|\mathcal{G}(x) - g\|^2. \tag{2.61}$$

Using now the following Taylor expansion of second order for $\mathcal{F}(x)$ around the present estimate x^k, given by

$$\mathcal{F}(x) \approx \mathcal{F}(x^k) + \left(\frac{\partial \mathcal{F}(x^k)}{\partial x} + \frac{1}{2}(x - x^k)^T \frac{\partial^2 \mathcal{F}(x^k)}{\partial x^2} \right)(x - x^k), \tag{2.62}$$

a quadratic approximation for $\mathcal{F}(x)$ is obtained and equated to zero for finding its minimiser

$$\frac{\partial \mathcal{F}(x)}{\partial x} = \frac{\partial \mathcal{F}(x^k)}{\partial x} + (x - x^k)^T \frac{\partial^2 \mathcal{F}(x^k)}{\partial x^2} = 0.$$

Assuming that the Hessian matrix $\partial \mathcal{F}^2 / \partial x^2$ is invertible we obtain the Newton-Raphson iteration

$$x^{k+1} = x^k - \left(\frac{\partial^2 \mathcal{F}(x^k)}{\partial x^2} \right)^{-1} \frac{\partial \mathcal{F}(x^k)}{\partial x}. \tag{2.63}$$

Let us use the notation $\mathcal{C}_{(k)}$ (the Jacobian matrix of $\mathcal{G}$) and $\mathcal{D}_{(k)}$ for the following quantities

$$\mathcal{C}_{(k)} = \frac{\partial \mathcal{G}(x^k)}{\partial x},$$
$$\mathcal{D}_{(k)} = -\left(\sum_{j=1}^{M} (g_j - F_j(x^k) \frac{\partial^2 F_j}{\partial x^2}) \right).$$

Then the gradient $\partial \mathcal{F} / \partial x \in \mathbb{R}^N$ is given by

$$\frac{\partial \mathcal{F}(x^k)}{\partial x} = -2\mathcal{C}_{(k)}^T (g - \mathcal{G}(x^k)), \tag{2.64}$$

and the Hessian matrix $\partial \mathcal{F}^2 / \partial x^2 \in \mathbb{R}^{N \times N}$ by

$$\frac{\partial^2 \mathcal{F}(x^k)}{\partial x^2} = 2\mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + 2\mathcal{D}_{(k)}, \tag{2.65}$$

Using now Equations (2.64) and (2.65), the iteration formula (2.63) gets the form

$$x^{k+1} = x^k + \left(\mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + \mathcal{D}_{(k)}\right)^{-1} \mathcal{C}_{(k)}^T (g - \mathcal{G}(x^k)), \tag{2.66}$$

For more details about this derivation see [107].

Notice that different approximations of the term $\mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + \mathcal{D}_{(k)}$ in (2.66) lead to a range of popular iterative approximation methods. The simplest one is the *steepest descent method* which approximates $\mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + \mathcal{D}_{(k)}$ with the identity matrix. The *Gauss-Newton method* approximates this term by $\mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + \mathcal{D}_{(k)} \approx \mathcal{C}_{(k)}^T \mathcal{C}_{(k)}$, that is, when $\partial^2 F_j / \partial x^2 \approx 0$ or the residual $\|\mathcal{G}(x) - g\| \approx 0$. *Levenberg-Marquardt method* approximates $\mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + \mathcal{D}_{(k)} \approx \mathcal{C}_{(k)}^T \mathcal{C}_{(k)} + \alpha I$, where α is a positive regularizer parameter.

Convergence properties of the above schemes have been intensively investigated in the literature, see for example [102]. Let us assume that the nonlinear problem (2.53) is solvable and that the regularizer parameter $\alpha := \alpha(\delta)$, where δ is the noise level, $\delta^2/\alpha \to 0$, when both α and δ tends to zero. Then the minimizer of the Tikhonov functional (2.56) converges to a solution of the nonlinear problem (2.53). One important convergence result is the following [102]. Let $x^\dagger$ be the minimal norm solution of (2.53) and assume that the following equation holds

$$x^\dagger - x^0 = \left((\mathcal{G}'_{x^\dagger})^* \mathcal{G}'_{x^\dagger}\right)^\nu u,$$

where $1/2 \leq \nu \leq 1$ and $\|u\|$ sufficiently small. Then, the rate of convergence of a Tikhonov regularization is given by

$$\|x_\alpha - x^\dagger\| = \mathcal{O}(\delta^{\frac{2\nu}{2\nu+1}}).$$

As for linear discrete operators, one can consider the *generalized Tikhonov regularization for the discrete nonlinear problem*. It is given by the minimization of the following functional

$$\mathcal{F}_\alpha(x) = \|\mathcal{G}(x) - g\|^2 + \alpha G(x), \tag{2.67}$$

where G is a nonnegative functional and $\alpha > 0$ is the regularization parameter. To obtain the iteration formula for the approximation of the minimizer of (2.67), we proceed in a

similar manner as in the derivation of (2.66). Firstly, the gradient

$$\frac{\partial \mathcal{F}(x^k)}{\partial x} = -2\mathcal{C}_{(k)}^T(g - \mathcal{G}(x^k)) + \alpha g_k^G, \tag{2.68}$$

and the Hessian matrix

$$\frac{\partial^2 \mathcal{F}(x^k)}{\partial x^2} = 2\mathcal{C}_{(k)}^T\mathcal{C}_{(k)} + 2\mathcal{D}_{(k)} + \alpha\mathcal{H}_k^G, \tag{2.69}$$

are calculated where

$$g_k^G = \frac{\partial G(x^k)}{\partial x},$$
$$\mathcal{H}_k^G = -\frac{\partial^2 G(x^k)}{\partial x^2}.$$

Then, using eq. (2.68) and (2.69), we obtain the iteration rule

$$x^{k+1} = x^k + \left(\mathcal{C}_{(k)}^T\mathcal{C}_{(k)} + \mathcal{D}_{(k)} + \frac{1}{2}\alpha\mathcal{H}_k^G\right)^{-1}\left(\mathcal{C}_{(k)}^T(g - \mathcal{G}(x^k)) - \frac{1}{2}\alpha g_k^G\right). \tag{2.70}$$

The Gauss-Newton regularized method assumes $\mathcal{D}_k \approx 0$, resulting in the modified iteration rule

$$x^{k+1} = x^k + \left(\mathcal{C}_{(k)}^T\mathcal{C}_{(k)} + \frac{1}{2}\alpha\mathcal{H}_k^G\right)^{-1}\left(\mathcal{C}_{(k)}^T(g - \mathcal{G}(x^k)) - \frac{1}{2}\alpha g_k^G\right). \tag{2.71}$$

2.3 Iterative reconstruction techniques

Since optical or fluorescence tomography mathematically represent *nonlinear* imaging modalities, iterative reconstruction or regularization techniques are of special importance. We have discussed above already the well-known Newton-Raphson technique and some of its variants. In this section, we will outline a few additional iterative techniques which are of particular importance in this thesis. These are the Kaczmarz (K) technique, the Landweber (L), Landweber-Kaczmarz (LK) and Loping-Landweber-Kaczmarz (L-LK) methods.

2.3.1 Kaczmarz method

Let $\mathcal{A}_j : H \to H$ be linear bounded operators where H and H_j, $j = 1, \ldots, M$ denote separable Hilbert spaces, and let $b_j \in \mathrm{Ran}(\mathcal{A}_j)$ be given. Then we consider the following problems

$$\mathcal{A}_j x = b_j, \quad 1 \leq j \leq M. \tag{2.72}$$

Let furthermore X_j be affine subspaces of H defined by

$$X_j = \{x \in H_1 \mid \mathcal{A}_j x = b_j\}, \tag{2.73}$$

and $\mathcal{P}_j : H_1 \to X_j$ orthogonal projectors onto these subspaces X_j. The sequential projections are defined as $\mathcal{P} := \mathcal{P}_M \mathcal{P}_{M-1} \cdots \mathcal{P}_2 \mathcal{P}_1$. For an initial guess x^0, the Kaczmarz sequence $\{x^k\} \subset H$ is defined recursively as

$$x^{k+1} = \mathcal{P} x^k. \tag{2.74}$$

The following theorem gives sufficient conditions of convergence for solution of equation (2.30) using the Kaczmarz sequence as an iteration approximation method.

Theorem 11 *If $X = \bigcap_{j=1}^{M} X_j$ is not empty. Then the Kaczmarz sequence $\{x^k\}$ converges to the least-squares minimum norm solution of the equation (2.30), i.e.,*

$$\lim_{k \to \infty} x^k = x, \quad \mathcal{A}x = b, \quad x \perp \ker(\mathcal{A}).$$

When applied in the X-ray tomography's context, The Kaczmarz method is also known as Algebraic Reconstruction Technique (ART). For a finite-dimensional linear problem it can be formulated as follows. Let $A : \mathbb{R}^N \to \mathbb{R}^M$ be a linear mapping, then $A \in \mathbb{R}^{M \times N}$ can be decomposed in mappings $A_j : \mathbb{R}^N \to \mathbb{R}^{l_j}$ which are assumed to be surjective for the following exposition, i.e., the blocks A_j have full row rank. Then A has the following form

$$A = \begin{bmatrix} A_1 \\ \vdots \\ A_l \end{bmatrix} \in \mathbb{R}^{M \times N} \quad A_j \in \mathbb{R}^{l_j \times N},$$

where $l_1 + l_2 + ... + l_s = M$. Furthermore, the linear problem (2.30) can be written as a system of equations

$$A_j x = b_j, \quad A_j \in \mathbb{R}^{l_j \times N}, \; b_j \in \mathbb{R}^{l_j}. \tag{2.75}$$

Therefore, the orthogonal projection P_j to the hyperplanes X_j defined in (2.73) is explicitly given by

$$P_j x = x + A_j^T \left(A_j A_j^T\right)^{-1} (b_j - A_j x). \tag{2.76}$$

The general Kaczmarz method allows for a more general form of projections P_j given in (2.76) by

$$P_{j\omega} x = x + \omega A_j^T \left(A_j A_j^T\right)^{-1} (b_j - A_j x).$$

where $0 < \omega < 2$ is a relaxation parameter [64, 101]. In practice, a priori information is used to select the initial element x^0 of the sequence (2.74). In the most popular form of of the Kaczmarz method found in the literature the blocks A_j are considered as $A_j : \mathbb{R}^N \to \mathbb{R}^1$ and $y \in \mathbb{R}^1$, that is, $l_j = 1$. In this case, the blocks A_j are reduced to rows of the matrix $A \in \mathbb{R}^{M \times N}$. A cycle in this method consists of M consecutive steps, that is, the subindex k of x^k in the following algorithm can be substituted by $[k]$, where $[k] := (k \mod M)$. Figure 2.1 depicts a sequence of steps of this algorithm for two lines r_1 and r_2 in $2D$, that is, for a matrix $A \in \mathbb{R}^{2\times 2}$.

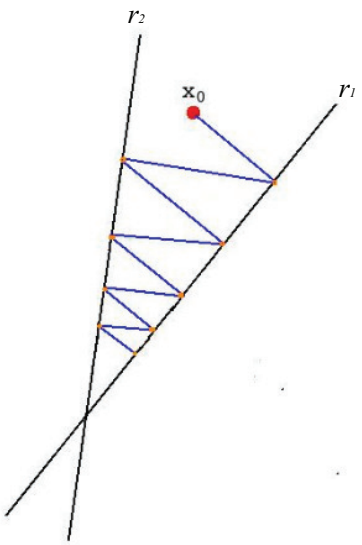


Figure 2.1: Kaczmarz method in two dimensions.

The explicit algorithm for the Kaczmarz method can be formulated as follows.

Algorithm
Initialization
Set $k = 0, x^0 = 0$;
Repeat until convergence
$z^0 = x^k$
for $j = 1 : M$
$z^j = z^{j-1} + A_j^T (A_j A_j^T)^{-1} (b_j - A_j^T z^{j-1})$
end
$x^{k+1} = z^M;\ k \leftarrow k + 1$
end

Table 2.1: Kaczmarz method.

2.3.2 Landweber, Landweber-Kaczmarz and Loping-LK methods

Let $S \subset H$, where H is a Hilbert space. A set S is called an *invariant set* for an operator $T : H \to H$ if $T(S) \subset S$. A vector $x \in H$ is called a *fixed point* of T if the image $T(x)$ is equal to x, that is, $T(x) = x$. The operator T is called a *contraction* on S if there exists a positive real number $\kappa < 1$ such that for all $x, z \in S$,

$$\|T(x) - T(z)\| < \kappa \|x - z\|. \tag{2.77}$$

Furthermore, T is called a *nonexpansive mapping* if there exists a positive real number $\kappa \leq 1$ such that for all $x, z \in S$ eq. (2.77) is satisfied. The Landweber method is based on the following fixed point theorem.

Theorem 12 *Let H be a Hilbert space and $S \subset H$ a closed invariant set for a contraction operator $T : H \to H$ in S. Then there exists an unique fixed point $x \in S$, that is, $T(x) = x$. This fixed point x can be calculated as the limit of the sequence (x^k) using the iteration*

$$x^{k+1} = T(x^k),$$

for a given but arbitrary value x^0.

The data vector b in equation (2.30) can be written as

$$b = Pb + (1 - P)b,$$

where

$$Pb \in \overline{\mathrm{Ran}(\mathcal{A})}, \qquad (1-P)b \in (\mathrm{Ran}\,\mathcal{A})^{\perp},$$

where P denotes the projection operator defined in (2.36). Since $(\mathrm{Ran}\,\mathcal{A})^{\perp} = \ker(\mathcal{A}^*)$ as stated in the SVD Theorem 6, applying the adjoint operator $\mathcal{A}^*$ to eq. (2.30) yields

$$\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*Pb + \mathcal{A}^*(1-P)b = \mathcal{A}^*b,$$

that is, one obtains the normal equation

$$\mathcal{A}^*\mathcal{A}x = \mathcal{A}^*b. \tag{2.78}$$

Therefore, after applying $\mathcal{A}^*$ to eq. (2.30), one has filtered out the component vector $(1-P)b \in (\mathrm{Ran}\,\mathcal{A})^{\perp}$ of b which is not in the range of $\mathcal{A}$. Now, we can see that if the normal equation (2.78) holds, we obtain

$$x = x + \beta\mathcal{A}^*(b - \mathcal{A}x),$$

for all $\beta \in \mathbb{R}$. Defining the operator T as the right part of the above equation

$$\mathcal{T}(x) := x + \beta\mathcal{A}^*(b - \mathcal{A}x), \tag{2.79}$$

we obtain

$$x = \mathcal{T}(x).$$

Therefore, the solution of the normal equation (2.78) is a fixed point for the operator $\mathcal{T}$ defined as above. If we assume now that $\mathcal{T}$ is a contraction, we can apply an iteration procedure $x^{k+1} = \mathcal{T}(x^k)$ to approximate the solution of (2.78), that is,

$$x^{k+1} = x^k + \beta\mathcal{A}^*(b - \mathcal{A}x^k), \quad k \in \mathbb{N}, \tag{2.80}$$

with $\beta > 0$ as a relaxation parameter. Equation (2.80) is known as the *Landweber iteration*, also often called *Landweber method* . Some a priori information is used to initialize this iteration. The Landweber method is very closely related to other methods such as the

van Cittert method, Cimmino's method, the simultaneous iterative reconstruction technique SIRT, and Gerchberg-Papoulis method [63]. Kaczmarz method is also often used in the "source-type adjoint field" technique known in the engineering community where it is required to calculate the adjoint operator of $\mathcal{A}$ [58, 102].

For example in a finite-dimensional case, Cimmino's iteration is written as

$$x^{k+1} = x^k + \tau D A^T (b - Ax^k), \quad k \in \mathbb{N},$$

where $D = \mathrm{diag}(d_i)$. The individual elements are defined as

$$d_i = \begin{cases} \dfrac{1}{M} \dfrac{1}{\|A_i|^2} & A_i \neq \mathbf{0}, \\ 0 & A_i = \mathbf{0}, \end{cases}$$

where A_i^T are the rows of the matrix $A \in \mathbb{R}^{M \times N}$.

In general, the Landweber, Cimmino and van Cittert methods belong to a more general class of *Landweber-type methods* of the form

$$x^{k+1} = x^k + \tau \mathcal{A}^* \mathcal{D} (b - \mathcal{A}x^k), \quad k \in \mathbb{N},$$

where $\mathcal{D}$ is a symmetric positive definite operator.

In the finite-dimensional setting, the matrix T is a contraction if $0 < \beta < 2/\lambda_1^2$, in which case the iteration sequence $\{x^k\}$ converges to the solution $x \in (\ker(\mathcal{A}))^\perp$ of the normal equation (2.78). In the more general infinite-dimensional setting, if $\mathcal{A}$ is compact, there is no guarantee that the Landweber iteration converges since the normal equation (2.78) may not be solvable and the operator $\mathcal{T}$ in (2.79) may not be a contraction for ill-posed problems. In these cases, if $0 < \beta < 2/\|\mathcal{A}^*\mathcal{A}\|$, then the operator $I - \beta \mathcal{A}^* \mathcal{A}$ is nonexpansive and one can still apply the Landweber iteration (2.48) and truncate this sequence using a convenient stopping criterion. For ill-posed problems usually the operator $I - \beta \mathcal{A}^* \mathcal{A}$ is not a contraction because the eigenvalues of $\mathcal{A}^* \mathcal{A}$ accumulate at the origin [63]. If $\mathcal{T}$ is compact and $b \in \mathrm{Ran}(\mathcal{A}) \oplus (\mathrm{Ran}(\mathcal{A}))^\perp$, then the Landweber iteration (2.48) converges uniformly [110].

The Landweber and the steepest descent iteration methods (in their standard form) are generally known to be slow in convergence. For this reason, modified versions have

been proposed as for example to use a non-monotone gradient method with the *Barzilai-Borwein (BB) step size criterion*. The difference between the three techniques (i) a non-monotone gradient method, (ii) the steepest descent and (iii) the Landweber method is essentially the selection of step size τ^k in each iteration when calculating

$$x^{k+1} = x^k - \tau^k \nabla \mathcal{F}(x^k) \tag{2.81}$$

for the nonlinear problem (2.55). In fact, the Landweber method is a particular case of the steepest descent method with fixed $\tau^k = \tau$ for the entire iteration scheme. In the steepest descent method, τ^k is usually determined in each step using a line search, i.e.,

$$\tau^k = \arg\min_{\tau>0} \mathcal{F}(x^k - \tau \nabla \mathcal{F}^k),$$

where $\nabla \mathcal{F}^k = \nabla \mathcal{F}(x^k)$. In the BB method [15, 43, 71, 167, 171], the Hessian is emulated by τI, that is

$$\tau I \approx \frac{\nabla \mathcal{F}^k - \nabla \mathcal{F}^{k-1}}{x^k - x^{k-1}},$$

where I is the identity. The last expression is solved in a least squares sense since it may not have a solution. Then one obtains

$$\tau^k = \arg\min_{\tau} \|\tau(x^k - x^{k-1}) - \nabla \mathcal{F}^k - \nabla \mathcal{F}^{k-1}\|^2, \tag{2.82}$$

which gives rise to the first choice of the step size τ^k for the BB method, namely

$$\tau^k_{BB1} = \frac{s^T_{k-1} r_{k-1}}{s^T_{k-1} s_{k-1}}, \quad k = 2, 3, ... \tag{2.83}$$

Here

$$s_k = x^{k+1} - x^k,$$
$$r_k := \nabla \mathcal{F}^{k+1} - \nabla \mathcal{F}^k,$$

where $\tau_1 > 0$ is an arbitrary starting value. If the symmetric form of (2.82) is used, that is,

$$\tau^k = \arg\min_{\tau} \|(x^k - x^{k-1}) - \tau(\nabla \mathcal{F}^k - \nabla \mathcal{F}^{k-1})\|^2,$$

one obtains the second choice of the step size τ^k for the BB method, namely

$$\tau_{BB2}^k = \frac{s_{k-1}^T r_{k-1}}{r_{k-1}^T r_{k-1}}, \quad k = 2, 3.... \tag{2.84}$$

Assuming that $\mathcal{G}$ is Fréchet differentiable in the problem (2.53), the nonlinear Landweber iteration for it is defined by

$$x^{k+1} = x^k - \tau(\mathcal{G}'(x^k))^*(\mathcal{G}(x^k) - g), \quad k \in \mathbb{N}, \tag{2.85}$$

where $\tau > 0$ is a relaxation parameter. The discrepancy principle can be used as a stopping rule for this method, that is, the scheme is stopped after k_* iterations until the following inequality holds

$$\|g_\delta - \mathcal{G}(x_\delta^{k_*})\| \leq \kappa\delta < \|g_\delta - \mathcal{G}(x_\delta^k)\|, \quad 0 \leq k < k_*, \tag{2.86}$$

where δ is the level noise of (2.54), and $\kappa > 0$ is chosen conveniently. For regularized linear problems using the Landweber iteration, the Morozov's discrepancy principle has been used successfully with $\kappa > 1$ [102].

If we assume that the nonlinear problem (2.53) has a solution locally, i.e., in a closed ball with centre at x^0 and radius 2ρ denoted by $B_{2\rho}(x^0) \subset H_1$, local convergence of (2.85) is achieved if the following two conditions are satisfied. Firstly, the Fréchet derivative $\mathcal{G}'$ must be satisfied

$$\|\mathcal{G}'(x)\| \leq 1, \quad x \in B_{2\rho}(x^0). \tag{2.87}$$

Condition (2.87) can be achieved scaling eq. (2.53) appropriately with a relaxation parameter τ. Alternatively, instead of scaling eq. (2.53), one can add the relaxation parameter τ in the iteration scheme, as in iteration (2.85). Secondly, the following *local tangential cone condition* must be satisfied

$$\|\mathcal{G}(x) - \mathcal{G}(\tilde{x}) - \mathcal{G}'(x)(x - \tilde{x})\| \leq \eta\|\mathcal{G}(x) - \mathcal{G}(\tilde{x})\|, \tag{2.88}$$

with $\eta < 1/2$ and $x, \tilde{x} \in B_{2\rho}(x^0)$, assuming that the nonlinear problem (2.53) has a solution $x_* \in B_\rho(x^0)$ and the conditions (2.87) and (2.88) hold. Let now be $x_\delta^k \in B_\rho(x_*)$.

Then, if the following condition holds

$$\|g_\delta - \mathcal{G}(x_\delta^k)\| > 2\frac{1+\eta}{1-2\eta}\delta, \tag{2.89}$$

x_δ^{k+1} is a better approximation of x_* than x_δ^k, where both $x_\delta^k, x_\delta^{k+1} \in B_\rho(x_*) \subset B_\rho(x^0)$. This also implies that $x_\delta^{k_*}$ converges to a solution of the problem (2.53), where k_* is selected according to the stopping rule (2.86). If moreover $\ker(\mathcal{G}'(x^\dagger)) \subset \ker(\mathcal{G}'(x))$ for all $x \in B_\rho(x^\dagger)$, then x_{k_*} converges to $x^\dagger$ when $\delta \to 0$.

In view of condition (2.89) and the stopping rule (2.86), the parameter κ must be chosen such that

$$\kappa > 2\frac{1+\eta}{1-2\eta} > 2 \tag{2.90}$$

for obtaining local convergence of the Landweber iteration.

Similar to linear problems, the convergence rate for nonlinear problems using Landweber regularization is quite slow. For a discussion of convergence rates and more details of the nonlinear Landweber iteration and its modified versions, we refer to [102].

Given a set of M data, equation (2.53) can be written as a set of M equations of the form

$$\mathcal{G}_i(x) = g_i, \quad i = 0, \ldots, M-1, \tag{2.91}$$

where $\mathcal{G}_i : D_i \subset H_1 \to H_2$, and D_i being the corresponding domains of the functions $\mathcal{G}_i$. Furthermore, $g_i \in H_{2,i}$, where the $H_{2,i}$ are (possibly different) spaces such that $H_2 = \bigcup_{i=0}^{M} H_{2,i}$.

Standard methods for the solution of (2.91) are based on rewriting (2.91) as a single equation $\mathcal{G}(x) = y$, where the operator $\mathcal{G}$ is defined by $\mathcal{G} := 1\sqrt{M}(\mathcal{G}_0, \ldots, \mathcal{G}_{M-1})$ and the data vector g by $g = 1/\sqrt{M}(g_0, \ldots, g_{M-1})$. In practice, we have measurements $y_{i,\delta}$ which are typically corrupted with some noise, and satisfying

$$\|g_{i,\delta} - g_i\| \leq \delta_i, \quad i = 0, ..., M-1, \tag{2.92}$$

where δ_i is the noise level of each measurement. We denote δ by $\delta := (\delta_0, ..., \delta_{M-1})$. Applying the nonlinear Landweber iteration (2.85), one obtains

$$x_\delta^{k+1} = x_\delta^k - \sum_{i=0}^{M-1} \tau_i(\mathcal{G}_i'(x^k))^*(\mathcal{G}_i(x^k) - g_{i,\delta}), \quad k \in \mathbb{N}, \tag{2.93}$$

Adding a Kaczmarz cycle into the nonlinear Landweber iteration (2.85), we obtain

$$x_\delta^{k+1} = x_\delta^k - \tau(\mathcal{G}'_{[k]}(x^k))^*(\mathcal{G}_{[k]}(x^k) - g_{[k],\delta}), \quad k \in \mathbb{N}, \tag{2.94}$$

where $[k] := (k \mod M)$. Iteration (2.94) is known as the *Landweber-Kaczmarz method (LK)* which is a convergent regularization technique as proved in [27]. The *loping-Landweber-Kaczmarz method (L-LK)* is a combination of the Landweber iteration with a variant of a cycle of the Kaczmarz method given by

$$x_\delta^{k+1} = x_\delta^k - \omega_k\tau(\mathcal{G}'_{[k]}(x^k))^*(\mathcal{G}_{[k]}(x^k) - g_{[k],\delta}), \quad k \in \mathbb{N}. \tag{2.95}$$

Here, ω_k is a bang-bang relaxation parameter defined as

$$\omega_k := \begin{cases} 1 & \text{if } \|\mathcal{G}_{[k]}(x_\delta^k - g_{[k],\delta})\| \geq \kappa\delta_{[k]}, \\ 0 & \text{otherwise,} \end{cases} \tag{2.96}$$

where $\delta_{[k]}$ is given as in (2.92), and the parameter $\kappa > 0$ satisfies (2.90), i.e, $\kappa > 2$, with $\eta < 1/2$. We mention that the Kaczmarz method has been combined with other methods such as the Steepest Descent and Levenberg-Marquardt methods [16, 47].

If it is assumed that the nonlinear problem (2.53) has a solution in a closed ball $B_{2\rho}(x^0)$, the operators $\mathcal{G}_i$ are Fréchet differentiable in $B_{2\rho}(x^0)$ such that conditions (2.87) and the local tangential cone condition (2.88) are satisfied for every operator $\mathcal{G}_i$, i.e.,

$$\|\mathcal{G}_i'(x)\|_{H_{2,i}} \leq 1, \quad x \in B_{2\rho}(x^0), \quad x \in B_{2\rho}(x^0) \subset \bigcap_{i=0}^{M-1} D_i,$$

and

$$\|\mathcal{G}_i(z) - \mathcal{G}_i(x) - \mathcal{G}_i'(x)(z-x)\|_{H_{2,i}} \leq \eta\|\mathcal{G}_i(z) - \mathcal{G}_i(x)\|_{H_{2,i}}, \quad \forall x, z \in B_\rho(x^0), \tag{2.97}$$

respectively, with $\eta < 1/2$.

The stopping criterion for this method is as follows. The iteration should be stopped when for the first time all x_δ^k are the same within a cycle, that is, when the relaxation

parameters ω_k are equal to zero within a cycle. In other words, when

$$\|\mathcal{G}_i(x_\delta^k) - g_{i,\delta}\| \leq \kappa\delta_i \quad i = 0, ..., M-1, \tag{2.98}$$

the iteration is stopped at

$$k_* := \arg\min\{lM \in \mathbb{N} : x_\delta^{lM} = x_\delta^{lM+1} = ... = x_\delta^{lM+M-1}\}, \tag{2.99}$$

or

$$k_* := \arg\min\{lM \in \mathbb{N} : \omega_{lM} = \omega_{lM+1} = ... = \omega_{lM+M-1} = 0\}.$$

Condition (2.98) is satisfied if (2.90) holds. In addition, let k_* be determined according to (2.99). Then the element $x_\delta^{k_*}$ of the L-LK iteration converges to a solution of (2.85). Moreover, if

$$\ker(\mathcal{G}'(x^\dagger)) \subset \ker(\mathcal{G}'(x)) \quad \text{for all} x \in B_\rho(x^\dagger), \quad i = 0, ..., M-1,$$

then $x_\delta^{k_*}$ converges to the minimal norm solution $x^\dagger$ when $\delta \to 0$ [86, 102].

It is noted that the L-LK method reduces to the LK method for noise-free data ($\delta = 0$), that is, $\omega_k = 1$ for each k. The L-LK method is a locally convergent regularization technique if it is assumed that (2.53) has a solution, that is if the Fréchet derivative is bounded (2.87), and the local tangential cone condition given by (2.97) holds, as it is proved in [16, 27]. Another way to speed up the convergence of Kaczmarz method or in particular the LK or L-LK methods is to iterate formulas (2.74),(2.94) or (2.95) in a random fashion. We also mention that more variants of the Kaczmarz method can be found in [64].

Chapter 3

Imaging reconstructions using the Landweber-Kaczmarz and level set methods

In this chapter, we discuss in more depth the inverse problem of Diffuse Optical Tomography (DOT), together with some solution techniques developed for this application. First, in section 3.1 the inverse problem in DOT is outlined and discussed. In particular, in section 3.3 the shape reconstruction problem in DOT is formulated in preparation for a later chapter where the level set (LS) method for shape reconstruction is applied to DOT. In section 3.2, we give a short review of the level set method using calculus of variations as in [141]. A formula and algorithm are presented in section 3.4 to obtain the so-called *smooth gradient* of the misfit data functional. In section 3.5, we briefly review a method suggested in [144] for solving the simultaneous shape and contrast value inverse problem in DOT, by using the RTE instead of its diffusive approximation. In section 3.6, we describe in detail the discretization of the forward and inverse problem based on the RTE as used in this thesis. Finally, in the last section, results of our numerical experiments aiming at single and simultaneous reconstructions of absorption and scattering coefficients in DOT are presented.

3.1 Inverse problem in DOT

We define the space U of states u, the space P of coefficients μ, b, and the space Z of data as follows

$$U = L^2\big(\Omega \times S^{n-1} \times [0,T]\big) \qquad P = L^2(\Omega) \qquad Z = L^2\big(\Omega \times [0,T]\big), \tag{3.1}$$

with the standard inner products in L^2. Note that the adjoint space of U is U itself. The natural spaces in the framework of the RTE are Banach L^1-spaces for the states u and sources q and Banach L^∞-spaces for the coefficients μ and b because of their physical significance. The following L^1-norm of the function of $u(x,\theta;t_0)$ corresponds to the total number of photons inside Ω at a given time t_0

$$\mathfrak{N}(t_0) = \int_\Omega \int_{S^{n-1}} u(x,\theta;t_0)\, d\theta dx.$$

Nevertheless, we have chosen to work with the above Hilbert spaces instead of the Banach spaces because the regularization formulas are simpler in this setting [112, 143]. Also, deriving the adjoint operator and the smooth scheme of the gradient of the cost functional defined further below is more straightforward using Hilbert spaces. We mention that for L^2-spaces in the framework of the RTE there is no physical interpretation of the L^2-norm of u. Because of this physical significance of the L^1-norm, many authors have decided to choose the states of the density of photons u in L^1, whereas for L^2 there is no physical interpretation of the L^2-norm of u. Nevertheless, some authors have chosen the space L^2 since some calculations are simplified in that setup. For further details about the RTE and transport equations see [33, 45, 50, 62, 75, 109]. For a deeper discussion of function spaces for the RTE see [45].

Detectors are positioned in DOT along the boundary $\partial\Omega$. These record measurements for each source q_i, $i = 0, \ldots, M-1$. The measurement operators $\mathcal{G}_i : P \times P \to Z$ in this technique quantify the outgoing flux across the boundary $\partial\Omega$ known as the exitance and defined by

$$\mathcal{G}_i[\mu, b](x_r, t_r) := \int_{S_+^{n-1}} u_i(x_r, \theta, t_r)\nu(x) \cdot \theta \, d\theta \quad \text{on} \quad \partial\Omega \times [0,T], \tag{3.2}$$

where x_r and t_r denote the receiver position and receiving time, respectively, and u_i is the

solution of the RTE (2.1) for each source q_i, $i = 0, \ldots, M-1$. Here S_+^{n-1} denotes the subset of direction vectors $\theta \in S^{n-1}$ for which $\nu(x) \cdot \theta > 0$. The emitters of light of DOT in general can be light emitting diodes, halogen bulbs, lasers while the detectors can be photodiodes, photomultipliers and phototransistors [106].

It can be stated the inverse problem in DOT as follows: Given the scattering function η and the recorded data (3.2) for sources q_i, $i = 0, \ldots, M-1$, find the location-dependent absorption coefficients μ and the scattering coefficients b inside the domain Ω. We mention that the reconstruction of the absorption coefficient μ is often considered more important than the reconstruction of the scattering coefficient b in medical diagnosis, since the content of the oxi-haemoglobin is linked with tumours detection [69]. However, in order to correctly determine μ we need to know b as well.

We define the following residual operators $\mathcal{R}_i : P \times P \to Z$ for given sources q_i as

$$\mathcal{R}_i[\mu, b](x_r, t_r) = \mathcal{G}_i[\mu, b](x_r, t_r) - \tilde{\mathcal{G}}_i(x_r, t_r), \tag{3.3}$$

where $\tilde{\mathcal{G}}_i$ are some given "physical true" measurements which correspond to the "true" absorption and scattering coefficients $\tilde{\mu}$ and $\tilde{b}$ respectively. Then, the residual operator is zero for the "true" coefficients $\tilde{\mu}$, $\tilde{b}$, that is, $\mathcal{R}_i[\tilde{\mu}, \tilde{b}](x_r, t_r) = 0$ for each i-th source if the uncertainty of the measurements is equal to zero. Another way to formulate the inverse problem in DOT is to determine a pair of coefficients $(\bar{\mu}, \bar{b})$ such that the residual operator $\mathcal{R}_i$ vanishes at $(\bar{\mu}, \bar{b})$ for each i, i.e.,

$$\mathcal{R}_i[\bar{\mu}, \bar{b}] = 0, \quad i = 0, \ldots, M-1. \tag{3.4}$$

Standard methods for the solution of (3.4) are based on rewriting (3.4) as a single equation

$$\mathcal{R}[\bar{\mu}, \bar{b}] = 0. \tag{3.5}$$

where the operator $\mathcal{R}$ is defined by $\mathcal{R} := 1\sqrt{M}(\mathcal{R}_0, \ldots, \mathcal{R}_{M-1})^T$, that is, the i-th row of $\mathcal{R}$ is $\mathcal{R}_i$. Typically the nonlinear problem (3.4) is written as the following least squares form

$$\mathcal{J}(\mu, b) := \frac{1}{2} \sum_{i=0}^{M-1} \|\mathcal{J}_i(\mu, b)\|_Z^2; \quad \mathcal{J}_i(\mu, b) := \mathcal{R}_i[\mu, b], \tag{3.6}$$

Then, the inverse optical problem is to find μ, b which minimise the misfit functional $\mathcal{J}$

$$\min_{(\mu,b)\in P\times P} \mathcal{J}(\mu, b). \tag{3.7}$$

Since problem (3.7) is ill-posed, it requires regularization. A Tikhonov regularization scheme is typically implemented here, but we will discuss alternative options below.

The residual operator defined in (3.3) is Fréchet differentiable and its explicit expression is stated in the next theorem which is proved in [50].

Theorem 13 *Let $\delta\mu, \delta b$ be perturbations of μ, b respectively. Assuming that the pair (μ, b) is admissible, then the mapping (3.3) is Fréchet-differentiable, and its Fréchet derivative $\mathcal{R}_i'[\mu, b] : P \times P \to Z$ at $[\mu, b]$ is given by*

$$\mathcal{R}_i'[\mu, b]\begin{pmatrix}\delta\mu\\ \delta b\end{pmatrix} = \int_{S^{n-1}} \nu(x)\cdot w_i(x_r, \theta, t_r)\, d\theta$$

where w_i solves the following transport equation

$$\frac{\partial w_i}{\partial t} + \theta\cdot\nabla w_i(x,\theta,t) + (\mu(x)+b(x))\, w_i(x,\theta,t) - b(x)\int_{S^{n-1}} \eta(\theta\cdot\theta')w_i(x,\theta,t)d\theta' \\ = Q_{\delta\mu}(x,\theta,t) + Q_{\delta b}(x,\theta,t) \quad \textit{in } X,$$

with initial condition

$$w_i(x,\theta,0) = 0 \quad \textit{in } \Omega\times S^1,$$

and boundary condition

$$w_i(x,\theta,t) = 0 \quad \textit{on } \partial X_-,$$

where the "scattering sources" are defined by

$$Q_{\delta\mu}(x,\theta,t) = -u_i(x,\theta,t)\delta\mu,$$
$$Q_{\delta b}(x,\theta,t) = -u_i(x,\theta,t)\delta b + \delta b\int_{S^{n-1}} \eta(\theta\cdot\theta')u_i(x,\theta',t)\, d\theta',$$

where u_i is the solution of (2.1) for a source q_i.

Notice that, strictly speaking, the proof of this theorem makes use of the natural Banach spaces, but that the explicit expressions for calculating derivatives (if they exist) given above are independent of this choice. The physical interpretation of the above theorem is that the perturbations of the coefficients $\delta\mu$ and δb create a virtual scattering source $Q_{\delta\mu} + Q_{\delta b}$ which gives rise to a density of virtual ("marked") photons w_i named "secondary particles" which travel to the receivers through the unperturbed medium. Then, the Fréchet derivative $\mathcal{R}'$ quantifies the exitance of w_i at detector positions x_r.

Let ξ be a vector in the data space Z. Then the *adjoint linearized residual operator* $R_i'[\mu, b]^* : Z \to P \times P$ is defined by

$$\left\langle \mathcal{R}_i'[\mu, b] \begin{pmatrix} \delta\mu \\ \delta b \end{pmatrix}, \xi \right\rangle_Z = \left\langle \begin{pmatrix} \delta\mu \\ \delta b \end{pmatrix}, \mathcal{R}_i'[\mu, b]^* \xi \right\rangle_{P\times P}. \tag{3.9}$$

An analytical expression of the adjoint of the linearized residual operator applied to the vector ξ denoted by $\mathcal{R}'[\mu, b]^*\xi$ and defined in (3.9) is given in the next theorem [9, 53].

Theorem 14 *Let $z \in U^*$ and $\xi \in Z$, the adjoint equation of the RTE is given by*

$$-\frac{\partial z}{\partial t} - \theta \cdot \nabla z + (\mu(x) + b(x))z(x, \theta, t) - b(x) \int_{S^{n-1}} \eta(\theta \cdot \theta')z(x, \theta', t)\, d\theta' = 0 \quad \textit{in } X, \tag{3.10a}$$

with initial condition

$$z(x, \theta, T) = 0 \quad \textit{on } \Omega \times S^{n-1}, \tag{3.10b}$$

and boundary condition

$$z(x, \theta, t) = \xi \quad \textit{on } \partial X_+. \tag{3.10c}$$

Then, the adjoint operator $\mathcal{R}_i'[\mu, b]^*$ is linear and given by

$$\mathcal{R}_i'[\mu, b]^*\xi = \begin{pmatrix} \mathfrak{L}_1(u_i, z_i) \\ \mathfrak{L}_2(u_i, z_i) \end{pmatrix} \tag{3.11}$$

$$\mathfrak{L}_1(u_i, z_i) := -\int_0^T \int_{S^{n-1}} u_i(x, \theta, t) z_i(x, \theta, t)\, d\theta dt, \tag{3.12a}$$

$$\mathfrak{L}_2(u_i, z_i) := \int_0^T \int_{S^{n-1}} \left(\int_{S^{n-1}} \eta(\theta \cdot \theta') u_i(x, \theta', t)\, d\theta' - u_i(x, \theta, t) \right) z_i(x, \theta, t)\, d\theta dt, \tag{3.12b}$$

where ξ is applied uniformly in all θ directions with $\theta \cdot \nu > 0$ and u_i, z_i solves eq. (2.1) and eq. (3.10) with sources q_i and ξ respectively. The term ξ is considered physically as the "adjoint" source on the boundary. It produces 'adjoint particles' that travel backwards in time into Ω according to equation (3.10), which is the reason why the adjoint equation (3.10) is often called "backtransport" equation. The expressions $\mathfrak{L}_1(u_i, z_i)$ and $\mathfrak{L}_2(u_i, z_i)$ are known as the sensitivity functions of DOT based on the RTE for a source-receiver pair corresponding to a source q_i and given receiver postion x_r and time t_r with $\xi(x, t) = \delta(x - x_r)\delta(t - t_r)$, where δ is the Dirac delta distribution [54]. In DOT, it has been shown in [54] that sensitivity functions disperse over a big region and have large values near to the source and receiver positions. For more information about sensitivity analysis in DOT, see [57].

Since the nonlinear problem (3.7) is ill-posed and highly underdetermined [8, 30, 53], regularization techniques are typically applied to solve it. Considering firstly the Landweber (L) method to solve the minimization problem (3.7), we obtain

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k \\ b^k \end{pmatrix} - \tau \nabla \mathcal{J}(\mu^k, b^k), \tag{3.13}$$

where τ is a relaxation parameter defined by

$$\tau := \begin{pmatrix} \tau_\mu & 0 \\ 0 & \tau_b \end{pmatrix}, \quad \tau_\mu, \tau_b > 0, \tag{3.14}$$

and $\nabla \mathcal{J}$ denotes the gradient of the cost functional $\mathcal{J}$ defined in (3.6), which is given by

$$\nabla \mathcal{J}(\mu^k, b^k) = \sum_{i=0}^{M-1} \nabla \mathcal{J}_i(\mu^k, b^k). \tag{3.15}$$

Furthermore, $\nabla \mathcal{J}_i$ denotes the gradient of the cost functional $\mathcal{J}_i$ defined in (3.6), which is

given by

$$\nabla \mathcal{J}_i(\mu^k, b^k) = \mathcal{R}_i'[\mu^k, b^k]^* \big(\mathcal{R}_i[\mu^k, b^k]\big), \tag{3.16}$$

where i denotes the number of the source q_i. Now, we define the i-th adjoint source as

$$\xi_i := \mathcal{R}_i[\mu, b]. \tag{3.17}$$

We have made a step further in our algorithms to approximate the adjoint source ξ_i in equation (3.17) as follows

$$\xi_i = \begin{cases} \mathcal{R}_i[\mu, b], & |x_0 - p_i| > M_F, \\ 0, & \text{otherwise}, \end{cases} \tag{3.18}$$

where $M_F \in \mathbb{R}$ and p_i is the position of the source q_i. Using this approximation, we do not take into account information near to the source p_j, which does not contain information about the inclusions deeply embedded inside the region Ω. Using the adjoint operator (3.11) and setting $\xi = \xi_i$ with ξ_i as in (3.17), $\nabla \mathcal{J}_i$ is given by

$$\nabla \mathcal{J}_i(\mu^k, b^k) = \begin{pmatrix} \mathfrak{L}_1(u_i, z_i) \\ \mathfrak{L}_2(u_i, z_i) \end{pmatrix}, \tag{3.19}$$

where u_i solves (2.1) with source q_i and z_i solves (3.10) with adjoint source $\xi = \xi_i$. Therefore, the updates for the absorption and scattering coefficients using the Landweber iteration (3.13) are given by

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k \\ b^k \end{pmatrix} - \tau \sum_{i=0}^{M-1} \begin{pmatrix} \mathfrak{L}_1(u_i, z_i) \\ \mathfrak{L}_2(u_i, z_i) \end{pmatrix}. \tag{3.20}$$

Adding a Kaczmarz cycle to the Landweber method results in the Landweber-Kaczmarz (LK) formula to solve the problem (3.7). We obtain

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k \\ b^k \end{pmatrix} - \tau \nabla \mathcal{J}_{[k]}(\mu^k, b^k), \tag{3.21}$$

with $\nabla \mathcal{J}_{[k]}(\mu^k, b^k)$ as given in (3.16) and $[k] := (k \mod M) \in \{0, \ldots, M-1\}$. Using the definition (3.17) for a source $q_{[k]}$ and the residual operator equation (3.3), the adjoint

boundary source $\xi_{[k]}$ is given by

$$\xi_{[k]} := \mathcal{R}_{[k]}[\mu^k, b^k]. \tag{3.22}$$

Then, $\nabla \mathcal{J}_{[k]}$ is

$$\nabla \mathcal{J}_{[k]}(\mu^k, b^k) = \begin{pmatrix} \mathfrak{L}_1(u_{[k]}, z_{[k]}) \\ \mathfrak{L}_2(u_{[k]}, z_{[k]}) \end{pmatrix}, \tag{3.23}$$

where $u_{[k]}$ is the solution of (2.1) for the source $q_{[k]}$ and $z_{[k]}$ is the solution of (3.10) for the adjoint source $\xi_{[k]}$ given by (3.22). Using equation (3.22) and the adjoint operator (3.11), the update formulas of the optical coefficients using the LK iteration formula (3.21) are

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k \\ b^k \end{pmatrix} - \tau \begin{pmatrix} \mathfrak{L}_1(u_{[k]}, z_{[k]}) \\ \mathfrak{L}_2(u_{[k]}, z_{[k]}) \end{pmatrix}. \tag{3.24}$$

On the other hand, if the loping-Landweber-Kaczmarz (L-LK) method is chosen to solve the problem (3.7), we obtain

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k \\ b^k \end{pmatrix} - \omega_k \tau \nabla \mathcal{J}_{[k]}(\mu^k, b^k), \tag{3.25}$$

where ω_k are the relaxation parameters introduced in (2.96), i.e.,

$$\omega_k := \begin{cases} 1 & \text{if } \|\mathcal{R}_{[k]}(\mu^k, b^k)\|_Z \geq \kappa \delta_{[k]}, \\ 0 & \text{otherwise,} \end{cases} \tag{3.26}$$

where $\nabla \mathcal{J}_{[k]}$ as defined in (3.16) with $[k] := (k \mod M)$. Here, δ_i denotes the noise level defined by

$$\|\mathcal{G}_i^\delta - \tilde{\mathcal{G}}_i\|_Z \leq \delta_i \quad i = 0, \ldots, M-1, \tag{3.27}$$

where $\tilde{\mathcal{G}}_i$ are the exact data and $\mathcal{G}_i^\delta$ the noisy data, and $\kappa > 0$. Finally, using equation (3.22) and the adjoint operator (3.11), the update formulas for the optical coefficients using the L-LK iteration formula (3.25) are given by

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k \\ b^k \end{pmatrix} - \omega_k \tau \begin{pmatrix} \mathfrak{L}_1(u_{[k]}, z_{[k]}) \\ \mathfrak{L}_2(u_{[k]}, z_{[k]}) \end{pmatrix}, \tag{3.28}$$

where $u_{[k]}$ is the solution of (2.1) with the source $q_{[k]}$ and $z_{[k]}$ is the solution of (3.10) with the adjoint source $\xi_{[k]}$.

We have reviewed the convergence of the LK and L-LK methods above in subsection 2.3.2. Now we formulate the conditions which we assume to hold in order to achieve the convergence of the LK and L-LK methods applied to the DOT problem.

Let $\sigma = [\mu, b]$, then the LK method is locally convergent if the following assumptions are satisfied. We will assume that the nonlinear problem (3.4) has a solution in a closed ball $B_{2\rho}(\sigma^0)$ with radius 2ρ and centred at σ^0, that the operators $\mathcal{R}_i$ are Fréchet differentiable in $B_{2\rho}(\sigma^0)$, and that the following property holds

$$\|\mathcal{R}_i'(\sigma)\|_Z \leq 1, \quad \sigma \in B_{2\rho}(\sigma^0) \subset \bigcap_{i=0}^{M-1} D_i, \tag{3.29}$$

and property (2.87) is satisfied using a relaxation parameter τ in the LK iteration (3.13). In addition, we assume that the following "local tangential cone condition" holds

$$\|\mathcal{R}_i(\varsigma) - \mathcal{R}_i(\vartheta) - \mathcal{R}_i'(\sigma)(\varsigma - \sigma)\|_Z \leq \lambda \|\mathcal{R}_i(\varsigma) - \mathcal{R}_i(\vartheta)\|_Z, \quad \forall \vartheta, \varsigma \in B_\rho(\sigma^0), \tag{3.30}$$

where $\lambda < 1/2$.

As stopping criterion for the LK method we apply the following rule. The iteration should be stopped when for the first time, all σ_k^δ are the same within a cycle, that is, when the relaxation parameters ω_k are equal to zero within a cycle, that is, if

$$\|\mathcal{G}_i[\sigma] - \mathcal{G}_i^\delta\| \leq \kappa \delta_i, \quad i = 0, \ldots, M-1, \tag{3.31}$$

the iteration is stopped at

$$k_* := \arg\min\left\{ lM \in \mathbb{N} : \sigma_\delta^{lM} = \sigma_\delta^{lM+1} = \cdots = \sigma_\delta^{lM+M-1} \right\}, \tag{3.32}$$

that is, k_* is the smallest multiple of M such that

$$\omega_{k_*} = \omega_{k_*+1} = \cdots = \omega_{k_*+M-1} = 0.$$

Condition (2.98) is satisfied if

$$\kappa > 2\frac{1+\lambda}{1-2\lambda} > 2. \tag{3.33}$$

In addition, let k_* be determined according to (2.99). Then the element $\sigma_{k_*}^\delta$ of the L-LK

iteration converges to a solution of (3.13). Moreover, if

$$\ker(\mathcal{R}'(\sigma^\dagger)) \subset \ker(\mathcal{R}'(\sigma)) \quad \forall \quad \sigma \in B_\rho(\sigma^\dagger), \quad i = 0, \dots, M-1,$$

then $\sigma^\delta_{k_*}$ converges to the minimal norm solution $\sigma^\dagger$ when $\delta \to 0$ [86, 102].

First, we describe the *Landweber iteration scheme* in form of a Pseudo-Code below.

Algorithm 1: Landweber

Input: σ^0

Output: σ^I

for $k := 1, \dots, I$ **do**

  for $i := 0, \dots, M-1$ **do**

    Compute the gradient $\nabla \mathcal{J}_i(\sigma^k)$ given by (3.19) ;

  end

  Update inhomogeneity by σ^{k+1} using (3.13) ;

end

Here, u_i solves (2.1) with σ^k and with source q_i, and z_i solves (3.10) with σ^k and with adjoint source ξ_i given by (3.17).

Next, we describe the *LK reconstruction scheme* for the reconstruction of absorption and scattering coefficients μ, b in form of a Pseudo-code.

Algorithm 2: LK

Input: σ^0

Output: $\sigma^{I(M-1)}$

for $k := 0, \dots, I(M-1)$ **do**

  Compute the gradient $\nabla \mathcal{J}_{[k]}(\sigma^k)$ given by (3.23);

  Update σ^{k+1} using (3.21) ;

end

Here $u_{[k]}$ solves (2.1) with σ^k and source $q_{[i]}$, and $z_{[k]}$ solves (3.10) with σ^k and adjoint

source $\xi_{[k]}$ given by (3.17).

This form of the LK algorithm is used in our numerical experiments, for which the results are shown further below. In order to obtain updates for the absorption and scattering coefficients μ^k and b^k, we have to solve a forward problem given by the RTE (2.1) and a backward equation (3.10) with the corresponding sources $q_{[k]}$ and $\xi_{[k]}$ respectively. This is done in a cyclic way until the algorithm terminates in a similar way as in the Kaczmarz method.

Finally, the *l-LK algorithm* reads in form of a Pseudo-Code.

Algorithm 3: l-LK

Input: σ^0

Output: $\sigma^{I(M-1)}$

for $k := 0, \ldots, I(M-1)$ **do**

Compute the gradient $\nabla \mathcal{J}_{[k]}(\sigma^k)$ given by (3.23) ;

Update σ^{k+1} using (3.25) ;

end

3.2 Level set method: calculus of variations approach

Level set methods, also known as the method of geodesic active contours or interfaces, are suitable for modelling the propagation of contours as first shown in 1988 in the paper of Osher and Sethian [128]. These and related techniques have been applied successfully in many areas such as: shape recovery, grid generation, geometry, fluid mechanics, combustion, solidification, device fabrication or shape design, morphing, object tracking/image sequence analysis in images, stereo vision, shape from shading, mathematical morphology, free boundary value problems, colour image segmentation, medical image segmentation, and others. For an overview see for example [151]. The success of the level set methods is mainly due to their ability to capture topological changes in an automatic fashion and due to their flexibility during the propagation of the contours using implicit representations. In our situation of DOT, the level set method provides an efficient and accurate mean for the recovery of the geometry of hidden obstacles.

It is assumed that the coefficient distribution for the sought physical quantity is described by

$$\sigma(x) = \begin{cases} \sigma_i & x \in \Upsilon, \\ \sigma_e & x \in \Omega \backslash \Upsilon, \end{cases} \tag{3.34}$$

where $\Upsilon \subset \Omega$ is a unknown region which might have complicated geometry. Notice that usually discontinuities occur at the interfaces $\partial\Upsilon$. The functions σ_i are arbitrary functions which depend only on the spatial variable x. For simplicity we will assume that the function value σ_i inside the region Υ is a constant known values, and σ_e correspond to background values. We assume that the region Υ can be written as

$$\Upsilon = \bigcup_{j=1}^{N} \Upsilon_j, \quad \Upsilon_j \cap \Upsilon_{j'} = \emptyset \quad \text{for} \quad j \neq j',$$

where the number of components N are finite and not known a-priori, so their topologies are quite arbitrary. Topological changes of the components of Υ need to be handled during the reconstruction process since we do not know a-priori the true number of components, that is, those components may merge or split, such that the number of components N of Υ will change during the shape evolution. In such situations, parameterized models for the shape Υ are cumbersome since re-parameterizations are needed in order to correctly track these topological changes. Level set methods offer a suitable alternative to handle this kind of scenario since it is not necessary to re-parameterize the shapes Υ.

Given a piecewise constant or Lipschitz-continuous function $\varphi : \Upsilon \to \mathbb{R}$, we call φ a *level set representation of the region* Υ if

$$\begin{cases} \varphi(x) \leq 0 & x \in \Upsilon, \\ \varphi(x) > 0 & x \in \Omega \backslash \Upsilon, \end{cases} \tag{3.35}$$

Then, the distributions of absorption and scattering coefficients are represented by the level set function φ, according to

$$\sigma(x) = \begin{cases} \sigma_i & \varphi(x) \leq 0, \\ \sigma_e & \varphi(x) > 0, \end{cases} \tag{3.36}$$

Let $B_\epsilon(x_0)$ be the ball of radius ϵ with centre at x_0;

$$B_\epsilon(x_0) = \{x \in \Omega : |x - x_0| < \epsilon\}.$$

The boundary of Υ is denoted by $\partial\Upsilon$, and is described by

$$\partial\Upsilon = \{x \in \Upsilon : \quad \forall\epsilon > 0 \quad \exists\, x_1, x_2 \in B_\epsilon(x), \quad \varphi(x_1) > 0 \text{ and } \varphi(x_2) < 0\}\,,$$

Assuming that $\nabla\varphi$ is well defined and positive along the boundary $\partial\Upsilon$, the boundary $\partial\Upsilon$ can also be represented as

$$\partial\Upsilon = \{x \in \Omega, \quad \varphi(x) = 0\}.$$

Assuming that the boundary at point x moves outwards by δx, this movement will change the points $x \in \partial\Upsilon$ to the new positions $x' = x + \delta x$, that is, the boundary $\partial\Upsilon$ will be deformed into the new boundary $\partial\Upsilon'$, see fig. 3.1.

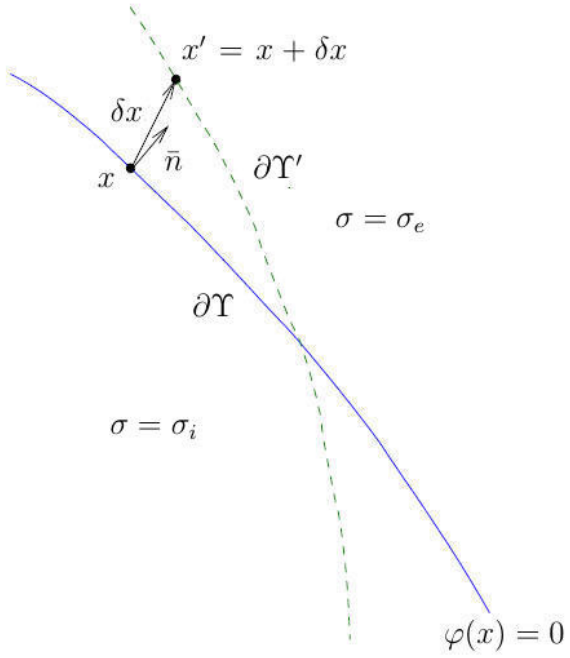


Figure 3.1: Geometry of the region Ω.

We can see in fig. 3.1 that the outward normal at x is

$$\bar{n}(x) = \frac{\nabla\phi(x)}{|\nabla\phi(x)|}.$$

If one takes the variation of the equation $\phi(x) = 0$, one obtains

$$\delta\phi + \nabla\phi \cdot \delta x = 0. \tag{3.37}$$

We will compute the variation of the coefficient distribution $\sigma(x)$ due to an infinitesimal deformation in the following. Consider the inner product of $\delta\sigma$ with a test function ς

$$\langle \delta\sigma, \varsigma \rangle_{\Upsilon} = \int_{\Upsilon} \delta\sigma(x)\varsigma(x)\, dx = \int_{\Upsilon\cap\Upsilon'} \delta\sigma(x)\varsigma(x)\, dx, \tag{3.38}$$

Since Υ is infinitesimal, we can reduce the area integral to the following line integral

$$\langle \delta\sigma, \varsigma \rangle_{\Upsilon} = \int_{\delta\Upsilon} (\sigma_e - \sigma_i)\delta x \cdot \bar{n}\varsigma(x)\, ds(x), \tag{3.39}$$

From eq. (3.38) and (3.2), we get that the distribution $\delta\sigma(x)$ defined on the entire domain and concentrated on $\delta\Upsilon$ is given by

$$\delta\sigma(x) = (\sigma_e - \sigma_i)\delta x \cdot \bar{n}\hat{\delta}_{\partial\Upsilon}(x), \tag{3.40}$$

where $\hat{\delta}_{\partial\Upsilon}(x)$ is the Dirac delta distribution concentrated on the boundary $\partial\Upsilon$. Then, $\delta\sigma(x)$ is interpreted as a surface measure on $\delta\Upsilon$. From eq.(3.37) $\delta\phi = -\nabla\phi \cdot \delta x$, then eq. (3.40) can be written as

$$\delta\sigma(x) = (\sigma_i - \sigma_e)\frac{\delta\phi(x)}{|\nabla\phi(x)|}\hat{\delta}_{\partial\Upsilon}(x). \tag{3.41}$$

We will use the analytical expression of the distribution of the absorption coefficient $\delta\mu$ given in (3.41) resulted from the Santosa's level set approach the next section, and analogous expression for the distribution of the scattering coefficient δb.

3.3 Shape reconstruction problem in DOT

A combination of the level set technique and the L, LK and L-LK methods will be described in this section. The shape reconstruction problem in DOT can be formulated as follows. Also here it is more convenient to work with L^2-spaces for the density of photons u and its adjoint density z, and for the absorption and scattering coefficients μ, b rather

than L^1-spaces for the same reasons as outlined in the last section. U, P, Z will denote the spaces of states u, of coefficients μ, b, and of the data, respectively, as defined in the last section with the same standard inner products in L^2. We add the space of level set functions which is $\Phi := L^2(\Omega)$.

The goal of the shape inverse problem is to find the spatial distribution Υ of some coefficients inside a region Ω. In the shape reconstruction approach, we assume that the distributions of the absorption and scattering coefficients are described by (similarly as in the last section 3.2)

$$\mu(x) = \begin{cases} \mu_i & x \in \Upsilon_1, \\ \mu_e & x \in \Omega \backslash \Upsilon_1, \end{cases} \qquad b(x) = \begin{cases} b_i & x \in \Upsilon_2, \\ b_e & x \in \Omega \backslash \Upsilon_2, \end{cases} \tag{3.42}$$

where $\Upsilon_1, \Upsilon_2 \subset \Omega$ are unknown regions which might have complicated geometry. Notice that usually discontinuities occur at the interfaces $\partial\Upsilon_1, \partial\Upsilon_2$. The functions μ_i are arbitrary functions which depend only on the spatial variable x. For simplicity we will assume that the function values μ_i inside the regions Υ_1, Υ_2 are constant known values, and μ_e correspond to background values. We assume that the regions Υ_1, Υ_2 can be written as

$$\Upsilon_1 = \bigcup_{j=1}^{N_1} \Upsilon_{1,j}, \quad \Upsilon_{1,j} \cap \Upsilon_{1,j'} = \emptyset \quad \text{for} \quad j \neq j',$$
$$\Upsilon_2 = \bigcup_{j=1}^{N_2} \Upsilon_{2,j}, \quad \Upsilon_{2,j} \cap \Upsilon_{2,j'} = \emptyset \quad \text{for} \quad j \neq j',$$

where the number of components N_1, N_2 are finite and not known a-priori, so their topologies are quite arbitrary. Topological changes of the components of Υ_1, Υ_2 need to be handled during the reconstruction process since we do not know a-priori the true number of components, that is, those components may merge or split, such that the number of components N_1, N_2 of Υ_1, Υ_2 will change during the shape evolution. In such situations, parameterized models for the shape Υ are cumbersome since re-parameterizations are needed in order to correctly track these topological changes. Level set methods offer a suitable alternative to handle this kind of scenario since it is not necessary to re-parameterize the shapes Υ_1, Υ_2. The distributions of absorption and scattering coeffi-

cients are represented by the level set functions ϕ and ψ, respectively, according to

$$\mu(x) = \begin{cases} \mu_i & \phi(x) \leq 0, \\ \mu_e & \phi(x) > 0, \end{cases} \qquad b(x) = \begin{cases} b_i & \psi(x) \leq 0, \\ b_e & \psi(x) > 0. \end{cases} \tag{3.43}$$

Here the indices i and e stand for the interior and the exterior of domains of interest Υ_1 and Υ_2, respectively. We define the characteristic function $\chi_D : \Omega \to \{0, 1\}$ as usual for a given shape $D \subset \Omega$ as

$$\chi_D(x) = \begin{cases} 1, & x \in D, \\ 0, & x \in \Omega \backslash D. \end{cases}$$

We will assume that $\partial D \subset \Omega$ for simplicity. The level set representation of a shape is defined as follows.

Definition 15 *Let be $\Upsilon \subset \Omega$ a domain. A function $\phi : \Upsilon_1 \to \mathbb{R}, \phi \in \Phi$ is called a level set representation of Υ_1 if it holds*

$$\varphi_\phi(x) = \chi_{\Upsilon_1}(x),$$

where $\varphi : \Omega \to \{0, 1\}$ is defined as

$$\varphi_\phi(x) = \begin{cases} 1, & \phi(x) \leq 0, \\ 0, & \phi(x) > 0. \end{cases}$$

In the same way, a function $\psi : \Upsilon_2 \to \mathbb{R}, \psi \in \Phi$ is called a level set representation of Υ_2 if

$$\varphi_\psi(x) = \chi_{\Upsilon_2}(x).$$

Each of the level set functions ϕ and ψ determine a unique regions which we denote by $\Upsilon_1[\phi]$ and $\Upsilon_2[\psi]$ respectively. Let $B_{\epsilon_1}(x_0)$ be the ball of radius ϵ_1 with centre at x_0;

$$B_{\epsilon_1}(x_0) = \{x \in \Omega : |x - x_0| < \epsilon\}.$$

The boundaries of Υ_1 and Υ_2 are denoted by $\partial\Upsilon_1$ and $\partial\Upsilon_2$, respectively, and are described

by

$$\partial\Upsilon_1 = \{x \in \Upsilon_1 : \quad \forall \epsilon_1 > 0 \quad \exists\, x_1, x_2 \in B_{\epsilon_1}(x), \quad \phi(x_1) > 0 \text{ and } \phi(x_2) < 0\}\,,$$
$$\partial\Upsilon_2 = \{x \in \Upsilon_2 : \quad \forall \epsilon_2 > 0 \quad \exists\, x_1, x_2 \in B_{\epsilon_2}(x), \quad \psi(x_1) > 0 \text{ and } \psi(x_2) < 0\}\,,$$

Assuming that $\nabla\phi$ and $\nabla\psi$ are well defined and positive along the boundaries $\partial\Upsilon_1$ and $\partial\Upsilon_2$ respectively, the boundaries $\partial\Upsilon_1$ and $\partial\Upsilon_2$ can also be represented as

$$\partial\Upsilon_1 = \{x \in \Omega, \quad \phi(x) = 0\}, \qquad \partial\Upsilon_2 = \{x \in \Omega, \quad \psi(x) = 0\}.$$

We define now the following operators Λ_1, Λ_2 which map level set functions to coefficient distributions.

$$\Lambda_1(\phi)(x) = \begin{cases} \mu_i, & \phi(x) \leq 0, \\ \mu_e, & \phi(x) > 0, \end{cases} \qquad \Lambda_2(\psi)(x) = \begin{cases} b_i, & \psi(x) \leq 0, \\ b_e, & \psi(x) > 0. \end{cases} \tag{3.44}$$

Using the above definition, we obtain

$$\mu(x) = \Lambda_1(\phi)(x) = \mu_i \phi_\phi(x),$$
$$b(x) = \Lambda_2(\psi)(x) = b_i \phi_\psi(x).$$

Using the residual operators defined in (3.3), the forward operator $T_i : \Phi \to Z$ is defined for each source q_i as the composition of Λ and $\mathcal{R}_i$,

$$T_i(\phi, \psi) = \mathcal{R}_i(\Lambda_1(\phi), \Lambda_2(\psi)), \quad i = 0, \ldots, M-1. \tag{3.45}$$

The aim of the shape reconstruction problem is to find a pair of level set functions $\tilde{\phi}, \tilde{\psi}$ such that

$$T_i(\tilde{\phi}, \tilde{\psi}) = 0, \qquad i = 0, \ldots, M-1. \tag{3.46}$$

Typically, the above problem is rewritten in the following least-squares form

$$\mathcal{J}(\phi) := \frac{1}{2} \sum_{i=0}^{M-1} \|\mathcal{J}_i(\phi, \psi)\|_Z^2; \quad \mathcal{J}_i(\phi, \psi) := T_i(\phi, \psi),$$

where $\mathcal{J}$ is the misfit of the data functional. Applying the LK formula (2.94) to the shape problem in DOT (3.46), we obtain

$$\begin{pmatrix} \phi^{k+1} \\ \psi^{k+1} \end{pmatrix} = \begin{pmatrix} \phi^{k} \\ \psi^{k} \end{pmatrix} - \tau \nabla \mathcal{J}_{[k]}(\phi^k, \psi^k). \tag{3.47}$$

In (3.47), τ is a relaxation parameter defined by

$$\tau := \begin{pmatrix} \tau_\phi & 0 \\ 0 & \tau_\psi \end{pmatrix}, \quad \tau_\phi, \tau_\psi > 0, \tag{3.48}$$

and $\nabla \mathcal{J}_i(\phi^k, \psi^k)$ is given by

$$\nabla \mathcal{J}_i(\phi^k, \psi^k) = T'_{[k]}(\phi^k, \psi^k)^*(T_{[k]}(\phi^k, \psi^k)), \tag{3.49}$$

where i denotes the number of the source q_i.

We will need an expression for the adjoint of the linearised forward operator T'_i in order to calculate (ϕ^{k+1}, ψ^{k+1}) given in (3.47). We mention that, in the following, we have used the Level Set approach as suggested by Santosa [141] combined with the analytical expressions for the inverse problem in DOT as provided in section 3.2.

To start with, the linearised forward operator $T'_i[\phi, \psi] : \Phi \to Z$, for an infinitesimal change $(\delta\phi, \delta\psi)$, is given by

$$T'_i[\phi, \psi] \begin{pmatrix} \delta\phi \\ \delta\psi \end{pmatrix} = \mathcal{R}'_i[\Lambda_1(\phi), \Lambda_2(\psi)] \begin{pmatrix} \Lambda'_1[\phi]\delta\phi \\ \Lambda'_2[\psi]\delta\psi \end{pmatrix}, \tag{3.50}$$

where

$$\begin{pmatrix} \Lambda'_1[\phi]\delta\phi(x) \\ \Lambda'_2[\psi]\delta\psi(x) \end{pmatrix} = -\begin{pmatrix} [\mu_i - \mu_e]\delta\phi|\nabla\phi|^{-1}\hat{\delta}_{\partial\Upsilon_1}(x) \\ [b_i - b_e]\delta\psi|\nabla\psi|^{-1}\hat{\delta}_{\partial\Upsilon_2}(x) \end{pmatrix}.$$

Here, $\hat{\delta}_{\partial\Upsilon_1}(x)$ and $\hat{\delta}_{\partial\Upsilon_2}(x)$ are the Dirac delta distribution on $\partial\Upsilon_1$ and $\partial\Upsilon_2$ respectively [52]. These Dirac delta distributions are approximated using "narrowbands" or "narrow-tubes" denoted by $\chi_{\phi,d_1}(x), \chi_{\psi,d_2}(x)$ as follows

$$\hat{\delta}_{\partial\Upsilon_1}(x) \approx \chi_{\phi,d_1}(x), \qquad \hat{\delta}_{\partial\Upsilon_2}(x) \approx \chi_{\psi,d_2}(x),$$

where $\chi_{\phi,d_1}(x)$ is defined as [96]

$$\chi_{\phi,d_1}(x) = \begin{cases} 1 & \forall x \in \Omega \quad \text{s.t.} \quad |x_0 - x| < d_1 \quad \text{and} \quad \phi(x_0) = 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.51}$$

In the latter expression, $d_1 \in \mathbb{R}$ indicates the width of the narrowband. The narrowband $\chi_{\psi,d_2}(x)$ is defined analogously. A different choice for approximating the Dirac delta functions $\hat{\delta}_{\partial\Upsilon_1}(x), \hat{\delta}_{\partial\Upsilon_2}(x)$ is presented in [40, 129], or the narrowtube proposed in [130]. Let now $\xi \in Z$ be the adjoint operator of T' given by

$$T_i'[\phi,\psi]^*\xi = (\Lambda_1'[\phi], \Lambda_2'[\psi])^*\mathcal{R}_i'[\Lambda_1(\phi), \Lambda_2(\psi)]^*\xi,$$

where $\Lambda_1'[\phi]^* : P \to \Phi$, $\Lambda_2'[\psi]^* : P \to \Phi$ are the adjoint operators of $\Lambda_1'[\phi]$ and $\Lambda_2'[\psi]$. These are given by

$$\begin{pmatrix} (\Lambda_1'[\phi]^*\delta\phi)(x) \\ (\Lambda_2'[\psi]^*\delta\psi)(x) \end{pmatrix} = -\begin{pmatrix} [\mu_i - \mu_e]|\nabla\phi|^{-1}\delta\phi\hat{\delta}_\Gamma(x) \\ [b_i - b_e]|\nabla\psi|^{-1}\delta\psi\hat{\delta}_\Gamma(x) \end{pmatrix}.$$

Using the approximation of the Dirac delta distribution $\hat{\delta}_\Gamma(x)$ (3.51) to the above equation, we obtain

$$\begin{pmatrix} (\Lambda_1'[\phi]^*\delta\phi)(x) \\ (\Lambda_2'[\psi]^*\delta\psi)(x) \end{pmatrix} = -\begin{pmatrix} [\mu_i - \mu_e]|\nabla\phi|^{-1}\chi_{\phi,d}(x) \\ [b_i - b_e]|\nabla\psi|^{-1}\chi_{\psi,d}(x) \end{pmatrix}.$$

Now, we define the adjoint source as

$$\xi_{[k]} := T_{[k]}(\phi^k, \psi^k). \tag{3.52}$$

We have made a similar approximation to the adjoint source ξ in equation (3.17)

$$\xi_i = T_i(\phi^k, \psi^k) = \begin{cases} T_i(\phi^k, \psi^k), & |x - p_i| > M_F, \\ 0, & \text{otherwise}, \end{cases} \tag{3.53}$$

where $M_F \in \mathbb{R}$. Using equation (3.52) and setting $\xi = \xi_i$, and the adjoint operator equation (3.11), we obtain

$$\nabla\mathcal{J}_{[k]}(\phi^k, \psi^k) = \begin{pmatrix} (\mu_i - \mu_e)|\nabla\phi^k|^{-1}\chi_{\phi,d}(x)\mathfrak{L}_1(u_{[k]}, z_{[k]}) \\ (b_i - b_e)|\nabla\psi^k|^{-1}\chi_{\psi,d}(x)\mathfrak{L}_2(u_{[k]}, z_{[k]}) \end{pmatrix}, \tag{3.54}$$

where $u_{[k]}$ solves (2.1) with source $q_{[k]}$ and $z_{[k]}$ solves (3.10) with adjoint source $\xi = \xi_{[k]}$ defined in (3.52). Since the values $|\nabla\phi(x)|$ and $|\nabla\psi(x)|$ may be small in (3.54), the level set functions ϕ and ψ are scaled by factors such that their extreme values stay fixed in order to gain stability in our formula (3.54). Doing this obviously does not change the domains Υ_1 and Υ_2 represented by the level set function ϕ and ψ, respectively. We go one step further and assume that the norms $|\nabla\phi(x)|$ and $|\nabla\psi(x)|$ can be approximated along the boundary by a constant ρ_1 and ρ_2 respectively. Doing this improves stability in (3.54), even though the updates are not marching in the 'optimal' direction. With these modifications, equation (3.49) is now written as

$$\nabla\mathcal{J}_{[k]}(\phi^k,\psi^k) = \begin{pmatrix} (\mu_i-\mu_e)\rho_1^{-1}\chi_{\phi,d}(x)\mathfrak{L}_1(u_{[k]},z_{[k]}) \\ (b_i-b_e)\rho_2^{-1}\chi_{\psi,d}(x)\mathfrak{L}_2(u_{[k]},z_{[k]}) \end{pmatrix}, \tag{3.55}$$

We have scaled the level set functions at each iteration using scaling factors $c_\phi^k, c_\psi^k \in \mathbb{R}$ to keep the extreme values of the level set functions bounded, thus equation (3.47) becomes now

$$\begin{pmatrix} \phi^{k+1} \\ \psi^{k+1} \end{pmatrix} = C\left(\begin{pmatrix} \phi^k \\ \psi^k \end{pmatrix} - \nabla\mathcal{J}_{[k]}(\phi^k,\psi^k)\right), \tag{3.56}$$

where $C := \begin{pmatrix} c_\phi^k & 0 \\ 0 & c_\psi^k \end{pmatrix}$. As already mentioned, by doing this rescaling procedure the domains $\Upsilon_{1,k}, \Upsilon_{2,k}$ remain unchanged. Finally, the absorption and scattering coefficient distributions are obtained using (3.44)

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \Lambda_1(\phi^{k+1}) \\ \Lambda_2(\psi^{k+1}) \end{pmatrix}. \tag{3.57}$$

Instead of using arbitrary initial guesses for $(\Upsilon_{1,0},\phi^0)$ and $(\Upsilon_{2,0},\psi^0)$ in the iteration formula of the combination of the LK and LS methods in algorithm 4, the LK method (explained in the subsection 2.3.2) can be used during the first few (here 9) sweeps of the routine to provide such an initial guess which helps accelerate the reconstruction process as proposed in the two-step method [55]. Therefore, instead of using the LK-LS method during the whole reconstruction iteration, we apply the LK method in the first stage of the algorithm 4. The LK method yields satisfactory results for the positions of the obstacles during the early sweeps, but neither satisfactory shapes nor good approximations for their values. Therefore, we implemented the LK-LS method for the rest of the iteration process to obtain better approximations for the shapes of the obstacles. The incorporation of this

step in algorithm 4 accelerates the convergence of the reconstruction process by avoiding long-range shape evolutions. In the simulations presented further below, the convergence is more than ten times faster compared to using arbitrary initial guesses (certainly, this factor depends on the choice of these alternative guesses) as shown in figure 3.10.

We will now describe the reconstruction scheme which uses in the first stage the algorithm 4 for obtaining the initial regions together with their corresponding level set functions $(\Upsilon_{1,0}, \phi^0)$ and $(\Upsilon_{2,0}, \psi^0)$. These will be used as the initial guess of the second stage of this algorithm. Let γ_{LS} be a threshold value $0 < \gamma_{ls} < 1$ (in our numerical experiment we have used $\gamma_{LS}) = 0.9$. We define the parameter

$$\mu^{LS} := \gamma_{LS} \max_{x\in\Omega} \mu^{LK}(x), \qquad b^{LS} := \gamma_{LS} \max_{x\in\Omega} b^{LK}(x),$$

where $\mu^{LK}(x)$ and $b^{LK}(x)$ are given by the LK method. For the level sets zero L^0_ϕ, L^0_ψ of the level set functions ϕ^0, ψ^0 we require

$$\mathfrak{H}^0_\phi = \left\{x \in \Omega |\ \ \mu^{LK}(x) = \mu^{LS}\right\}, \qquad \mathfrak{H}^0_\psi = \left\{x \in \Omega |\ \ b^{LK}(x) = b^{LS}\right\}.$$

The level set functions are now defined as

$$\phi^0(x) = \mu^{LS} - \mu^{LK}(x), \qquad \psi^0(x) = b^{LS} - b^{LK}(x).$$

Hence, using (3.44), the domains μ_0 and b_0 are given by

$$\begin{pmatrix} \mu^0 \\ b^0 \end{pmatrix} = \begin{pmatrix} \Lambda_1(\phi^0) \\ \Lambda_2(\psi^0) \end{pmatrix}.$$

A Pseudo-code for this combination of the LK and a LS methods for shape reconstruction is described in the algorithm 4.

Algorithm 4: LK and LS methods

Input: $(\Upsilon_{1,0}, \mu^0, \phi^0)$, $(\Upsilon_{2,0}, b^0, \psi^0)$ Initial Guess provided by algorithm 2 (LK)

Output: $\Upsilon_{1,I(M-1)} = \Upsilon_1[\phi^{I(M-1)}]$, $\Upsilon_{2,I(M-1)} = \Upsilon_2[\psi^{I(M-1)}]$

for $k := 0, \ldots, I(M-1)$ **do**

Compute the gradient $\nabla \mathcal{J}_{[k]}(\phi^k, \psi^k)$ given by (3.55) ;

Update (ϕ^{k+1}, ψ^{k+1}) using (3.56); % update of the level set functions ϕ, ψ;

Update (μ^{k+1}, b^{k+1}) using (3.57); % update of the coefficient μ and b;

end

In the above algorithm, $u_{[k]}$ solves (2.1) with μ^k, b^k and source $q_{[k]}$, and $z_{[k]}$ solves (3.10) with μ^k, b^k and source $\xi_{[k]}$ given by (3.52).

3.4 Smooth gradient

In the case of noisy data, we will apply an additional regularizer to gradient of the misfit functional $\mathcal{J}_i$. Doing this the boundaries of the obstacles will be smoothed but stability will be gained and artefacts will tend to disappear. This smoothing application will also be used in Sparsity reconstructions obtained in chapter 5 as suggested in [41, 79, 96, 98] for obtaining more accurate results.

We will derive a formula for calculating the smooth gradient of $\mathcal{J}_i$ which will be used in our main algorithm of this chapter as well as the algorithms presented in chapter 5. Based on this formula, we will present an algorithm to compute the smooth gradient as presented in [84]. The distribution of the coefficient $\sigma = (\mu, b)$ will be in $H^1(\Omega)$ instead of $L^2(\Omega)$, where H^1 stands for the Sobolev space defined by

$$H^1(\Omega) = \{\sigma \in L^2(\Omega), \partial_x \sigma, \partial_y \sigma \in L^2(\Omega)\}.$$

Here, $\partial_x \sigma$, $\partial_y \sigma$ are weak partial derivatives. Therefore, we restrict the absorption and scattering coefficients to a subset of $L^2(\Omega)$. Instead of using the standard inner product of $H^1(\Omega)$, i.e.,

$$\langle \sigma_1, \sigma_2 \rangle_{H^1} = \langle \sigma_1, \sigma_2 \rangle_{L^2} + \langle \nabla \sigma_1, \nabla \sigma_2 \rangle_{L^2},$$

with its induced norm

$$\|\sigma\|_{H^1}^2 = \|\sigma\|_{L^2}^2 + \|\nabla\sigma\|_{L^2}^2.$$

we use the inner product

$$\langle\sigma_1, \sigma_2\rangle_{H^1} = \langle\sigma_1, \sigma_2\rangle_{L^2} + \gamma\,\langle\nabla\sigma_1, \nabla\sigma_2\rangle_{L^2}\,. \tag{3.58}$$

The norm of the space $H^1(\Omega)$ is the one induced by its inner product (3.58). If we denote the space $H^1(\Omega)$ by $\hat{P}$, and letting $\xi \in Z$ and $\delta\sigma$ the perturbation of σ, we have

$$\langle\mathcal{R}_i'[\sigma]\delta\sigma, \xi\rangle_Z = \langle\delta\sigma, \mathcal{R}_i'[\sigma]^*\xi\rangle_P = \langle\delta\sigma, \mathcal{R}_i'[\sigma]^s\xi\rangle_{\hat{P}}\,, \tag{3.59}$$

where P and Z are the spaces of coefficients and data defined in (3.1), respectively, and $\mathcal{R}_i'[\sigma]^s\xi$ denotes the image of ξ under application of the adjoint residual operator with respect to the recently defined weighted inner product (3.58), mapping into the smaller space $\hat{P}$ than P. Then, the smooth gradient $\nabla\mathcal{J}_i^s$ is defined by

$$\langle\delta\sigma, \nabla\mathcal{J}_i^s(\sigma)\rangle_{\hat{P}} = \langle\delta\sigma, \nabla\mathcal{J}_i(\sigma)\rangle_P\,. \tag{3.60}$$

Integrating by parts on the left-hand side of (3.60), we have

$$(I - \gamma\Delta)\nabla\mathcal{J}_i^s(\sigma) = \nabla\mathcal{J}_i(\sigma), \tag{3.61a}$$

$$\nabla\mathcal{J}_i^s(\sigma)\cdot\nu = 0, \tag{3.61b}$$

where I and Δ are the Identity and the Laplacian operator, respectively, in (3.61a), and (3.61b) is the homogeneous Neumann boundary condition on $\partial\Omega$. An approximative scheme for solving the differential equation given by (3.61) is derived next.

Let $\varpi = \nabla\mathcal{J}_i(\sigma)$ and let $\mathfrak{F}$ be the functional

$$\mathfrak{F}(\vartheta) = \frac{1}{2}\left(\gamma\|\nabla\vartheta\|_{L^2}^2 + \|\vartheta - \varpi\|_{L^2}^2\right). \tag{3.62}$$

The first term of (3.62) penalises the roughness of ϑ, while the second term penalises the misfit between the gradient direction $\nabla\mathcal{J}_i$ and the "smooth" proposed term ϑ. Calculating

the gradient direction of $\mathfrak{F}$ results in

$$\nabla\mathfrak{F} = (I - \gamma\Delta)\vartheta - \varpi,$$

with the Neumann boundary condition $\nabla\vartheta = 0$ on $\partial\Omega$. If ϑ is a minimum of (3.62), then $\nabla\mathfrak{F} = 0$, that is

$$(I - \gamma\Delta)\vartheta = \varpi, \tag{3.63}$$

which is equivalent to (3.61a). Applying a descent direction method in [119, 162] to (3.62) with $\vartheta_0 = \varpi$, results in the iteration

$$\vartheta^{n+1} = \vartheta^n + s[\gamma\Delta\vartheta^n + (\varpi - \vartheta^n)], \tag{3.64}$$

where s as the step-size parameter. The above iteration can be re-written as

$$\frac{\vartheta^{n+1} - \vartheta^n}{s} = \gamma\Delta\vartheta^n + (\varpi - \vartheta^n), \tag{3.65}$$

which is equivalent to the discretization of the following heat equation with time-dependent heating source $\varpi - v$ and $\delta t = s$

$$v_t = \gamma\Delta v + (\varpi - v), \tag{3.66}$$

$$v(0) = \varpi. \tag{3.67}$$

We mention that an alternative to the above smoothing regularization is to convolve the gradient $\varpi = \nabla J_i(\sigma)$ with a Gaussian kernel of variance $\varrho > 0$

$$f_\varrho(x) = \frac{1}{4\pi\varrho}\exp\Big(-\frac{|x|^2}{4\varrho}\Big)$$

which produces the smoothed update

$$\vartheta = f_\varrho * \varpi = \int f_\varrho(x - y)\varpi(y)dy$$

Again, the smoothed update ϑ is computed by solving the heat equation (3.66).

Next we provide a Pseudo-Code of a practical algorithm which solves (3.61a)-(3.61b).

Algorithm 5: Smooth Gradient

Input: $\varpi = \nabla \mathcal{J}_i[\sigma]$, where $\nabla \mathcal{J}_i$ given in (3.19), $\gamma > 0$

Output: $\nabla \mathcal{J}_i^s[\sigma] = \vartheta^I$

Initialisation $\vartheta^0 = \varpi$;

for $n := 1, \ldots, I-1$ **do**

$\vartheta^{n+1} = \vartheta^n + s[\gamma \Delta \vartheta^n + (\varpi - \vartheta^n)]$ Solve (3.64) with $\nabla \vartheta^{n+1} = 0$ on $\partial\Omega$;

end

Define $\vartheta^I = \alpha^{-1}\vartheta^N$ with $\alpha = \|\vartheta^N\|/\|\varpi\|$;

3.5 Shape and contrast value reconstructions in DOT

In order to reconstruct the values and shapes for both scattering and absorption coefficients individually or simultaneously for the inverse problem in DOT, we proceeded analogously as suggested in [7, 144]. This method is described now. Using the level set technique presented in section 4.2, the absorption and scattering coefficients μ and b are distributed in a domain Ω as in expressions (3.42) with their associated level set functions given in (3.43). Its last equations are mathematically equivalent to

$$\mu(\phi) = \mu_i\,(1 - H(\phi)) + \mu_e H(\phi) \qquad b(\psi) = b_i\,(1 - H(\psi)) + b_e H(\psi), \tag{3.68}$$

where H is the Heaviside function in $1D$. It has been suggested in [40, 129] to approximate the Heaviside function with some suitable smooth function since the Heaviside function itself is not differentiable. Given the residual operator defined in (3.3), the inverse problem is to find the level set functions ϕ and ψ and the interior coefficient values μ_i and b_i for which the following cost functional $\mathcal{J}_j$ is minimal

$$\mathcal{J}_j(\phi, \psi, \mu_i, b_i) = \mathcal{R}_j(\phi, \psi, \mu_i, b_i) \quad \text{for} \quad j = 0, \ldots, M-1.$$

We mention that we have used the index j in this section in order to denote the j-th number of the source q_j instead of the index i to avoid confusion with the subindex i, which denotes the interior of the distribution of the coefficients μ, b. The descent direction for the functional $\mathcal{J}_j$ is obtained by taking the total derivative of the functional $\mathcal{J}_j$ with

respect to the time t. By using the chain rule we obtain

$$\begin{aligned}\frac{d\mathcal{J}_j}{dt} = \frac{\partial \mathcal{J}}{\partial \mu}\frac{\partial \mu}{\partial \phi}\frac{d\phi}{dt} + \frac{\partial \mathcal{J}_j}{\partial b}\frac{\partial b}{\partial \psi}\frac{d\psi}{dt} \\ + \frac{\partial \mathcal{J}_j}{\partial \mu}\frac{\partial \mu}{\partial \mu_i}\frac{d\mu_i}{dt} + \frac{\partial \mathcal{J}_j}{\partial b}\frac{\partial b}{\partial b_i}\frac{db_i}{dt}.\end{aligned} \tag{3.69}$$

Denoting $\mathcal{R}_j(\phi, \psi, \mu_i, b_i)$ by $\mathcal{R}_j(\mu(\phi, \mu_i), b(\psi, b_i))$ gives

$$\frac{\partial \mathcal{J}_j}{\partial \mu}\delta\mu = \mathrm{Re}\left(\int_\Omega \mathcal{R}'_{j,\mu}[\mu, b]^* \mathcal{R}_j(\mu, b)\delta\mu \, dx\right), \tag{3.70}$$

where $\mathcal{R}'_{j,\mu}[\mu, b]^*$ is the adjoint of the Fréchet derivative of $\mathcal{R}_j(\mu, b)$ with respect the coefficient μ given by (3.11). Using equation (3.68), one obtains

$$\frac{\partial \mu}{\partial \phi} = (\mu_e - \mu_i)H'(\phi) \qquad \frac{\partial \mu}{\partial \mu_i} = 1 - H(\phi), \tag{3.71a}$$

$$\frac{\partial b}{\partial \psi} = (b_e - b_i)H'(\psi) \qquad \frac{\partial b}{\partial b_i} = 1 - H(\psi), \tag{3.71b}$$

where $H'(\phi)$ and $H'(\psi)$ are the $1D$ Dirac delta distributions. Defining the following forcing terms

$$f_\phi(x,t) = \frac{d\phi}{dt}, \quad f_\psi(x,t) = \frac{d\psi}{dt}, \quad g_\mu(t) = \frac{d\mu_i}{dt}, \quad g_b(t) = \frac{db_i}{dt}, \tag{3.72}$$

and using (3.70) and (3.71), equation (3.69) results in

$$\begin{aligned}\frac{d\mathcal{J}_j}{dt} = \mathrm{Re}\Bigg[\int_\Omega \left[(\mu_e - \mu_i)H'(\phi)\mathcal{R}'_{j,\mu}[\mu, b]^*\mathcal{R}_j(\mu, b)\right] f_\mu \, dx \\ + \int_\Omega \left[(b_e - b_i)H'(\psi)\mathcal{R}'_{j,b}[\mu, b]^*\mathcal{R}_j(\mu, b)\right] f_b \, dx \\ g_\mu \int_\Omega (1 - H(\phi))\, \mathcal{R}'_{j,\mu}[\mu, b]^*\mathcal{R}_j(\mu, b)\, dx \\ + g_b \int_\Omega (1 - H(\psi))\, \mathcal{R}'_{j,b}[\mu, b]^*\mathcal{R}_j(\mu, b)\, dx\Bigg],\end{aligned} \tag{3.73}$$

since the terms g_μ and g_b only depend on t. Using the fact that $H'(\phi)$ and $H'(\psi)$ are positive, the steepest descent direction terms f and g of equation (3.72) are now given as

$$f_{\phi,[k]} = -(\mu_e - \mu_i)\mathcal{R}'_j[\mu, b]^*\mathcal{R}_j(\mu, b), \tag{3.74a}$$

$$f_{\psi,[k]} = -(b_e - b_i)\mathcal{R}'_j[\mu, b]^*\mathcal{R}_j(\mu, b), \tag{3.74b}$$

and

$$g_\mu = -\int_{\Upsilon_1} \mathcal{R}'_{j,\mu}[\mu, b]^* \mathcal{R}_j(\mu, b)\, dx, \tag{3.75a}$$

$$g_b = -\int_{\Upsilon_2} \mathcal{R}'_{j,b}[\mu, b]^* \mathcal{R}_j(\mu, b)\, dx. \tag{3.75b}$$

Applying the LK formula (2.94) into the forcing terms (3.74) and into forcing parameters (3.75), one obtains

$$\begin{pmatrix} \phi^{k+1} \\ \psi^{k+1} \end{pmatrix} = \begin{pmatrix} \phi^k + \tau_\phi f_{\phi,[k]} \\ \psi^k + \tau_\psi f_{\psi,[k]} \end{pmatrix}, \tag{3.76}$$

where

$$f_{\phi,[k]} = -\nabla \mathcal{J}_{[k]}(\mu, b), \tag{3.77a}$$

$$f_{\psi,[k]} = -\nabla \mathcal{J}_{[k]}(\mu, b), \tag{3.77b}$$

with initial values $\hat{\mu}^0 := \mu_b + \epsilon_1$ and $\hat{b}^0 := b_b + \epsilon_2$. Here μ_b and b_b are the values of the background for μ and b in Ω respectively, and $\epsilon_1, \epsilon_2 \geq 0$, and

$$\begin{pmatrix} \mu^{k+1} \\ b^{k+1} \end{pmatrix} = \begin{pmatrix} \mu^k + \tau_\mu (g_\mu)_{[k]} \\ b^k + \tau_b (g_b)_{[k]} \end{pmatrix}, \tag{3.78}$$

where

$$(g_\mu)_{[k]} = -\int_{\Upsilon_1} \nabla \mathcal{J}_{[k]}(\mu, b)\, dx, \tag{3.79a}$$

$$(g_b)_{[k]} = -\int_{\Upsilon_2} \nabla \mathcal{J}_{[k]}(\mu, b)\, dx. \tag{3.79b}$$

Furthermore, $\nabla \mathcal{J}_{[k]}(\phi), \nabla \mathcal{J}_{[k]}(\psi)$ are given by (3.49) with $\hat{\mu} = \mu_i$, $\hat{b} = b_i$ and $\mu_b = \mu_e$, $b_b = b_e$ and τ_ϕ, τ_ψ defined in (3.48); Also $\nabla \mathcal{J}_{[k]}(\mu), \nabla \mathcal{J}_{[k]}(b)$ are given by (3.23) and τ_μ, τ_b defined in (3.14) for the absorption and scattering coefficient respectively.

For noisy data we have used the smooth gradient $\nabla \mathcal{J}^s_{[k]}(\mu, b)$ (as explained in the previous section) instead of the gradients $\nabla \mathcal{J}_{[k]}(\mu)$, $\nabla \mathcal{J}_{[k]}(b)$ in equations (3.76) and (3.78).

The just described shape and contrast value reconstruction algorithm for the absorption and scattering coefficients μ, b now reads in Pseudo-code form as follows.

Algorithm 6: LK and LS methods with contrast value.

Input: $(\Upsilon_{1,[0]}, \mu_{[0]}, \phi_{[0]}), (\Upsilon_{2,[0]}, b_{[0]}, \psi_{[0]})$ Initial guess

Output: $\Upsilon_{1,I(M-1)} = \Upsilon_1[\phi_{I(M-1)}]$, $\Upsilon_{2,I(M-1)} = \Upsilon_1[\psi_{I(M-1)}]$ and $b_{I(M-1)}$,

for $k := 0, \ldots, I(M-1)$ **do**

Compute the gradient $\nabla \mathcal{J}_{[k]}(\mu_k, b_k)$ given by (3.55) ;

Compute the smooth gradient $\nabla \mathcal{J}^s_{[k]}(\mu^k, b_k)$ given by algorithm 5 ;

Update ϕ^{k+1} and ψ^{k+1} using (3.76);

Update μ^{k+1} and b^{k+1} using (3.78);

end

Here $u_{[k]}$ solves (2.1) with μ^k and source $q_{[k]}$, and $z_{[k]}$ solves (3.10) with μ^k and source $\xi_{[k]}$ given by (3.52), respectively. As shown in the Pseudo-code, we have set an initial level set function as our initial guess in order to start the shape and contrast value reconstruction process. For this we have chosen a "signed distance function" for some simple initial shape.

3.6 Discretization of the forward and inverse problems

Below we will consider in more details the following evolution equation

$$\varphi_t + \mathcal{V} \cdot \nabla \varphi(x(t), t) = 0. \tag{3.80}$$

where the quantity $\mathcal{V}$ needs to be determined from techniques discussed above using the RTE. Here we will start deriving some general discretization schemes which can be used as well in the discretization of the RTE.

To start with, the spatial derivatives of the above equation can be approximated using finite differences techniques [146]. We obtain a first-order accurate forward difference scheme

$$\frac{\partial \varphi}{\partial x} \approx \frac{\varphi^{n+1} - \varphi^n}{\Delta x},$$

denoted as $D^{+x}\phi$, or alternatively a first-order accurate backward differences scheme

$$\frac{\partial \varphi}{\partial x} \approx \frac{\varphi^n - \varphi^{n-1}}{\Delta x},$$

denoted as $D^{-x}\varphi$,. We also can derive a second-order accurate central differences scheme

$$\frac{\partial \varphi}{\partial x} \approx \frac{\varphi^{n+1} - \varphi^{n-1}}{2\Delta x},$$

denoted as $D^{0x}\phi$ [146]. Considering the vector field $\mathcal{V}$ in $2D$, i.e., $\mathcal{V} = (v, w)$ and denoting $t^{n+1} = t^n + \Delta t$, $\varphi^n = \varphi(t^n)$, $\varphi^{n+1} = \varphi(t^{n+1})$, $v^n = v(t^n), w^n = w(t^n)$, after applying the forward Euler method for the time variable discretization of (4.33) in $2D$, one obtains

$$\frac{\varphi^{n+1} - \varphi^n}{\Delta t} + v^n \varphi_x^n + w^n \varphi_y^n = 0. \tag{3.81}$$

Approximating the spatial derivatives φ_x and φ_y by using forward or backward differences in eq. (3.81) is not suitable [126]. Firstly, we start by addressing the computation of the term $v^n \varphi_x^n$. The technique used to evaluate this term can be applied independently to the term $w^n \varphi_y^n$. For this purpose, we firstly treat the one-dimensional version of (3.81) where we obtain

$$\frac{\varphi^{n+1} - \varphi^n}{\Delta t} + v^n \varphi_x^n = 0. \tag{3.82}$$

Since φ also depends on x, and v can depend on x, at a specific point x_i, the above eq. is rewritten as

$$\frac{\varphi_i^{n+1} - \varphi_i^n}{\Delta t} + v_i^n (\varphi_x)_i^n = 0, \tag{3.83}$$

where $\varphi_i^n := \varphi(x_i, t_n)$, $(\varphi_x)_i^n := \varphi_x(x_i, t^n)$ and $v_i^n := v(x_i, t^n)$. In the upwinding or upwind differencing scheme one decides on the choice of a discretization of the spatial derivative of (3.82) depending on the sign of v. Upwind methods approximate derivatives by taking into account the directions into which the characteristic information is marching [146]. If $v_i > 0$, using the method of characteristics, φ_x is approximated with D^{-x}, and φ_x is approximated with D^{+x} if $v_i < 0$. A finite difference approximation to a linear partial differential equation is consistent if the approximation error converges to zero as $\Delta t \to 0$, $\Delta x \to 0$. "A finite difference approximation to a linear partial differential equation is convergent if and only if it is both consistent and stable" according to Lax-Richtmyer [126]. The combination of the forward Euler time discretization with the

upwind differencing is a consistent finite difference approximation to the equation (3.82), see [126], and stability is obtained if the following Courant-Friedrichs-Levy condition (CFL) holds [146],

$$\Delta t < \frac{\Delta x}{\max\{|v|\}}. \tag{3.84}$$

Here the maximum value $\max\{|v|\}$ is taken over the entire grid. In practice, the CFL condition (3.84) is satisfied by a CFL number $0 < \alpha < 1$ such that

$$\Delta t \left(\frac{\max\{|v|\}}{\Delta x} \right) = \alpha. \tag{3.85}$$

In $2D$, the CFL condition is written as

$$\Delta t < \frac{\min\{\Delta x, \Delta y\}}{\max\{|v| + |w|\}}.$$

In practice, the above inequality is enforced by choosing a CFL number $0 < \alpha < 1$ such that

$$\Delta t \left(\frac{\max\{\|\mathcal{V}\|\}}{\min\{\Delta x, \Delta y\}} \right) = \alpha.$$

We used a finite differences discretization of the RTE in $2D$ in our numerical simulations. For treating the RTE numerically in $3D$, see [45, 56, 60, 75]. We discretized the angular variable $\theta \in S^1$ using a variant of the discrete ordinates method, that is, using I equidistant directions

$$\theta_i = (\cos\vartheta_i, \sin\vartheta_i), \quad \vartheta_i = i\frac{2\pi}{I}, \quad i = 1, \ldots, I.$$

The time variable t is treated as follows

$$t_l = l\Delta t, \quad l = 0, 1, \ldots, L,$$

where Δt is some suitably chosen positive time-step. We denote the density of photons $u_i^{(l)}$ and the source $q_i^{(l)}$ corresponding to each direction θ_i at time t_l by

$$u_i^{(l)}(x, y) = u(x, y, \theta_i, t_l), \qquad q_i^{(l)}(x, y) = q(x, y, \theta_i, t_l),$$

where x, y are the spatial variables. Let $\theta_k := (\cos\vartheta_k, \sin\vartheta_k)$, $\vartheta_k = k2\pi/I$, $k = 1, \ldots, I$

and S_k be the sectors of the unit circle defined by

$$S_k = \left\{ \theta' = (\cos\vartheta', \sin\vartheta') \in S^1 |\ \frac{-\pi}{I} < \vartheta_k - \vartheta' \leq \frac{\pi}{I} \right\},$$

and let $\eta_{k,i}$ be the integration weights corresponding to the scattering phase function η defined by

$$\eta_{k,i} := \int_{S_k} \eta(\theta_i \cdot \theta')\, d\theta'.$$

Then the conservation condition (2.3) is satisfied. We approximate the scattering integral of equation (2.5) by

$$\begin{aligned} \int_{S^1} \eta(\theta_i \cdot \theta) u(\theta') d\theta' &= \sum_{k=1}^{I} \int_{S_k} \eta(\theta_i \cdot \theta') u(\theta')\, d\theta' \\ &\approx \sum_{k=1}^{I} u(\theta_k) \int_{S_k} \eta(\theta_i \cdot \theta')\, d\theta' = \sum_{k=1}^{I} u(\theta_k) \eta_{k,i}. \end{aligned}$$

An alternative method for discretizing the angular variable is the P_n method [33], or using Legendre polynomials or wavelets as a basis to calculate the integral term in the RTE (2.1) [75]. Owing to the lack of a naturally occurring scattering function in 2D DOT, we have used for convenience a $2D$-adaptation of the $3D$ H-G function (2.5) as the scattering phase function η, which has been designed to share important features of the $3D$ counterpart (e.g. the presence of a parameter between 0 and 1 steering the directional distribution of the scattering) [56]. We mention that we also have experimented with a H-G function in $2D$ given by (2.4) in a specific numerical experiment using the combination of the LK and LS methods with no significant discrepancy in the results of the shape reconstruction compared with using this $2D$ adaptation of the $3D$ H-G scattering function. Notice also that even the $3D$ H-G function is just one (arbitrary but convenient) synthetic model to be used in DOT, which might have to be adjusted to real data once they become available.

Using now the first-accurate forward differences scheme in time at t_l in (2.1) we obtain the discretized time-marching rule

$$\frac{u_i^{(l+1)} - u_i^{(l)}}{\Delta t} + \cos\vartheta_i \frac{\partial u_i^{(l)}}{\partial x} + \sin\vartheta_i \frac{\partial u_i^{(l)}}{\partial y} + [\mu(x,y) + b(x,y)] u_i^{(l)} - b(x,y) \sum_{k=1}^{I} \eta_{k,i} u_k^{(l)} = q_i^{(l)}.$$

The following part of the above equation

$$\frac{u_i^{(l+1)} - u_i^{(l)}}{\Delta t} + \cos\vartheta_i \frac{\partial u_i^{(l)}}{\partial x} + \sin\vartheta_i \frac{\partial u_i^{(l)}}{\partial y} = 0,$$

is equal to the discretization of eq. (3.80) if one corresponds $u_i \to \phi$, $\cos\vartheta_i \to v$, $\sin\vartheta_i \to w$; $\mathcal{V} = (v, w)$, therefore this part of the above eq. is discretized by the upwinding differences scheme explained at the beginning of this section. We discretize the spatial derivatives of $u_i^{(l)}$ using the upwinding scheme in a uniform grid with step length Δx as explained in subsection 4.2.2. Hence, one obtains

$$\frac{\partial u_i^{(l)}(x,y)}{\partial x} = \begin{cases} \frac{u_i^{(l)}(x,y) - u_i^{(l)}(x-\Delta x,y)}{\Delta x} & \cos\varphi_i \geq 0, \\ \frac{u_i^{(l)}(x+\Delta x,y) - u_i^{(l)}(x,y)}{\Delta x} & \cos\varphi_i < 0. \end{cases} \tag{3.86}$$

The derivative of $u_i^{(l)}$ with respect to y is discretized in an analogous fashion. In order to achieve a stable and physically valid numerical scheme, the CFL condition (3.85) must hold with regards to stability, and the following condition must hold

$$\Delta t < \frac{1}{\max\{\mu(x,y) + b(x,y)\}},$$

which also takes into account the physical values of μ and b in the tissue. Otherwise, the density $u_i^{(l+1)}$ in (3.86) can become negative resulting in destabilizing oscillations. In general, this physical CFL condition for Δt is stricter than the above mentioned numerical CFL condition. Considering both conditions results in choosing a very small value for Δt which is very expensive for our actual numerical scheme, since in order to compute the update for the absorption and scattering coefficients in either the Landweber method (3.13), the Landweber-Kaczmarz method (3.21), or the Loping-LK method (3.25), or updating their correspondent level set functions using (3.47), it is needed to calculate integrals over time in (3.12). This needs to be done *after* the forward and adjoint transport solutions have been obtained, which would require us to store a large amount of time-steps (predicted by the CFL condition) of these solutions in memory. For this reason, even though observing the physical CFL condition for the numerical propagation of photons, for the independent integration part calculating Frechet derivatives we have just stored every multiple of twelve of Δt for calculating these integrals. The backtransport problem

(3.10) is discretized analogously to the forward problem (2.1), with the difference that the signal now travels backwards in time and space, which needs to be taken into account in the upwind scheme for the spatial derivatives. We mention that an alternative spatial discretization for structured grids is the diamond scheme and its variants, see [13, 60, 75] and the finite element, finite volume and the discontinuous Galerkin methods for unstructured grids, see [1, 62, 75, 153]. For a numerical scheme for a transport regime using the RTE to model the propagation of the photons instead of a diffusive transport regime as considered in this dissertation, see [13].

3.7 Numerical Results

In all the numerical simulations presented in this dissertation, the data are synthetically generated, that is, they were not collected from real experimental data. We used the same discretization for both the forward model to create the synthetic data and for the reconstruction process. Doing this is considering as an "inverse crime". Therefore, we have added pointwise noise to the synthetic data. We mention that we have used the code of [52, 53] for the forward problem, which has been validated against Monte Carlo solver simulation.

Computational setup of our numerical experiments.

Our domain Ω is a square of $5 \times 5\ cm^2$ divided into 2500 pixels which each having a width of $1 \times 1\ mm^2$. We have embedded a so-called "clear layer" close to the boundary which simulates the cerebrospinal fluid (CFL) region of a head. The location and the value of the absorption and scattering coefficients of this clear layer are assumed to be known a priori. The total time T of the numerical photon flow considered in the forward or inverse problem using the RTE is 100 s, that is, our data are recorded in the interval of time $[0, 20s]$. As mentioned in section 2.1, the speed of the light in the medium was normalized to $c = 1$ cm/s, thus, a photon can propagate 20 cm in this interval of time. A time step is equal to 0.2 s when considering 20 discrete time steps of the propagation of the photons. As we mentioned in chapter 2, we have used a $2D$ adaptation of the $3D$ Henyey-Greenstein function (2.5) as the scattering function [56] with $g = 0.9$ for single and simultaneous reconstructions of absorption and scattering coefficients. The RTE of the forward and inverse problems were discretized using the finite differences

method discussed in section 3.6. The angular directional variable θ was discretized into 12 direction vectors which are equidistantly distributed over the unit circle S^1. We use 16 smeared-out (instead of delta-like) sources with width 3 mm or 3 pixels. These sources are emitting an ultra-short laser pulse at time $t = 0$ into the medium, in the direction perpendicular to the boundary. These sources are located at the boundary of the medium, four sources equidistantly 1 cm at each side of the boundary of Ω between 1-4 cm, such that on the corners there are not sources. The reason of this arrangement is because our domain is square and we have embedded a clear layer near to the boundary. Thus, the photons from the sources on the corners would be almost completely channelled through the clear layer. Detectors are positioned along the boundary $\partial\Omega$ which record the outgoing flux given by (3.2), and measurements are taken at receiver positions of length 1 cm along the boundary apart from the source position. The parameter values used in the different experiments are listed in the following tables.

	$\mu[cm^{-1}]$	$b[cm^{-1}]$
Background	0.1	100
Inclusion	0.5	100
Clear layer	0.01	0.1

Table 3.1: Experimental set-up of the absorption coefficient reconstruction.

	$\mu[cm^{-1}]$	$b[cm^{-1}]$
Background	0.1	100
Inclusion	0.1	150
Clear layer	0.01	0.1

Table 3.2: Experimental set-up of the scattering coefficient reconstruction.

	$\mu[cm^{-1}]$	$b[cm^{-1}]$
Background	0.1	100
Inclusion	0.5	150
Clear layer	0.01	0.1

Table 3.3: Experimental set-up of the simultaneous reconstruction of the absorption and scattering coefficients.

Further discussion of details of our numerical setup.

The choice of the receiver positions is done keeping in mind that receivers close to the source do not have important information of the inside of the medium. Moreover, if a clear layer is added to our reconstruction problem, early photons that arrive at receiver position x_r are channelled through this layer. Therefore, both sensitivity functions $\mathfrak{L}_1$ and $\mathfrak{L}_2$ have huge values inside the clear layer and small values outside the clear layer. For this reason, only late photons who arrive at the receiver position x_r are considered in our numerical experiment design. Notice that by taking into account only late photons, the resolution of the reconstruction decreases. Only photons which arrive at the detectors between $15 - 20s$ are considered as part of the data, because only late photons provide important information of the inside of the region Ω. Early photons are likely to have channelled through the clear layer as pointed out in section 3.1.

The relaxation parameters τ_μ, τ_b for the Landweber, LK and L-LK methods without sparsity regularization were manually selected to retrieve a feasible reconstruction. Detectors are located along all the boundary $\partial\Omega$ which record the outgoing flux given by (3.2). The results of shape and contrast value reconstruction presented in this section are similar with those obtained in [7, 144], where the authors used the diffusive approximation equation. Table 3.1 lists the physical values used in the simulation for the contrast value μ reconstruction. Figures 3.2-3.6 were obtained using algorithm 6.

Details of the different numerical experiments.

We have made the following shape and contrast value reconstructions of the absorption and scattering coefficients in DOT using the combination of the LK and LS methods with the algorithms presented in this chapter. Firstly, we present some reconstructions of the μ absorption coefficient, then for the b scattering coefficient, and finally the simultaneous reconstruction for μ and b. Two absorption obstacles of different size are reconstructed in the *first experiment* which simulates haematoma, and two scattering obstacles of the same size are reconstructed in the *second experiment*. Finally, two absorption and one scattering inclusion are reconstructed simultaneously in the last experiment.

Discussion of the individual numerical experiments.

Figure 3.2 shows the simultaneous reconstruction of contrast and shape for the absorption coefficient μ only when adding 5% Gaussian noise to the data. Figure 3.3 shows the simultaneous reconstruction of shape and contrast for the scattering parameter b only when adding 5% Gaussian noise to the data.

Figures 3.4-3.6 show the simultaneous reconstruction of shape and contrast value for both parameters μ and b from the same data set. Figure 3.4 shows the corresponding evolution for the μ coefficient. Figure 3.5 shows the evolution for the b coefficient. Figure 3.6 shows the evolution of the contrast values and the residual norm.

Figures 3.7-3.8 show the simultaneous reconstruction of shapes and contrast values of both the absorption parameter μ and scattering parameter b using the two-step shape reconstruction method described in algorithm 4, with the LK method used in the first stage and the combination of the LK and the Level Set method in the second stage.

Figure 3.9 shows the simultaneous shape reconstruction for μ and b coefficients using the algorithm 4. Figure 3.10 shows the comparison of the norm of the residuals using the two-step method using in the first stage only the LK method and in the second stage the LK-LS method on the one hand, and only using the LK-LS method on the other hand. As it can be seen in figure 3.10, the use of the two-step method is much more efficient than using only the LK-LS method. We mention that for single or individual reconstructions, one of the coefficients of μ, b is assumed to be known, in which case it does not vary over the iterations during the whole reconstruction process.

We want to comment further on the reconstructions shown in figure 3.5. During the simultaneous reconstruction of two different physical parameters, the so-called cross-talk is an undesirable side-effect where scattering (absorption) distributions of inclusions are obtained at positions where absorption (scattering) inclusions are located, respectively. Much more substantial cross-talk is obtained for the scattering inclusion reconstructions in these simultaneous reconstructions as it can be seen in figure 3.5. This phenomenon is due to the ill-posedness of the inverse problem, that is, distinct scattering and absorption spatial distributions can give almost the same measurement's values on the boundary of the tissue. Although there has been a huge effort to overcome this phenomenon in frequency-domain systems, time-resolved systems still present a substantial cross-talk [138, 145].

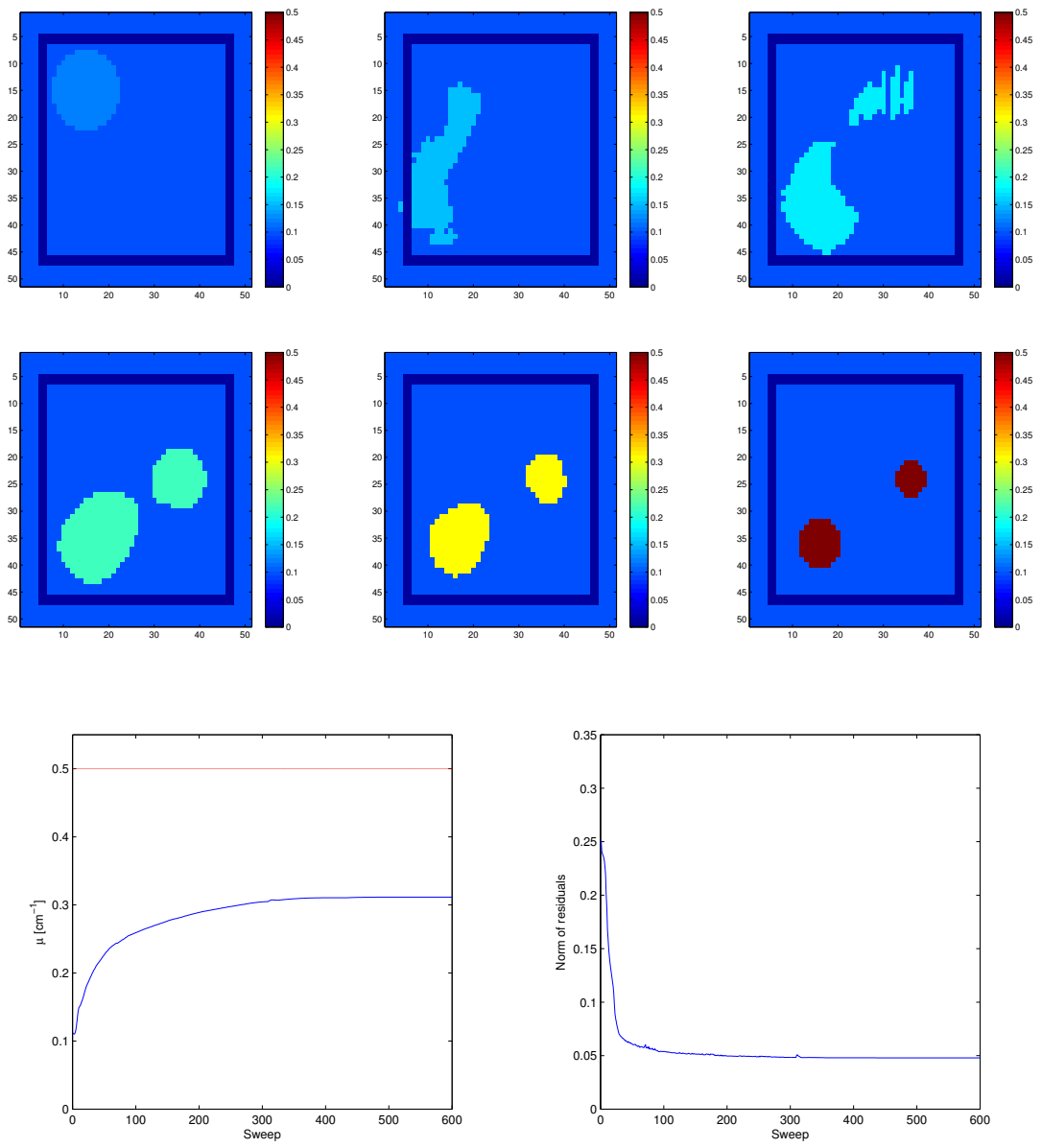


Figure 3.2: Contrast value reconstruction of μ. Top row from left to right: initial guess; reconstruction after 10 sweeps; final reconstruction after 20 sweeps. Middle row from left to right: reconstruction after 40 sweeps; final reconstruction; true objects. Bottom left: contrast value of μ; bottom right: norm of residuals of the cost.

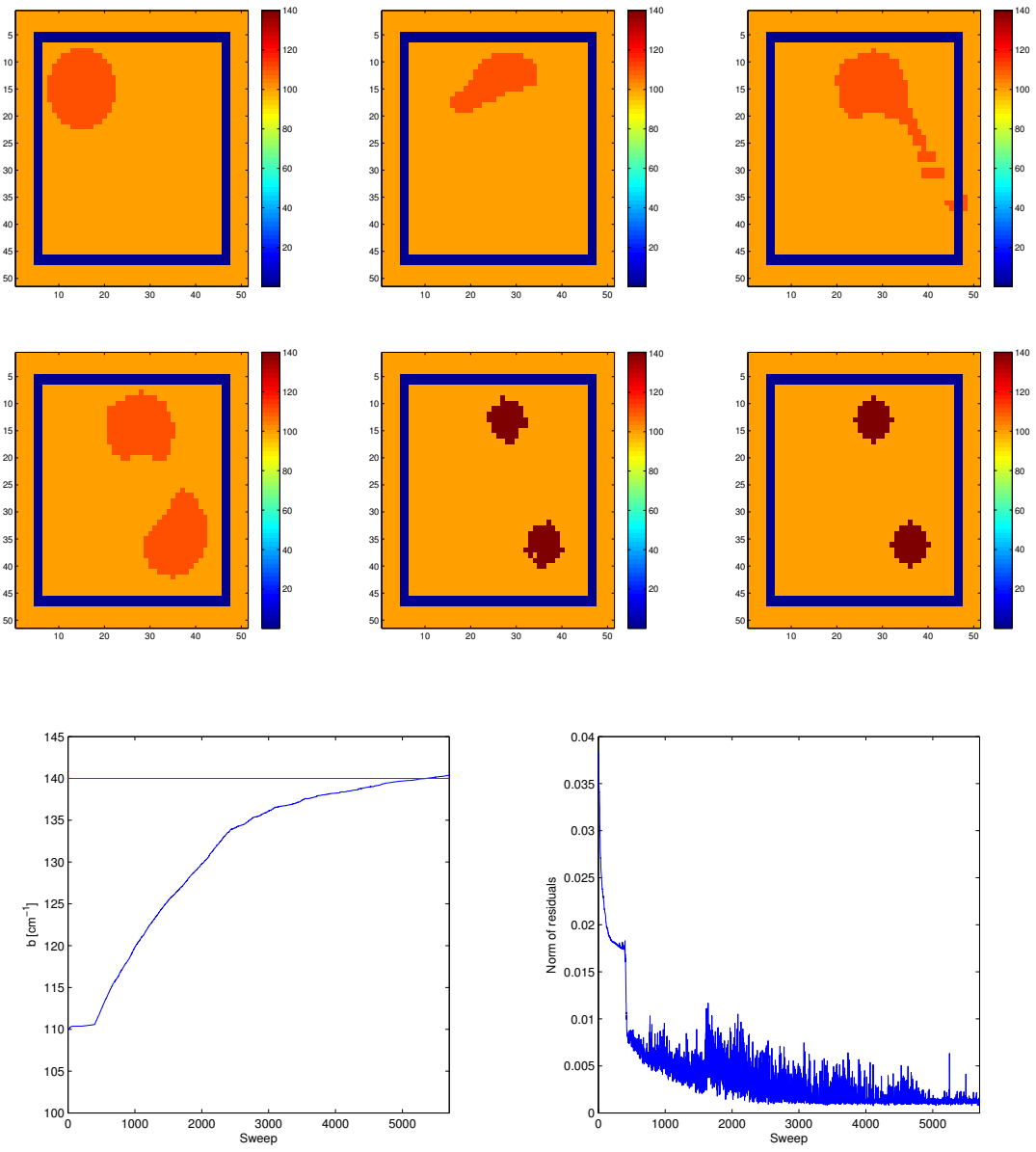


Figure 3.3: Contrast value reconstruction of b. Top row from left to right: initial guess; reconstruction after 200 sweeps; final reconstruction after 400 sweeps. Middle row from left to right: reconstruction after 450 sweeps; final reconstruction after 5800 sweeps; true objects. Bottom left: contrast value of b; bottom right: norm of residuals of the cost.

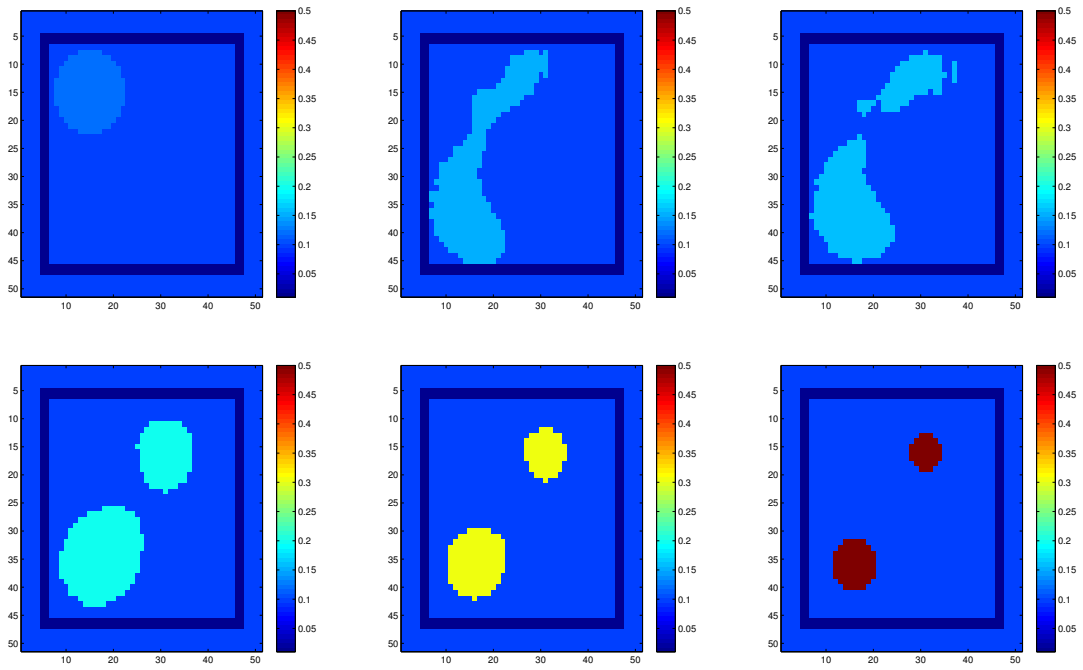


Figure 3.4: Simultaneous contrast value reconstruction of μ. Top row from left to right: initial guess; reconstruction after 140 sweeps; final reconstruction after 160 sweeps. Bottom row from left to right: reconstruction after 400 sweeps; final reconstruction; true objects.

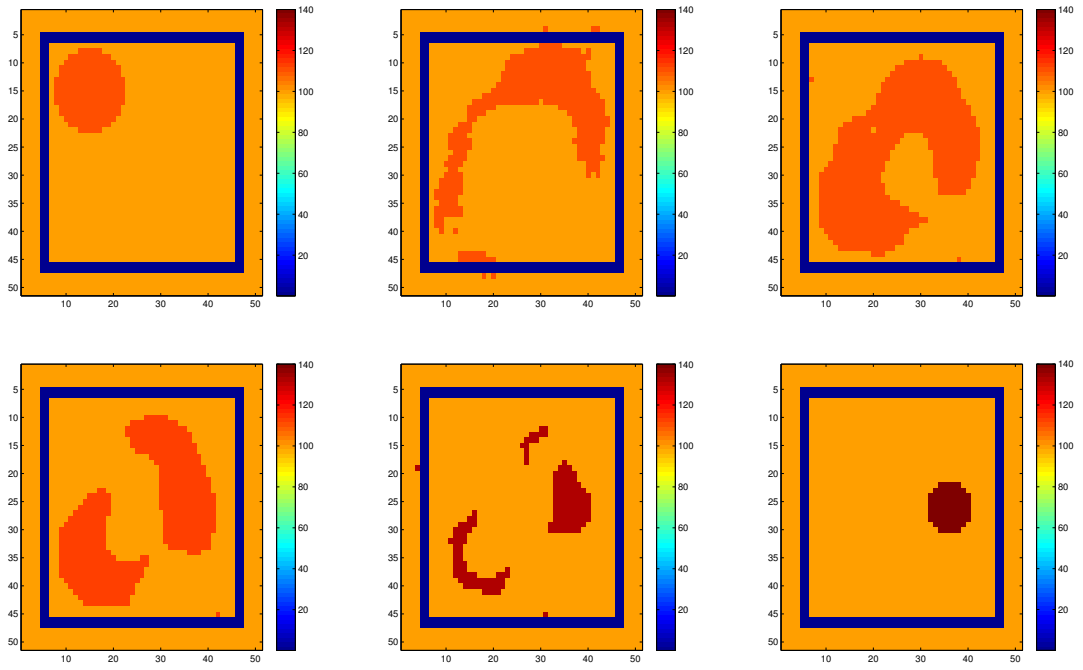


Figure 3.5: Simultaneous contrast value reconstruction of b. Top row from left to right: initial guess; reconstruction after 120 sweeps; final reconstruction after 200 sweeps. Bottom row from left to right: reconstruction after 300 sweeps; final reconstruction; true objects.

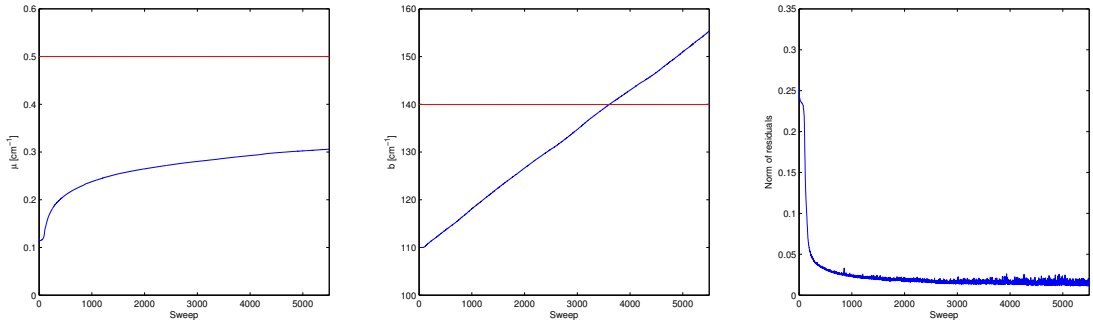


Figure 3.6: Simultaneous contrast value reconstruction of μ and b. From left to right: the contrast value of μ; the contrast value of b; norm of residuals of the cost.

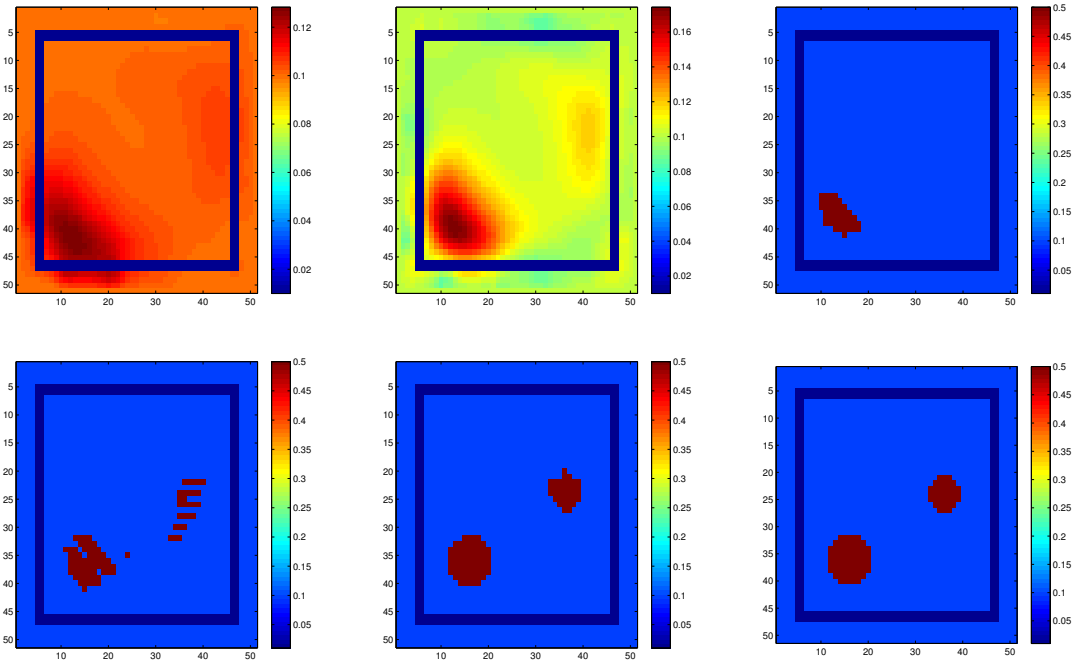


Figure 3.7: Simultaneous shape reconstruction of μ. Top row from left to right: reconstruction after 1 sweep using only the LK method; reconstruction after 9 sweeps using only the LK method; initial shape of the LK-LS reconstruction. Bottom row from left to right: reconstruction after 20 sweeps using the LK-LS method; final reconstruction after 100 sweeps using the LK-LS method; true objects.

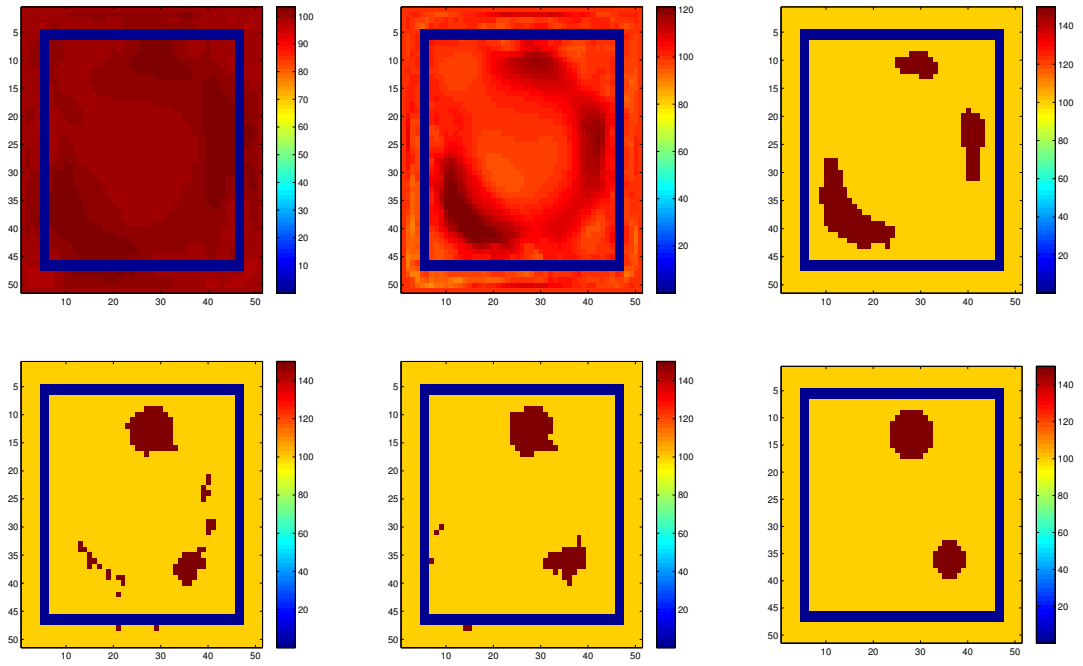


Figure 3.8: Simultaneous shape reconstruction of b. Top row from left to right: reconstruction after 1 sweep using only the LK method; reconstruction after 9 sweeps using only the LK method; initial shape of the LK-LS reconstruction. Bottom row from left to right: reconstruction after 50 sweeps using the LK-LS method; final reconstruction after 100 sweeps using the LK-LS method; true objects.

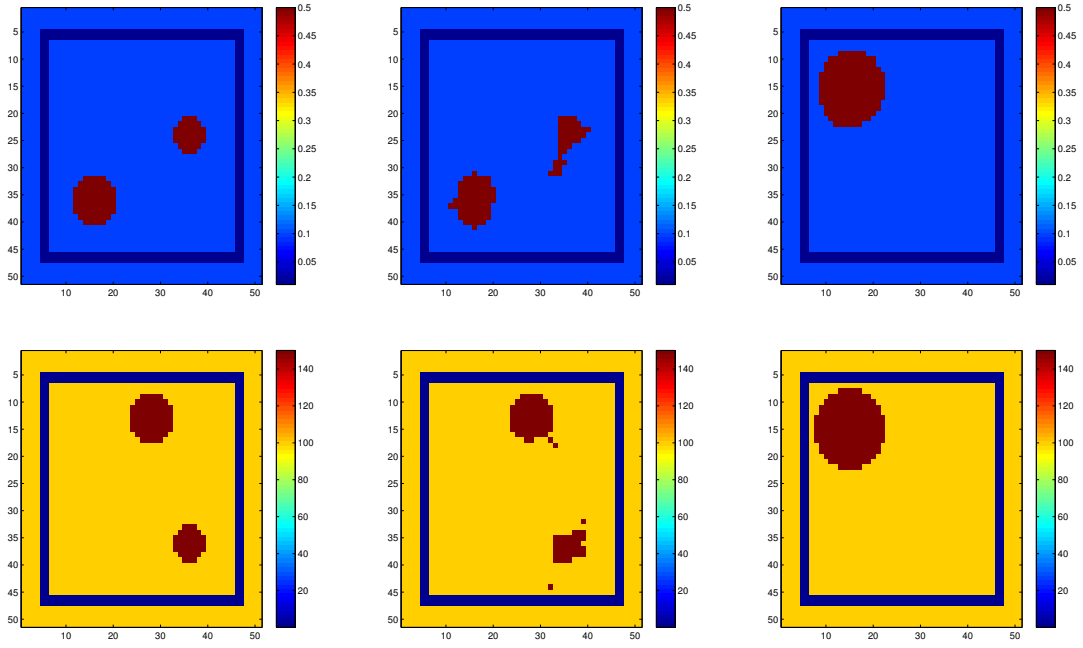
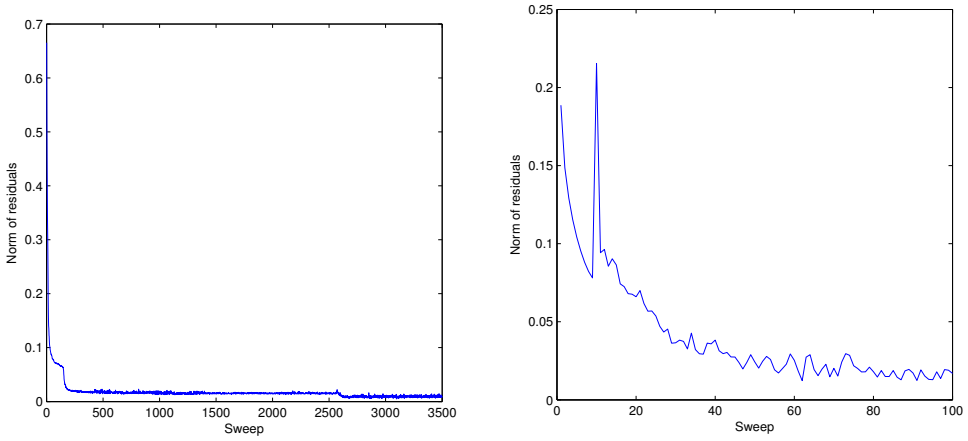


Figure 3.9: Simultaneous shape reconstruction of μ and b. Top row, absorption coefficient, from left to right: true object; final reconstruction; initial guess. Bottom row, scattering coefficient, from left to right: true object; final reconstruction; initial guess.



(a) (b)

Figure 3.10: Figure 3.10a shows the norm of residuals of the cost functional using the combination of the LK and LS methods during all the reconstruction process, figure 3.10b shows the norm of residuals of the cost functional using the LK method in the first stage and then using the combination of the LK and LS methods.

Chapter 4

Level set evolution by shape derivative for RTE

In section 4.1, the basic concept of the Shape Derivative is briefly reviewed. Then, we present a mathematical formulation for the specific shape derivative of the misfit functional $\mathcal{J}$ which appears in the diffuse optical tomography problem based on the RTE. Then we move on to more algorithmic issues. First, in section 4.2 the concept of a Hamilton-Jacobi equation is reviewed, which is closely related to the level set technique for shape reconstruction. Here we describe in particular the cases where the motion of the interfaces is driven by an external vector velocity field and by its mean curvature. Also, the level set evolution equation corresponding to a general motion is presented. In section 4.3 we then provide an algorithm for the shape reconstruction where only the absorption coefficient is assumed to be unknown. In contrast to earlier published work, we apply her the level set technique using directly the shape derivative of DOT computed for the RTE when calculating velocity functions for the interfaces. Finally, in the last section if this chapter, we present the results of some numerical experiments where we reconstruct the absorption coefficient using this new form of the shape derivative in the level set method.

4.1 Shape Derivative

In this section, it is assumed that the absorption coefficient is piecewise constant which takes two distinct values μ_e, μ_i and the scattering coefficient is constant in the whole domain Ω. Then, the domain Ω is split into two disjoint domains Υ_i and $\Upsilon_e := \Omega \backslash \Upsilon_i$,

and it is assumed that $\partial\Upsilon_i \cap \Gamma = \emptyset$. The domains $\Upsilon_i := \bigcup_{j=1} \Upsilon_{i,j}$ given as a guess of the locations of the true inclusions and $\Upsilon_e = \Omega \backslash \overline{\Upsilon}_i$ are depicted in figure 4.1. The background values of the absorption and scattering coefficients are denoted by μ_e, b_e.

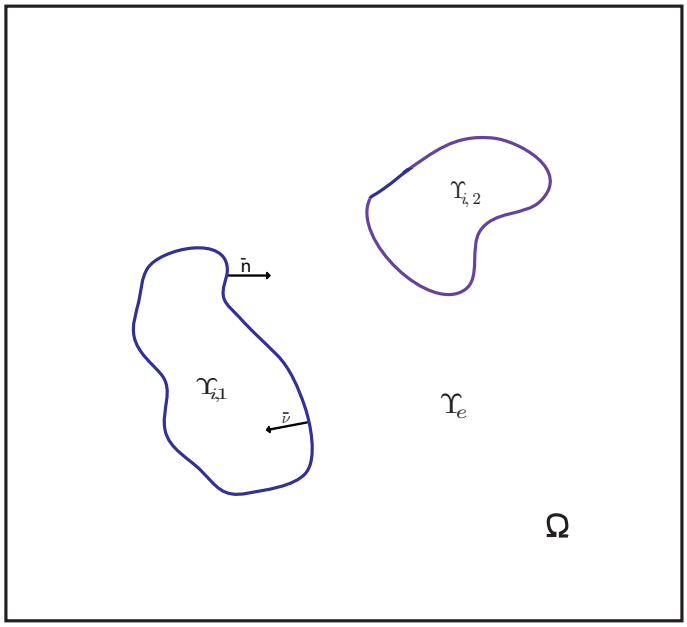


Figure 4.1: Geometry of the region Ω.

Sensitivity Analysis is applied to measure the changes of a certain region $\Upsilon_i \subset \Omega \subset \mathbb{R}^2$ with boundary $\partial\Upsilon$ under a vector velocity field $\mathcal{V}$. We assume that the background value of the absorption coefficient μ_e is known a priori. In this case, the inverse problem in DOT can be written as: minimise

$$\mathcal{J}_j(\Omega \backslash \overline{\Upsilon}_i) = \mathcal{R}_j[\Upsilon_e] = \int_{S^{n-1}} (u_j(\Upsilon_e) - \tilde{u}_j)\theta \cdot \bar{\nu}\, d\theta, \quad j = 0, \ldots, M-1, \tag{4.1}$$

where u_j is the solution of the RTE (2.1) with source q_j and $\tilde{u}_j$ denotes the physical true density for the same source q_j.

In order to calculate the shape derivative for diffusive optical tomography based on the RTE, the variational formulation of the RTE is given as follows. This misfit functional was defined in (3.6). As in section 2.1, it is assumed the following. Let be the space $\Omega \times S^{n-1} \times [0,T]$ denoted by X and the boundary $\partial X := \Gamma \times S^{n-1} \times [0,T]$, where Γ is the boundary of Ω, this boundary is assumed to be Lipchitz, which implies that the outward normal vector $\bar{\nu}$ exists almost everywhere on $\partial\Omega$. The boundary ∂X can be decomposed into an inflow part $\partial X_- := \{(x,\theta,t) \in \partial X, \theta \cdot \bar{\nu} < 0\}$ and an outflow part

∂X_+ where $\theta \cdot \bar{\nu} > 0$. Since the domain Ω is in two dimensions, i.e., $N = 2$, the direction vectors θ are in S^1. We have used the notation S^1 instead of S^{n-1} in order the expressions to be more compact. The function $u(x, \theta, t) \in X$ will be physically considered as the density of the photons travelling in the region X. In this chapter, it is considered the homogeneous RTE with inhomogeneous boundary conditions:

$$\frac{\partial u}{\partial t} + \theta \cdot \nabla u + (\mu(x) + b(x))\, u - b(x) \int_{S^1} \eta(\theta \cdot \theta') u(x, \theta', t)\, d\theta' = 0 \text{ in } X \tag{4.2a}$$

with initial condition

$$u(x, \theta, 0) = 0 \quad \text{in } \Omega \times S^1, \tag{4.2b}$$

with boundary condition

$$u(x, \theta, t) = q(x, \theta, t) \quad \text{on } \partial X_-, \tag{4.2c}$$

with boundary conditions on $\partial \Upsilon_i$

$$u^- = u^+, \tag{4.2d}$$

$$\mu_e \frac{\partial u^-}{\partial \bar{n}} = \mu_i \frac{\partial u^+}{\partial \bar{n}}, \tag{4.2e}$$

where q is the source term given by 2.2, u^- (u^+) denotes the limit values of u on $\partial \Upsilon_i$ from the exterior (interior) of the region Υ_i and μ is defined as follows

$$\mu(x) = \begin{cases} \mu_i(x) & \text{if } x \in \Upsilon_i, \\ \mu_e(x) & \text{if } x \in \Upsilon_e. \end{cases} \tag{4.3}$$

The boundary conditions given in (4.2d) are known as the continuity boundary conditions. An alternative way of describing laser sources at the boundary is to use an homogeneous boundary condition instead of (4.2c) and inhomogeneous source q term on the right side of equation (4.2a). Both expressions are equivalent [45, 53, 153]. In this chapter, we have opted to use L^1 spaces, instead of L^2 spaces, because we will use some theoretical results of the following map traces (defined below) to obtain a formula for the shape derivative of a functional $\mathcal{J}$ when the shape derivative of $\mathcal{J}$ depends on the tangential divergence of

a vector field ($\mathrm{div}_\Gamma\, \mathcal{V}$) defined later. To the best of our knowledge, only these map traces theorems exist for L^1 for the RTE. Then, we define the following space

$$U = \{u | u \in L^1(X),\ (\partial_t u + \theta \cdot \nabla u) \in L^1(X), u|_{\partial X_-} \in L^1(\partial X_-),$$
$$u(x,\theta,0) = 0 \text{ on } \Omega \times S^1, u(x,\theta,t) = q \text{ on } \partial X_-\},$$

with the norm defined as

$$\|u\|_U := \|u\|_{L^1(X)} + \|\partial_t u + \theta \cdot \nabla u\|_{L^1(X)}.$$

Now function spaces $L^1(\partial X_\pm, |\theta \cdot \bar{\nu}| d\theta dt d\Omega)$ on $\partial X_\pm$ are defined to be spaces of functions v on $\partial X_\pm$ such that the norm $|v|_{L^1(\partial X)}$ is finite with respect to the measure $|\theta \cdot \bar{\nu}|\, d\theta dt d\Omega$. Defining the space W as

$$W = \big\{w | w \in L^\infty(X),\ (\partial_t w + \theta \cdot \nabla w) \in L^\infty(X), w|_{\partial X_+} \in L^\infty(\partial X_+, \theta \cdot \bar{\nu}\, d\theta dt d\Omega),$$
$$w(x,\theta,T) = 0 \text{ on } \Omega \times S^1\big\}.$$

As stated in [5, 35, 36, 39, 45], if $v \in U$, the map traces $\gamma_\pm : U \to L^1(\partial X_\pm, |\theta \cdot \bar{\nu}|\, d\theta dt d\Omega)$ defined by $\gamma_\pm(v) := v|_{\partial X_\pm}$, are continuous and subjective. It is also assumed that $q \in L^1(\partial X_\pm, |\theta \cdot \bar{\nu}|\, d\theta dt d\Omega)$ for the inhomogeneous boundary condition given in (4.2c). Once the trace map has been stated, we can apply the Green's formula. In addition, we will use these results for the traces to obtain a formula for the shape derivative of a functional $\mathcal{J}$ when the shape derivative of $\mathcal{J}$ depends on the tangential divergence of a vector field ($\mathrm{div}_\Gamma\, \mathcal{V}$) defined later. We start multiplying the equation (4.2a) by a function $w \in W$. Then, this equation is integrated over the angular directions $\theta \in S^1$, time and spatial domain Ω, that is, on the set X. Then, one obtains

$$\int_\Omega \int_0^T \int_{S^1} \left[\frac{\partial u}{\partial t} w + \theta \cdot \nabla u + (\mu + b) u w - b \int_{S61} \eta(\theta \cdot \theta') u w\right] d\theta dt d\Omega = 0. \tag{4.4}$$

Integrating by parts over time variable, the first term in equation (4.4), one obtains

$$\int_\Omega \int_0^T \int_{S^1} \frac{\partial u}{\partial t} w d\theta d\Omega dt = \int_\Omega \int_{S^1} [u(x,\theta,T) w(x,\theta,T) - u(x,\theta,0) w(x,\theta,0)]\, d\theta d\Omega$$
$$- \int_\Omega \int_0^T \int_{S^1} \frac{\partial w}{\partial t} u\, d\theta dt d\Omega.$$

Using the initial condition of the RTE (4.2b), and since $w \in W$, it holds $w(x,\theta,T) = 0$, therefore

$$\int_\Omega \int_0^T \int_{S^1} \frac{\partial u}{\partial t} w \, d\theta dt d\Omega = \int_\Omega \int_0^T \int_{S^1} -\frac{\partial w}{\partial t} u \, d\theta dt d\Omega. \tag{4.5}$$

Applying Green's formula on the second term of (4.4) over the spatial domain Ω, one obtains

$$\begin{aligned} \int_\Omega \int_0^T \int_{S^1} w(\theta \cdot \nabla u) d\theta dt d\Omega = &- \int_\Omega \int_0^T \int_{S^1} (\theta \cdot \nabla w) u \, d\theta dt d\Omega \\ &+ \int_\Gamma \int_0^T \int_{S^1} wu(\theta \cdot \bar{\nu}) \, d\theta dt d\Gamma. \end{aligned} \tag{4.6}$$

To complete the derivation of the variational formulation of the RTE, it is typical to define the even-odd parity functions and decompose uniquely the function u in terms of these even-odd parts as in [61] and [1]. Instead of doing this, one can continue following the next split as in [153], that is, it can be defined the positive $(\theta \cdot \bar{\nu})_+$ and negative $(\theta \cdot \bar{\nu})_-$ parts of $(\theta \cdot \bar{\nu})$ as follows

$$(\theta \cdot \bar{\nu})_+ = \begin{cases} (\theta \cdot \bar{\nu}), & (\theta \cdot \bar{\nu}) \geq 0, \\ 0, & (\theta \cdot \bar{\nu}) < 0. \end{cases} \tag{4.7}$$

$$(\theta \cdot \bar{\nu})_- = \begin{cases} 0 & (\theta \cdot \bar{\nu}) \geq 0, \\ -(\theta \cdot \bar{\nu}), & (\theta \cdot \bar{\nu}) < 0. \end{cases} \tag{4.8}$$

Substituting $\theta \cdot \bar{\nu}$ by (4.7) and (4.8), the second term on the right hand side of equation (4.6) can be written as

$$\int_\Gamma \int_0^T \int_{S^1} wu(\theta \cdot \bar{\nu}) \, d\theta dt d\Omega = \int_\Gamma \int_0^T \int_{S^1} wu\big((\theta \cdot \bar{\nu})_+ - (\theta \cdot \bar{\nu})_-\big) \, d\theta dt d\Gamma.$$

Using the boundary condition of the RTE (4.2c) and equation (2.2), the above term can be written as

$$\begin{aligned} \int_\Gamma \int_0^T \int_{S^1} wu\big((\theta \cdot \bar{\nu})_+ - (\theta \cdot \bar{\nu})_-\big) \, d\theta dt d\Gamma = &\int_\Gamma \int_0^T \int_{S^1} wu(\theta \cdot \bar{\nu})_+ \, d\theta dt d\Gamma \\ &- \int_\Gamma \int_0^T \int_{S^1} wq(\theta \cdot \bar{\nu})_- \, d\theta d\Gamma dt. \end{aligned}$$

Finally, using Equations (4.5)-(4.9), equation (4.4) can be written as

$$\begin{aligned}\int_\Omega \int_0^T \int_{S^1} \Big[-\frac{\partial w}{\partial t} - \theta\cdot\nabla w + (\mu+b)w - b\int_{S^1}\eta(\theta\cdot\theta')w\Big]u\, d\theta dt d\Omega \\ + \int_\Gamma \int_0^T \int_{S^1} wu(\theta\cdot\bar{\nu})_+\, d\theta dt d\Gamma \\ = \int_\Gamma \int_0^T \int_{S^1} wq(\theta\cdot\bar{\nu})_-\, d\theta dt d\Gamma.\end{aligned}$$

In order to present the weak formulation of the RTE in DOT, the following definitions are given. Let $\mathfrak{a}(\Upsilon_i; u, w)$ be the bilinear form defined by

$$\begin{aligned}\mathfrak{a}(\Upsilon_i; u, w) = &\int_{\Upsilon_e}\int_0^T\int_{S^1}\Big[-\frac{\partial u}{\partial t}w - (\theta\cdot\nabla u)w + (\mu_e+b)uw \\ &- b\int_{S^1}\eta(\theta\cdot\theta')uw\Big]\, d\theta dt d\Upsilon + \int_{\partial\Upsilon_e}\int_0^T\int_{S^1} wu(\theta\cdot\bar{\nu})_+\, d\theta dt ds \\ &+ \int_{\Upsilon_i}\int_0^T\int_{S^1}\Big[-\frac{\partial u}{\partial t}w - (\theta\cdot\nabla u)w + (\mu_i+b)uw \\ &- b\int_{S^1}\eta(\theta\cdot\theta')uw\Big]\, d\theta dt d\Upsilon + \int_{\partial\Upsilon_i}\int_0^T\int_{S^1} wu(\theta\cdot\bar{\nu})_+\, d\theta dt ds.\end{aligned} \tag{4.9}$$

Note that $\mathfrak{a}(\Upsilon_i; u, w)$ is symmetric. Let be $\mathfrak{l}(w)$ the linear form given by

$$\mathfrak{l}(w) = \int_\Gamma\int_0^T\int_{S^1} wq(\theta\cdot\bar{\nu})_-\, d\theta dt ds. \tag{4.10}$$

Thus, the variational problem of the RTE is to find $u \in U$, such that

$$\mathfrak{a}(\Upsilon_i; u, w) = \mathfrak{l}(w) \quad \forall w \in W. \tag{4.11}$$

The shape derivative is defined as in [31, 32, 34, 49, 67, 68, 88, 92, 120, 121, 147]; let be a domain $\Upsilon \subset \mathbb{R}^2$ under a vector velocity field $\mathcal{V} : \mathbb{R}^2 \to \mathbb{R}^2$ which in Mechanics is the Spatial Velocity or Eulerian velocity [148]. The relation between the positions x of the particles of Υ at time ε with their positions ("particles") X at time $\varepsilon = 0$ is given by

$$x = x(X, \varepsilon),$$

then the velocity $\mathcal{V}$ generates a motion described by $\dot{x}(\varepsilon, X) = \mathcal{V}(X)$ with initial condition $x(X, 0) = X$. The description above can be characterised as follows, let $\mathfrak{P}_\varepsilon : \Upsilon \to$

$\mathbb{R}^2$ be a perturbation function of the identity function given by $\mathfrak{P}_\varepsilon = I + \varepsilon\mathcal{V}$, where I is the identity map. Then the perturbed domain denoted by Υ_ε is given by $\Upsilon_\epsilon = \mathfrak{P}_\varepsilon(\Upsilon)$, i.e.,

$$\Upsilon_\epsilon = \Upsilon + \varepsilon\mathcal{V}.$$

The shape derivative of a functional $\mathcal{J}(\Omega\backslash\Upsilon)$ with respect to the ε perturbation caused by the velocity $\mathcal{V}$ denoted by $D\mathcal{J}(\Omega\backslash\Upsilon)\cdot\mathcal{V}$ is defined as

$$D\mathcal{J}(\Omega\backslash\Upsilon)\cdot\mathcal{V} = \lim_{\varepsilon\to 0}\frac{\mathcal{J}(\Omega\backslash\Upsilon_\varepsilon) - \mathcal{J}(\Omega\backslash\Upsilon)}{\varepsilon} = \frac{d}{d\varepsilon}\mathcal{J}(\mathfrak{P}_\varepsilon(\Upsilon))\Big|_{\varepsilon=0}. \tag{4.12}$$

Using the variational formulation of the RTE (4.11), the calculation of the shape derivative (4.12) can be written as [120]

$$\begin{cases} \text{Calculate,} & \frac{d}{d\varepsilon}\mathcal{J}(\mathfrak{P}_\varepsilon(\Upsilon))|_{\varepsilon=0}, \\ \text{Subject to,} & \mathfrak{a}(\Upsilon; u, w) = \mathfrak{l}(w) \quad \forall w \in W. \end{cases}$$

This derivative corresponds to the spatial or Eulerian description of the displacement of the position x. Using the chain rule, the shape derivative results

$$D\mathcal{J}(\Omega\backslash\Upsilon)\cdot\mathcal{V} = \frac{dJ(u(\Upsilon_\varepsilon))}{du}\cdot\frac{du(\Upsilon_\varepsilon)}{d\varepsilon}\Big|_{\varepsilon=0}.$$

The derivative of u with respect to ε denoted by $\dot{u} = \frac{du_\varepsilon}{d\varepsilon}|_{\varepsilon=0}$ is difficult to calculate in general because there exists a problem for each direction of change of Υ. There exists a technique which is used to avoid calculating $\dot{u}$. Considering the domain Υ_i as previously defined. Let be the Lagrangian functional given by

$$\mathcal{L}(\Upsilon_{i,\varepsilon}; \lambda, v_\varepsilon) = \mathcal{J}(v_\varepsilon) + Re\left[\mathfrak{a}(\Upsilon_{i,\varepsilon}; \lambda, v_\varepsilon) - \mathfrak{l}(\lambda)\right],$$

where $\mathrm{Re}[z]$ denotes the real part of a complex number z and $\lambda \in W$ is a Lagrange multiplier. Note that the functional $\mathcal{J}$ of u_ε is equal to the Lagrangian evaluated in $v_\varepsilon = u_\varepsilon$, i.e.,

$$\mathcal{J}(u_{i,\varepsilon}) = \mathcal{L}(\Upsilon_{i,\varepsilon}; \lambda, u_\varepsilon) \quad \forall\lambda.$$

Thus, using the above observation, the Shape Derivative can be calculated as follows [34]

$$D\mathcal{J}(\Omega\backslash\Upsilon_i)\cdot\mathcal{V} = \frac{d}{d\varepsilon}\mathcal{L}(\Upsilon_{i,\varepsilon};\lambda,u_\varepsilon)\Big|_{\varepsilon=0} \quad \forall\lambda.$$

Since the bilinear form $\mathfrak{a}(\Upsilon_{i,\varepsilon};\lambda,u_\varepsilon)$ is linear on u and $\frac{d}{d\varepsilon}\,\mathrm{Re}[f(\varepsilon)] = \frac{d}{d\varepsilon}f(\varepsilon)$ for real value functions f, using the chain rule the derivative of the Lagrangian with respect ε is as follows

$$\begin{aligned}\frac{d}{d\varepsilon}\mathcal{L}(\Upsilon_{i,\varepsilon};\lambda,u_\varepsilon)\Big|_{\varepsilon=0} = \frac{d\mathcal{J}(u(\Upsilon_{i,\varepsilon}))}{du}\cdot\frac{du(\Upsilon_{i,\varepsilon})}{d\varepsilon}\Big|_{\varepsilon=0} + \frac{d}{d\varepsilon}\left(\mathfrak{a}(\Upsilon_{i,\varepsilon};\lambda,u)-\mathfrak{l}(\lambda)\right)\Big|_{\varepsilon=0}\\ + \mathfrak{a}\left(\Upsilon_{i,\varepsilon};\lambda,\frac{du(\Upsilon_{i,\varepsilon})}{d\varepsilon}\Big|_{\varepsilon=0}\right).\end{aligned} \tag{4.13}$$

Let w be

$$w = \frac{du(\Upsilon_{i,\varepsilon})}{d\varepsilon}\Big|_{\varepsilon=0},$$

the Lagrange multiplier $\lambda\in U$ is chosen such that the addition of the first and third terms on the right side of (4.13) is equal to zero

$$\mathfrak{a}^*(\Upsilon_{i,\varepsilon};w,\lambda) + \frac{dJ(u(\Upsilon_{i,\varepsilon}))}{du}\cdot w = 0 \quad \forall w\in W,$$

where $\mathfrak{a}^*(\Upsilon_{i,\varepsilon};w,\lambda) = \mathfrak{a}(\Upsilon_{i,\varepsilon};\lambda,w)\ \forall w\in W$. Doing this, $\dot{u}$ does not need to be calculated. The last equation is the adjoint equation. Thus, it is needed to obtain the solutions of the forward and adjoint problem to compute the shape derivative. Assuming that $\Upsilon\cap\Gamma=\emptyset$, it holds

$$\frac{d}{d\varepsilon}\mathfrak{l}(\lambda)\Big|_{\varepsilon=0} = 0,$$

since the linear form $\mathfrak{a}(\lambda)$ does not depend on the region $\Upsilon_{i,\varepsilon}$, actually $q=0$ on $\partial\Upsilon_{i,\varepsilon}$, which will be denoted by $\partial\Upsilon_{i,\varepsilon}$, therefore $\mathfrak{l}(w)=0$. Besides, it is needed to compute the second term of the right hand of (4.13). Let λ be the solution of the following adjoint equation of the RTE (3.10) denoted as z. Let $z\in W$ and $\xi\in Z$, the adjoint equation of

the RTE is given by

$$\frac{\partial z}{\partial t} - \theta \cdot \nabla z + (\mu(x) + b(x))z(x,\theta,t) - b(x)\int_{S^1} \eta(\theta \cdot \theta')z(x,\theta',t)\,d\theta' = 0 \text{ in } X \tag{4.14a}$$

with initial condition

$$z(x,\theta,T) = 0 \quad \text{on } \Omega \times S^1, \tag{4.14b}$$

and boundary condition

$$z(x,\theta,t) = \xi(x,t) \quad \text{on } \partial X_+, \tag{4.14c}$$

and boundary conditions on $\partial\Upsilon_i$

$$z^- = z^+, \tag{4.14d}$$

$$\mu_e \frac{\partial z^-}{\partial \bar{n}} = \mu_i \frac{\partial z^+}{\partial \bar{n}}, \tag{4.14e}$$

Where $\xi(x,t)$ as in (3.17) is applied uniformly in all θ directions with $\theta \cdot \nu > 0$, and z^- (z^+) denotes the limit of u from the interior (exterior) of the region D, and μ is defined as in (4.3). Thus, the shape derivative for diffusive optical tomography based on the RTE is given by

$$D\mathcal{J}(\Omega \backslash \Upsilon_{i,\varepsilon}) \cdot \mathcal{V} = \frac{d}{d\varepsilon}\left(\mathfrak{a}(\Upsilon_{i,\varepsilon}; z, u)\right)\Big|_{\varepsilon=0}. \tag{4.15}$$

Using the bilinear form ((4.9), for $z \in W$, one obtains

$$\begin{aligned}
\mathfrak{a}(\Upsilon_{i,\varepsilon}, u, z) = &\int_{\Upsilon_{e,\varepsilon}} \int_0^T \int_{S^1} \Big[-\frac{\partial z}{\partial t} - \theta \cdot \nabla z + (\mu_e + b)w - b\int_{S^1} \eta(\theta \cdot \theta')z\Big] u\, d\theta dt ds \\
&+ \int_{\partial\Upsilon_{i,\varepsilon}} \int_0^T \int_{S^1} z^+ u^+ (\theta \cdot \bar{\nu})_+ d\theta dt d(\partial\Upsilon_{i,\varepsilon}) \\
&+ \int_{\Upsilon_{i,\varepsilon}} \int_0^T \int_{S^1} \Big[-\frac{\partial z}{\partial t} - \theta \cdot \nabla w + (\mu_i + b)z - b\int_{S^1} \eta(\theta \cdot \theta')z\Big] u\, d\theta dt ds \\
&+ \int_{\partial\Upsilon_{i,\varepsilon}} \int_0^T \int_{S^1} z^- u^- (\theta \cdot \bar{\nu})_+\, d\theta dt d(\partial\Upsilon_{i,\varepsilon}).
\end{aligned}$$

Since z solves the equation (4.14), the integrals of the first and third terms of the above expression are equal to zero. Finally, since $u^+(u_-)$ and $z^+(z_-)$ satisfy the expressions (4.2d) and (4.14d) on the boundary $\partial\Upsilon_{i,\varepsilon}$ respectively, the resultant expression of the bilinear form is given by

$$\mathfrak{a}(\Upsilon_{i,\varepsilon}, u, z) = (\mu_i - \mu_e) \int_{\partial\Upsilon_{i,\varepsilon}} \int_0^T \int_{S^1} zu(\theta \cdot \bar{\nu})_+ \, d\theta dt d(\partial\Upsilon_{i,\varepsilon}). \tag{4.16}$$

As in [3, 26, 90], substituting the bilinear and linear form for the RTE (4.9) and (4.10) respectively, the shape derivative simplifies to

$$D\mathcal{J}(\Omega\backslash\Upsilon_i) \cdot \mathcal{V} = \frac{d}{d\varepsilon} \left(\int_{\partial\Upsilon_{i,\varepsilon}} \left(\int_0^T \int_{S_{n-1}} (\mu_i - \mu_e) \, zu(\theta \cdot \bar{n})_+ \, d\theta dt \right) d(\partial\Upsilon_{i,\varepsilon}) \right) \Bigg|_{\varepsilon \to 0}, \tag{4.17}$$

In order to calculate the derivative of this term, some equalities are given, which are basically relations between the surface, volume and gradient elements of the configuration Υ_ε and the corresponding elements in the original configuration Υ. The following relations hold [68]

$$\begin{aligned} \frac{\partial}{\partial\varepsilon} u &= (\nabla \mathfrak{P}_\varepsilon)^{-T} \nabla u, \\ d\Omega_\varepsilon &= \det \, d\Omega, \\ d\Gamma_\varepsilon &= \det \nabla \mathfrak{P}_\varepsilon \left\| \nabla \mathfrak{P}_\varepsilon^{-T} \, \bar{n} \right\| \, d\Gamma. \end{aligned}$$

Using the relations above, the following equalities hold which their proofs can be found in [49, 88, 147].

$$\frac{d}{d\varepsilon} \nabla \mathfrak{P}_\varepsilon \Big|_{\varepsilon=0} = \nabla \mathcal{V}, \tag{4.18}$$

$$\frac{d}{d\varepsilon} (\partial_{x_\varepsilon} u) \Big|_{\varepsilon=0} = -\nabla \mathcal{V}^T \nabla u, \tag{4.19}$$

$$\frac{d}{d\varepsilon} (d\Omega_\varepsilon) \Big|_{\varepsilon=0} = \operatorname{div} d\Omega, \tag{4.20}$$

$$\frac{d}{d\varepsilon} (d\Gamma_\varepsilon) \Big|_{\varepsilon=0} = \operatorname{div}_\Gamma \mathcal{V} d\Gamma, \tag{4.21}$$

where

$$\operatorname{div}_\Gamma \mathcal{V} = \operatorname{div} \mathcal{V} - \bar{n} \cdot (\nabla \mathcal{V}) \bar{n}. \tag{4.22}$$

This last operator is called the tangential divergence of the vector field $\mathcal{V}$. Using the last relation in (4.18) in equation (4.17), it is obtained

$$D\mathcal{J}(\Omega\backslash\Upsilon_i)\cdot\mathcal{V}=\int_{\partial\Upsilon_i}\left(\int_0^T\int_{S_{n-1}}(\mu_i-\mu_e)\,zu(\theta\cdot\bar{n})_+\,d\theta dt\right)\mathrm{div}_{\partial\Upsilon_i}\,\mathcal{V}\,ds,$$

Given a general functional $\mathcal{F}$ defined by

$$\mathcal{F}(\Omega)=\int_\Gamma f d\Gamma,$$

where Γ is the boundary of Ω as before. There is a general result which states that if f is a sufficiently smooth function defined on X such that the traces exist and are integrable on Γ, then, the Shape Derivative of the functional $\mathcal{J}$ caused by a perturbation velocity $\mathcal{V}$ is given by [49, 147]

$$D\mathcal{F}(\Omega)\cdot\mathcal{V}=\int_{\partial\Upsilon}\left(\frac{\partial f}{\partial\bar{n}}\cdot\mathcal{V}+f\,\mathrm{div}_{\partial\Upsilon}\,\bar{n}\right)ds. \tag{4.23}$$

As stated in the section 4.1, if $v\in U$, the map traces $\gamma_\pm: U\to L^1(\partial X_\pm,|\theta\cdot\bar{\nu}|\,d\theta dtd\Omega)$ defined by $\gamma_\pm(v):=v|_{\partial X_\pm}$ exist, and are continuous and subjective. [5, 35, 36, 39, 45]. Therefore, applying the above theorem, we obtain eq. (4.23). Moreover, expression (4.23) is equivalent to

$$D\mathcal{F}(\Omega)\cdot\mathcal{V}=\int_{\partial\Upsilon}\left(\frac{\partial f}{\partial\bar{n}}+\kappa f\right)\mathcal{V}\cdot\bar{n}ds, \tag{4.24}$$

where κ is the mean curvature of Γ defined by $\kappa=\mathrm{div}(\bar{N})$ and $\bar{N}$ is the unitary extension of the normal vector $\bar{n}$. The equivalence of eq. (4.23) and eq. (4.24) is proved in [49, 147] and given in [90]. Taking into account the expression in (4.17), let be the functional $\mathcal{J}$ given by

$$\mathcal{J}(\Omega\backslash\Upsilon_i)=\int_{\partial\Upsilon}\left(\int_0^T\int_{S_{n-1}}zu(\theta\cdot\bar{n})_+\,d\theta dt\right)d(\partial\Upsilon_i), \tag{4.25}$$

the trace map of the function f given by

$$f=\int_0^T\int_{S_{n-1}}(\mu_i-\mu_e)\,zu(\theta\cdot\bar{n})_+\,d\theta dt,$$

exists and is integrable since $u\in U$ and $z\in W$, therefore using formula (4.24) for the expression found for the shape derivative in (4.15), with f given as above, and using the Lebesgue's Dominated Convergence Theorem, it is obtained that the Shape Derivative is

given by

$$D\mathcal{J}(\Omega\backslash\Upsilon_i)\cdot\mathcal{V}=\int_{\partial\Upsilon}\int_0^T\int_{S_{n-1}}\left(\frac{\partial \mathfrak{f}}{\partial\bar{n}}+\kappa\mathfrak{f}\right)\mathcal{V}\cdot\bar{n}\,d\theta dtd(\partial\Upsilon_i), \tag{4.26}$$

where

$$\mathfrak{f}:=(\mu_i-\mu_e)\,zu(\theta\cdot\bar{n})_+. \tag{4.27}$$

4.2 Level set method: Front tracking by a Hamilton-Jacobi type equation

The material presented in this section is standard and can be found in [126, 146]. We will need these concepts in the following section where we incorporate shape derivatives in the level set evolution for shape reconstruction.

4.2.1 Eulerian and Lagrange formulations

In this subsection, the "Eulerian" and "Lagrange" formulations of the motion of a hypersurface $\Gamma(t)$ along its normal vector with speed V are described. In the Eulerian formulation the underlying coordinate system remains fixed. Consider an N dimensional function $\phi(x,t):\mathbb{R}^N\to\mathbb{R}$ and let $\Gamma(t)$ be the zero level set of ϕ, that is $\Gamma(t=0)=\{x\mid\phi(x,t=0)=0\}$. Then the hypersurface $\Gamma(t)$ is an $n-1$ dimensional function, in other words, it has co-dimension one. This hypersurface is known as the interface of ϕ, which is a closed curve or surface which separates domains of $\mathbb{R}^{2,3}$ with nonzero areas or volumes in $\mathbb{R}^{2,3}$, respectively [146]. Let $\phi(x,t=0)=\pm d$ where d is the distance from $x\in\mathbb{R}^N$ to $\Gamma(t=0)$, and the plus (minus) sign is chosen if the point x is outside (inside) the hypersurface $\Gamma(t=0)$. Then, the zero level set function is obtained for $d=0$. Let $|\nabla\phi|:=\sqrt{\phi_x^2+\phi_y^2}$ be the norm of the gradient of ϕ. The normal vector $\bar{n}$ to the hypersurface $\Gamma(t=0)$ is defined as $\bar{n}:=\nabla\phi/|\nabla\phi|$ such that $|\nabla\phi|\neq 0$. We have

$$x_t\cdot\hat{n}=V, \tag{4.28}$$

since the function $\Gamma(t)$ moves along its normal direction with speed V. Given $\phi(x, t = 0)$, the function $\phi(x, t)$ satisfies

$$\phi(x(t), t) = 0 \tag{4.29}$$

along its zero level set, that is, the zero level set of the function $\phi(x, t)$ coincides with the interface $\Gamma(t)$ at an arbitrary time t. For illustrating this situation, figure 4.2 shows a cartoon with the following scenarios. Top left: an initial circle is shown at time zero, together with the circle at a later time in bottom left. The top right shows the associated initial level set function ϕ at time zero. The bottom images show the associated level set function at two different times. At any arbitrary time, the actual interface is given by the zero level set of the level set function ϕ.

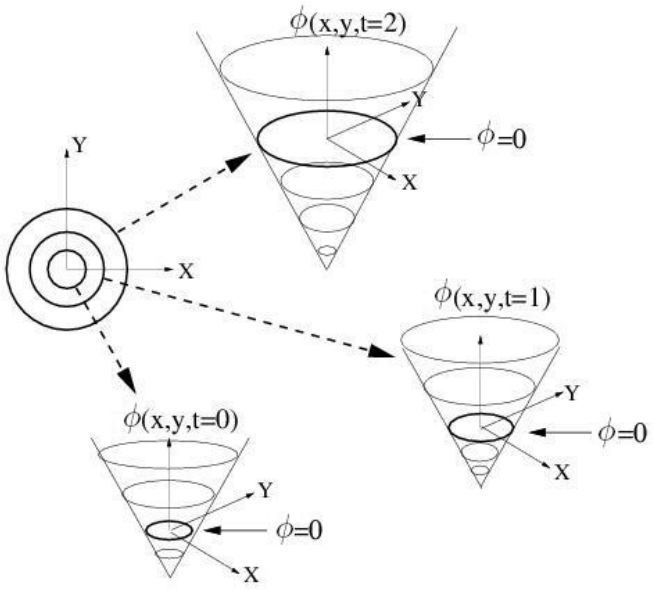


Figure 4.2: Level set function.

After applying the chain rule to (4.29), we get

$$\phi_t + \nabla\phi(x(t), t) \cdot x_t(t) = 0. \tag{4.30}$$

Using (4.28), we obtain the initial value evolution equation for ϕ, namely

$$\phi_t + V\,|\nabla\phi| = 0. \tag{4.31a}$$

$$\text{given} \quad \phi(x, t = 0), \tag{4.31b}$$

This is known as the level set equation of ϕ or the Eulerian representation of the level set [126]. This equation is also called a convection (or advection) equation or the *G*-equation in the framework of combustion processes [146]. For certain forms of the speed function V, a standard Hamilton-Jacobi equation is obtained. The hypersurface Γ can change topology, break and merge. Since $\bar{n}$ and $\nabla\phi$ point into the same direction, $\nabla\phi \cdot \bar{t} = 0$ for any vector $\bar{t}$ which is tangent to the curve $\Gamma(t)$, implying that the tangential velocity components vanish when they are substituted into the level set equation (4.31). The Eulerian formulation corresponds to the level set method which is used in this dissertation.

The Lagrangian formulation of the propagation of the front Γ is as follows [126, 146]. Let $\Gamma(t)$ be a smooth, closed curve in $\mathbb{R}^{2,3}$ which moves under a vector field $\mathcal{V}$, and let x be a point which belongs to the curve Γ as it is shown in the figure 4.3. Then, the motion of the curve Γ is described by the following ordinary differential equation

$$\frac{dx}{dt} = \mathcal{V}, \tag{4.32}$$

which is known as the Lagrangian formulation of the interface motion $\Gamma(t)$. Therefore, equation (4.30) can be written as

$$\phi_t + \nabla\phi(x(t), t) \cdot \mathcal{V} = 0. \tag{4.33}$$

If $\mathcal{V} := V\bar{n} + W\bar{t}$ is plugged into the above equation, where $\bar{t}$ is a tangential vector of ϕ and W is the tangential component, one obtains eq. (4.31), since $\nabla\phi \cdot \bar{t} = 0$ for any tangential vector $\bar{t}$.

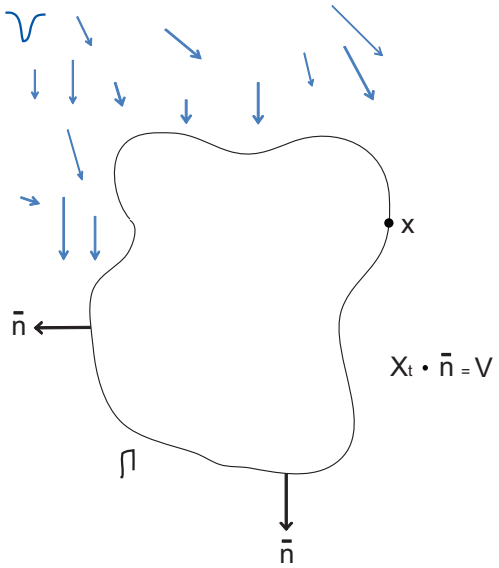


Figure 4.3: Lagrange formulation of the level set equation.

In the Lagrangian formulation, the interface evolution is described explicitly by the points x, whereas in the Eulerian formulation the evolution of the interface is described implicitly by the evolution of a function ϕ.

4.2.2 Motion under an external velocity field

Central differences can also be used for spatial derivatives in (3.82) instead of the upwind differencing. This scheme is not stable if one uses the forward Euler time discretization and the standard CFL condition with $\Delta t \approx \Delta x$. However, it can achieve stability by using a smaller time step $\Delta t \approx (\Delta x)^2$ or by using the third-order accurate Runge-Kutta method or by adding a fictitious dissipation to the level set equation (4.31) [126]. Upwinding with the forward Euler time discretization is typically used instead of using central differences with these three options mentioned above.

Since the level set equation (4.31) belongs to the following general type of Hamilton-Jacobi equations,

$$\phi_t + \mathcal{H}(\nabla\phi) = 0, \tag{4.34}$$

where $\mathcal{H}$ is a function known as the Hamiltonian, $\mathcal{H}(\nabla\phi) = V|\nabla\phi|$, which in general depends on ϕ_x, ϕ_y in $2D$, numerical schemes such as Lax-Friedrichs, Roe-Fix, Godunov, Hamilton-Jacobi ENO and WENO can be applied to solve (4.31). For more details of these schemes, see [126, 146].

The upwind difference scheme for (4.31) in $2D$ discussed in section 3.6 is the following. If the speed function V is convex, the level set equation (4.31) in $2D$ can be approximated by the following first order convex scheme

$$\phi_{ij}^{n+1} = \phi_{ij}^{n} - \Delta t[\max(V_{ij},0)\nabla^{+} + \min(V_{ij},0)\nabla^{-}], \tag{4.35}$$

where

$$\nabla^{+} = \max(D_{ij}^{-x},0)^2 + \min(D_{ij}^{+x},0)^2 + \max(D_{ij}^{-y},0)^2 + \min(D_{ij}^{+y},0)^2, \tag{4.36a}$$

$$\nabla^{-} = \max(D_{ij}^{+x},0)^2 + \min(D_{ij}^{-x},0)^2 + \max(D_{ij}^{+y},0)^2 + \min(D_{ij}^{-y},0)^2, \tag{4.36b}$$

and the subindices $(i,j) \in \mathbb{Z}^2$ in ϕ_{ij}^{n} stand for the spatial coordinates $(i\Delta x, j\Delta y)$, and n stands for the time variable $t_n = n\Delta t$, that is

$$\phi_{ij}^{n} = \phi(i\Delta x, j\Delta y, n\Delta t).$$

Here the notation $D_{ij}^{+x} := D^{+x}\phi_{i,j}^{n}$ (and so forth) has been used, and D^{+x} (and so forth) as introduced in section 3.6. If the speed function V is nonconvex, the Lax-Friedrichs' scheme can be used

$$\begin{aligned}\phi_{ij}^{n+1} = \phi_{ij}^{n} - \Delta t\Bigg[&\mathcal{H}\left(\frac{D_{ij}^{-x}+D_{ij}^{+x}}{2}, \frac{D_{ij}^{-y}+D_{ij}^{+y}}{2}\right) \\ &- \frac{1}{2}\alpha_v(D_{ij}^{+x}-D_{ij}^{-x}) - \frac{1}{2}\alpha_w(D_{ij}^{+y}-D_{ij}^{-y})\Bigg],\end{aligned} \tag{4.37}$$

where $\alpha_{\phi_x}, \alpha_{\phi_y}$ are bounds on the partial derivative of the Hamiltonian $\mathcal{H} = V|\nabla\phi|$, i.e.,

$$\alpha_{\phi_x} := \max|\mathcal{H}_{\phi_x}| = \max\left|\frac{V\phi_x}{|\nabla\phi|}\right| \qquad \alpha_{\phi_y} = \max|\mathcal{H}_{\phi_y}| = \max\left|\frac{V\phi_y}{|\nabla\phi|}\right|.$$

The maximum values are taken here on the Cartesian mesh. If ϕ is a distance function ($|\nabla\phi| = 1$), $\alpha_{\phi_x} = V\phi_x, \alpha_{\phi_y} = V\phi_y$.

Instead of using the first or second order finite differences directly to approximate the

normal vector $\hat{n}$, we can instead use the more accurate approximation given by

$$\begin{aligned} n_{ij}^{*} &= \frac{(\phi_x, \phi_y)}{(\phi_x^2 + \phi_y^2)^{1/2}} \\ &= \frac{(D_{ij}^{+x}, D_{ij}^{+y})}{((D_{ij}^{+x})^2 + (D_{ij}^{+y})^2)^{1/2}} + \frac{(D_{ij}^{-x}, D_{ij}^{+y})}{((D_{ij}^{-x})^2 + (D_{ij}^{+y})^2)^{1/2}} \\ &+ \frac{(D_{ij}^{+x}, D_{ij}^{-y})}{((D_{ij}^{+x})^2 + (D_{ij}^{-y})^2)^{1/2}} + \frac{(D_{ij}^{-x}, D_{ij}^{-y})}{((D_{ij}^{-x})^2 + (D_{ij}^{-y})^2)^{1/2}}, \end{aligned} \tag{4.38}$$

which averages these forward and backward finite differences in both directions x, y to avoid sudden changes on the direction of the normal vector $\hat{n}$ at the corners. The expression (4.38) is normalised to obtain the normal vector, that is, $n_{ij} = n_{ij}^{*}/|n_{ij}^{*}|$. It is noted that if one of the denominators of (4.38) is zero, this term is not taken into account in the calculation of the normal vector n_{ij}^{*}.

4.2.3 Motion Involving Mean Curvature

This brief subsection is included since the analytical expression of the shape derivative of DOT based on the RTE found in chapter 4 depends on the Mean Curvature of the interface Γ. In this subsection, the motion of the interface $\Gamma(t)$ under a scalar diffusive field $\mathcal{B}$ is described that depends on the curvature of the interface Γ. In other words, $\mathcal{B} = -B\kappa$, where $B \in \mathbb{R}$ and κ is the curvature of the interface $\Gamma(t)$ defined by

$$\kappa := \nabla \cdot \frac{\nabla \phi}{|\nabla \phi|} = \frac{\phi_{xx}\phi_y^2 - 2\phi_x\phi_y\phi_{xy} + \phi_{yy}\phi_x^2}{\left(\phi_x^2 + \phi_y^2\right)^{3/2}}. \tag{4.39}$$

Substituting $\mathcal{B} = -B\kappa$ instead of the vector field $\mathcal{V}$ into eq. (4.31), we have

$$\phi_t = B\kappa|\nabla\phi|, \tag{4.40}$$

If $B > 0$, the interface moves in the direction of concavity, so that circles decrease. On the other hand, if $B < 0$, the interface moves in the direction of convexity, so that the circles increase [126]. Upwind discretization can not be applied to the level set equation when it is under the motion of the curvature of the interface since $B\kappa\,|\nabla\phi|$ is a parabolic term. If ϕ is a signed distance function, that is, if $|\nabla\phi| = 1$, equation (4.30) becomes the

heat equation

$$\phi_t = B\Delta\phi, \tag{4.41}$$

which is a parabolic equation. In this case, difference schemes for the heat equation can be applied to (4.41). In general, parabolic equations are discretized using central differencing since where the domain of dependence includes information from all spatial directions. Therefore the discretization of eq. (4.41) is as follows

$$\phi_{ij}^{n+1} = \phi_{ij}^{n} - \Delta t B\kappa_{ij}^{n}((D_{ij}^{0x})^2 + (D_{ij}^{0y})^2)^{1/2}, \tag{4.42}$$

where D_{ij}^{0x}, D_{ij}^{0y} are defined as $D_{ij}^{0x} := D^{0x}\phi_{i,j}^{n}$, $D_{ij}^{0y} := D^{0y}\phi_{i,j}^{n}$ and κ_{ij}^{n} is the central difference approximation to the curvature definition given in (4.41). The CFL condition of (4.41) in $2D$ using central differencing for the spatial term $\Delta\phi$ combined with the forward Euler time discretization is

$$\Delta t\left(\frac{2B}{(\Delta x)^2} + \frac{2B}{(\Delta y)^2}\right) < 1.$$

4.2.4 General Motion

A motion of the region Ω can be mathematically governed by an advective term V which corresponds to the movement in the normal direction to the front, and a diffusive term $B\kappa$ which corresponds to a decreasing or increasing behaviour of the front. Thus, considering both, the advective and the diffusive terms, the level set equation is given by

$$\phi_t + V|\nabla\phi| - B\kappa|\nabla\phi| = 0. \tag{4.43}$$

The equation (4.43) is discretized considering firstly the advective part, which is discretized using the first order convex scheme given in (4.35), and secondly the diffusive part, which is discretized using the central differencing scheme (4.42). Thus, the first order numerical scheme in $2D$ for eq. (4.43) is written as

$$\phi_{ij}^{k+1} = \phi_{ij}^{k} + \Delta t\begin{bmatrix} -\{\max(V_{ij}^{k},0)\nabla^{+} + \min(V_{ij}^{k},0)\nabla^{-}\} \\ +B\kappa_{i,j}^{k}((D_{ij}^{0x})^2 + (D_{ij}^{0y})^2)^{1/2} \end{bmatrix}. \tag{4.44}$$

Equation (4.53) must satisfy the following combined CFL condition

$$\Delta t \left(\frac{|V|}{\Delta x} + \frac{|V|}{\Delta y} + \frac{2B}{(\Delta x)^2} + \frac{2B}{(\Delta y)^2} \right) < 1.$$

In our simulation we have used a variant of the above CFL condition given in [26, 73, 90]. The interval Δt must satisfy the following two inequalities. The first one is needed because of the advective part of the Hamilton-Jacobi-type equation, and the second one because of the diffusive part of the Hamilton-Jacobi-type equation. These inequalities are

$$\Delta t \left(\frac{\max\limits_{(x,y)\in\Omega} |V|}{\min(\Delta x, \Delta y)} \right) < \frac{1}{2\sqrt{2}}, \tag{4.45a}$$

$$\Delta t \left(\frac{|B|}{\min((\Delta x)^2, (\Delta y)^2)} \right) < 1. \tag{4.45b}$$

A narrowband technique that was explained in section 3.3 and used already in chapter 3, will be used in this chapter again. In addition to the narrowband technique we also use the velocity extension technique, which is described as follows. The advective term V in (4.43) of equation (4.31) is extended off the boundary Γ to the value V_{ext} in the whole domain Ω such that the extended value V_{ext} satisfies

$$V_{\text{ext}}\Big|_{\Gamma} = V\Big|_{\Gamma} \quad \text{and} \quad \nabla V_{\text{ext}} \cdot \nabla \phi = 0,$$

that is, V_{ext} is constant along the normal curves to Γ. To achieve this, the following differential equation is solved for the function q until a stationary state is reached [38, 73, 114, 129]

$$q_\tau + S(\phi) \frac{\nabla \phi}{|\nabla \phi|} \cdot \nabla q = 0 \text{ in } \Omega \times \mathbb{R}^+, \tag{4.46a}$$

$$q(x, y, 0) = p(x, y, t), \quad (x, y) \in \Omega. \tag{4.46b}$$

Here S is the approximate sign function defined by

$$S(w) = \frac{w}{\sqrt{w^2 + (|\nabla w|\Delta x)^2}}, \tag{4.47}$$

and $p(x, y, t)$ is the initial condition defined by

$$p(x, y, t) = \begin{cases} V & \text{if } (x, y) \in \Gamma, \\ 0 & \text{if } (x, y) \notin \Gamma. \end{cases}$$

Then, the extended velocity V_{ext} is defined by the stationary state of q

$$V_{\text{ext}} = \lim_{\tau \to \infty} q(x, y, \tau).$$

Equation (4.46) is solved using the following upwind scheme at each iteration n as given in (4.44)

$$\begin{aligned} q_{ij}^{m+1} = q_{ij}^{m} - \Delta\tau[&\max(S_{ij}(n_{ij}^{x}), 0)D^{-x}q_{ij}^{m} + \min(S_{ij}(n_{ij}^{x}, 0))D^{+x}q_{ij}^{m} \\ &+ \max(S_{ij}(n_{ij}^{y}), 0)D^{-y}q_{ij}^{m} + \min(S_{ij}(n_{ij}^{y}, 0))D^{-y}q_{ij}^{m}], \end{aligned} \tag{4.48}$$

where $S_{ij} = S(\phi_{ij}^{m})$ and $\Delta\tau$ is a time step such that $t^{m} = m\Delta\tau$. The initial value q^{0} is equal to V on the grid points with a distance to Γ less than or equal to $\min(\Delta x, \Delta y)$ and equal to zero elsewhere [74, 130].

The solution of the level set equation (4.43) must be reinitialised at periodical intervals of time t for numerical stability and accuracy because the level set function should not be too flat or too steep at any time. Here we enforce that the level set function ϕ remains a distance function [74, 114, 129, 146]. To achieve this, the following differential equation is solved at some intervals of time t until to a stationary state is reached

$$\varphi_{\tau} + S(\phi)(|\nabla\varphi| - 1) = 0 \text{ in } \Omega \times \mathbb{R}^{+}, \tag{4.49a}$$

$$\varphi(x, y, 0) = \phi(x, y, t), \quad (x, y) \in \Omega, \tag{4.49b}$$

where S is the signed function given in (4.47), such that $\phi(x, y, t) = \lim_{\tau \to \infty} \varphi(x, y, \tau)$. The reinitialisation equation (4.49) is solved using the following upwind scheme (see [130])

$$\begin{aligned} \varphi_{ij}^{k+1} = \varphi_{ij}^{k} &- \frac{\Delta\tau}{\Delta x} S_{ij}^{+} \left(\sqrt{(a^{+})^{2} + (b^{-})^{2} + (c^{+})^{2} + (d^{-})^{2}} - 1 \right) \\ &- \frac{\Delta\tau}{\Delta x} S_{ij}^{-} \left(\sqrt{(a^{-})^{2} + (b^{+})^{2} + (c^{-})^{2} + (d^{+})^{2}} - 1 \right), \end{aligned}$$

where a, b, b, d are defined as follows

$$a = D^{-x}\varphi_{ij}^k \qquad b = D^{+x}\varphi_{ij}^k \tag{4.50}$$

$$c = D^{-y}\varphi_{ij}^k \qquad d = D^{+y}\varphi_{ij}^k \tag{4.51}$$

and S_{ij} stands for $S(\phi_{ij}^k)$ and $(x)^+ = \max(x, 0)$, $(x)^- = \min(x, 0)$. Alternatively, we can also use the Godunov scheme given by

$$\begin{aligned}\varphi_{ij}^{k+1} = \varphi_{ij}^k - \frac{\Delta\tau}{\Delta x} S_{ij}^+ \left(\sqrt{\max[(a^+)^2, (b^-)^2] + \min[c^+)^2, (d^-)^2]} - 1\right) \\ - \frac{\Delta\tau}{\Delta x} S_{ij}^- \left(\sqrt{\max[(a^-)^2, (b^+)^2] + \min[c^-)^2, (d^+)^2]} - 1\right).\end{aligned} \tag{4.52}$$

where a, b, c, d as defined in (4.50). We have used the Godunov scheme (4.52) for re-initializing the level set function ϕ in our numerical experiments.

We summarize by providing the outline of the algorithm to solving eq. (4.43) in Pseudo Code form.

- Initialise the level set function ϕ to be a signed distance function on Γ.
- Reinitialise ϕ using (4.52) at some specific intervals of time t.
- Compute the extension velocity V_{ext} from V using (4.48).
- Advance ϕ in time t using (4.44).

We point out that we have only reinitialised the level set function ϕ in situations where we have extended the velocity V to the whole region Ω. We have used a narrowband technique described in section 3.3 in our simulations.

4.3 Incorporating the shape derivative into the level set method

In practical implementations, the interface $\partial\Upsilon_i$ is replaced by a narrow transition zone with smoothly varying parameters. We will use the mathematical expression of the general motion of a region described in subsection 4.2.4. The motion of the domain Υ_i under

a normal velocity speed V and a mean curvature term $B\kappa$ is considered, which mathematically is expressed by the evolution of the level set equation

$$\phi_t + V|\nabla\phi| - B\kappa|\nabla\phi| = 0, \tag{4.53}$$

where

$$V = -\int_0^T \int_{S_{n-1}} \frac{\partial \mathfrak{f}}{\partial \bar{n}}\, d\theta dt, \tag{4.54}$$

and

$$B = \int_0^T \int_{S_{n-1}} \mathfrak{f}\, d\theta dt. \tag{4.55}$$

Here $\mathfrak{f}$ is defined as in (4.27). Equation (4.53) is known as the advection-diffusion equation, where the advective part is linked with the hyperbolic equation

$$\phi_t + V|\nabla\phi| = 0, \tag{4.56}$$

and the diffusive part is linked with the parabolic equation

$$\phi_t = B\kappa\Delta\phi, \tag{4.57}$$

where it has been assumed that the level set function ϕ is a signed distance function, i.e., $|\nabla\phi| = 1$. We keep this assumption in the following (which is justified as long as we re-initialize sufficiently often).

In chapter 3 we used the level set method following Santosa's approach, which is deduced from calculus of variations. In contrast to that, in this chapter we have considered the level set approach where the evolution follows a Hamilton-Jacobi-type equation (4.31) driven by shape derivatives. In particular, in this chapter the advective part V of the level set evolution equation (4.31) is based on our results of section 3, that is,

$$V = \frac{\partial}{\partial \bar{n}}\Big(-(\mu_i - \mu_e)(\theta \cdot \nu)_- \nabla J_j(\mu)\Big).$$

In contrast to that, in chapter 3 the "advective" part was given as

$$V = -(\mu_i - \mu_e)\nabla J_j(\mu).$$

Thirdly, the level set eq. (4.53) now contains a curvature term $B\kappa$, whereas when using the level set approach in chapter 3 we did not use such a curvature term. We have compared both Level Set approaches in our numerical results further below.

Some further numerical details.

Equation (4.53) is solved numerically by using the first order convex speed method given in (4.44), or the combination of the nonconvex method given in (4.37) and the numerical method for the evolution of ϕ involving a mean curvature as given in (4.42).

The term $\partial_n \mathfrak{f}$ in (4.54) is calculated as follows.

$$\frac{\partial \mathfrak{f}}{\partial \bar{n}} = \nabla \mathfrak{f} \cdot \bar{n} = [\nabla u z(\theta \cdot \bar{n})_+ + u \nabla z(\theta \cdot \bar{n})_+ + uz \nabla \left((\theta \cdot \bar{n})_+\right)] \cdot \bar{n}.$$

All the derivatives for the gradient terms of the above expression are calculated using central differences. The normal vector is calculated using (4.38).

We again have combined the Landweber-Kaczmarz technique with the level set evolution presented in this chapter for the reconstruction of the absorption coefficient.

The following Pseudo-Code outlines the basic algorithm for shape reconstruction of the absorption coefficient in DOT using the results of this section.

Algorithm 7: Shape reconstruction using the LS method with shape derivative

Input: $\Upsilon_0, \mu^0, \phi^0$ Initial guess

Output: $\Upsilon_{I(M-1)}, \mu^{I(M-1)}, \phi^{I(M-1)}$

for $k := 0, \ldots, I(M-1)$ **do**

Compute V and B given by (4.54) and (4.55) respectively;

Compute ϕ^{k+1} given by (4.44);

Update μ^{k+1} using (3.57);

end

Here $u_{[k]}$ solves (2.1) with μ^k and source $q_{[k]}$, and $z_{[k]}$ solves (3.10) with μ^k and adjoint source $\xi_{[k]}$ given by (3.52) and κ given by (4.39).

Next, we provide a Pseudo-Code of the algorithm that corresponds to the same situation as above but with an extension of the velocity V and corresponding re-initialisation of the level set function ϕ to remain a distance function.

Algorithm 8: Shape reconstruction using the LS method with shape derivative and extended V.

Input: $\Upsilon_0, \mu^0, \varphi^0$

Output: $\Upsilon_{I(M-1)}, \mu^{I(M-1)}, \phi^{I(M-1)}$

for $k := 0, \ldots, I(M-1)$ **do**

Compute V and B given by (4.54) and (4.55) respectively;

Compute the extension V_{ext} in Ω, using the scheme (4.48);

Reinitialisation of ϕ^{k+1} at periodical intervals of time using the scheme (4.52);

Compute ϕ^{k+1} given by (4.44);

Update μ^{k+1} using (3.57);

end

4.4 Numerical Results

We have performed shape reconstructions regarding the absorption coefficient using the shape derivative presented in section 4.1 via the level set technique. Here we have used the same numerical experimental setup as in section 3.7, using the values of the Table 3.1, except that we have suppressed the clear layer near to the boundary only in this section for the purpose of checking our new shape reconstruction technique. The following figures show the results of our numerical experiments.

To start with, Figures 4.4 and 4.5 show the shape reconstruction for the coefficient μ in DOT using algorithm 7 for one inclusion and two inclusions, respectively.

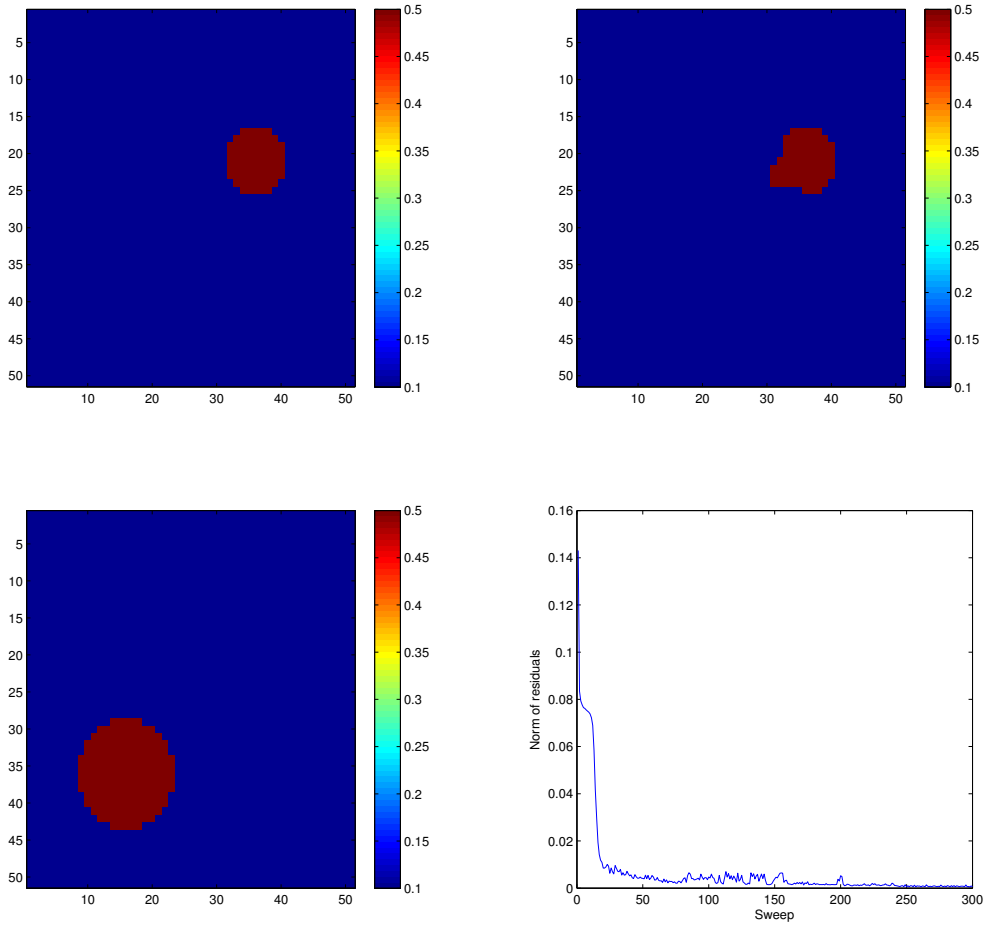


Figure 4.4: Shape reconstruction for μ using algorithm 7 for one obstacle. Top left: true object; top right: final reconstruction. Bottom left: initial guess; bottom right: norm of residuals of the cost.

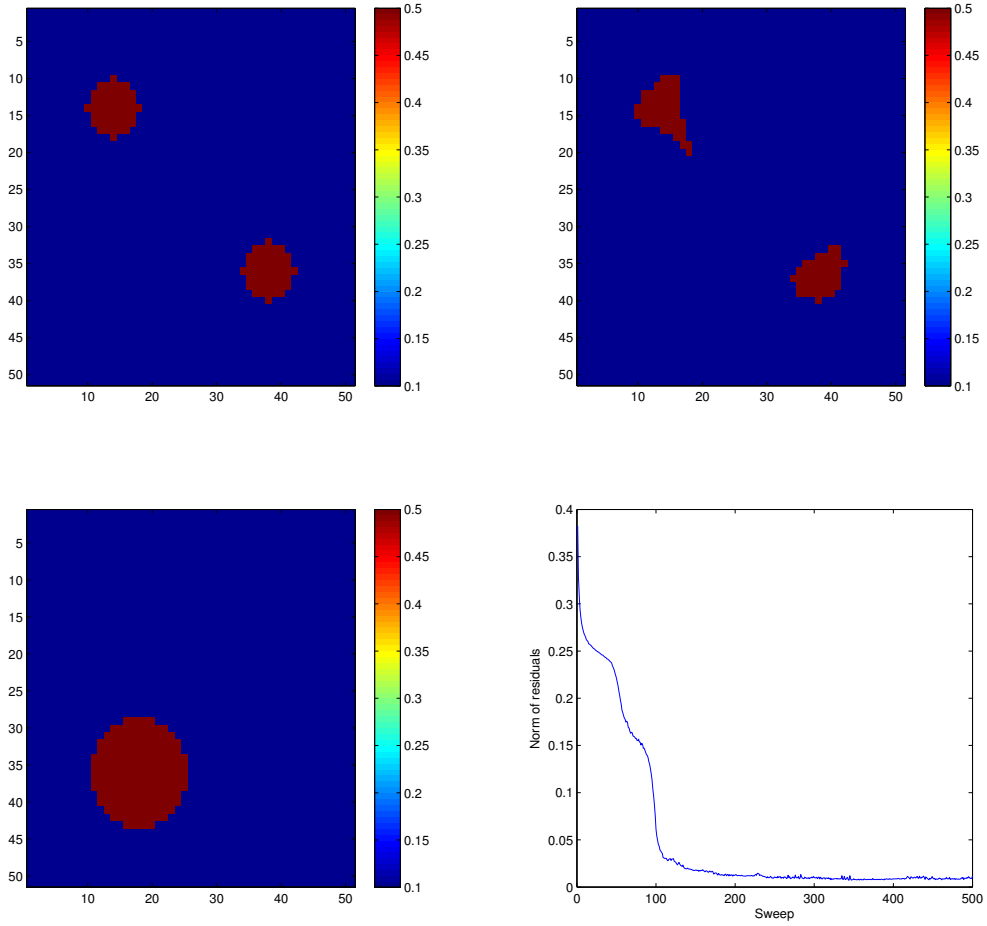


Figure 4.5: Shape reconstruction for μ using algorithm 7 for two obstacles. Top left: true object; top right: final reconstruction. Bottom left: initial guess; bottom right: norm of residuals of the cost.

Figures 4.6 and 4.7, on the other hand, show the shape reconstruction for the coefficient μ in DOT using algorithm 8 for one inclusion and two inclusions, respectively. This means, in these reconstructions we have used the extension velocity and a repeated re-initialization technique.

In both Algorithms 7 and 8, we have used a "signed distance function" as our initial guess for the level set function to start the shape reconstruction process. We observe that the results of the shape reconstruction schemes for the absorption coefficient in DOT based on the RTE using the shape derivative compare favourably with those obtained in [55, 56], where the authors used the Landweber-Kaczmarz method as it can be seen in figure 4.8.

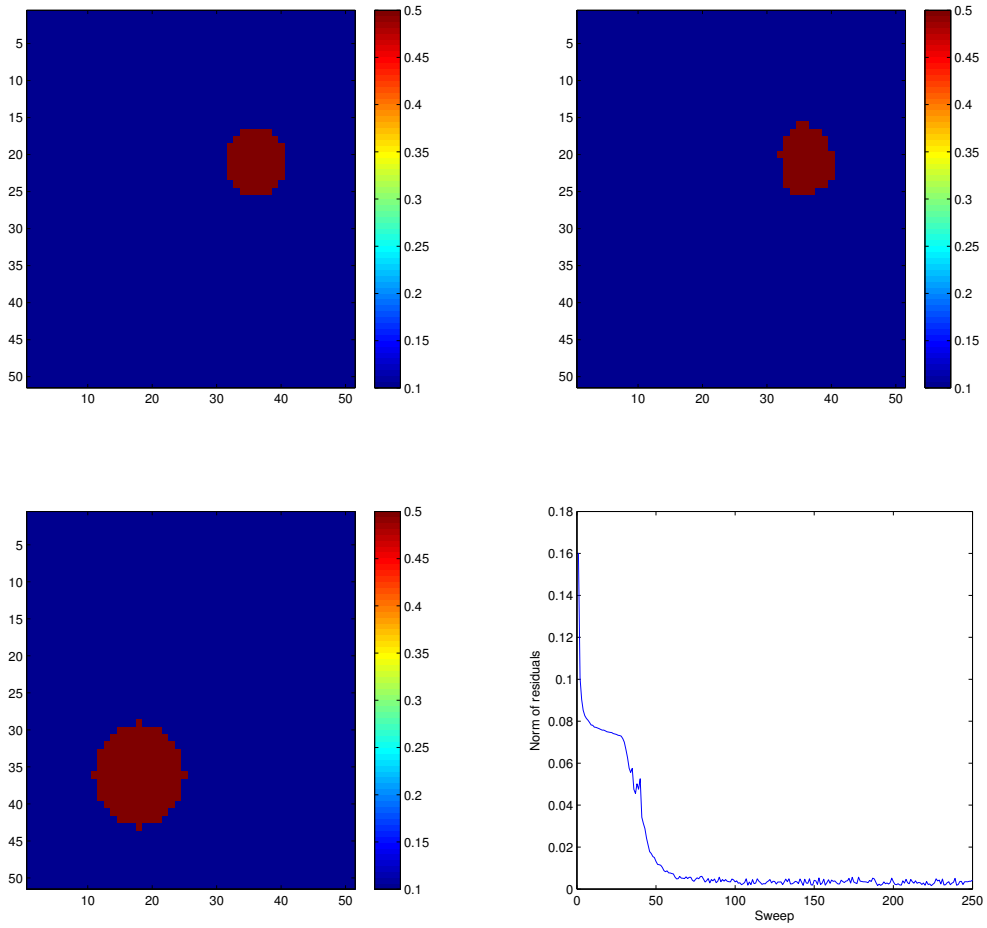


Figure 4.6: Shape reconstruction for μ using algorithm 8 with V_{ext} for one obstacle. Top left: true object; top right: final reconstruction. Bottom left: initial guess; bottom right: norm of residuals of the cost.

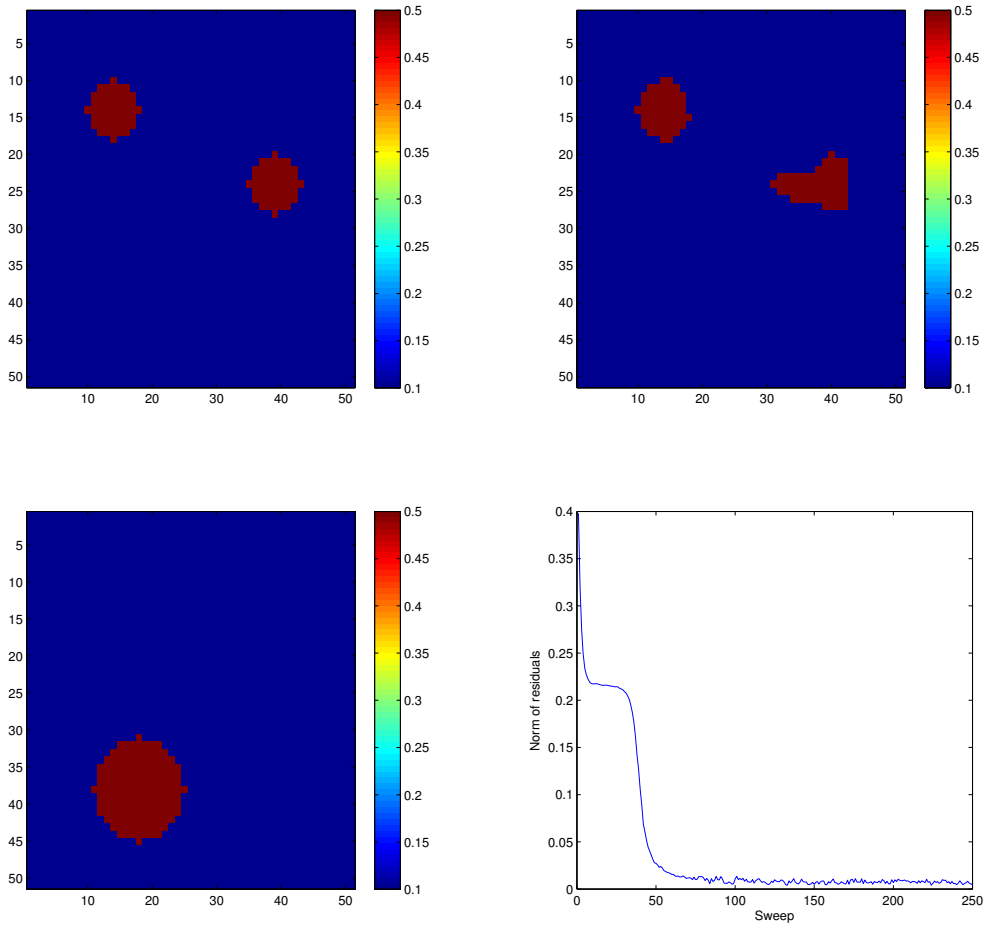


Figure 4.7: Shape reconstruction for μ using algorithm 8 with V_{ext} for two obstacles. Top left: true object; top right: final reconstruction. Bottom left: initial guess; bottom right: norm of residuals of the cost.

Figure 4.8 shows the comparison between the reconstruction using the shape derivative derived in this chapter versus the technique discussed in the previous chapter 3 with algorithm 4 using an optimization based scheme for calculating shape sensitivities. More specifically, figure 4.8 shows the evolution of the residual norm over iteration index for both of these different algorithms.

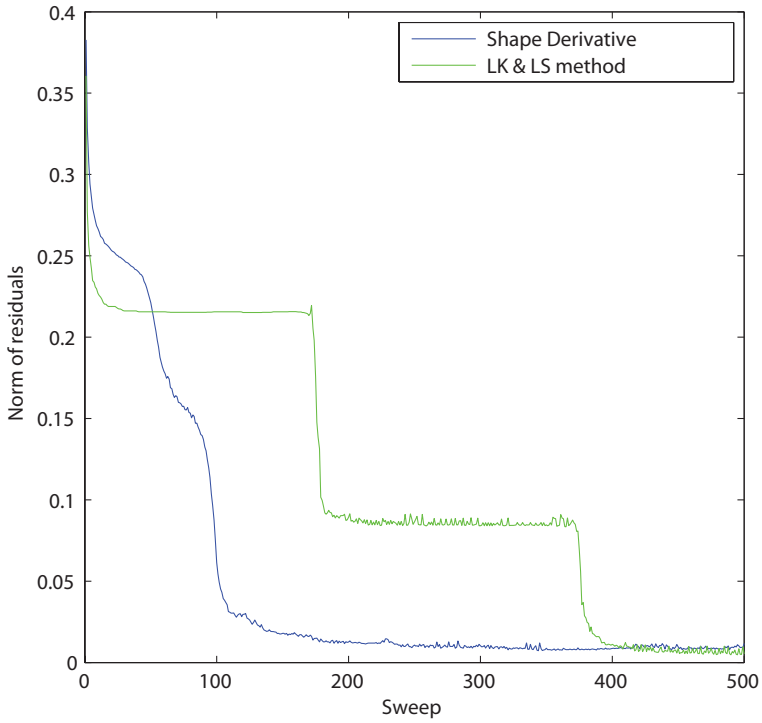


Figure 4.8: Norm of residuals of the misfit functional of the μ absorption reconstruction using the shape derivative and the LK method.

Chapter 5

Sparsity regularization

In this chapter we will apply the concept of sparsity regularization to the DOT problem. First, the basic concept of sparsity regularization is briefly discussed. In particular, we define and discuss compressibility and the so-called Bregman distance. Then the linearised Bregman iteration method is presented. In section 5.3, expressions for the sparsity reconstruction in a linear inverse problem are discussed. Our DOT inverse problem is nonlinear, so in section 5.4, the corresponding newly developed algorithms of sparsity reconstruction as used in this thesis are presented. Finally, in the last section the results of numerical experiments are discussed which illustrate the performance of these new numerical schemes in some situations.

5.1 Basic concepts of sparsity and compressive sensing

The inverse problem in DOT is underdetermined. The main difficulty in the inverse problem in DOT is its instability which requires some form of regularization. In this chapter we present a novel method for recovering the shape and texture of μ using Sparsity Regularization. To start with, we will give some definitions and notations. We will first focus on sparsity in linear problems, which is closely related to Compressive Sensing. Later on we will also discuss some concepts for nonlinear problems. For more details we refer to [46, 72] and the references therein.

We define the ℓ_p-norm of a vector $x \in \mathbb{R}^N$ for $0 < p \leq \infty$ as

Definition 16

$$\|x\|_p = \begin{cases} \left(\sum_{j=1}^{N} |x_j|^p \right)^{1/p} & \text{if } 0 \leq p < \infty, \\ \max_{j \in 1,\ldots,N} |x_j| & \text{if } p = \infty. \end{cases}$$

Then the standard inner product in $\mathbb{R}^N$ can be defined by

$$\langle x, z \rangle = z^T x = \sum_{j=1}^{N} x_j z_j,$$

which induces the ℓ_2-norm, $\|x\|_2 = \sqrt{\langle x, x \rangle}$. The ℓ_p-norm is indeed a norm for $1 \leq p \leq \infty$, while for $0 < p < 1$ it is a 'quasi-norm'. For $p = 0$ the often called 'l_0-norm' is defined by

$$\|x\|_0 = |\text{supp}(x)|,$$

where $supp(x) := \{j : x_j \neq 0\}$ is the support of a vector x, and $|supp(x)|$ denotes the cardinality of $supp(x)$.

Notice that the ℓ_0-norm is in fact

$$\lim_{p \to 0} \|x\|_p^p = |\text{supp}(x)|,$$

and is not even a quasinorm. Fig. shows the unit sphere, i.e., $\{x \in \mathbb{R}^2 : \|x\|_p = 1\}$ for different values of p.

We will now provide the definition of k-sparse which plays a central role in sparsity or compressive sensing.

Definition 17 *A vector $x \in \mathbb{R}^N$ is called* k-sparse *if $\|x\|_0 \leq k$. The set of k-sparse vectors is denoted by Σ_k for $k \in \{1, 2, ...N\}$.*

Next we define the concept of compressibility.

Definition 18 *A vector x is called* compressible *if the best k-term approximation of a vector x decreases fast in k, where the best k-term approximation error of a vector $x \in \mathbb{R}^N$ in ℓ_p is given by*

$$\sigma_k(x)_p = \inf_{z \in \Sigma_k} \|x - z\|_p.$$

The key idea in sparsity reconstruction is to reconstruct directly a compressed vector x, which (roughly) can be achieved by keeping only the k largest entries of that vector and

setting to zero the rest of its entries. This approximation problem is nonlinear because picking the largest entries depends of the vector itself. Image and signal compression techniques such as MP3, MPEG and JPEG are based on this process of keeping only the largest entries of a signal.

Another important feature to keep in mind when choosing the ℓ_p norm for reconstruction is illustrated in fig. 5.1. Suppose that we want to approximate a point $x \in \mathbb{R}^2$ by using a point in a $1D$ affine space A (representing a general linear relationship) by solving the following ℓ_p-minimisation problem

$$\hat{x} = \arg\min \|x - z\|_p \quad \text{s.t.} \quad Az = y. \tag{5.1}$$

Firstly, in order to obtain a sparse representation, scientists tried to solve problem (5.1) directly for the sparsest vector x, which corresponds to solving the l_0-optimisation problem

$$\min \|x\|_0 \text{ subject to } Ax = y.$$

However, this leads to an NP-hard problem in general. Thus, the next option was to try to solve the convex l_1-optimisation problem

$$\min \|x\|_1 \text{ subject to } Ax = y \tag{5.2}$$

instead. Different choices of p for the l_p-norm of the problem (5.1) will lead to different scenarios for obtaining the solution $\hat{x}$ of this problem. Fig. (5.1) shows a few different scenarios for $p = 1, 2, \infty, 1/2$. The closest point $\hat{x} \in A$ is found 'geometrically' by growing an ℓ_p sphere centred on $x = 0$ until it meets with the affine space (line) A. For small values of p, the closest point $\hat{x} \in A$ has sparse components, i.e., one of its coordinates is approximately zero; for the cases $0 < p \leq 1$ one of its coordinates of $\hat{x}$ is equal to zero, i.e., for those values of p, $\hat{x}$ practically is a sparse solution, while for large values of p the solution $\hat{x}$ tends to have more evenly distributed coordinate contributions; for the case $p = \infty$ its coordinates of $\hat{x}$ tend to be equally distributed.

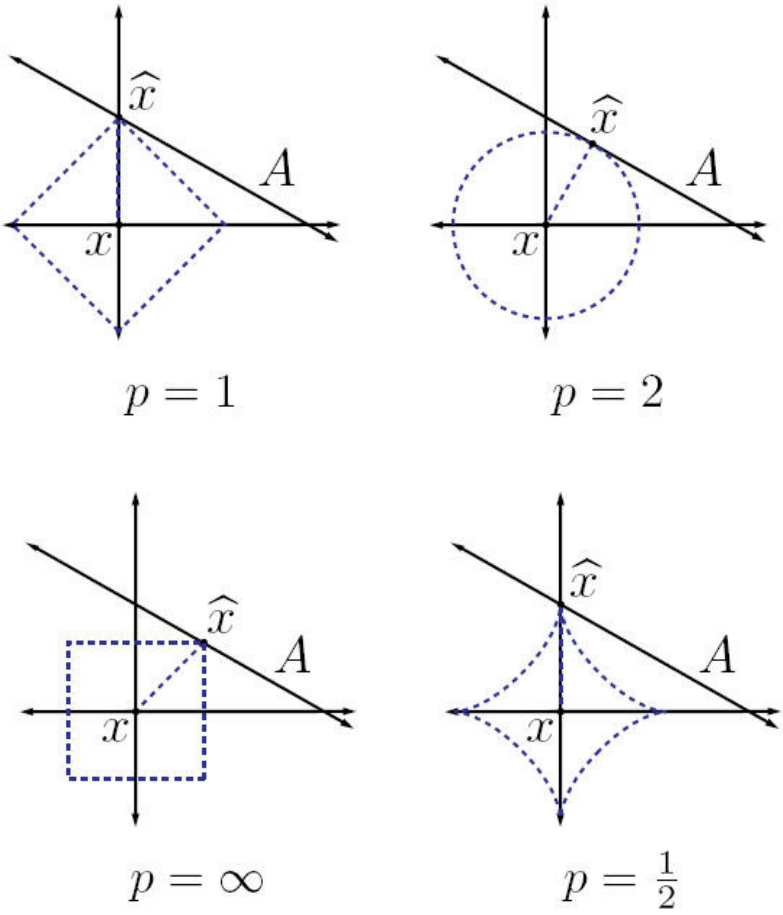


Figure 5.1: Spheres in L^p.

Next we present the following definition of the Null Space Property of order k which is of importance in the convex ℓ_1-minimisation problem as stated in the following Theorem.

For a subset $T \subset \{1, ..., N\}$, we denote by $x_T \in \mathbb{R}_N$ the vector which coincides with $x \in \mathbb{R}_N$ on the entries in T and is zero outside T. Also, $T^c := \{1, ..., N\} \backslash T$ denotes the complement of T.

Definition 19 *A matrix $A \in \mathbb{R}^{M \times N}$ is said to satisfy the* Null Space Property (NSP) of order k *with some constant $\gamma \in (0, 1)$ if*

$$\|x_T\|_1 \leq \gamma \|x_{T^c}\|_1$$

for all sets $T \subset \{1, ..., N\}, |T\| \leq k$ and for all $x \in \ker A$.

Theorem 20 *Let $x \in \mathbb{R}^n$ and x^* be a solution of the ℓ_1-optimisation problem (5.2), and assume that A satisfies the Null Space Property of order k with some constant $\gamma \in (0, 1)$.*

Then

$$\|x - x^*\|_1 \leq \frac{2(1+\gamma)}{1-\gamma}\sigma_k(x)_1.$$

In particular, if x is k-sparse then $x^* = x$.

It can also be proven that if all k-sparse x can be recovered from $Ax = y$ using the ℓ_1-minimisation approach, then the matrix A satisfies the null space property of order k with some constant $\gamma \in (0, 1)$. Since the NSP is difficult to verify in practice, the following equivalent 'restricted isometry property' is often tested instead, which also implies stability under some noise δ_k.

Definition 21 *A matrix A is said to satisfy the* restricted isometry property (RIP) of order *k if there exists a constant $\delta_k \in (0, 1)$ such that*

$$(1 - \delta_k)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta_k)\|z\|_2^2,$$

for all $z \in \Sigma_k$.

Theorem 22 *Assume that $A \in \mathbb{R}^{M\times N}$ satisfies the RIP of order $K = k + h$ with constant $\delta_K \in (0, 1)$. Then A satisfies the NSP of order k with constant $\gamma = \sqrt{k(1+\delta_K)/h(1-\delta_K)}$.*

If the vectors in the set $\{\varphi_j\}_{j=1}^N$ in $\mathbb{R}^N$ are linearly independent and span $\mathbb{R}^N$, then this set is called a *basis* for $\mathbb{R}^N$, i.e., for any $x \in \mathbb{R}^N$ there exists an unique sequence of coefficients $\{c_j\}_{j=1}^N$ such that

$$x = \sum_{j=1}^{n} \mathfrak{c}_j \varphi_j.$$

The above equation can be written in matrix form such that $x = \Phi\mathfrak{c}$, where the columns of Φ are given by the vectors φ_j and the components of c are given by the coefficients $\mathfrak{c}_j$ which are the coordinates of x. The definition of a frame (also called overcomplete dictionary) generalises the concept of a basis for a finite-dimensional spaces [46, 72] and is given next.

Definition 23 *Let $D < N$, then a set $\{\varphi_j\}_{j=1}^N$ in $\mathbb{R}^D$ is called a* frame *if the matrix $\Phi \in \mathbb{R}^{D\times N}$ with columns given by φ_j holds the following equation*

$$A\|x\|_2^2 \leq \|\Phi^T x\|_2^2 \leq B\|x\|_2^2,$$

for all vectors $x \in \mathbb{R}^D$*, with* $0 < A \leq B < \infty$.

Notice that in the above definition it is allowed to have a set $\{\varphi_j\}_{j=1}^N$ of possible linearly dependent vectors. Thus, the sequence of coefficients $\mathfrak{c}_i$ is not unique for the representation $x = \Phi\mathfrak{c}$ for $x \in \mathbb{R}^n$ as a linear combination of vectors φ_j of a frame. When $A = B = 1$, then Φ is called a *Parseval frame*. A frame is called *equal-norm*, if there exists some positive λ such that $\|\varphi_j\|_2 = \lambda$ for all $j = 1, ..., N$, and it is *unit-norm* if $\lambda = 1$. In the finite-dimensional case, A, B correspond to the smallest and largest eigenvalues of $\Phi\Phi^T$, respectively. The implied redundancy is an important component for providing a representation of data x, because there are infinitely many coefficients which satisfy $x = \Phi\mathfrak{c}$. The above frame condition leads to a *dual frame* $\tilde{\Phi}$ defined by

$$\Phi\tilde{\Phi}^T = \tilde{\Phi}\Phi^T = I.$$

The particular selection $\tilde{\Phi} = (\Phi\Phi^T)^{-1}\Phi$ is known as the *Moore-Penrose pseudoinverse* or *canonical dual frame*. This dual frame is well-defined since $A > 0$ implies Φ to have linearly independent rows, which ensures that $\Phi\Phi^T$ is invertible. Then, the coefficients c_D may be given by

$$\mathfrak{c}_D = (\Phi\Phi^T)^{-1}\Phi x,$$

which are the smallest coefficients, that is, $\|\mathfrak{c}_D\|_2 \leq \|\mathfrak{c}\|_2$ for all $\mathfrak{c}$ such that $x = \Phi\mathfrak{c}$.

Next we will present the property of a matrix A which is quite easy to verify in order that the ℓ_1-minimisation problem (5.2) has a solution. For doing this, we give the following definition.

Definition 24 *The* coherence $\vartheta(A)$ *of a matrix* A*, is the largest absolute inner product between any two columns* a_j, a_l *of* A*:*

$$\vartheta(A) = \max_{1 \leq j \leq l \leq n} \frac{|\langle a_j, a_l \rangle|}{\|a_j\|_2 \|a_l\|_2}.$$

If a matrix A has unit-norm columns and coherence $\vartheta(A)$, then A satisfies the RIP of order k with some consonant $\delta = (k-1)\vartheta(A)$ for all $k < 1/\vartheta(A)$ [46, 72].

In a Hilbert space H a frame $\{\varphi_j\}$ is a set of vectors for which there exist positive

constants A, B such that

$$A \sum_{j \in \mathbb{N}} |\langle x, \varphi_j \rangle|^2 \leq \|x\|^2 \leq B \sum_{j \in \mathbb{N}} |\langle x, \varphi_j \rangle|^2 , \quad \forall x \in H. \tag{5.3}$$

The infimum of all B and the supremum of all A that satisfy condition (5.3) are called the *frame bounds*. The *frame operator* $\Phi : H \to \ell_2$ is defined by

$$(\Phi x)_j = \langle \varphi_j, x \rangle.$$

By definition, the frame operator satisfies

$$A\|x\|^2 \leq \|\Phi x\|^2 \leq B\|x\|^2.$$

When $A = B$, then $\{\varphi_j\}$ is called a *tight frame*. This is the case when for all $x \in H$,

$$x = A \sum_{j \in \mathbb{N}} \langle x, \varphi_j \rangle \varphi_j.$$

Frames span the whole Hilbert space H, otherwise, $\ker(\Phi) \neq \{0\}$, therefore there would be some $x \neq 0$ such that $\|\Phi x\|^2 = 0 < A\|x\|^2$. Also, frames commonly are not linearly independent, and in particular do not have to be orthogonal, since an orthogonal frame satisfies by definition

$$\sum_{j \in \mathbb{N}} |\langle x, \varphi_j \rangle|^2 = \|x\|^2 \quad \forall x \in H.$$

Frames are typically "overcomplete" since they still span the whole space H even after some frame elements of $\{\varphi_j\}$ are removed. This is one of the reasons why a frame is considered a generalization of a basis. In the finite-dimensional case, a frame is a basis if and only if it is linearly independent [44]. In our application of DOT the perturbations of the absorption and scattering coefficients $\delta\mu, \delta b$ we will aim at identifying the expansion coefficients of these functions with respect to the basis $\{\varphi_k\}$, which is formed by the characteristic functions of 'pixels'.

We say that a signal x is *k-sparse* in the basis Φ if there exists a vector $\mathfrak{c} \in \mathbb{R}^N$ with only $k << N$ nonzero entries such that $x = \mathfrak{c}\Phi$. We call the set of indices corresponding to the nonzero entries the *support* of $\mathfrak{c}$ and denote it by $\mathrm{supp}(\mathfrak{c})$.

5.2 Linearised Bregman iteration

In this section we consider the following nonlinear minimisation problem

$$\min_{\sigma} \mathcal{J}(\sigma) + \alpha\mathcal{K}(\sigma), \tag{5.4}$$

where $\mathcal{J}$ and $\mathcal{K}$ might be non-differentiable convex functionals defined on a Hilbert space. We mention that problem (5.4) with $\mathcal{J}(\sigma) = \|A\sigma - d\|_2^2$ corresponds to a quadratic minimisation problem ('linear' regression) which is treated in many areas like Image Processing, Restoration and Denoising [29, 79, 125, 127, 166]. In order to obtain the linearised Bregmann iteration for solving the minimisation problem (5.4), both terms $\mathcal{J}(\sigma)$ and $\mathcal{K}(\sigma)$ are replaced by the following approximations as proposed in [113, 174]. Firstly, the penalty term $\mathcal{K}(\sigma)$ is substituted by its *Bregman distance* $D_{\mathcal{K}}^{p^k}(\sigma, \sigma^k)$ which is defined in [10, 79, 83, 125, 127, 174, 175] as follows

$$D_{\mathcal{K}}^{p}(\sigma, \upsilon) = \mathcal{K}(\sigma) - \mathcal{K}(\upsilon) - \langle p, \sigma - \upsilon\rangle, \quad p \in \partial\mathcal{K}(\upsilon). \tag{5.5}$$

Here $\partial\mathcal{K}(\upsilon)$ is called the *subdifferential* of $\mathcal{J}$ at υ, and is defined as

$$\partial\mathcal{K}(\upsilon) = \{p \in P^* : \mathcal{K}(\sigma) \geq \mathcal{K}(\upsilon) - \langle p, \sigma - \upsilon\rangle,\ \forall\sigma \in P\},$$

where P^* is the adjoint of P, with P being the space of coefficients σ defined in (3.1), and $\langle\cdot,\cdot\rangle$ stands for the standard duality product. The elements p of $\partial\mathcal{K}(\upsilon)$ are called *subgradients*. $\mathcal{K}$ is *subdifferentiable at* υ if $\mathcal{K}(\upsilon)$ is finite and the subdifferential $\partial\mathcal{K}(\upsilon)$ is nonempty [10]. The Bregman distance does not satisfy all the properties of a proper distance function, but it satisfies the non-negativity and identity properties, i.e., $D_{\mathcal{K}}^{p}(\sigma, \upsilon) \geq 0$ for all $\sigma \neq \upsilon$ and $D_{\mathcal{K}}^{p}(\sigma, \upsilon) = 0$ if and only if $\sigma = \upsilon$, respectively. It does not satisfy the symmetry property, i.e., $D_{\mathcal{K}}^{p}(\sigma, \upsilon) \neq D_{\mathcal{K}}^{p}(\upsilon, \sigma)$, nor the triangle inequality. However, the Bregman distance measures how close two elements $\tilde{\sigma}$ and υ are to each other. Assuming that w belongs to the segment which joins $\tilde{\sigma}$ and υ, we have $D_{\mathcal{K}}^{p}(\tilde{\sigma}, \upsilon) \geq D_{\mathcal{K}}^{p}(w, \upsilon)$. In more details, the Bregman distance estimates how far the point $\mathcal{K}(\tilde{\sigma})$ is to the tangent plane $\mathcal{K}(\upsilon) - \langle p, \sigma - \upsilon\rangle$. This concept is illustrated in figure (5.2).

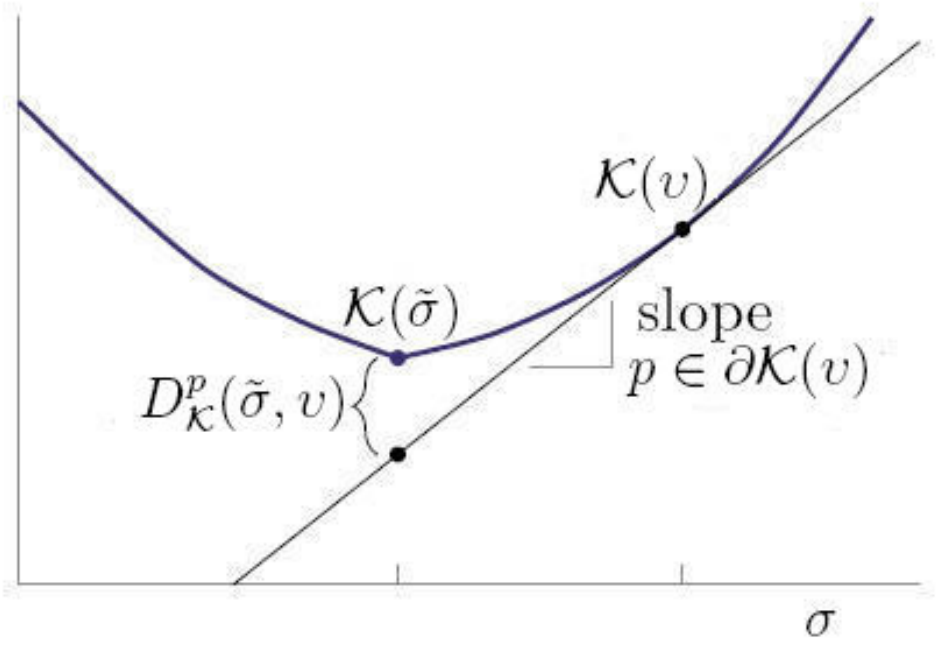


Figure 5.2: The Bregman distance $D^p_{\mathcal{K}}(\tilde{\sigma}, v)$.

Now, instead of solving the problem (5.4) directly, it is solved the following iterative scheme

$$\sigma^{k+1} = \min_{\sigma} \left\{ \mathcal{J}(\sigma) + \alpha D^{p^k}_{\mathcal{K}}(\sigma, \sigma^k) \right\}, \quad p^k \in \partial\mathcal{K}(\sigma^k), \tag{5.6}$$

which is called the *Bregman iteration*. This method minimises the Bregman distance $D^{p^k}_{\mathcal{K}}(\sigma, \sigma^k)$ instead of the function $\mathcal{K}$. It was noted in [29, 79] that $\mathcal{J}(\sigma^k)$ decreases monotonically. Since σ^{k+1} is the minimiser of (5.6), we have

$$\mathcal{J}(\sigma^{k+1}) + \alpha D^{p^k}_{\mathcal{K}}(\sigma^{k+1}, \sigma^k) \leq \mathcal{J}(\sigma^{k+1}) + \alpha D^{p^k}_{\mathcal{K}}(\sigma^k, \sigma^k).$$

Then,

$$\mathcal{J}(\sigma^{k+1}) \leq \mathcal{J}(\sigma^{k+1}) + \alpha D^{p^k}_{\mathcal{K}}(\sigma^{k+1}, \sigma^k) \leq \mathcal{J}(\sigma^{k+1}) + \alpha D^{p^k}_{\mathcal{K}}(\sigma^k, \sigma^k) = \mathcal{J}(\sigma^k).$$

Notice that if the term $\mathcal{J}(\sigma)$ is assumed to be differentiable, then its subdifferential $\partial\mathcal{J}(\sigma) = \{\nabla\mathcal{J}(\sigma)\}$, i.e., the gradient of $\mathcal{J}$ is the *only* subgradient of $\mathcal{J}$. In other words, if $\mathcal{J}$ is differentiable at σ there exists one and only one subgradient of $\mathcal{J}$ at σ, which is the gradient of $\mathcal{J}$ at σ. In this case the subdifferential of the right expression of (5.6) is

given by

$$\begin{aligned}\partial_\sigma \left(\mathcal{J}(\sigma) + \alpha D_{\mathcal{K}}^{p^k}(\sigma, \sigma^k)\right) &= \partial_\sigma \left(\mathcal{J}(\sigma) + \alpha(\mathcal{K}(\sigma) - \mathcal{K}(\sigma^k) - \langle p^k, \sigma - \sigma^k\rangle)\right) \\ &= \{\nabla \mathcal{J}(\sigma)\} + \alpha \partial \mathcal{K}(\sigma) - \{p^k\}.\end{aligned}$$

Since σ^{k+1} is a minimiser of (5.6), we have

$$\mathbf{0} \in \{\nabla \mathcal{J}(\sigma^{k+1})\} + \alpha \partial \mathcal{K}(\sigma^{k+1}) - \{p^k\},$$

where $\mathbf{0}$ is the zero element of P^*. This implies

$$p^k - \nabla \mathcal{J}(\sigma^{k+1}) \in \partial \mathcal{K}(\sigma^{k+1}).$$

Let $p^{k+1} = p^k - \nabla \mathcal{J}(\sigma^{k+1})$, then $p^{k+1} \in \partial \mathcal{K}(\sigma^{k+1})$. Thus, the Bregman iteration (5.6) becomes

$$\sigma^{k+1} = \min_\sigma \left\{\mathcal{J}(\sigma) + \alpha D_{\mathcal{K}}^{p^k}(\sigma, \sigma^k)\right\}, \tag{5.7a}$$

$$p^{k+1} = p^k - \nabla \mathcal{J}(\sigma^{k+1}). \tag{5.7b}$$

Now, the residuum $\mathcal{J}(\sigma)$ is substituted by its first-order Taylor approximation at σ^k and an ℓ_2-penalty term $\left\|\sigma - \sigma^k\right\|_2^2 / 2\tau$, i.e.,

$$\mathcal{J}(\sigma) = \mathcal{J}(\sigma^k) + \langle \nabla \mathcal{J}(\sigma^k), \sigma - \sigma^k\rangle + \frac{1}{2\tau}\left\|\sigma - \sigma^k\right\|_2^2,$$

where $\tau \geq 0$ denotes the step size which can be updated by a line search criterion in each iteration. The constant term $\mathcal{J}(\sigma^k)$ is subtracted from $\mathcal{J}(\sigma)$ and the constant term $\langle \nabla \mathcal{J}(\sigma^k), \nabla \mathcal{J}(\sigma^k)\rangle$ to $\mathcal{J}(\sigma)$ is added to obtain

$$\mathcal{J}(\sigma) = \frac{1}{2\tau}\left\|\sigma - \left(\sigma^k - \tau \nabla \mathcal{J}(\sigma^k)\right)\right\|_2^2.$$

Therefore, the minimisation problem (5.4) is solved as

$$\sigma^{k+1} = \min_\sigma \frac{1}{2\tau}\left\|\sigma - \left(\sigma^k - \tau \nabla \mathcal{J}(\sigma^k)\right)\right\|_2^2 + \alpha \mathcal{K}(\sigma). \tag{5.8}$$

Finally, the problem (5.4) is solved with the following *Bregman iteration* which is a combination of the Bregman iteration (5.7) and the fixed-point iteration scheme (5.8)

$$\sigma^{k+1} = \min_{\sigma} \frac{1}{2\tau} \left\| \sigma - \left(\sigma^k - \tau \nabla \mathcal{J}(\sigma^k) \right) \right\|_2^2 + \alpha D_{\mathcal{K}}^{p^k}(\sigma, \sigma^k), \tag{5.9a}$$

$$p^{k+1} = p^k - \frac{1}{\tau} \left(\sigma - (\sigma^k - \tau \nabla \mathcal{J}(\sigma^k)) \right). \tag{5.9b}$$

5.3 Linearised Bregman Iteration for the ℓ_1-minimisation: linear case

We turn our attention now to the problem (5.4) with $\mathcal{J}(\sigma) = \|A\sigma - d\|_2^2$ and penalty term $\mathcal{K}(\sigma) = \|\sigma\|_1$. Here the Bregman iteration (5.9) takes a specific expression as derived in [29, 174] and explained as follows. Since the function $\mathcal{K}(\sigma) = \|\sigma\|_1$ is not differentiable at the origin, it is replaced by the following continuously differentiable function $\mathcal{K}_\epsilon$

$$\mathcal{K}_\epsilon(\sigma) = \sum_{j=1}^{N} \vartheta_\epsilon \left(\sigma(j) \right), \tag{5.10}$$

where $\sigma \in \mathbb{R}^N$ and N denotes the number of disjoint pixels which form the domain Ω. The function $\mathcal{K}_\epsilon$ is the *Moreau-Yosida* C^1 *regularisation* for $\|\sigma\|_1$, also known as the *Huber-norm*. This function $\mathcal{K}_\epsilon$ is also Lipchitz differentiable, i.e.,

$$\|\nabla \mathcal{K}(\sigma) - \nabla \mathcal{K}(\upsilon)\|^2 \leq \gamma \langle \nabla \mathcal{K}(\sigma) - \nabla \mathcal{K}(\upsilon), \sigma - \upsilon \rangle, \quad \forall \sigma, \upsilon \in \mathbb{R}^N,$$

where γ is a positive constant, see figure 5.3. The function $\vartheta_\epsilon(\zeta)$ is the *Huber function* defined by

$$\vartheta_\epsilon(\zeta) = \begin{cases} \dfrac{\zeta^2}{2\epsilon}, & \text{if } |\zeta| \leq \epsilon, \\ |\zeta| - \dfrac{\epsilon}{2}, & \text{if } |\zeta| > \epsilon. \end{cases} \tag{5.11}$$

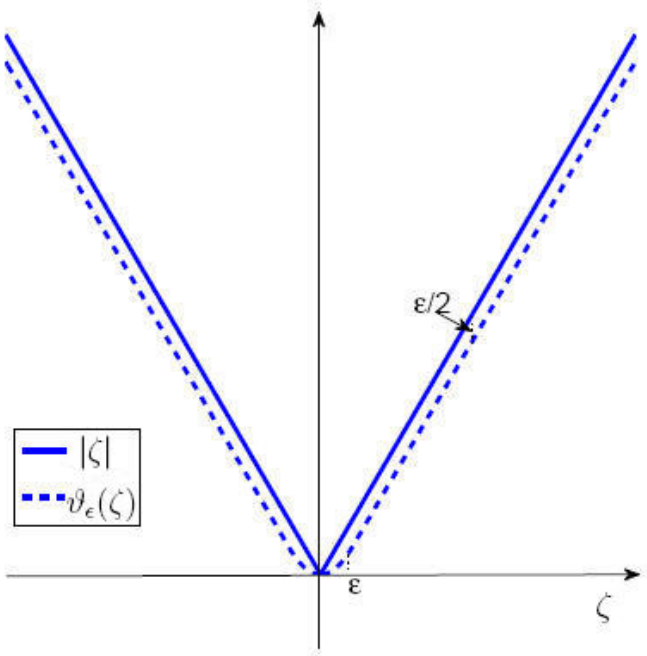


Figure 5.3: The approximation of ϑ_ϵ to the absolute value function.

Let now $\omega^k = \sigma^k - \tau A^T(A\sigma^k - d)$ and $\mathcal{K} = \mathcal{K}_\epsilon(\sigma)$ as defined in (5.10). Then the problem (5.9a) can be decoupled into the following N one-variable minimisation problems ($j = 1, \ldots, N$):

$$\sigma^{k+1}(j) = \min_\zeta \left\{ \alpha\Big(\vartheta_\epsilon(\zeta) - \vartheta_\epsilon\big(\zeta_k(j)\big) - p^k(j)\big(\zeta - \sigma^k(j)\big)\Big) + \frac{1}{2\tau}\big(\zeta - \omega^k(j)\big)^2 \right\}. \quad (5.12)$$

After minimising the equation (5.12) with respect to ζ and equating to zero, the solution becomes

$$\tau\alpha\vartheta'_\epsilon(\zeta) + \zeta = \omega_k + \tau\alpha p^k(j), \quad (5.13)$$

where the derivative of ϑ_ϵ is given by

$$\vartheta'_\epsilon(\zeta) = \begin{cases} \dfrac{\zeta}{\epsilon} & \text{if } |\zeta| \le \epsilon \\ \operatorname{sign}(\zeta) & \text{if } |\zeta| > \epsilon. \end{cases} \quad (5.14)$$

See figure 5.4.

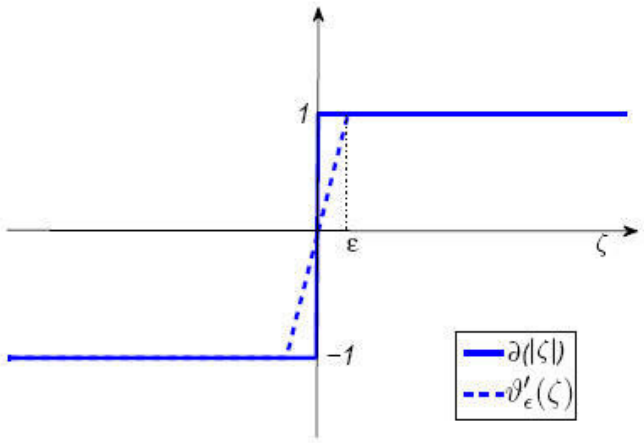


Figure 5.4: The approximation of ϑ'_ϵ to the subdifferential of the absolute value function.

The solution of the differential equation (5.13) is given by

$$\zeta = s_{\tau\alpha,\epsilon}\big(\omega^k + \tau\alpha p^k(j)\big), \tag{5.15}$$

where $s_{\lambda,\epsilon}(\varkappa)$ is defined as

$$s_{\lambda,\epsilon}(\varkappa) = \begin{cases} \frac{\epsilon}{\lambda+\epsilon}\varkappa, & \text{if } |\varkappa| \leq \lambda + \epsilon \\ \operatorname{sign}(\varkappa)\left(|\varkappa| - \lambda\right), & \text{if } |\varkappa| > \lambda + \epsilon. \end{cases} \tag{5.16}$$

This solution is unique because the problem (5.12) is strongly convex [29]. Let $\mathcal{S}_{\lambda,\epsilon}$ be defined by

$$\mathcal{S}_{\lambda,\epsilon}(\sigma) := \Big[s_{\lambda,\epsilon}\big(\sigma(1)\big), s_{\lambda,\epsilon}\big(\sigma(2)\big), \cdots, s_{\lambda,\epsilon}\big(\sigma(N)\big)\Big]^T, \tag{5.17}$$

with $s_{\lambda,\epsilon}(\varkappa)$ as defined in (5.16). Then, using the function given in (5.17), the linearised Bregman iteration (5.9) is formulated as

$$\sigma^{k+1} = \mathcal{S}_{\tau\alpha,\epsilon}\left(\sigma^k + \tau\alpha p^k - \tau A^T(A\sigma^k - d)\right), \tag{5.18a}$$

$$p^{k+1} = p^k - \frac{1}{\tau\alpha}\left(\sigma^{k+1} - \left(\sigma^k - \tau A^T(A\sigma^k - d)\right)\right), \tag{5.18b}$$

where $\sigma^0 = 0$, $p^0 = 0$. Let further $v^k = \sigma p^k + \frac{1}{\tau}\sigma^k$. Then the linearised Bregman iteration (5.18) becomes

$$\sigma^{k+1} = \mathcal{S}_{\tau\alpha,\epsilon}(\alpha v^{k+1}), \tag{5.19a}$$

$$v^{k+1} = v^k - A^T(A\sigma^k - d), \tag{5.19b}$$

where $\sigma^0 = 0,\ v^0 = 0$. When ϵ tends to zero (see figure 5.5 for illustration), the function $\mathcal{S}_{\lambda,\epsilon}(\zeta)$ defined in (5.17) tends to the *soft-thresholding function* $\mathcal{S}_\lambda(\zeta)$. This function is defined by

$$\mathcal{S}_\lambda(w) := \Big[s_\lambda\big(w(1)\big), s_\lambda\big(w(2)\big), \cdots, s_\lambda\big(w(N)\big)\Big]^T, \tag{5.20}$$

where

$$s_\lambda(\varkappa) = \begin{cases} 0, & \text{if } |\varkappa| \le \lambda, \\ \operatorname{sign}(\varkappa)\left(|\varkappa| - \lambda\right), & \text{if } |\varkappa| > \lambda. \end{cases} \tag{5.21}$$

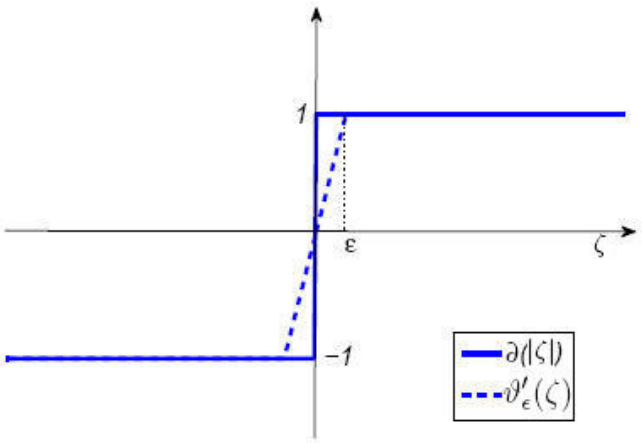


Figure 5.5: The approximation of $s_{\lambda,\epsilon}$ to the soft-thresholding function s_λ.

Therefore, when $\epsilon \to 0$, the iteration (5.19) becomes [29, 174]

$$\sigma^{k+1} = \mathcal{S}_{\tau\alpha}(\alpha v^{k+1}), \tag{5.22a}$$

$$v^{k+1} = v^k - A^T(A\sigma^k - d). \tag{5.22b}$$

A different and more general approach for obtaining equation (5.22) for a convex or non-convex functional $\mathcal{J}$ (instead of the matrix A) and a convex and non-differentiable function $\mathcal{K}$, together with a convergence analysis, is presented in [25]. The linearised Bregman iteration for the nonlinear problem (5.4) with the same penalty term $\mathcal{K}(\sigma) = \|\sigma\|_1$ analogous to (5.22) is obtained and explained in more details in the next section. A convergence analysis for the Bregman iteration for a nonlinear problem with a total variation penalty term $\mathcal{K}(\sigma)$ is found in [10].

5.4 Sparsity method

We are interested in "simple" geometries or descriptions of the absorption and scattering coefficients. Let $\delta\mu := \mu - \mu^0$ and $\delta b := b - b^0$ be the inhomogeneities of the unknown physical absorption and scattering distributions μ and b, respectively, where μ^0, b^0 are the background values which in DOT correspond to those of healthy tissue based on standard values for the types of tissue being tested. The "simple" geometries correspond to the fact that $\delta\mu, \delta b$ have sparse representation with respect to a specific basis or frame. Such a prior knowledge is reasonable and justifiable in cases where one looks for some "simple" inclusions (tumours, haematomas, etc) embedded in an uninteresting background. Let $\mathscr{A}$ be the admissible set of absorption and scattering coefficients given by

$$\mathscr{A} = \{(\mu, b) \in H^1(\Omega) \times H^1(\Omega) : c_0 \leq \mu \leq c_1, c_2 \leq b \leq c_3 \text{ a.e. } \Omega\},$$

where c_0, c_1, c_2, c_3 are known positive constants. Let $\sigma = (\mu, b)$ be as described in previous chapters. We consider the minimisation of the following functional

$$\mathcal{F}(\sigma) = \mathcal{J}(\sigma) + \alpha\Psi(\delta\sigma), \tag{5.23}$$

over the set $\mathscr{A}$, with $\mathcal{J}(\sigma)$ as in (3.6). The vector regularization parameter is $\alpha := (\alpha_\mu, \alpha_b)$, one parameter for μ and one for b as in the coefficient vector σ. The l_1 penalty term $\Psi(\delta\sigma) = \|\delta\sigma\|_1$ is then defined by

$$\Psi(\delta\sigma) = \|\delta\mu\|_1 + \|\delta b\|_1, \tag{5.24}$$

where the ℓ_1 norms of the coefficients μ, b are given as

$$\|\delta\mu\|_1 := \sum_{j\in\mathbb{N}} \left|\langle \delta\mu, \varphi_j\rangle_{H^1(\Omega)}\right| \qquad \|\delta b\|_1 := \sum_{j\in\mathbb{N}} \left|\langle \delta b, \varphi_j\rangle_{H^1(\Omega)}\right|. \tag{5.25}$$

Here $\{\varphi_j\}$ is an orthonormal basis or frame which can be a Fourier basis, wavelets or formed by the characteristic functions of pixels or voxels of the domain Ω. We mention that a novel method to approximate the solution of (5.23) when ℓ_p-norm of Ψ has been introduced for $0 < p \leq 1$ and for a smooth version of ℓ_0 [135].

Notice that the penalty term $\Psi(\sigma)$ is convex but not differentiable. We have used the latter one in our numerical simulations. Then, the term $\alpha\Psi(\delta\sigma)$ is given by

$$\begin{aligned}\alpha\Psi(\delta\sigma) &= \alpha \sum_{j\in\mathbb{N}} \left|\langle \delta\sigma, \varphi_j\rangle_{H^1(\Omega)\times H^1(\Omega)}\right| \\ &= \alpha_\mu \sum_{j\in\mathbb{N}} \left|\langle \delta\mu, \varphi_j\rangle_{H^1(\Omega)}\right| + \alpha_b \sum_{j\in\mathbb{N}} \left|\langle \delta b, \varphi_j\rangle_{H^1(\Omega)}\right|.\end{aligned}$$

We identify $\langle \delta\mu, \varphi_j\rangle_{H^1}, \langle \delta b, \varphi_j\rangle_{H^1}$ with the sequence of expansion coefficients with respect to the basis $\{\varphi_j\}$. We can see from the above expression that we impose a penalisation on sequences of "small" expansion coefficients, those with $|\langle \delta\mu, \varphi_j\rangle| < 1$ and $|\langle \delta b, \varphi_j\rangle| < 1$, while imposing lesser penalty on sequences of "large" expansion coefficients for which $|\langle \delta\mu, \varphi_j\rangle| \geq 1$ and $|\langle \delta b, \varphi_j\rangle| \geq 1$.

Next we will outline the existence proof of the minimiser of (5.23). Also here we will work with the Hilbert spaces (3.1). We then apply the following Theorem.

Theorem 25 *Let H_1, H_2, H be Hilbert spaces. If $R : H_1 \to H_2$ is continuous and weakly sequentially closed and $\Psi : H \to \mathbb{R}$ is proper, positive, weakly lower semicontinuous with bounded level sets, and α is positive. Then there exists a minimiser of $\mathcal{F}$.*

The operator $\Psi(\sigma)$ satisfies all the properties of the above Theorem, i.e., Ψ is weakly lower semicontinuous, proper, positive and has bounded sets as it has been proved in

[41]. The operator R_i is called *weakly sequentially closed* if for $\sigma^k \rightharpoonup \sigma$, then $\sigma \in \mathscr{A}$ and $R_i(\sigma^k) \rightharpoonup R_i(\sigma)$. We will assume that the residual operator $R_i(\sigma)$ defined in (3.3) for $i = 0, \ldots, M-1$, is continuous and weakly sequentially closed. Then, we can assure that there exists a minimiser of $\mathcal{F}$. The formal proof of Theorem 25 is beyond the scope of this dissertation.

We anticipate that we will use the Landweber-Kaczmarz method for solving the problem (5.23). There we will aim at minimizing

$$\mathcal{F}_i(\sigma) = \mathcal{J}_i(\sigma) + \alpha\Psi(\delta\sigma), \tag{5.26}$$

over the same set $\mathscr{A}$, where i correspond to the ith source q_i. Let $\nabla\mathcal{J}_i$ be as in (3.19). We will use the notation $\nabla\mathcal{J}_i = (\nabla\mathcal{J}_{i,\mu}, \nabla\mathcal{J}_{i,b})$, i.e.,

$$\nabla\mathcal{J}_i = \begin{pmatrix} \nabla\mathcal{J}_{i,\mu} \\ \nabla\mathcal{J}_{i,b} \end{pmatrix} = \begin{pmatrix} \mathfrak{L}_1(u_i, z_i) \\ \mathfrak{L}_2(u_i, z_i) \end{pmatrix}.$$

Using the smooth gradient $\nabla\mathcal{J}_i^s$ described in section 3.4, the functional $\mathcal{F}_i(\sigma) = \mathcal{F}_i(\sigma^0 + \delta\sigma)$ can be locally approximated by the following *surrogate functional*

$$\mathcal{F}_i(\sigma^0+\delta\sigma) \approx \mathcal{F}_i(\sigma^0+\delta\sigma^k)+\left\langle \delta\sigma-\delta\sigma^k, \nabla\mathcal{J}_i^s(\sigma^k)\right\rangle_{L^2(\Omega)}+(\tau)^{-1}\frac{1}{2}\|\delta\sigma-\delta\sigma^k\|^2_{L^2(\Omega)}+\alpha\|\delta\sigma\|_1, \tag{5.27}$$

where τ is some step size to be discussed later. This surrogate functional is convex and separable in $\delta\sigma$, as in the proxy function approach [25, 44, 98]. The minimisation of this proxy problem is (up to an unimportant constant) equivalent to

$$\delta\sigma^{k+1} = \arg\min_{\delta\sigma} (\tau)^{-1}\frac{1}{2}\|\delta\sigma - \left(\delta\sigma^k - \tau\nabla\mathcal{J}_i^s(\sigma^k)\right)\|^2_{L^2(\Omega)} + \alpha\|\delta\sigma\|_1. \tag{5.28}$$

The following *soft-thresholding approach* was proposed for minimizing this surrogate functional $\mathcal{F}_i$ by [25, 44, 137]

$$\begin{aligned} \delta\sigma^{k+1} &= \boldsymbol{\mathcal{S}}_{\tau\alpha}\Big(\delta\sigma^k - \tau\nabla\mathcal{J}_i(\sigma^k)\Big), \\ &= \boldsymbol{\mathcal{S}}_{\tau\alpha}\Big(\delta\mu - \tau_\mu\mathcal{J}_{i,\mu}(\mu^k, b^k), \delta b - \tau_b\mathcal{J}_{i,b}(\mu^k, b^k)\Big), \end{aligned} \tag{5.29}$$

where $\boldsymbol{\mathcal{S}}_{\tau\alpha}$ is the soft-thresholding function defined by

$$\boldsymbol{\mathcal{S}}_{\tau\alpha}(\varkappa, \varkappa) = \Big(\big(\mathcal{S}_{\alpha_\mu \tau_\mu}(\varkappa)\big)_j, \big(\mathcal{S}_{\alpha_b \tau_b}(\varkappa)\big)_j \Big). \tag{5.30}$$

Here $\mathcal{S}_\lambda$ is the *soft-thresholding function* as defined in (5.20), which can also be written component-wise in the form

$$\big(\mathcal{S}_\lambda(\varkappa)\big)_j = \begin{cases} 0, & \text{if } |\varkappa_j| \leq \lambda, \\ \operatorname{sign}(\varkappa_j)\left(|\varkappa_j| - \lambda\right), & \text{if } |\varkappa_j| > \lambda, \end{cases} \tag{5.31}$$

or, in a more compact way, as

$$\big(\mathcal{S}_\lambda(\varkappa)\big)_j = \operatorname{sign}(\varkappa_j) \max\{|\varkappa_j| - \lambda, 0\}, \tag{5.32}$$

which is also called a *shrinkage function*. This operator is typically used for promoting sparsity on $\delta\sigma$. The update of $\delta\sigma$ (5.29) is composed of two stages. The first stage corresponds to the Landweber-Kaczmarz method with step size τ, whereas the second stage takes care of the promotion of sparsity using the shrinkage operator. A direct application of the gradient $\nabla\mathcal{J}_i$ in this context does not retrieve accurate reconstructions because of its insufficient regularity properties [78, 98]. Therefore, we have used the smooth gradient $\mathcal{J}_i^s$ as discussed in section 3.4. This is why we have agreed on $H^1(\Omega)$ being the space of coefficients μ, b. The convergence of the Landweber method with constant fixed step size is slow. In order to alleviate this drawback, Maass and Jin [78, 98] implemented the Barzilai-Borwein (BB) step size criterion instead, which reads $\tau^k := \left(\begin{smallmatrix} \tau_\mu^k & 0 \\ 0 & \tau_b^k \end{smallmatrix}\right)$ and is discussed in section 2.3.2; this yields formula (2.83), with τ_μ^k, τ_b^k given by

$$\tau_\mu^k = \frac{\left\langle \delta\mu^k - \delta\mu^{k-1}, \nabla\mathcal{J}^s(\sigma^k) - \nabla\mathcal{J}^s(\sigma^{k-1}) \right\rangle_{H^1(\Omega)}}{\|\delta\mu^k - \delta\mu^{k-1}\|^2_{H^1(\Omega)}}, \tag{5.33a}$$

$$\tau_b^k = \frac{\left\langle \delta b^k - \delta b^{k-1}, \nabla\mathcal{J}^s(\sigma^k) - \nabla\mathcal{J}^s(\sigma^{k-1}) \right\rangle_{H^1(\Omega)}}{\|\delta b^k - \delta b^{k-1}\|^2_{H^1(\Omega)}}. \tag{5.33b}$$

Since with the Barzalai-Borwein criterion the functional $\mathcal{F}_i$ does not automatically decrease monotonically, the following condition was implemented on the step size τ^k to assure monotonous decrease of $\mathcal{F}$ [78, 98]. We use the step size given by the BB rule given by (5.33) as initial guess, and reduce it until the following weak monotonicity

condition on the functional $\mathcal{F}_i(\sigma)$ is achieved

$$\mathcal{F}_i\big(\sigma^0 + \boldsymbol{\mathcal{S}}_{\tau^k\alpha}\big(\delta\sigma^k - \tau^k\nabla\mathcal{J}_i^s(\sigma^k)\big)\big) \leq \max_{k-m+1\leq l\leq k}\mathcal{F}_i(\sigma^l) - \tau^k\frac{\varepsilon}{2}\|\boldsymbol{\mathcal{S}}_{\tau^k\alpha}\big(\delta\sigma^k - \tau^k\nabla\mathcal{J}_i^s(\sigma^k)\big) - \delta\sigma^k\|_{H^1}^2. \tag{5.34}$$

where $m \in \mathbb{N}$ and $\varepsilon = (\varepsilon_\mu, \varepsilon_b)^T$, with $\varepsilon_\mu, \varepsilon_b \in \mathbb{R}$. We mention that the first τ^k on the right hand side of the above inequality is considered as $\tau^k = (\tau_\mu^k, \tau_b^k)$ in order that the product $\tau\varepsilon$ is well defined. A strict monotonicity condition holds for $m = 1$. We have fixed the value $m = 2$ in our reconstructions. In our implementation, we use the step length given by the BB rule as the initial guess at each iteration and decrease it geometrically by a factor r until condition (5.34) holds. Also, the initial step length is constrained to $(\tau_{\min}, \tau_{\max})$ and the iteration is terminated when τ^k falls below $\tau_{\min}$ and when the iteration is deemed to stagnate. Therefore, we calculate $\delta\sigma^{k+1}$ using equation (5.33) with the smooth gradient $\nabla\mathcal{J}_i^s$ instead of $\nabla\mathcal{J}_i$. Moreover, the step size τ^k given by (5.33) until the monotonicity condition (5.34) is satisfied, i.e.,

$$\delta\sigma^{k+1} = \boldsymbol{\mathcal{S}}_{\tau\alpha^k}(\delta\sigma^k - \tau^k\nabla\mathcal{J}_i^s(\sigma^k)). \tag{5.35}$$

Basically, this method is composed of two parts, the first one is the application of the Landweber-Kaczmarz method given by $\tau^k\nabla\mathcal{J}_i^s(\sigma^k)$, and the second part is the sparsity promotion by the shrinkage operator $\boldsymbol{\mathcal{S}}_{\tau^k\alpha}$. One can use the Landweber and the L-LK methods instead of the LK method for this update.

A Pseudo code for the reconstruction scheme using the Landweber method and sparsity regularization is described in algorithm 9.

Algorithm 9: Landweber using sparsity regularization

Input: $\sigma^0, \delta\sigma^0 = 0, \alpha, \tau_{\min}, \tau_{\max}, m$

Output: $\sigma^{I(M-1)}$

for $k := 0, \dots, I$ **do**

  for $i := 0, \dots, M-1$ **do**

    Compute the gradient $\nabla \mathcal{J}_i(\sigma^k)$ given by (3.16);

    Compute the smooth gradient $\nabla \mathcal{J}_i^s(\sigma^k)$ using algorithm 5 ;

  end

  Compute $\nabla \mathcal{J}^s(\sigma^k) = \sum_{i=0}^{M-1} \nabla \mathcal{J}_i^s(\sigma^k)$;

  Calculate τ^k given by BB criterion (5.33) holding (5.34);

  Update inhomogeneity $\delta\sigma^{k+1}$ using (5.35);

  Calculate $\sigma^{k+1} = \sigma^0 + \delta\sigma^{k+1}$;

end

Here u_i solves (2.1) with σ^k and source q_i, and z_i solves (3.10) with σ^k and adjoint source ξ_i given by (3.17) for the computation of $\nabla \mathcal{J}_i(\sigma^k)$.

A Pseudo Code for the reconstruction scheme using the LK method and sparsity regularization is described in algorithm 10.

Algorithm 10: LK using sparsity regularization

Input: $\sigma^0, \delta\sigma^0 = 0, \alpha, \tau_{\min}, \tau_{\max}, m$

Output: $\sigma^{I(M-1)}$

for $k := 0, \dots, I(M-1)$ **do**

  Compute the gradient $\nabla \mathcal{J}_{[k]}(\sigma^k)$ given by (3.19);

  Compute the smooth gradient $\nabla \mathcal{J}_{[k]}^s(\mu^k)$ using algorithm 5 ;

  Calculate the step size τ^k given by BB criterion (5.33) holding condition (5.34);

  Update inhomogeneity $\delta\sigma^{k+1} = \boldsymbol{\mathcal{S}}_{\tau\alpha^k}(\delta\sigma^k - \tau^k \nabla \mathcal{J}_{[k]}^s(\sigma^k))$;

  Calculate $\sigma^{k+1} = \mu^0 + \delta\sigma^{k+1}$;

end

Here u_k solves (2.1) with σ^k and source q_i, and z_i solves (3.10) with σ^i and adjoint source ξ_i given by (3.17) for the computation of $\nabla \mathcal{J}_{[k]}(\sigma^k)$.

Our suggested alternative way for minimising the functional (5.23) over the set A is explained next. Instead of updating the inhomogeneity $\delta\sigma^k$ using the soft-thresholding function defined in (5.21) in our sparsity method, we update firstly the value of the coefficient σ^k using the gradient descent direction of $\mathcal{J}$ and the soft-thresholding operator as indicated in the following

$$\sigma^{k+1} = \boldsymbol{\mathcal{S}}_{\alpha_1\tau^k}(\sigma^k - \tau^k \nabla \mathcal{J}^s(\sigma^k)), \tag{5.36}$$

where $\boldsymbol{\mathcal{S}}$ as defined in (5.30), with the following "semi-shrinkage" function

$$\mathcal{S}_\lambda(\varkappa) = \begin{cases} \sigma^0, & |\varkappa| \leq \lambda, \\ |\varkappa| - \lambda, & |\varkappa| > \lambda. \end{cases} \tag{5.37}$$

Here $\mathcal{S}_\lambda$ is the soft-thresholding function as defined in (5.20), which, as before, can be written component-wise as

$$\big(\mathcal{S}_\lambda(\varkappa)\big)_j = \operatorname{sign}(\varkappa_j)\max\{|\varkappa_j| - \lambda, 0\}. \tag{5.38}$$

Moreover, $\alpha_1 := \sigma^0 + \alpha$, with $\alpha << \alpha_1$. The parameter α is used in the monotonicity condition (5.34) for calculating τ^k.

A Pseudo Code for this alternative algorithm is given below.

Algorithm 11: LK using our sparsity method

Input: $\sigma^0, \alpha, (\alpha_1 = \sigma^0 + \alpha), \tau_{\min}, \tau_{\max}, m$

Output: $\sigma^{I(M-1)}$

for $k := 0, \ldots, I(M-1)$ **do**

Compute the gradient $\nabla\mathcal{J}_{[k]}(\sigma^k)$ given by (3.19);

Compute the smooth gradient $\nabla\mathcal{J}^s_{[k]}(\sigma^k)$ using algorithm 5 ;

Calculate τ^k using the BB criterion (5.33) holding condition (5.34) using α;

Update the coefficient by σ^{k+1} and using (5.36);

end

For completeness we mention that a sparse finite element method, based however on the pure absorption RTE presented in Subsection 2.1.2, was studied in [168].

5.5 Numerical Results

We have carried out the following numerical experiments targeting image reconstructions of the absorption and scattering coefficients in DOT with sparsity regularization. We have been using the algorithms 9,10, and 11 as presented in this chapter. In the following, we explain in more details what has been done and how the algorithms performed in different situations.

To start with, we present some reconstructions of the absorption coefficient μ. Afterwards we will also consider reconstructions of the scattering coefficient b. Finally, the simultaneous reconstruction of μ and b is attempted. In the first numerical experiment, two obstacles in the absorption coefficient, both having the same size, are reconstructed from the data. These can for example represent haematomas. In the second experiment, two scattering obstacles with different sizes are reconstructed from the gathered data. Finally, two absorption and two scattering inclusions are reconstructed simultaneously in the last experiment using the LK technique with and without sparsity regularization.

The maximum number of iterations I was taken to be 400, the reduction factor for the step size τ^k was 0.6, and the fixed number m to preserve the monotonicity in (5.34) was 2. The parameter s used in (5.34) was 1.0×10^{-5}, the regularization parameters $\alpha = (\alpha_\mu, \alpha_b)$ of expression (5.23) (as defined at the beginning on section 5.4) of the shrinkage thresholding function $\mathcal{S}_{\tau\alpha^k}$ for all reconstructions was manually selected to give feasible reconstructions. For the purpose of comparison, we present also the reconstructions of the absorption coefficient μ and the scattering coefficient b individually, and simultaneously, using the Landweber and Landweber-Kaczmarz techniques as a benchmark to compare our current sparsity imaging reconstructions with. We have used the same numerical experimental setting as in section 3.7. We model an 'idealized' human head by a domain of size 5×5 cm^2. It is composed by a homogeneous background and an embedded clear layer, which simulates the cerebrospinal fluid of the head. As we mentioned in chapter 2, we have used a 2D-adapted version of the $3D$ Henyey-Greenstein function (2.5) as the scattering function with $g = 0.9$ for the individual reconstructions, and $g = 0.6$ for the simultaneous reconstruction of absorption and scattering coefficients values given in Table 5.1. These values are in the range of biomedical applications [142]. Saratoon et al. [142] have pointed out recently that for those values of the absorption and scattering coefficients, it is needed to use at least 16 vector velocity directions in the discretization

of the angular variable θ in order to obtain accurate results. We opted to use only 12 directions because in many results as in [52, 53, 54, 144] we have seen that we are able to retrieve acceptable reconstructions from simulated (noisy) data. Certainly, when real data are used, possibly a more refined discretization of the directional variable will be necessary, in particular in regions close to the boundary and inside the clear layers. On the other hand, we do not expect that the fundamental behaviour of our algorithm will depend on the exact number of directions chosen in the angular discretization. In our proof-of-concept experiments we put therefore more emphasis on numerical efficiency and simplicity.

	$\mu[cm^{-1}]$	$b[cm^{-1}]$
Background	1.0	10
Inclusion	3.0	30
Clear layer	0.1	0.1

Table 5.1: Experimental set-up for simultaneous absorption and scattering reconstruction using sparsity regularization

We observed in our numerical experiments that the evolution of the functional $\mathcal{F}$ presents a similar behaviour as the evolution of the cost $\mathcal{J}$ or the norm of residuals, such that we have only presented the graphs of the latter ones in our figures. A similar observation is reported in [41, 78, 98].

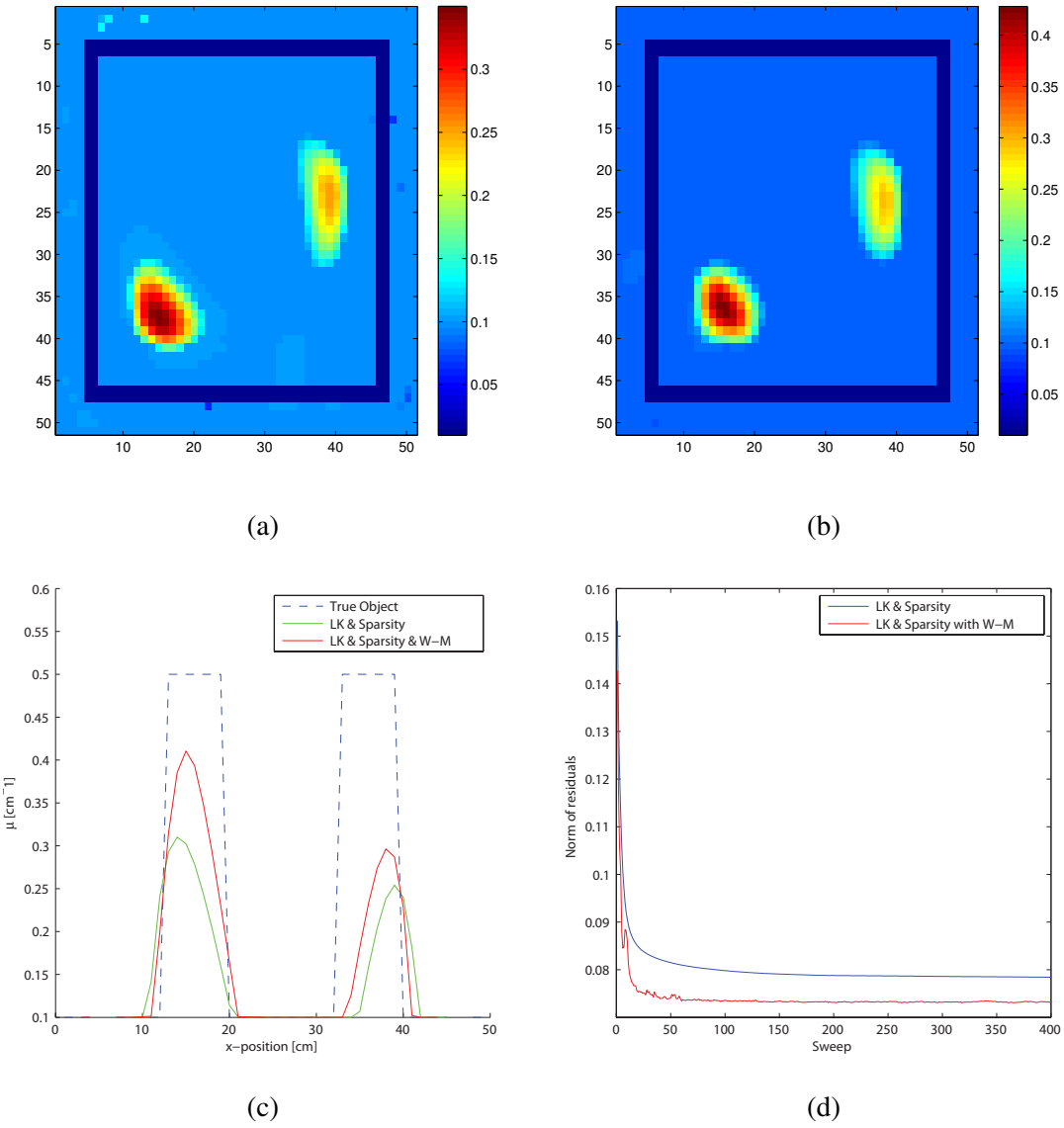


Figure 5.6: Figures 5.6a and 5.6b shows the reconstruction images using the algorithm 10 with Weak Monotonicity and without Weak Monotonicity condition respectively. Figure 5.6c and figure 5.6d show the cross sections and the norm of the residuals of the imaging reconstruction using the LK method with sparsity regularization with Weak Monotonicity and without the Weak Monotonicity condition respectively.

The noisy data are generated by adding pointwise Gaussian noise to the absorption and scattering exact data according to $\mu_\delta^k = \mu^k + \varepsilon\varrho|\mu^k|$, and $b_\delta^k = b^k + \varepsilon\varrho|b^k|$, where ϱ is the standard normal random variable, δ refers to the level of noise, ε is the relative noise level, and $|\mu^k|, |b^k|$ denotes the absolute value of μ^k and b^k respectively. To measure the accuracy of an approximation μ_δ of the absorption coefficient, we compute the L^2-norm error $e = \|\mu_\delta - \bar{\mu}\|_{L^2}$, where $\bar{\mu}$ is the true physical coefficient. In an analogous way, we compute the L^2-norm error for the scattering coefficient b. Figure 5.6 shows the comparison of the image reconstruction of the absorption coefficient μ using the algorithm

10 with the Monotonicity condition given in (5.34) on the one hand, and without using the Monotonicity condition on the other hand. A qualitative and quantitative difference of using the weak monotonicity condition (5.34) with the BB step size criterion is shown in this figure. As it can be observed, the convergence indicated by the red line is faster than the blue one. One needs to keep in mind in this comparison that the execution time of the program is increased using the monotonicity condition. Also, we see that the reconstruction using the weak monotonicity condition is more accurate with respect to the true one. These results indicate that the application of the combined BB step size criterion with the weak monotonicity condition provides quite accurate reconstructions.

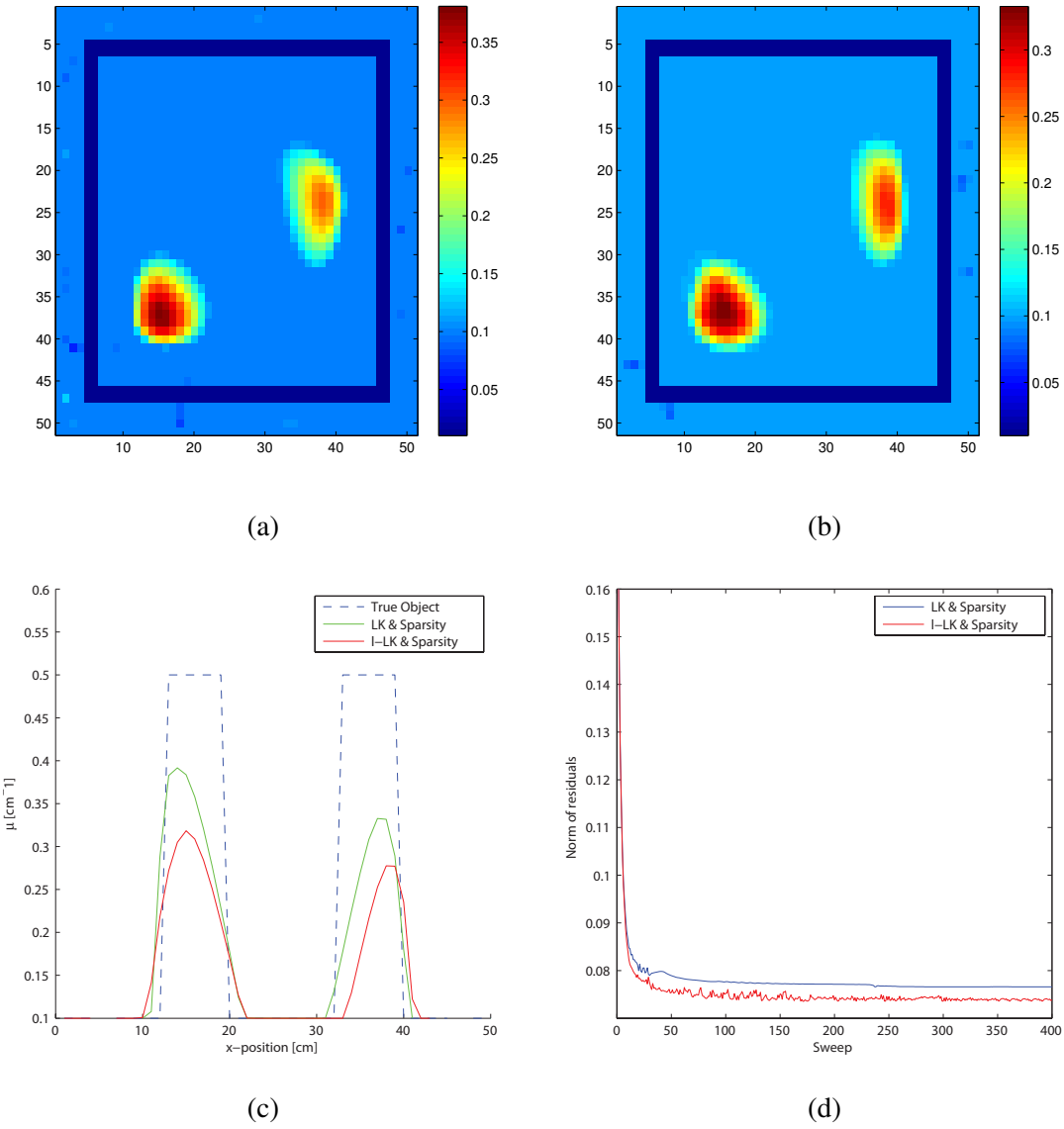


Figure 5.7: Figures 5.7a and 5.7b shows the reconstruction images using using the algorithm 10 with the L-LK variant without the Weak Monotonicity in both cases. Figure 5.7c and figure 5.7d show the cross section and the residuals of the LK with sparsity regularization without versus L-LK and sparsity regularization respectively without the Weak Monotonicity condition for both cases.

Figure 5.7 shows the comparison of the reconstruction of the absorption coefficient μ using algorithm 10 (the LK method and sparsity regularization) with the reconstruction obtained using the same algorithm 10, but instead of computing $\nabla \mathcal{J}_{[k]}(\sigma^k)$ using the LK method, we use now the L-LK method, and without the Monotonicity in the BB step size criterion condition in both cases.

Figure 5.8 shows the comparison of the reconstruction of the absorption coefficient μ using the algorithm 10 versus using the L-LK method for the computation of $\nabla \mathcal{J}_{[k]}(\sigma^k)$ (and the Monotonicity condition in both cases). A significant difference of using the LK

versus the L-LK method is seen in figure 5.7, while no significant difference of using one instead of the other method is seen when the weak monotonicity condition is used as step size criterion in each method as it can be seen in figure 5.8.

We have also tested the sparsity regularization algorithms 9 and 10 with the variant of the L-LK method, and 11 described in this chapter. As a benchmark we have used algorithm 2 (i.e., without sparsity regularization) . All these reconstructions are compared in figure 5.9. In practice, one would choose the algorithm 10 without using the weak monotonicity condition, since it retrieves acceptable reconstructions and is the fastest of those compared in this chapter.

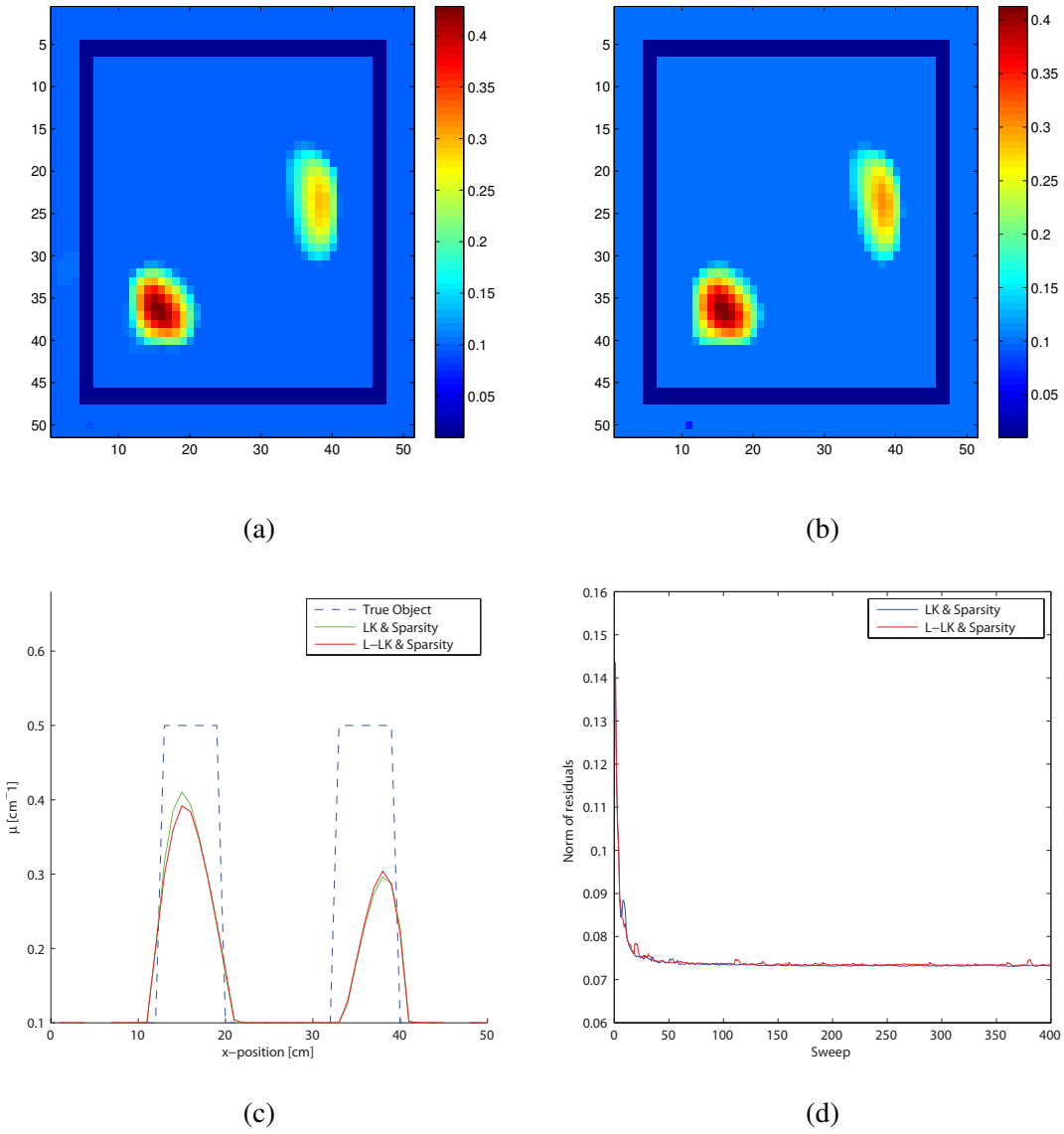
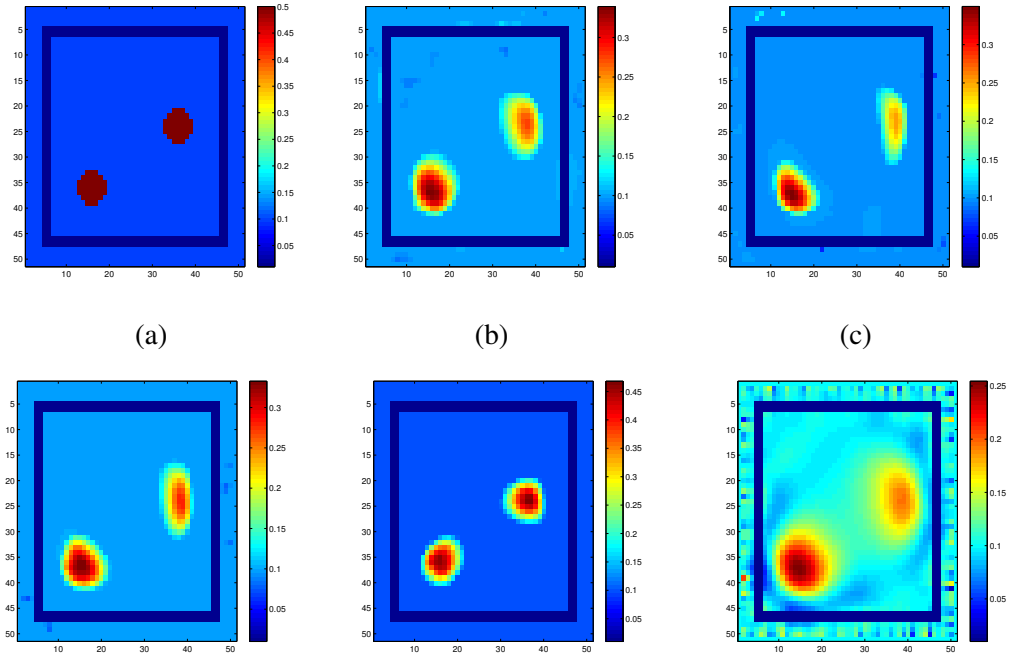


Figure 5.8: Figures 5.8a and 5.8b shows the reconstruction image using algorithm 10 (with the L-LK variant) and with the Weak Monotonicity in both cases. Figure 5.8c and figure 5.8d show the cross section and the residuals of the LK with sparsity regularization versus L-LK and sparsity regularization respectively with the Weak Monotonicity condition respectively for both cases.

Profiles in the x-direction are displayed in figure 5.10. Shown are the norm of the residuals, and the L^2-norm errors of the image reconstruction using the L, LK, L-LK method with sparsity regularization, the LK method with the own sparsity regularization and with only LK method without sparsity regularization, respectively. Figure 5.12 shows the robustness of the sparsity regularization without the weak monotonicity condition. As it can be seen in these last two figures, our proposed sparsity method behaves well for a large range of values of the regularization parameter α.



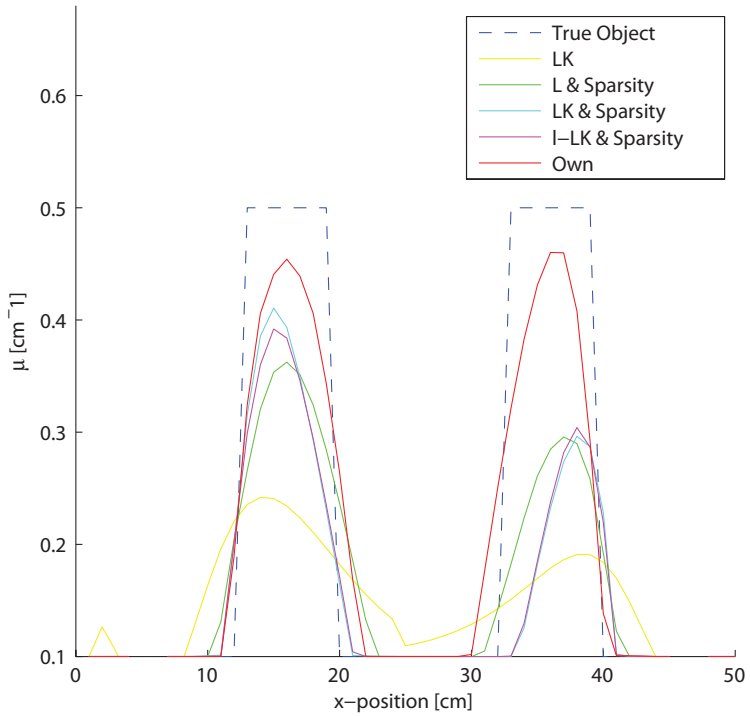
(a) (b) (c)

(d) (e) (f)

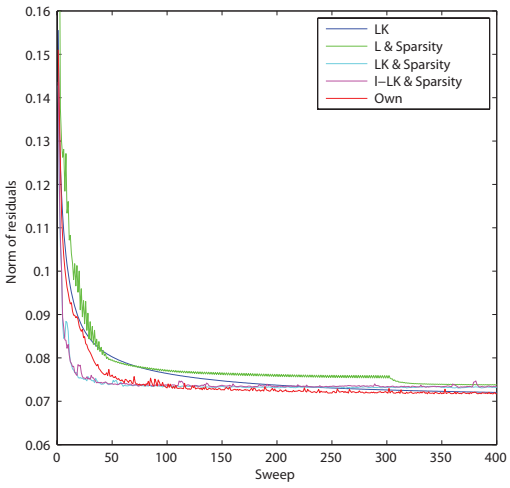
Figure 5.9: Figure 5.9a shows the true object, fig. 5.9b shows the reconstruction using algorithm 9, fig. 5.9c shows the reconstruction using algorithm 10, fig. 5.9d shows the reconstruction using the algorithm 10 with the L-LK variant. Figure 5.9e shows the reconstruction using the algorithm 11. Figure 5.9f shows the reconstruction using the algorithm 2 (only the LK method without sparsity regularization.

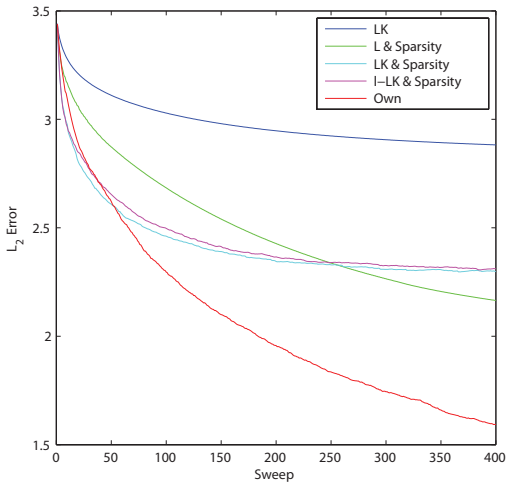
Figure 5.13 compares the reconstruction of the scattering coefficient using the algorithm 10, i.e., the LK method with sparsity regularisation, with our sparsity regularization proposed here using algorithm 11 versus using algorithm 2, i.e., only the LK method without sparsity regularisation. It also shows the profiles in the y-direction and the norm of the residuals of the reconstruction using the LK method with sparsity regularization, the LK method with our sparsity regularization, and with only LK method without sparsity regularization.

A simultaneous reconstruction of absorption and scattering coefficients using the algorithm 10 and 2 are presented in Figures 5.14 and 5.15, respectively. An accurate localization and estimate of the inclusions of the absorption and scattering coefficients using sparsity regularization are observed here compared to the exact ones, and compared to the smooth reconstruction using only the LK method without sparsity regularisation, see Figures 5.9, 5.10 and 5.13, respectively.



(a)





(b) (c)

Figure 5.10: Figure 5.10a shows the cross section of the final reconstructions using the methods of figure 5.9. Figure 5.10b shows their correspondent residuums. Figure 5.10c shows their correspondent L^2-errors using the methods of figure 5.9.

The location and estimate of the inclusion is well achieved using sparsity regularization for the simultaneous reconstruction of the absorption and scattering coefficients in comparison with using only the LK method, see figure 5.14 and 5.15. Tables 3.1 and 3.2 list the physical values used in the reconstruction of the absorption coefficient μ and

scattering coefficient b only. Table 5.1 lists the physical values of μ and b used in the simultaneous reconstruction of absorption and scattering coefficients as suggested in [142], using the LK method with sparsity regularization as shown in figure 5.14. Table 3.3 lists the physical values of μ and b used in the simultaneous reconstruction of absorption and scattering coefficients using only the LK method without sparsity regularization as shown in figure 5.15.

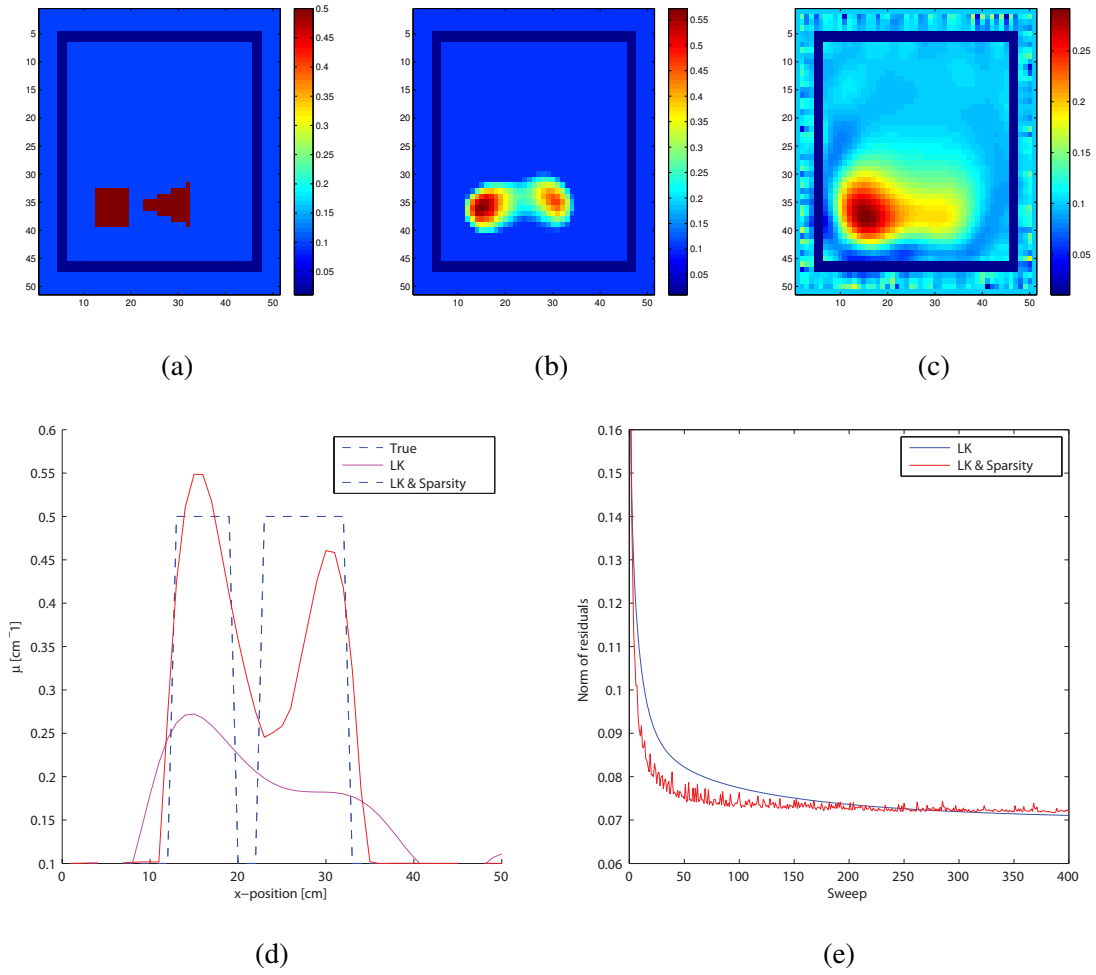


Figure 5.11: Figure 5.11a shows the true object, then reconstruction : 5.11b using algorithm 11, 5.11c using algorithm 2. Figure 5.11d shows the cross section of the final reconstructions using algorithms 10 and 2. Figure 5.11e shows their correspondent costs using LK method and LK method with sparsity regularization.

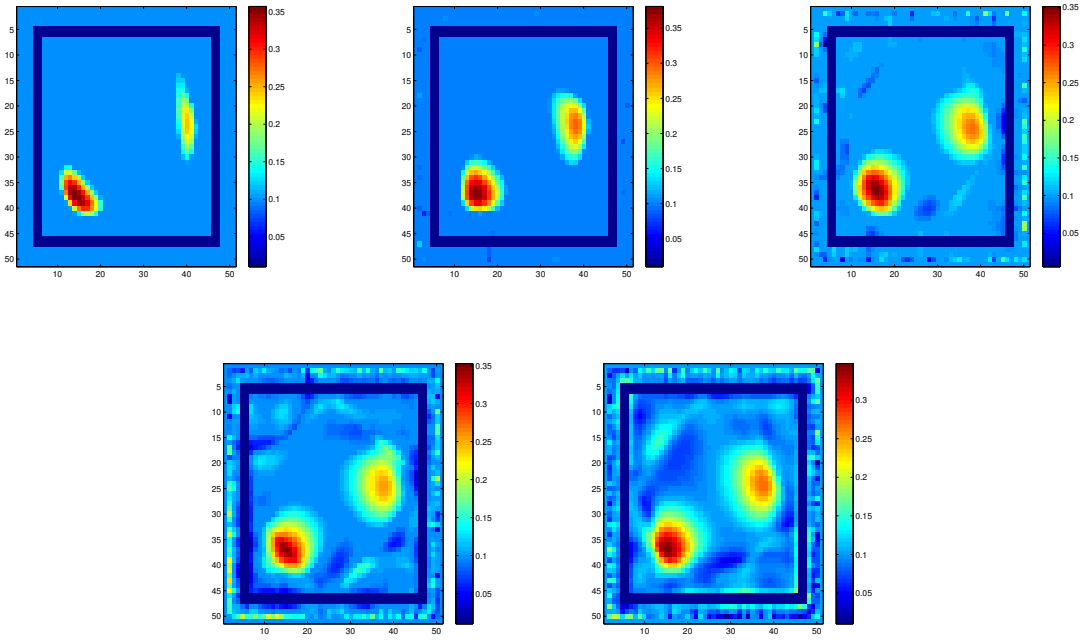


Figure 5.12: Reconstruction of absorption coefficient using algorithm 10 with various regularizer α values. Top row from left to right: $\alpha = 2 \times 10^{-4}$, $\alpha = 8 \times 10^{-5}$ and $\alpha = 1 \times 10^{-5}$. Bottom left: $\alpha = 5 \times 10^{-6}$, bottom right: $\alpha = 1 \times 10^{-6}$.

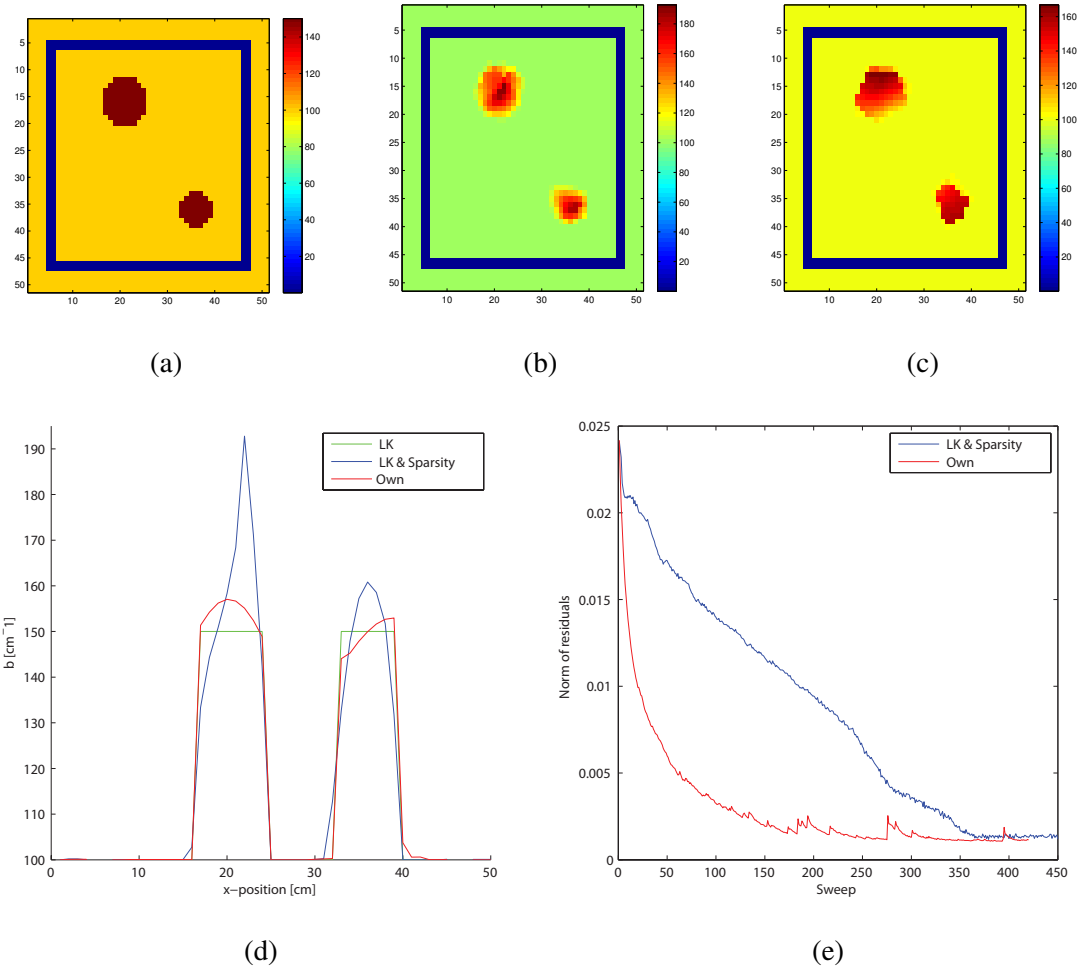


Figure 5.13: Figure 5.13a shows the true object, then figure 5.13b shows the reconstruction using using the algorithm 10, figure 5.13c shows the reconstruction using the algorithm 11. Figure 5.13d shows the cross section of the final reconstructions using only LK method, using LK with sparsity regularization and the LK with the own sparsity method. Figure 5.13e shows their correspondent residuums using LK method with the sparsity regularization and LK method with the own sparsity regularization.

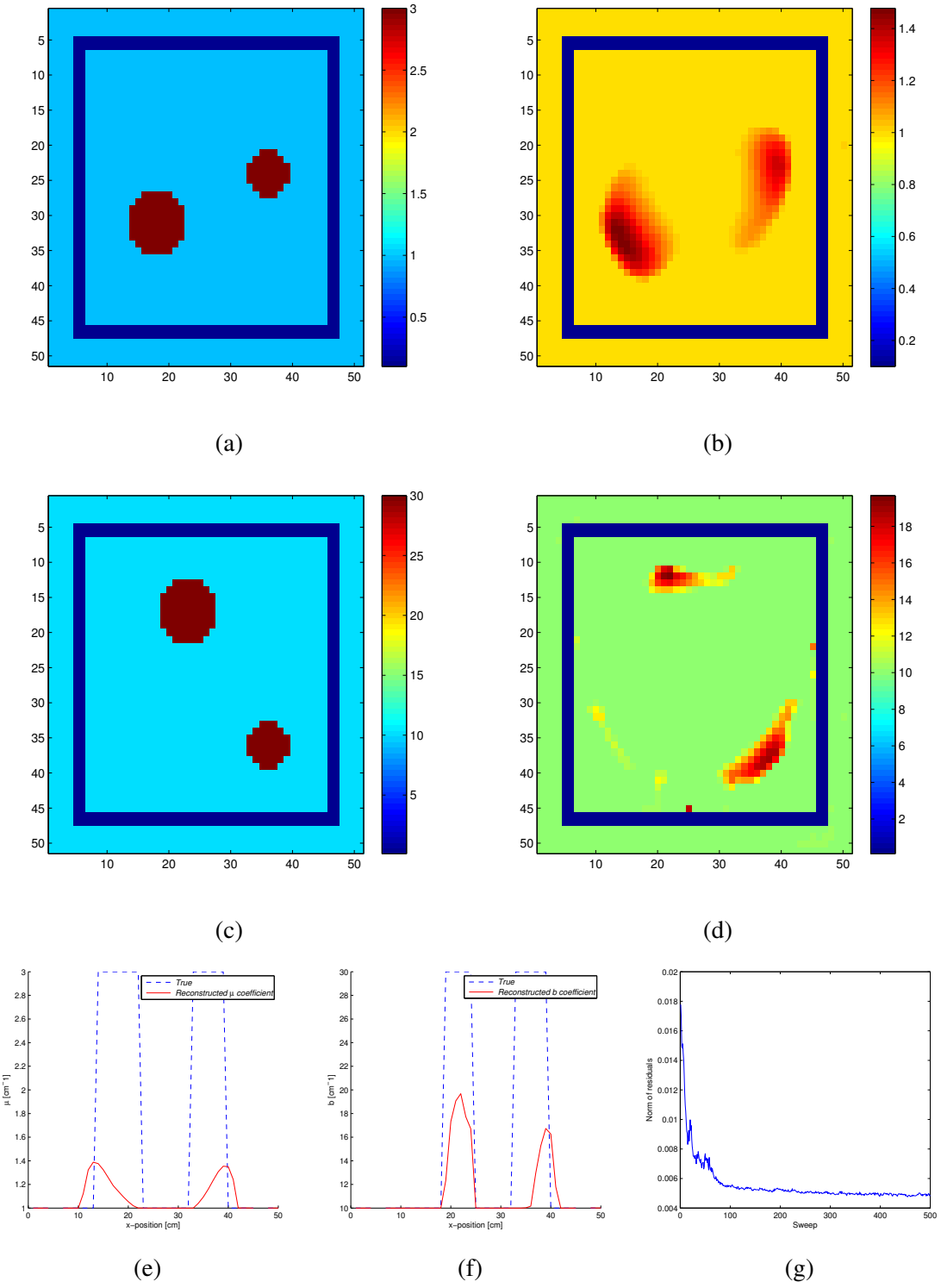


Figure 5.14: Figure 5.14a shows the μ true object, figure 5.14b shows the μ reconstruction using the algorithm 10, figure 5.14c shows the b true object, figure 5.14d shows the b reconstruction using LK method and sparsity regularization. Figure 5.14e shows the cross section of μ using the LK with sparsity regularization. Figure 5.14f shows the cross section of b using the LK with sparsity regularization. Figure 5.14g shows the residual of the simultaneous reconstruction of μ and b using the LK method and sparsity regularization.

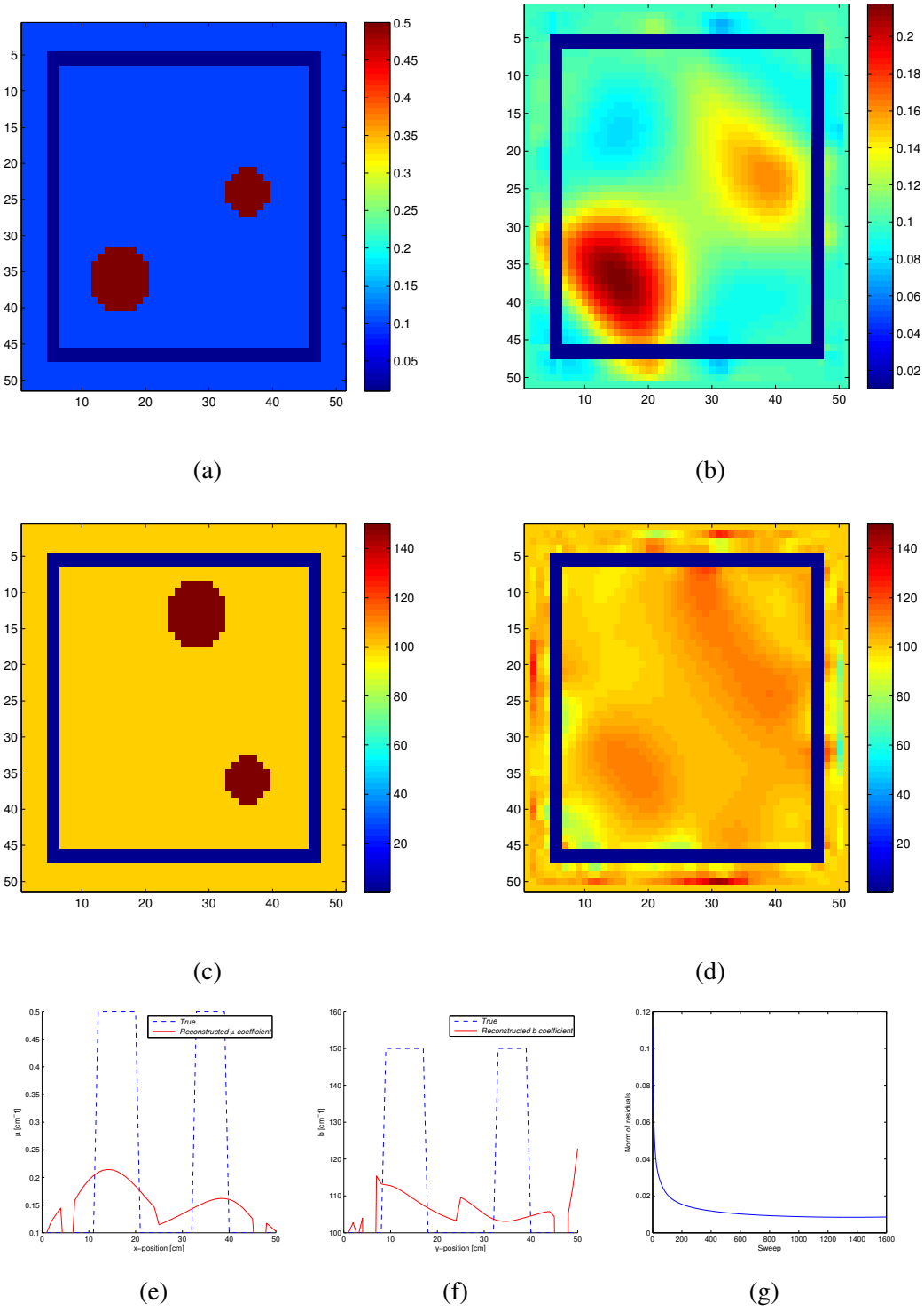


Figure 5.15: Figure 5.15a shows the μ true object, figure 5.15b shows the μ reconstruction using only the LK method, figure 5.15c shows the b true object, figure 5.15d shows the b reconstruction using only the LK method. Figure 5.15e shows the cross section of μ using only the LK. Figure 5.15f shows the cross section of b using only the LK method. Figure 5.15g shows the residual of the simultaneous reconstruction of μ and b reconstruction using only the LK method.

Chapter 6

Total variation regularization

In this chapter, first some basic concepts of the already classical total variation regularization are briefly reviewed. In section 6.1, the definition of total variation of a function is presented and discussed. Numerical algorithms for practically minimizing total variation functionals, however, are still object of active research. Therefore, in section 6.2 a more recent approach based on the Split Bregman Iteration is presented, which was also adopted for our numerical experiments applied in this thesis to the special case of DOT. Then, in section 6.5 some details of the corresponding discretization in one and two space dimensions is briefly presented. Afterwards, the lagged diffusivity algorithm is described in section 6.4 which was as well implemented in this thesis for the special case of DOT. In section 6.3 the *limited-memory BFGS with line search* technique is formulated which is used in our numerical experiments. Finally, some computational results of our numerical experiments are discussed in the last section.

6.1 Introduction

In this chapter we will consider the following Tikhonov-type functional

$$\mathcal{F}(\sigma) = \min_{\sigma} \mathcal{J}(\sigma) + \beta \mathcal{K}(\sigma), \tag{6.1}$$

where $\mathcal{J}(\sigma)$ is as defined in (3.6) and $\mathcal{K}(\sigma)$ is the penalty term which can be a convex and differentiable function and α is the regularizer parameter [162]. Let $\Omega \in \mathbb{R}^2$ and $\sigma \in L^1$,

then the *total variational* of σ is defined by

$$\mathcal{K}(\sigma) = \sup_{\varphi \in \mathscr{B}} \left\{ \int_{\Omega} \sigma \operatorname{div} \varphi \, dx \right\}, \tag{6.2}$$

where the space of test functions is given by

$$\mathscr{B} = \left\{ \varphi \in C_0^1(\Omega, \mathbb{R}^2) \mid |\varphi(x)|_{L^\infty} \leq 1 \quad \forall x \in \Omega \right\},$$

and $C_0^1(\Omega, \mathbb{R}^2)$ denotes the space of vector-valued functions $\varphi = (\varphi_1, \varphi_2)$ whose components φ_j are continuously differentiable and having compact support on Ω.

The Sobolev space $W^{1,1}$ denotes the closure of $C_0^1(\Omega)$ with respect to the norm

$$\|\upsilon\|_{1,1} = \int_{\Omega} \left[|\upsilon| + \sum_{j=1}^{2} \left| \frac{\partial \upsilon}{\partial x_j} \right| \right].$$

If we have $\sigma \in W^{1,1}$, then the total variation can be written as

$$\mathcal{K}(\sigma) = \int_{\Omega} |\nabla \sigma|. \tag{6.3}$$

where $|\nabla \sigma|$ is the Euclidean norm of the weak gradient $\nabla \sigma \in \mathbb{R}^2$ of the distributional derivative of σ. The expression (6.3) is also known as the *total variation semi-norm* of σ, denoted by $\|\sigma\|_{TV}$. Also, the *bounded variation (BV) semi-norm* is endowed with $\|\sigma\|_{BV} = \|\sigma\|_1 + \|\sigma\|_{TV}$ [172]. Since the total variation regularization $\mathcal{K}(\sigma)$ is the surface integral over the scalar field $|\nabla \sigma|$, Vogel gave an interpretation of the total variation regularization as measuring the changes or jumps on the lateral surfaces of the function σ on the domain Ω [162]. Thus, the total variations penalise the oscillations of $\sigma(x)$. Blocky reconstructions are obtained in many linear problems such as image denoising and deblurring after total variation regularization is applied. The classical Tikhonov regularization smooths out blocky structures, whereas total variation regularization is used to obtain satisfactory reconstructions of obstacles which have sharp features.

6.2 Split Bregman method

Osher and Goldstein consider in [83] the ℓ_1 regularization problem (6.1) with $\mathcal{K}(\sigma) = \|\Phi(\sigma)\|_1$, i.e.,

$$\min_{\sigma} \mathcal{J}(\sigma) + \|\Phi(\sigma)\|_1, \tag{6.4}$$

where $\|\Phi(\sigma)\|_1$ is considered as a penalty term, and $J(\sigma)$ is assumed to be a convex function. Also, they assume that Φ is differentiable. The key idea of the *Split Bregman method* is to de-couple the ℓ_1 ($\|\Phi(\sigma)\|_1$) and ℓ_2 ($\mathcal{J}(\sigma)$) terms of equation (6.4). Letting $w = \Phi(\sigma)$, the minimisation problem (6.4) is equivalent to the constrained problem

$$\min_{\sigma} \mathcal{J}(\sigma) + \|w\|_1 \quad \text{s.t.} \quad w = \Phi(\sigma). \tag{6.5}$$

Notice that the two terms in (6.5) do not depend directly on each other any longer, that is, these two terms have been split, which is the basic idea of the Split Bregman method. The unconstrained version of (6.5) with the substitution $w = \nabla\sigma$ is

$$\min_{\sigma,w} \mathcal{J}(\sigma) + \|w\|_1 + \frac{\beta}{2}\|w - \nabla\sigma\|_2^2 \tag{6.6}$$

If the Bregman iteration (5.7) is applied to (6.6) with $\mathcal{K}(\sigma) = \|w\|_1$, one obtains

$$(\sigma^{k+1}, w^{k+1}) = \min_{\sigma,w} D_{\mathcal{J}}^{p_\sigma^k}(\sigma, \sigma^k) + D_{\mathcal{K}}^{p_w^k}(w, w^k) + \frac{\beta}{2}\|w - \nabla\sigma\|_2^2, \tag{6.7a}$$

$$p_\sigma^k = p_\sigma^k - \beta\nabla^T(w^{k+1} - \nabla\sigma^{k+1}), \tag{6.7b}$$

$$p_w^k = p_w^k - \beta(w^{k+1} - \nabla\sigma^{k+1}), \tag{6.7c}$$

where $D_{\mathcal{J}}^{p_\sigma^k}(\sigma, \sigma^k)$, $D_{\mathcal{K}}^{p_w^k}(w, w^k)$ denotes the Bregman distances as defined in eq. (5.5), $\nabla^T := -\operatorname{div}$ denotes the adjoint of ∇ and the Laplacian is defined by $\Delta = -\nabla^T\nabla$. The minimisation problem (6.7) can be solved alternating the variable σ and w separately while keeping the other variable fixed, which results in a scheme presented in [17, 79, 83, 163, 165, 166, 172, 175, 176], namely

$$\sigma^{k+1} = \min_{\sigma} D_{\mathcal{J}}^{p_\sigma^k}(\sigma, \sigma^k) + \frac{\beta}{2}\left\|w^k - \nabla\sigma\right\|_2^2, \tag{6.8a}$$

$$p_\sigma^k = p_\sigma^k - \beta\nabla^T(w^{k+1} - \nabla\sigma^{k+1}), \tag{6.8b}$$

$$w^{k+1} = \min_{w} \; D_{\mathcal{K}}^{p_w^k}(w, w^k) + \frac{\beta}{2} \left\| w - \nabla\sigma^{k+1} \right\|_2^2, \tag{6.9a}$$

$$p_w^k = p_w^k - \beta(w^{k+1} - \nabla\sigma^{k+1}), \tag{6.9b}$$

Then, equation (6.8) and (6.9) address the ℓ_2 term $\mathcal{J}(\sigma)$ and ℓ_1 term $\mathcal{K}(\sigma)$ of the problem (6.4), respectively. Finally, the Split Bregman iteration is obtained by writing Equations (6.8) and (6.9) in the following form as shown in [83]

$$\sigma^{k+1} = \min_{\sigma} \; \mathcal{J}(\sigma) + \frac{\beta}{2} \left\| w^k - \nabla\sigma \right\|_2^2, \tag{6.10a}$$

$$w^{k+1} = \min_{w} \|w\|_1 + \frac{\beta}{2} \left\| w - \nabla\sigma^{k+1} \right\|_2^2. \tag{6.10b}$$

We mention that some authors have used an auxiliary variable v to enforce the constraint equality $w = \nabla\sigma$, which has the effect that the split Bregman formula (6.10) consists of three equations; however, the resulting expression with this auxiliary variable is equivalent to (6.10). The solution of the w-subproblem (6.10b) for fixed σ is given by

$$w^{k+1} = \bar{\mathcal{S}}_{1/\beta} \left(\nabla\sigma^{k+1} \right), \tag{6.11}$$

where $\bar{\mathcal{S}}_{1/\beta}$ is the *vectorial soft-thresholding function* defined for $\varkappa \in \mathbb{R}^2$ as

$$\bar{\mathcal{S}}_{\gamma}(\varkappa) = \begin{cases} \max\left(\|\varkappa\|_2 - \gamma, 0 \right) \dfrac{\varkappa}{\|\varkappa\|_2}, & \text{if } \varkappa \neq \mathbf{0}, \\ 0, & \text{if } \varkappa = \mathbf{0}. \end{cases}$$

Let $\mathcal{T}(\sigma)$ be equal to the right hand side of (6.10a), i.e.,

$$\mathcal{T}(\sigma) = \mathcal{J}(\sigma) + \frac{\beta}{2} \left\| w^k - \nabla\sigma \right\|_2^2. \tag{6.12}$$

Then the solution of the σ-subproblem (6.10a) for fixed w can be found by calculating the gradient of (6.10a) with respect to σ, which results in

$$\nabla\mathcal{T}(\sigma) = \nabla\mathcal{J}(\sigma) + \beta\nabla^T \left(w^k - \nabla\sigma \right). \tag{6.13}$$

Hence, the steepest descent and LK methods without inexact line search can be implemented in order to solve the σ-minimisation subproblem (6.10a). This results in

$$\sigma^{k+1} = \sigma^k - \tau\Big(\nabla J_{[k]}(\sigma^k) + \beta\nabla^T\left(w^k - \nabla\sigma^k\right)\Big), \tag{6.14a}$$

$$\nabla\sigma^k \cdot \bar{\nu} = 0 \quad \text{on } \Gamma, \tag{6.14b}$$

where $\nabla J_{[k]}(\sigma^k)$ is given by (3.19) with $[k] := (k \mod M)$ and M is the number of sources q_j.

A suitable selection of the regularizer parameter β is essential in order the Split Bregman iteration (6.10) numerically converges rapidly. For large values of β, the effect of the σ-subproblem (6.14) increases because the Laplacian term $\Delta\sigma$ generates heavy spatial interaction, and w has more influence on the solution σ because the w solution of (6.11) ($\nabla\sigma$) is shrank by small number $1/\beta$. While for small values of β, the behaviour of both $\sigma\backslash w$-subproblems is opposed. Therefore, β should be neither small nor rather large for rapid convergence.

We have worked with the following functional equivalent to (6.6) as suggested in [149]

$$\min_{\sigma,w} \frac{1}{\alpha}\mathcal{J}(\sigma) + \|w\|_1 + \beta\|\nabla\sigma - w\|_2^2,$$

which makes is easier to find the proper choice of values of the parameters α and β. Then, the expression for $\mathcal{T}(\sigma)$ is

$$\mathcal{T}(\sigma) = \frac{1}{\alpha}\mathcal{J}(\sigma) + \frac{\beta}{2}\left\|w^k - \nabla\sigma\right\|_2^2, \tag{6.15}$$

and the σ-subproblem to be solved becomes

$$\sigma^{k+1} = \sigma^k - \tau\left(\frac{1}{\alpha}\nabla J_{[k]}(\sigma^k) + \beta\nabla^T\left(w^k - \nabla\sigma^k\right)\right), \tag{6.16a}$$

$$\nabla\sigma^k \cdot \bar{\nu} = 0 \quad \text{on } \Gamma, \tag{6.16b}$$

where $\nabla J_{[k]}(\sigma^k)$ is given by (3.19) with $[k] := (k \mod M)$ and M is the number of sources q_j.

To accelerate the rate of convergence of the solution of (6.16), we start with small values of α and increase them until some acceptable convergence criterion is satisfied, as

described in the algorithm 13. Then, we use the continuation scheme on the parameter β in algorithm 13 to increase the rate of convergence of the algorithm 12 as tested in [149]. We start with small values of β at the beginning of the algorithm, and then increase them by a fixed number after a fixed number of iterations until a stopping criterion regarding the L^2 error norm of the residuum $\mathcal{J}(\sigma)$ is satisfied. This Split Bregman Iteration (6.10) without continuation has been used in [79, 113, 176]. Both schemes have been used in the numerical experiments presented further below in this chapter. Finally, we mention that a convergence analysis of the Split Bregman iteration for the nonlinear problem of DOT would be necessary, but it is out of the scope of this dissertation.

A Pseudo Code of the algorithm of the split Bregman iteration for total variation regularization reads as follows.

Algorithm 12: Split Bregman iteration for total variation regularization

Input: α, β, initial guess: $\sigma^0, w^0 = 0$

Output: $\sigma^{I(M-1)}$

for $k := 0, \ldots, I(M-1)$ **do**

Solve the w-subproblem (6.11);

Solve the σ-subproblem (6.16) ;

end

A Pseudo Code for the Split Bregman iteration using a continuation scheme for the total variation is given next.

Algorithm 13: Split Bregman iteration with continuation scheme for TV

Input: α, β, initial guess: $\sigma^0, w^0 = 0$ and $\beta_{\max}, \epsilon, l$

Output: $\sigma^{I(M-1)}$

while $\beta \leq \beta_{\max}$ **do**

  $\hat{\sigma} \leftarrow \sigma + 2\epsilon$;

  while $\|\sigma - \hat{\sigma}\| > \epsilon$ *or after* l *iterations* **do**

    Solve the w-subproblem (6.11);

    Solve the σ-subproblem (6.16);

  end

  $\hat{\sigma} \leftarrow \sigma$;

  $\beta = 2 \cdot \beta$;

end

6.3 Limited-memory BFGS algorithm

We present in this section the Limited-memory-BFGS formula which is used in the next section when employing a quasi-Newton step in the algorithm of the Lagged Diffusivity Method.

To start with, we state the Taylor theorem in the form of [119] which says that a function f can be approximated up to some degree of accuracy under some conditions on the function f

Theorem 26 *Assuming that $f : \mathbb{R}^N \to \mathbb{R}$ is twice continuously differentiable, then*

$$f(\sigma + p) \approx f(\sigma) + p^T \nabla f(\sigma) + \frac{1}{2} p^T \nabla^2 f(\sigma + tp) p, \tag{6.17}$$

for some $t \in (0, 1)$. Moreover, the following equation holds

$$\nabla f(\sigma + p) = \nabla f(\sigma) + \int_0^1 \nabla^2 f(\sigma + tp) p \, dt \tag{6.18}$$

for some $t \in (0, 1)$.

Let $M_k(p)$ be the Taylor series expansion of $f(\sigma^k + p)$ up to second order,

$$M_k(p) := f(\sigma^k) + p^T \nabla f(\sigma^k) + \frac{1}{2} p^T \nabla^2 f(\sigma^k) p.$$

Then, the Newton direction is the vector p which minimises the function $M_k(p)$ with respect to p, i.e.,

$$p^k = -\left(\nabla^2 f^k\right)^{-1} \nabla f^k, \tag{6.19}$$

where the notation $f^k = f(\sigma^k)$ has been used. An approximation B_k can be used instead of the true Hessian $\nabla^2 f^k$ which is derived using (6.18). After adding the term $\nabla^2 f(\sigma)p$ to (6.18), one obtains

$$\nabla f(\sigma + p) = \nabla f(\sigma) + \nabla^2 f(\sigma)p + \int_0^1 \left[\nabla^2 f(\sigma + tp) - \nabla^2 f(\sigma)\right] p \, dt. \tag{6.20}$$

The order of the second term on the right hand side of (6.20) is $o\left(\|\sigma^{k+1} - \sigma^k\|\right)$ since f is assumed twice continuously differentiable. Thus, letting $\sigma = \sigma^k$ and $p = \sigma^{k+1} - \sigma^k$, we obtain

$$\nabla f^{k+1} = \nabla f^k + \nabla^2 f^{k+1}(\sigma^{k+1} - \sigma^k) + o\left(\|\sigma^{k+1} - \sigma^k\|\right).$$

Then it follows

$$\nabla^2 f^{k+1}(\sigma^{k+1} - \sigma^k) \approx \nabla f^{k+1} - \nabla f^k, \tag{6.21}$$

provided σ^{k+1} is sufficiently close to σ^k. By setting

$$\begin{aligned} B_{k+1} &:= \nabla^2 f^{k+1}, \\ s_k &:= \sigma^{k+1} - \sigma^k, \\ r_k &:= \nabla f^{k+1} - \nabla f^k, \end{aligned}$$

the equation (6.21) can be written as

$$B_{k+1} s_k = r_k. \tag{6.22}$$

Therefore, using the approximation B_k of the true Hessian, equation (6.19) is written as

$$p^k = -B_k^{-1} \nabla f^k, \tag{6.23}$$

with the update iteration rule

$$\sigma^{k+1} = \sigma^k + \tau^k p^k. \tag{6.24}$$

The step size τ^k is chosen to be the step size of the inexact line search of the Newton method. In our numerical experiments, we have used Brent's Method given in [136] to perform the inexact line search. As stated in [119], equation (6.22) holds only if the following so-called *curvature condition* is satisfied

$$s_k^T r^k > 0. \tag{6.25}$$

This curvature condition holds when the following Wolfe conditions are satisfied

$$f(\sigma^k + \tau^k p^k) \leq f(\sigma^k) + c_1 \tau^k (\nabla f^k)^T p^k, \tag{6.26a}$$

$$\left(\nabla f(\sigma^k + \tau^k p^k)\right)^T p^k \geq c_2 \tau^k (\nabla f^k)^T p^k, \tag{6.26b}$$

with $0 < c_1 < c_2 < 1$. Both Wolfe conditions (6.26) are satisfied if the first Wolfe condition (6.26a) holds and using the below backtracking scheme [119], which is given by the below stated algorithm. We mention that, in practice, we have added a counter in the statement of the "while" cycle of the backtracking scheme such that this cycle ends with no more than 6 iterations.

We state now the Pseudo-code of the Backtracking Line-Search.

Algorithm 14: Backtracking Line Search

Input: $\bar{\tau} > 0$ given by the inexact line search using Brent's method and $0 < \iota, c_1 < 1$

Output: $\tau^k := \tau$

Set $\tau = \bar{\tau}$;

while $f(\sigma^k + \tau p^k) > f(\sigma^k) + c_1 \tau (\nabla f^k)^T p^k$ **do**

$\tau = \iota \tau$;

end

Putting $H_k := B_k^{-1}$, equation (6.24) becomes

$$\sigma^{k+1} = \sigma^k - \tau^k H_k \nabla f^k,$$

where we are using the Sherman-Morrison-Woodbury (SMW) update formula for H_k

$$H_{k+1} = V_k^T H_k V_k + \gamma_k s_k s_k^T, \tag{6.27}$$

with

$$\gamma_k = \frac{1}{r_k^T s_k}, \qquad V_k = I - \gamma_k r_k s_k^T. \tag{6.28}$$

We assume that $\sigma = \sigma_{i,j}$ is defined on an equispaced mesh in two dimensions, $\{(x_i, y_i) \mid x_i = i\Delta x, y_j = j\Delta y, i = 1, \ldots, N_x, \ j = 1, \ldots, N_y\}$. However, the last expression is not convenient to use because $H_k \in \mathbb{R}^{N_x \times N_y}$ is generally very large. Thus, a modified version of the SMW formula is often used to overcome this inconvenience which uses only a number of m vectors s_k and r_k at each iteration. The optimal choice of m is problem-dependent, however, selecting m between 5 and 10 has given satisfactory convergence results as reported in [119, 149]. The approximation to the inverse of the Hessian H_k is updated by (6.27), using m pairs s_i, y_i, starting from an initial inverse Hessian approximation H_k^0 which we assume to be in iteration k

$$H_k^0 = \rho_k I, \tag{6.29}$$

where

$$\rho_k = \frac{s_{k-1}^T r_{k-1}}{r_{k-1}^T r_{k-1}}.$$

Then, at each iteration $k > m$, the oldest pair of s_i, r_i,i.e., s_{k-m}, r_{k-m} is discarded from storage.

A Pseudo Code of this process is explained in the following algorithm given by Broyden, Fletcher, Goldfarb, and Shanno.

Algorithm 15: L-BFGS

Input: $\nabla f(\sigma^k)$, H_k^0 and $m \in \mathbb{Z}^+$

Output: $H_k \nabla f(\sigma^k)$

$q = \nabla f(\sigma^k)$;

for $i = k-1, k-2, \ldots, k-m$ **do**

$\lambda_i = \gamma_i s_i^T q$;

$q = q - \lambda_i r_i$;

end

$\zeta = H_k^0 q$;

for $i = k-m, k-m+1, \ldots, k-m$ **do**

$\varpi = \gamma_i r_i^T \zeta$;

$\zeta = \zeta + s_i(\lambda_i - \varpi)$;

end

$H_k \nabla f(\sigma^k) = \zeta$

Finally, a Pseudo Code of the Limited-memory BFGS Method with inexact line search is given in the next algorithm.

Algorithm 16: L-BFGS Method with inexact line search

Input: σ^k and integer $m \in \mathbb{Z}^+$

Output: σ^{k+1}

repeat

H_k^0 given by (6.29) ;

$p_k = -H_k \nabla f(\sigma^k)$ using algorithm 15 ;

Compute τ_{k+1} using algorithm 14 ;

$\sigma^{k+1} = \sigma^k + \tau^k p^k$;

until ***Convergence***;

6.4 Lagged Diffusivity

We will used the lagged diffusivity approach presented in [79, 162] to solve the minimisation problem (6.1). After taking minus times the gradient of (6.1), one obtains

$$-\nabla\mathcal{F}(\sigma) = -\nabla\mathcal{J}(\sigma) + \alpha\nabla\cdot\frac{\nabla\sigma}{|\nabla\sigma|}.$$

The descent direction method for solving (6.1) can be viewed as the solution of the following diffusive equation with homogeneous Neumann boundary conditions

$$\begin{aligned}\frac{\partial\sigma}{\partial t} &= -\nabla\mathcal{J}(\sigma) + \alpha\nabla\cdot\frac{\nabla\sigma}{|\nabla\sigma|}\\ \nabla\sigma\cdot\bar{\nu} &= 0 \quad \text{on } \Gamma,\end{aligned} \tag{6.30}$$

where $\bar{\nu}$ is the normal vector on Γ and $|\cdot|$ stands for the Euclidean norm. The discretized time step Δt can be set to a fixed number for this method or equal to the variable step size τ if a line search is computed.

The following scheme for solving the diffusion equation (6.30) numerically is presented in [79].

$$\begin{aligned}\sigma_{i,j}^{n+1} = \sigma_{i,j}^{n+1} + \Delta t\Bigg[&-\nabla\mathcal{J} + \alpha\nabla_x^-\left(\frac{\nabla_x^+\sigma_{i,j}^n}{\sqrt{(\nabla_x^+\sigma_{i,j}^n)^2 + (\mathfrak{m}(\nabla_x^+\sigma_{i,j}^n, \nabla_x^-\sigma_{i,j}^n)^2}}\right)\\ &+\alpha\nabla_y^-\left(\frac{\nabla_y^+\sigma_{i,j}^n}{\sqrt{(\nabla_y^+\sigma_{i,j}^n)^2 + (\mathfrak{m}(\nabla_y^+\sigma_{i,j}^n, \nabla_y^-\sigma_{i,j}^n)^2}}\right)\Bigg], \quad i,j = 1,...,N-1,\\ \sigma_{0,j}^n = \sigma_{1,j}^n, \quad \sigma_{N,j}^n = \sigma_{N-1,j}^n, \quad &\sigma_{i,0}^n = \sigma_{i,1}^n, \quad \sigma_{i,N}^n = \sigma_{i,N-1}^n, \quad i,j = 0,...,N,\end{aligned} \tag{6.31}$$

where the third row in (6.31) represents the homogeneous Neumann boundary condition, and $\nabla_x^+, \nabla_x^-, \nabla_y^+, \nabla_y^-$ denote the forward and backward differences in the x and y directions, respectively. Moreover, $\mathfrak{m}(a,b)$ denotes the minmod function defined by

$$\mathfrak{m}(a,b) = \left(\frac{\text{sign}(a) + \text{sign}(b)}{2}\right)\min\left(|a|,|b|\right).$$

In case that σ is locally constant, the quotients of the first line in (6.31) are computed as equal to 0, therefore $\sigma_{i,j}^{n+1} = \sigma_{i,j}^n$.

The second term of the right hand side of (6.30), which is minus times the gradient of the regularizing term

$$-\nabla\mathcal{K}(\sigma) = \nabla\cdot\frac{\nabla\sigma}{|\nabla\sigma|}, \tag{6.32}$$

can be interpreted as the curvature of the level curve of the function $\sigma(x)$ at the point x as pointed it out by Osher and Tsai [129]. If equation (6.30) is multiplied by $|\nabla\sigma|$ as Osher and Tsai suggest in [129], the behaviour of the diffusion equation (6.30) is equivalent to the behaviour of the evolution of the following level set equation in the variable σ

$$\partial_t\sigma = \left(\nabla\cdot\frac{\nabla\sigma}{|\nabla\sigma|}\right)|\nabla\sigma|, \tag{6.33}$$

which evolves under a mean curvature term for the level set function σ. The solution of equation (6.33) approaches the solution of (6.32) as long as the gradient of σ is nonzero. Then, it an interpretation can be given of the removal of noise after applying a total variation regularization method as follows; the level curves around noise artefacts have high curvatures which shrink at speed of the mean curvature and eventually disappear. Osher and Tsai in [129] also proposed a scheme for discretizing the curvature term

$$\kappa = \nabla\cdot\frac{\nabla\sigma}{|\nabla\sigma|}. \tag{6.34}$$

as follows. Firstly, the norm of the gradient $|\nabla\sigma|$ is in the denominator of the mean curvature, which can be approximately zero when the function σ is nearly constant. Hence it is approximated by

$$|\nabla\sigma|_\epsilon = (|\nabla\sigma|^2+\epsilon^2)^{1/2}. \tag{6.35}$$

Doing this replacement, the total variation defined in (6.3) is differentiable at the origin. The curvature $\kappa_{i,j}$ is discretized by

$$\kappa_{i,j} = \frac{n^x_{i+1/2,j}-n^x_{i-1/2,j}}{\Delta x} + \frac{n^y_{i,j+1/2}-n^y_{i,j-1/2}}{\Delta y}, \tag{6.36}$$

where the components n^x, n^y of the normal vector $\bar{n}$ of the domain Ω are given by

$$n^x_{i\pm 1/2,j} := \frac{\nabla^{\pm}_x \sigma_{i,j}}{\sqrt{(\nabla^{\pm}_x \sigma_{i,j})^2 + \nabla^0_y (B^{\pm}_x \sigma_{i,j})^2 + \epsilon^2}},$$
$$n^y_{i,j\pm 1/2} := \frac{\nabla^{\pm}_y \sigma_{i,j}}{\sqrt{\nabla^0_x (B^{\pm}_y \sigma_{i,j})^2 + (\nabla^{\pm}_y \sigma_{i,j})^2 + \epsilon^2}},$$

where

$$B^{\pm}_x \sigma_{i,j} = \frac{\sigma_{i\pm,j} + \sigma_{i,j}}{2}, \quad \text{and} \quad B^{\pm}_y \sigma_{i,j} = \frac{\sigma_{i,j\pm 1} + \sigma_{i,j}}{2},$$

are the arithmetic averaging differences in the x and y direction, respectively, and ∇^0_x, ∇^0_y denote the central differences in the x and y direction, respectively. We have also used the definition of the curvature as given in [146], which is

$$\kappa = \frac{\sigma_{xx}\sigma_y^2 - 2\sigma_{xy}\sigma_x\sigma_y + \sigma_{yy}\sigma_x^2}{\left(\sigma_x^2 + \sigma_y^2\right)^{3/2}}. \tag{6.37}$$

We have discretized the curvature given above using central finite differences in both directions x and y. Other approaches for solving the problem (6.1) are suggested in [2, 17, 70, 162, 174], as follows. Since $|\nabla\sigma|$ is not differentiable at the origin as it was mentioned before, the total variation term $\mathcal{K}(\sigma)$ is replaced by

$$\mathcal{K}(\sigma) = \int_\Omega |\nabla\sigma|_\epsilon \, ds, \tag{6.38}$$

where $|\nabla\sigma|_\epsilon$ is given in (6.35). Also, the above total variation term can be approximated by

$$\mathcal{K}(\sigma) = \int_\Omega \vartheta(\nabla\sigma) \, ds, \tag{6.39}$$

where $\vartheta(\zeta)$ is a smooth and convex function which may be given by

$$\vartheta(\zeta) = \sqrt{\zeta^2 + \epsilon^2},$$

or by the Huber function given in (5.11), see figure (5.3), both with the property

$$\vartheta'(\zeta) > 0 \quad \text{whenever} \quad \zeta > 0. \tag{6.40}$$

We now calculate the gradient of the approximation of the total variation given in

(6.39) as follows. We assume that the function ϑ in (6.39) can be written as

$$\vartheta(\zeta) = \hat{\vartheta}(|\zeta|^2),$$

where $\hat{\vartheta} : \mathbb{R} \to \mathbb{R}$ is an increasing and differentiable function. Note that both possible candidates, the Huber function and $\vartheta = \sqrt{\zeta^2 + \beta}$, can be written in the above representation.

For a smooth function υ and $\tau \in \mathbb{R}$, we have

$$\frac{d}{d\tau}\mathcal{K}(\sigma + \tau\upsilon)\Big|_{\tau=0} = \int_\Omega \vartheta'\big(|\nabla\sigma|^2\big)(\nabla\sigma)^T\nabla\upsilon. \tag{6.41}$$

Integrating by parts, one obtains

$$\nabla\mathcal{K} = \mathcal{L}(\sigma)\sigma, \tag{6.42}$$

where $\mathcal{L}(\sigma)$ is a diffusion operator with Neumman boundary condition given by

$$\mathcal{L}(\sigma)\varsigma = \operatorname{div}\Big(\vartheta'\big(|\nabla\sigma|^2\big)\nabla\varsigma\Big), \quad x \in \Omega, \tag{6.43a}$$

$$(\nabla\varsigma)^T\bar{\nu} = 0, \quad x \in \Gamma. \tag{6.43b}$$

Note that the diffusion coefficient $\vartheta'\big(|\nabla\sigma|^2\big)$ is positive since ϑ is an increasing function.

For a smooth function υ, w and $\tau, \epsilon \in \mathbb{R}$, the Hessian of the approximation of the total variation given in (6.39) is

$$\begin{aligned}\frac{\partial^2}{\partial\tau + \partial\epsilon}\mathcal{K}(\sigma + \tau\upsilon\epsilon w)\Big|_{\tau,\epsilon=0} &= \int_\Omega \vartheta'\big(|\nabla\sigma|^2\big)(\nabla\upsilon)^T\nabla w + 2\vartheta''\big(|\nabla\sigma|^2\big)\big((\nabla\upsilon)^T\nabla\sigma\big)\\ &\qquad \big((\nabla\sigma)^T\nabla\upsilon\big),\\ &= (\nabla\upsilon)^T\Big[\vartheta'\big(|\nabla\sigma|^2\big)\mathcal{I} + 2\vartheta''\big(|\nabla\sigma|^2\big)\nabla\sigma(\nabla\sigma)^T\Big]\nabla w,\end{aligned} \tag{6.44}$$

where $\mathcal{I}$ denotes the identity map. Integration by parts yields

$$\operatorname{Hess}\mathcal{K}(\sigma)\varsigma = -\operatorname{div}\Big\{\Big[\vartheta'\big(|\nabla\sigma|^2\big)\mathcal{I} + 2\vartheta''\big(|\nabla\sigma|^2\big)\nabla\sigma(\nabla\sigma)^T\Big]\nabla\varsigma\Big\}, \quad x \in \Omega, \tag{6.45a}$$

$$(\nabla\varsigma)^T\bar{\nu} = 0, \quad x \in \Gamma. \tag{6.45b}$$

Using (6.42) and (6.45), the gradient and the Hessian of $\mathcal{F}(\sigma)$ are given by

$$\nabla \mathcal{F}(\sigma) = \nabla J(\sigma) + \alpha \mathcal{L}(\sigma)\sigma, \tag{6.46}$$

and

$$\operatorname{Hess} \mathcal{F}(\sigma) = \operatorname{Hess} \mathcal{J}(\sigma) + \alpha \mathcal{L}(\sigma) + \alpha \mathcal{L}'(\sigma)\sigma, \tag{6.47}$$

respectively, where $\nabla J(\sigma)$ is given in (3.15).

To start our description of the technique, the gradient of the approximation of the total variation given in (6.39) is discretized in one dimension as shown in [162] in order to understand better the discretization of it in two dimensions. Let $\sigma(x)$ be a function defined on $\mathbb{R}^1$ and given by $\sigma = (\sigma_1, \ldots, \sigma_N)$ with $\sigma_i = \sigma(x_i)$, $x_i = i\Delta x$. Let the approximation $\mathfrak{D}$ of the derivative be defined by

$$\mathfrak{D}\sigma = \frac{1}{\Delta x} \begin{bmatrix} \mathfrak{D}_1 \sigma \\ \vdots \\ \mathfrak{D}_{N-1}\sigma \end{bmatrix}, \tag{6.48}$$

where

$$\mathfrak{D}_i f = f_{i+1} - f_i \quad \forall i = 1, \ldots, N-1, \tag{6.49}$$

has the $N \times 1$ matrix representation, $\mathfrak{D}_i = [0, \ldots, 0, -1, 1, 0, \ldots, 0]$, and $\mathfrak{D}$ is the $(N-1) \times N$ matrix whose ith row is $\mathfrak{D}_i$.

The discretization of the total variation $\mathcal{K} : \mathbb{R}^N \to \mathbb{R}$ given in (6.39) is

$$\mathcal{K}(\sigma) = \sum_{i=1}^{N-1} \vartheta\big((\mathfrak{D}\sigma)^2\big) \tag{6.50}$$

Then, for any $\upsilon \in \mathbb{R}^N$ and $\tau \in \mathbb{R}$, the gradient of $\mathcal{K}(\sigma)$ is given by

$$\begin{aligned} \frac{d}{d\tau}\mathcal{K}(\sigma + \tau\upsilon)\Big|_{\tau=0} &= \sum_{i=1}^{N-1} \vartheta'\big((\mathfrak{D}\sigma)^2\big)(\mathfrak{D}_i\sigma)(\mathfrak{D}_i\upsilon) \\ &= (\mathfrak{D}\upsilon)^T \operatorname{diag}\left(\vartheta'(\sigma)\right)\mathfrak{D}\sigma \\ &= \Big\langle \mathfrak{D}^T \operatorname{diag}\left(\mathfrak{D}'(\sigma)\right)\mathfrak{D}\sigma, \upsilon \Big\rangle, \end{aligned} \tag{6.51}$$

where $\operatorname{diag}\left(\vartheta'(\sigma)\right)$ denotes the $(N-1)\times(N-1)$ diagonal matrix with elements $\vartheta'(\mathfrak{D}_i\sigma)^2$

in its diagonal, and $\langle \cdot, \cdot \rangle$ denotes the Euclidean inner product on $\mathbb{R}^{N-1}$. The gradient of $\mathcal{K}$ can be written as

$$\nabla \mathcal{K}(\sigma) = \mathcal{L}(\sigma)\sigma, \tag{6.52}$$

where $\mathcal{L}(\sigma)$ is a symmetric $(N-1) \times (N-1)$ matrix defined by

$$\mathcal{L}(\sigma) = \mathfrak{D}^T \operatorname{diag}(\vartheta'(\sigma))\mathfrak{D}. \tag{6.53}$$

$\mathcal{L}(\sigma)$, which is positive semidefinite if condition (6.40) is satisfied.

Then, for $v, w \in \mathbb{R}^N$ and $\tau, \epsilon \in \mathbb{R}$, the Hessian of $\mathcal{K}$ is calculated as follows

$$\begin{aligned}
\frac{\partial^2}{\partial\tau\partial\epsilon}\mathcal{K}(\sigma + \tau v + \epsilon w)\Big|_{\tau,\epsilon=0} &= \sum_{i=1}^{N-1} \vartheta'\left((\mathfrak{D}_i\sigma)^2\right)(\mathfrak{D}_i v)(\mathfrak{D}_i w) \\
&\quad + \sum_{i=1}^{N-1} \vartheta''\left((\mathfrak{D}_i\sigma)^2\right)(\mathfrak{D}_i\sigma)(\mathfrak{D}_i v)(\mathfrak{D}_i w)(2\mathfrak{D}_i\sigma) \\
&= \left\langle \left[\mathfrak{D}^T \operatorname{diag}\left(\vartheta'(\sigma)\right) + \operatorname{diag}(2(\mathfrak{D}\sigma)^2\vartheta''(\sigma))\right]\mathfrak{D}v, \mathfrak{D}w \right\rangle,
\end{aligned}$$

where $\operatorname{diag}(2(\mathfrak{D}\sigma)^2\vartheta''(\sigma))$ denotes the $(N-1) \times (N-1)$ diagonal matrix with elements $2(\mathfrak{D}_i\sigma)^2\vartheta''((\mathfrak{D}_i\sigma)^2)$ in its diagonal. Hence, the Hessian is given by

$$\operatorname{Hess}\mathcal{K}(\sigma) = \mathcal{L}(\sigma) + \mathcal{L}'(\sigma)\sigma, \tag{6.54}$$

where $\mathcal{L}(\sigma)$ as in (6.53) and

$$\mathcal{L}'(\sigma)\sigma = \mathfrak{D}^T \operatorname{diag}(2(\mathfrak{D}f)^2\vartheta''(\sigma))\mathfrak{D}.$$

Now we proceed to the description in two space dimensions, which is similar to the procedure described above for one dimension. We want to obtain the gradient and the Hessian of $\mathcal{K}(\sigma)$ in two dimensions. Assuming $\sigma = \sigma_{i,j}$ is defined on an equi-spaced mesh in two dimensions, $\{(x_i, y_i) \mid x_i = i\Delta x, y_j = j\Delta y, i = 1, \ldots, N_x,\ j = 1, \ldots, N_y\}$. Then, the total variation term $\mathcal{K} : \mathbb{R}^{N_x \times N_y} \to \mathbb{R}$ in two dimensions is given by

$$\mathcal{K}(\sigma) = \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} \vartheta\Big(([\mathfrak{D}_x\sigma]_{ij})^2 + ([\mathfrak{D}_y\sigma]_{ij})^2\Big),$$

where

$$[\mathfrak{D}_x\sigma]_{ij} = \sigma_{i,j} - \sigma_{i-1,j}, \quad 1 \le i \le N_x - 1, \quad 1 \le j \le N_y, \tag{6.55a}$$

$$[\mathfrak{D}_y\sigma]_{ij} = \sigma_{i,j} - \sigma_{i,j-1}, \quad 1 \le i \le N_x, \quad 1 \le j \le N_y - 1. \tag{6.55b}$$

Let $\mathfrak{D}$ be the discrete derivative

$$\mathfrak{D} = \begin{bmatrix} \mathfrak{D}_x \\ \mathfrak{D}_y \end{bmatrix} : \mathbb{R}^{N_x \times N_y} \to \mathbb{R}^{(N_x-1)\times(N_y-1)} \times \mathbb{R}^{(N_x-1)\times(N_y-1)} \tag{6.56}$$

where the components $\mathfrak{D}_x$ and $\mathfrak{D}_y$ denote the $(N_x-1)(N_y-1) \times N_xN_y$ matrices corresponding to the discrete gradients given in (6.55) on directions x and y, respectively. Let $\langle\langle f, g\rangle\rangle$ be the Frobenius inner product in $\mathbb{R}^{(N_x-1)\times(N_y-1)}$ which is defined by

$$\langle\langle f, g\rangle\rangle = \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} f_{i,j}g_{i,j}.$$

The adjoint of $\mathfrak{D}$, $\mathfrak{D}^* : \mathbb{R}^{(N_x-1)\times(N_y-1)} \times \mathbb{R}^{(N_x-1)\times(N_y-1)} \to \mathbb{R}^{N_x\times N_y}$ is minus times the discrete divergence operator

$$\left\{\mathfrak{D}^* \begin{bmatrix} v \\ w \end{bmatrix}\right\}_{ij} = v_{i-1,j} - v_{i,j} + w_{i,j-1} - w_{i,j}, \quad 2 \le i \le N_x, \quad 2 \le j \le N_y - 1. \tag{6.57}$$

Here we define $v_{1,j} = v_{N_x,j} = v_{i,N_y} = 0$ and $w_{i,1} = w_{i,N_y} = w_{N_x,j} = 0$. From the above definition of $\mathfrak{D}^*$, we have

$$\mathfrak{D}^* = \begin{bmatrix} \mathfrak{D}_x^T & \mathfrak{D}_y^T \end{bmatrix}.$$

Thus, the gradient of $\mathcal{K}$ is given by

$$\begin{aligned} \frac{d}{d\tau}\mathcal{K}(\sigma + \tau v)\Big|_{\tau=0} &= \sum_{i=1}^{N_x-1}\sum_{j=1}^{N_y-1} \vartheta'\Big([\mathfrak{D}_x\sigma]_{ij})([\mathfrak{D}_x v]_{ij}) + ([\mathfrak{D}_y\sigma]_{ij})([\mathfrak{D}_y v]_{ij})\Big], \\ &= \Big\langle\Big\langle \operatorname{diag}\left(\vartheta'(|\mathfrak{D}\sigma|)^2\right)\mathfrak{D}_x\sigma, \mathfrak{D}_x v\Big\rangle\Big\rangle + \Big\langle\Big\langle \operatorname{diag}\left(\vartheta'(|\mathfrak{D}\sigma|)^2\right)\mathfrak{D}_y\sigma, \mathfrak{D}_y v\Big\rangle\Big\rangle, \\ &= \Big\langle\Big\langle \mathfrak{D}^*\vartheta'(|\mathfrak{D}\sigma|^2)\mathfrak{D}\sigma, v\Big\rangle\Big\rangle, \end{aligned}$$

where $\vartheta'(|\mathfrak{D}\sigma|^2)$ denotes the $(N_x-1)(N_y-1) \times N_xN_y$ diagonal matrix with elements

$\vartheta'\left(\left|[\mathfrak{D}_x\sigma]_{ij}\right|^2 + \left|[\mathfrak{D}_y\sigma]_{ij}\right|^2\right)$ in its diagonal given by

$$\vartheta'(|\mathfrak{D}\sigma|^2) = \begin{bmatrix} \operatorname{diag}\left(\vartheta'(\sigma)\right) & 0 \\ 0 & \operatorname{diag}\left(\vartheta'(\sigma)\right) \end{bmatrix}.$$

Analogous to the case in one dimension we denote $\mathcal{L}(\sigma)$ by

$$\begin{aligned} \mathcal{L}(\sigma) &= \mathfrak{D}^*\vartheta'\left(|\mathfrak{D}\sigma|^2\right)\mathfrak{D}, \\ &= \begin{bmatrix} \mathfrak{D}_x^T & \mathfrak{D}_y^T \end{bmatrix} \begin{bmatrix} \operatorname{diag}\left(\vartheta'(\sigma)\right) & 0 \\ 0 & \operatorname{diag}\left(\vartheta'(\sigma)\right) \end{bmatrix} \begin{bmatrix} \mathfrak{D}_x \\ \mathfrak{D}_y \end{bmatrix}. \end{aligned} \tag{6.58}$$

Then, the gradient of $\mathcal{K}$ can be written as

$$\nabla\mathcal{K}(\sigma) = \mathcal{L}(\sigma)\sigma, \tag{6.59}$$

where $\mathcal{L}(\sigma)$ is the diffusion operator with Neumann boundary condition

$$\begin{aligned} \mathcal{L}(\sigma)\sigma &= -\nabla\cdot\left(\vartheta'(|\nabla\sigma|^2)\nabla\sigma\right) \\ &= -\frac{\partial}{\partial x}\left(\vartheta'(|\nabla\sigma|^2)\frac{\partial\sigma}{\partial x}\right) - \frac{\partial}{\partial y}\left(\psi'(|\nabla\sigma|^2)\frac{\partial\sigma}{\partial y}\right), \end{aligned} \tag{6.60}$$

with the diffusion coefficient

$$\vartheta'(|\nabla\sigma|^2) = \psi'\left(\left(\frac{\partial\sigma}{\partial x}\right)^2 + \left(\frac{\partial\sigma}{\partial y}\right)^2\right).$$

Since ϑ is convex, the diffusion coefficient $\vartheta'(|\mathfrak{D}\sigma|^2)$ is positive. Therefore the matrix $\mathcal{L}(\sigma)$ in (6.58) is symmetric and positive semi-definite. The Hessian is computed analogously as in the one dimensional case

$$\operatorname{Hess}\mathcal{K}(\sigma) = \mathcal{L}^T\psi'(|\mathcal{L}\sigma|^2)\mathcal{L} + \mathcal{L}^T\vartheta''(2|\mathfrak{D}\sigma|^2)(\mathfrak{D}\sigma)(\mathfrak{D}\sigma)^T\mathfrak{D}, \tag{6.61}$$

which can be written as in (6.61), with $\mathcal{L}(\sigma)$ defined in (6.58).

Finally, in two dimensions, using (6.59) and (6.61), the gradient and the Hessian of $\mathcal{F}(\sigma)$ are given by

$$\nabla\mathcal{F}(\sigma) = \nabla J(\sigma) + \alpha\mathcal{L}(\sigma)\sigma, \tag{6.62}$$

and

$$\operatorname{Hess} \mathcal{F}(\sigma) = \operatorname{Hess} \mathcal{J}(\sigma) + \alpha \mathcal{L}(\sigma) + \alpha \mathcal{L}'(\sigma)\sigma, \tag{6.63}$$

respectively, where $\nabla J(\sigma)$ is given in (3.15).

An application of the Newton method for the the minimisation problem (6.1) results in

$$\sigma^{k+1} = \sigma^k - \left(\operatorname{Hess}_{\mathcal{L}} \mathcal{F}(\sigma^k)\right)^{-1} \nabla \mathcal{F}(\sigma^k), \tag{6.64}$$

where $\operatorname{Hess}_{\mathcal{L}} \mathcal{F}(u_k)$ is given by (6.63).

Next we provide the Pseudo Code for the Lagged Diffusivity Fixed Point Method for total variation regularization as presented in [162].

Algorithm 17: Lagged diffusivity method for total variation regularization

Input: α, σ^0

Output: σ^I

for $k := 1, \ldots, I$ **do**

Compute the gradient $\nabla \mathcal{F}(\sigma^k)$ given in (6.62);

Update $\sigma^{k+1} = \sigma^k + \tau^k p^k$ using algorithm 16;

end

Here u_i solves (2.1) with source q_i, and z_i solves (3.10) with adjoint source ξ_i given by (3.17), and $p^k := -\left(\operatorname{Hess}_{\mathcal{L}} \mathcal{F}(\sigma_k)\right)^{-1} \nabla \mathcal{F}(\sigma^k)$.

For using the limited-memory L-BFGS algorithm 16 in the algorithm 17, we identify the gradient $\nabla \mathcal{F}(\sigma^k)$ with $\nabla f(\sigma^k)$ and $\left(\operatorname{Hess}_{\mathcal{L}} \mathcal{F}(\sigma_k)\right)^{-1} \nabla \mathcal{F}(\sigma^k)$ with $\mathcal{H}_k \nabla f(\sigma^k)$, as given in section 6.3.

6.5 Discrete Derivatives

Let $\sigma_{i,j}$ be defined on a uniformly-spaced grid in two dimensions $\{(x_i, y_i) \mid x_i = i\Delta x, y_j = j\Delta y,\ i = 1, \ldots, N_x,\ j = 1, \ldots, N_y\}$. Then the discrete gradient of $\sigma_{i,j}$ is given by $\nabla \sigma := \mathfrak{D}\sigma$ where $\mathfrak{D}\sigma$ is defined in (6.56).

The discrete divergence denoted by div or $\nabla \cdot$ is defined as $\operatorname{div} := -\mathfrak{D}^*$. Then, using

the definition (6.57) of the adjoint $\mathfrak{D}^*$ for v_{ij} and w_{ij}, we obtain

$$\left\{ \operatorname{div} \begin{bmatrix} v \\ w \end{bmatrix} \right\}_{ij} = v_{i,j} - v_{i-1,j} + w_{i,j} - w_{i,j-1}, \quad 2 \le i \le N_x, \quad 2 \le j \le N_y - 1. \tag{6.65}$$

The discrete Laplacian is defined as usual as $\Delta = \operatorname{div} \nabla = \nabla \cdot \nabla$. Then the Laplacian of σ is

$$\begin{aligned} \Delta\sigma &= \left\{ \operatorname{div} \begin{bmatrix} [\mathfrak{D}_x\sigma]_{ij} \\ [\mathfrak{D}_y\sigma]_{ij} \end{bmatrix} \right\}_{ij} \\ &= -\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{ij}\big) - \mathfrak{D}_x^*\big([\mathfrak{D}_y\sigma]_{ij}\big) \\ &= \sigma_{i+1,j} - 2\sigma_{i,j} + \sigma_{i-1,j} + \sigma_{i,j+1} - 2\sigma_{i,j} + \sigma_{i,j-1} \\ &= \sigma_{i+1,j} + \sigma_{i-1,j} - 4\sigma_{i,j} + \sigma_{i,j+1} + \sigma_{i,j-1}. \end{aligned}$$

We consider the following two types of boundary conditions for the discrete differences. Firstly, we consider the symmetric boundary condition given by

$$[\mathfrak{D}_x\sigma]_{ij} = \begin{cases} \sigma_{i+1,j} - \sigma_{i,j} & \text{if} \quad 1 \le i \le N_x - 1, \quad 1 \le j \le N_y, \\ 0 & \text{if } i = N_x, \end{cases} \tag{6.66}$$

and

$$[\mathfrak{D}_y\sigma]_{ij} = \begin{cases} \sigma_{i,j+1} - \sigma_{i,j} & \text{if} \quad 1 \le i \le N_x, \quad 1 \le j \le N_y - 1, \\ 0 & \text{if } j = N_y, \end{cases} \tag{6.67}$$

Secondly, we consider the periodic boundary condition given by

$$[\mathfrak{D}_x\sigma]_{ij} = \begin{cases} \sigma_{i+1,j} - \sigma_{i,j} & \text{if} \quad 1 \le i \le N_x - 1, \quad 1 \le j \le N_y, \\ \sigma_{1,j} - \sigma_{N_x,j} & \text{if } i = N_x, \end{cases} \tag{6.68}$$

and

$$[\mathfrak{D}_y\sigma]_{ij} = \begin{cases} \sigma_{i,j+1} - \sigma_{i,j} & \text{if} \quad 1 \le i \le N_x - 1, \quad 1 \le j \le N_y, \\ \sigma_{i,1} - \sigma_{i,N_y} & \text{if } j = N_y. \end{cases} \tag{6.69}$$

Using the symmetric boundary conditions defined in (6.66), the discrete gradient $\mathfrak{D}_x\sigma$

is represented by the following matrix

$$\begin{pmatrix} [\mathfrak{D}_x\sigma]_{1j} \\ [\mathfrak{D}_x\sigma]_{2j} \\ \vdots \\ [\mathfrak{D}_x\sigma]_{(N_x-1)j} \\ [\mathfrak{D}_x\sigma]_{N_xj} \end{pmatrix} = \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{pmatrix} \begin{pmatrix} \sigma_{1,j} \\ \sigma_{2,j} \\ \vdots \\ \sigma_{N_x-1,j} \\ \sigma_{N_x,j} \end{pmatrix}, \tag{6.70}$$

and similarly for $\mathfrak{D}_x\sigma$.

Using the symmetric boundary conditions (6.66), the discretization of $-\mathfrak{D}_x^*$ for the extreme points is

$$\begin{aligned} -\mathfrak{D}_x^*\sigma_{1,j} &= \sigma_{1,j} \\ -\mathfrak{D}_x^*\sigma_{N_x,j} &= \sigma_{N_x,j} - \sigma_{N_x-1,j} = \sigma_{N_x-1,j}, \end{aligned}$$

since $\sigma_{N_x,j}$ is equal to zero by the boundary conditions (6.66). Then, the negative adjoint $-\mathfrak{D}_x^*\sigma$ is given by

$$\begin{pmatrix} -[\mathfrak{D}_x\sigma]_{1j} \\ -[\mathfrak{D}_x\sigma]_{2j} \\ \vdots \\ -[\mathfrak{D}_x\sigma]_{(N_x-1)j} \\ -[\mathfrak{D}_x\sigma]_{N_xj} \end{pmatrix} = \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & \ddots & \ddots & & \\ & & -1 & 1 & \\ & & & -1 & 0 \end{pmatrix} \begin{pmatrix} \sigma_{1,j} \\ \sigma_{2,j} \\ \vdots \\ \sigma_{N_x-1,j} \\ \sigma_{N_x,j} \end{pmatrix}, \tag{6.71}$$

which is minus times the transpose of the matrix of equation (6.70). We proceed in a similar fashion for $-\mathfrak{D}_x^*\sigma$.

Finally, the discretization of the Laplacian $-\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{ij}\big)$ is

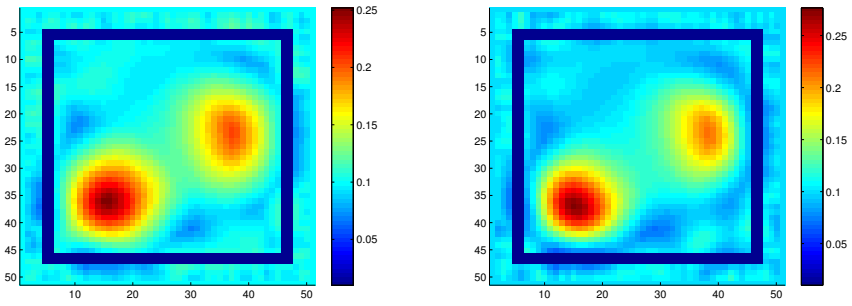
$$-\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{ij}\big) = \begin{cases} \sigma_{2,j} - \sigma_{1,j} & \text{if } i = 1, \\ \sigma_{i+1,j} - 2\sigma_{i,j} + \sigma_{i-1,j} & \text{if } i = 2, \dots, N_x - 1, \\ \sigma_{N_x-1,j} - \sigma_{N_x,j} f_{N-1} - f_N & \text{if } i = N_x, \end{cases}$$

$$\begin{pmatrix} -\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{1j}\big) \\ -\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{2j}\big) \\ \vdots \\ -\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{(N_x-1)j}\big) \\ -\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{N_xj}\big) \end{pmatrix} = \begin{pmatrix} -1 & 1 & & & \\ 1 & -2 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma_{1,j} \\ \sigma_{2,j} \\ \vdots \\ \sigma_{N_x-1,j} \\ \sigma_{N_x,j} \end{pmatrix}, \tag{6.72}$$

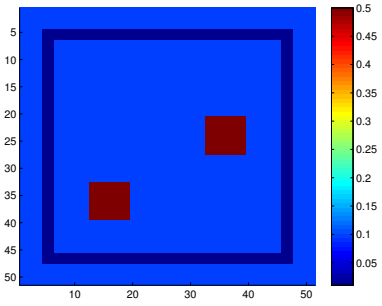
and similarly for $-\mathfrak{D}_y^*\big([\mathfrak{D}_y\sigma]_{ij}\big)$. Therefore, the discretization of the Laplacian of σ is $-\mathfrak{D}_x^*\big([\mathfrak{D}_x\sigma]_{ij}\big) - \mathfrak{D}_x^*\big([\mathfrak{D}_y\sigma]_{ij}\big)$. In the interior of the domain, this coincides with the 5-point stencil of the Laplacian $\sigma_{i+1,j} + \sigma_{i-1,j} - 4\sigma_{i,j} + \sigma_{i,j+1} + \sigma_{i,j-1}$.

6.6 Numerical Results

We have performed the following numerical experiments for testing our algorithms as derived above. They are aiming at reconstructing the absorption coefficient of the RTE in the application of DOT. We have used the same (simulated) experimental setting as in section 3.7.



(a) (b)



(c)

Figure 6.1: Figure 6.1c shows the true object, then figure 6.3b shows the imaging reconstruction of μ using the split Bregman method without continuation scheme; algorithm 12, figure 6.1a shows the imaging reconstruction of μ using the split Bregman method with continuation scheme; algorithm 13.

Figure 6.1 shows the reconstruction of the absorption coefficient μ using the the Split Bregman method. In particular, figure 6.1a shows the reconstruction of μ using algorithm 12, and figure 6.1b shows the reconstruction of μ using algorithm 13.

The evolution of the norm of the misfit functional during the iterations is shown in figure 6.2.

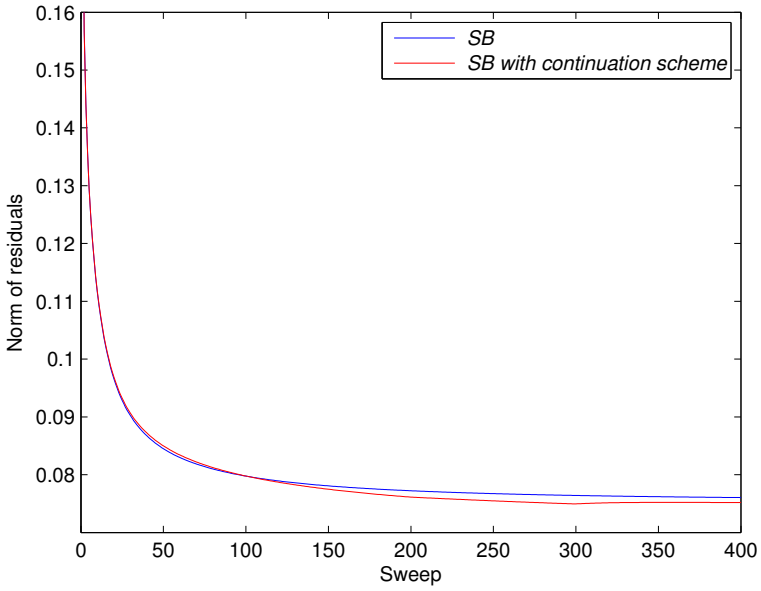
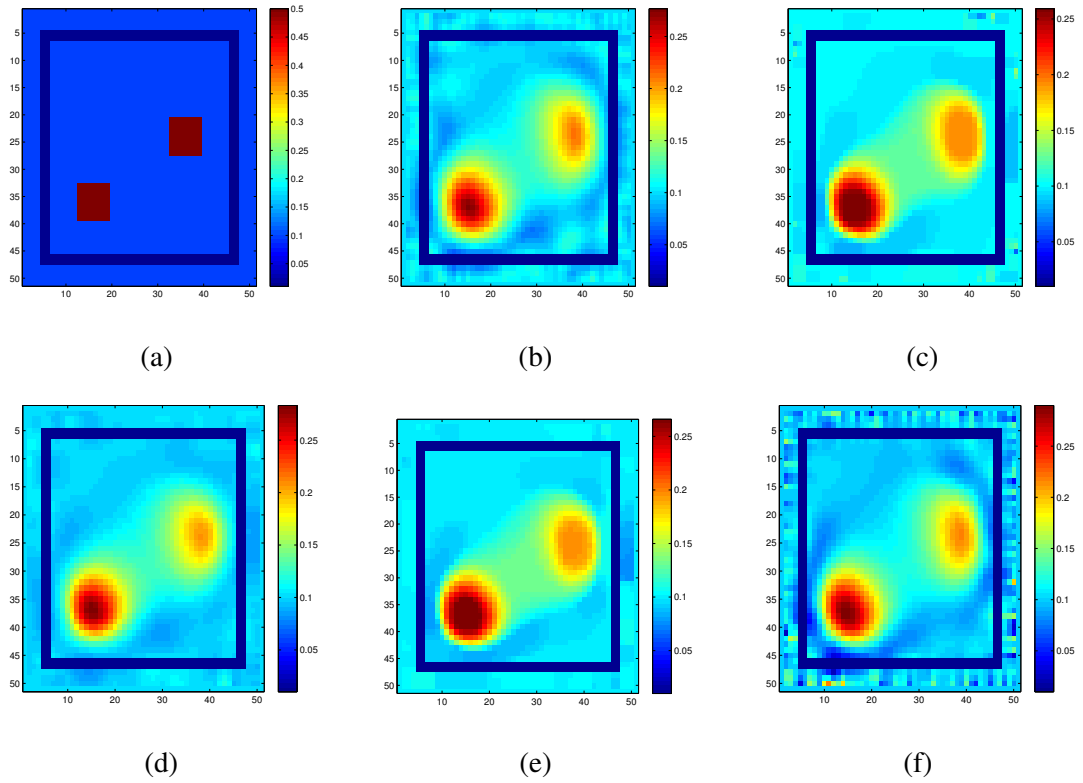


Figure 6.2: Norm of residuals of the misfit functional of the μ absorption reconstruction using the SB and the SB with continuation scheme.

Next we have tested the total variation regularization described in this chapter for blocky inclusions with different reconstruction methods such as split Bregman and lagged diffusivity with different modalities. The results are compared in figure 6.3. Figure 6.3f shows our benchmark using only the algorithm 2

Figure 6.3b shows the reconstruction of μ using the algorithm 12. The modalities of lagged diffusivity denoted by LD-1, LD-2 and LD-3 make use of the algorithm 17 according to the following descriptions. Figure 6.3c shows the LD-1 reconstruction which was obtained using algorithm 17 (lagged diffusivity method) without using the limited-memory BFGS scheme, that is, approximating the Hessian as a multiple of the Identity. Figure 6.3d LD-2 was obtained using algorithm 17 without using the limited-memory BFGS scheme, that is, approximating the Hessian by a multiple of the Identity, and instead of using the discretization of the total variation penalty term by (6.59), here we used the discretization of the total variation penalty term by (6.36). Finally, figure 6.3e shows the

LD-3 reconstruction which was made using the lagged diffusivity algorithm 17.



(a) (b) (c)

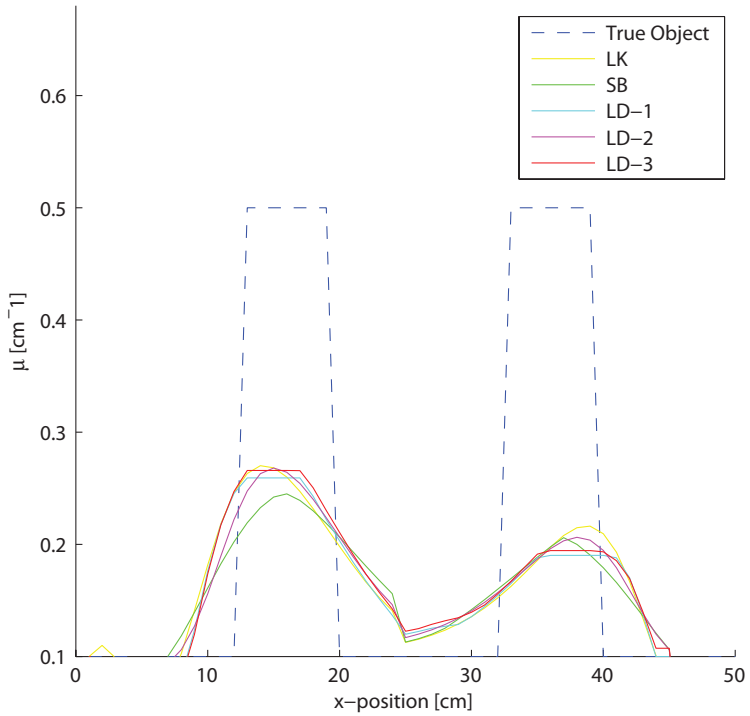
(d) (e) (f)

Figure 6.3: Figure 6.3a shows the true object, figure 6.3b shows the imaging reconstruction of μ using the split Bregman method with continuation scheme; algorithm 13, figure 6.3c shows the imaging reconstruction using the lagged diffusivity (LD-1); algorithm 17 without using the limited-memory BFGS scheme, figure 6.3d shows the imaging reconstruction of μ using the lagged diffusivity with the scheme discretization (6.36); (LD-2); figure 6.3e shows the imaging reconstruction of μ using the lagged diffusivity with L-BFGS scheme (LD-3); algorithm 17, figure 6.3f shows the imaging reconstruction of μ using the LK method; algorithm 2.

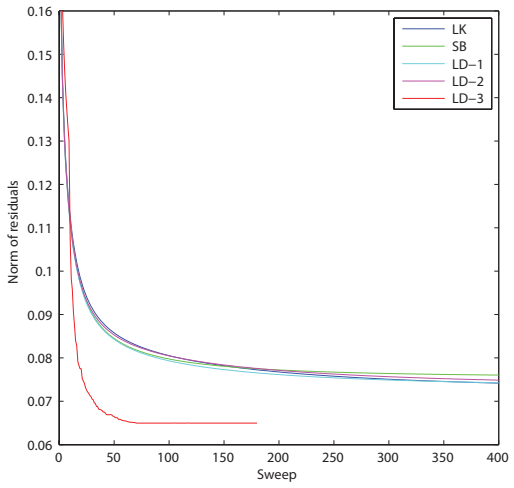
Profiles in the x-direction, the norm of the residuals, and the L^2-norm errors of the imaging reconstruction using the split Bregman, the Lagged Diffusivity with different modalities, and with only the LK method without total variation regularization, are shown in figure 6.4. Note that the reconstructions for the Lagged Diffusivity LD-3 and using the L-BFGS formula for the Hessian correspond to the 100-th sweep of the algorithm 17.

The values used for the reconstruction of the absorption coefficient in this chapter are displayed in Table 3.1.

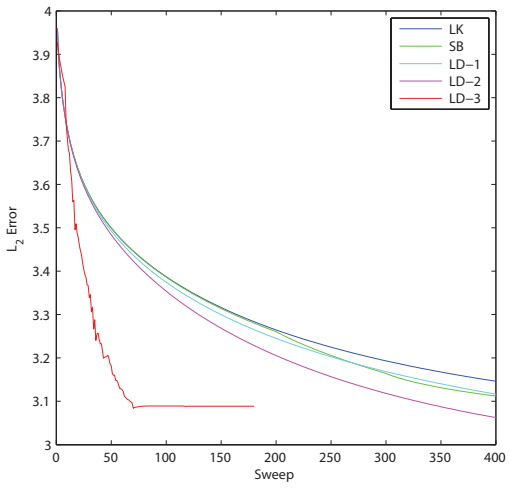
Finally, figure 6.5 shows our numerical results which were obtained by using the Landweber method with the L-BFGS scheme but without total variation regularization, algorithm 16.



(a)



(b)



(c)

Figure 6.4: Figure 6.4a shows the cross section of the final reconstructions using the methods presented in this section(SB,LD-1, LD-2 and LD-3 and the benchmark LK). Figure 6.4b and 6.4c shows their correspondent residuums and L_2-norm errors respectively.

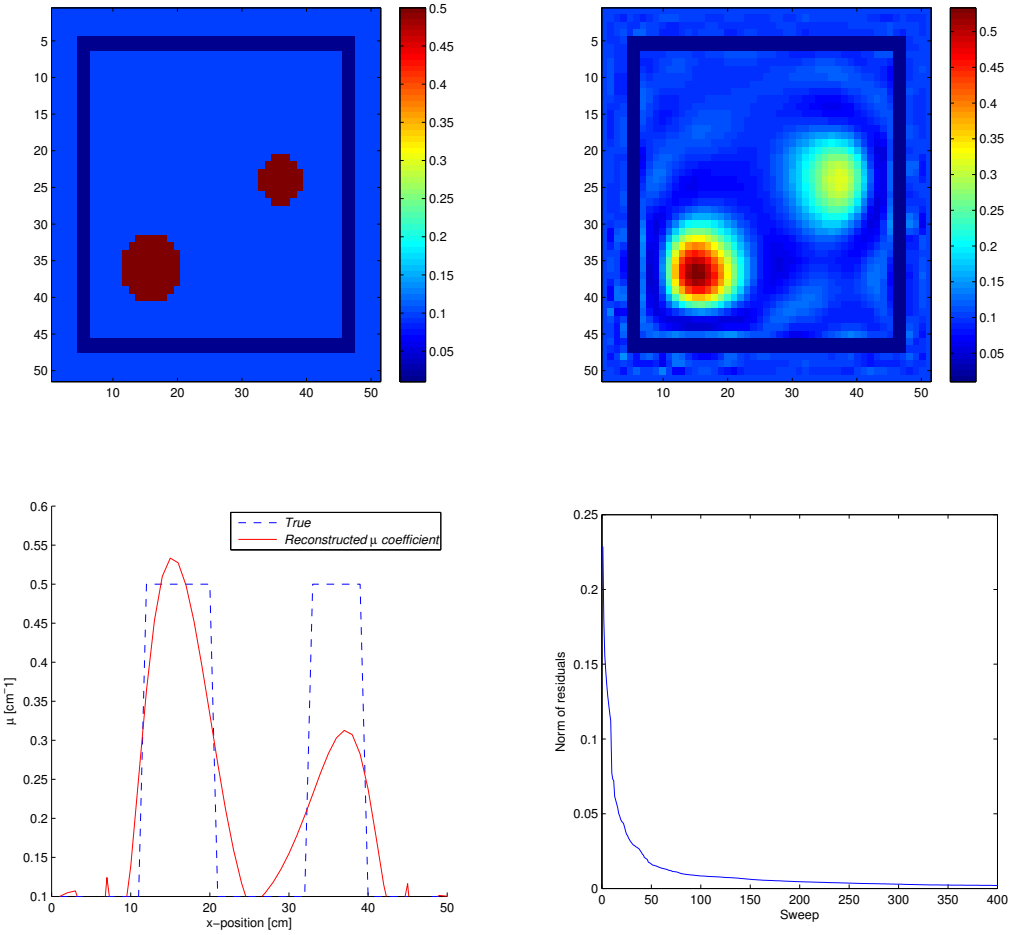


Figure 6.5: Shape reconstruction for σ using the Landweber method; algorithm 1 with the L-BFGS formula. Top left: true object; top right: final reconstruction. Bottom left: profile taken in the x-direction; bottom right: norm of residuals of the cost.

Chapter 7

Conclusions and Future Work

In this chapter we provide some conclusions and outline some possible directions for future work.

In this thesis, some novel non-standard image reconstruction techniques for Diffuse Optical Tomography were introduced and discussed. The main novelty in this thesis lies in the combination of quite recent mathematical regularization schemes (in particular the level set shape formulation, the sparsity regularization and the total variation regularization) with the still difficult to handle radiative transfer equation for the application to DOT.

The goal of this thesis was proof-of-concept (rather than dealing with real data), which means we wanted to show that these advanced techniques can indeed be combined, and wanted to provide strategies on how to deal with several practical complications which occur in this specific combination of advanced regularization schemes with a demanding and expensive discretizations of the RTE forward model. In order to be able to deal with this combination and test a significant amount of different algorithms, we restricted ourselves to a two-dimension physical setup. However, we believe that all techniques and results presented in this thesis generalize in a straightforward way to the more realistic three-dimensional setup, except of the fact that computation time might still be high with current computing systems.

In chapter 3 of this thesis, the inverse and shape reconstruction problems in DOT were formulated. We presented an individual and simultaneous shape and contrast value reconstruction scheme for the optical properties of tissue in DOT by using the Landweber-Kaczmarz (LK) method and the level set approach following the pioneering idea of San-

tosa in [141]. The reconstruction of the shape and the contrast value of the absorption and scattering coefficients have been achieved quite successfully with this still very competitive technique. The results are similar to those that were obtained in [7, 144] .

Then, in chapter 4, we tried to apply a slightly different approach to shape reconstruction by making use of expressions for shape sensitivities. For this purpose, analytical expressions for the shape derivative in DOT based on the RTE were obtained, and their implementation as part of the level set method were tested by using a Hamilton-Jacobi-type shape evolution approach. Several variants of the shape reconstruction problem in DOT were discussed here, using a narrowband technique and an extended velocity field in the whole domain. In addition, one of these variants is compared with our implementation of the shape reconstruction using algorithm 4. Successful reconstructions of the absorption coefficient in DOT using this technique based on shape sensitivities were obtained, and compare favourably to the shape reconstructions using the more traditional technique from chapter 3. Indeed, in some experiments, the shape of the inclusions were retrieved even faster using the shape derivative method with extended velocity field than using algorithm 4.

In chapter 5 we have then continued by introducing a novel reconstruction technique based on sparsity regularization. We have shown that the location and an estimate of the contrast value of the absorption and scattering coefficients can be obtained reliably with such a method. In addition, we have observed that in many situations sparsity regularization reconstructions converge faster and provide better resolution than Tikhonov-type (L/LK) reconstructions. Also, often sparsity reconstructions provide a more accurate resolution for two closely-spaced inclusions than Tikhonov-type reconstructions.

On the more technical side, the combination of the Barzilai-Borwein step size criterion with the monotonicity condition (5.34) works quite efficiently as step size criterion. Here, our results indicate that the loping-Landweber-Kaczmarz method can be more efficient than the classical Landweber-Kaczmarz method if the weak monotonicity condition is not included in the sparsity Algorithms. However, the reconstructions are not significantly different if the weak monotonicity condition holds when using sparsity regularization.

Then, in chapter 6 we were addressing image reconstruction from Total Variation functionals. In particular, the Lagged Diffusivity method in combination with the L-BFGS formula for the absorption coefficient reconstruction was shown to be a quickly converg-

ing method in terms of number of iterations, even though it consumes more time than the other total variation regularization methods used in this thesis. We observed in our numerical experiments that total variation regularization reconstructions are able to successfully retrieve the location and an estimate of the contrast value of the absorption and scattering coefficients of two-dimensional experiments. Total variation regularization schemes also often converge faster and provide better resolution than Tikhonov-type (L/LK) reconstructions. Our implementation of the split Bregman method yielded acceptable denoised reconstructions in terms of the location and estimate of the shape of the inclusions, and in those experiments performed in this thesis it provided a more satisfactory reconstruction than those obtained with only using the LK method.

Overall we observed that sparsity and total variation regularization schemes do retrieve successfully the location and an estimate of the contrast value of the absorption and scattering coefficients of two-dimensional experiments. Moreover, sparsity and total variation regularization reconstructions have been shown to often converge faster and provide better resolution than Tikhonov-type (L/LK) reconstructions.

The cross talk phenomenon, which was mentioned at the beginning of this thesis and usually occurs in the simultaneous reconstructions for time-systems in DOT, is still an open problem. In addition, as we have pointed out in chapter 5, the reconstructions of μ and b are considered accurate if we use at least 16 discrete directions for the angular variable θ, as pointed out in [142]. We did not notice this effect in our numerical experiments since we were not dealing with real data. However, we are aware of this limitation and advise readers who want to apply our techniques to real data to use a well-designed numerical setup for their practical situations. On the other hand, we believe that our results are overall quite insensitive to the particular choice of forward modelling simulator, since all the formulation are based on the continuous description of the inverse problems using PDEs instead of matrix algebra. Only at the last stage we have discretized the reconstruction techniques, such that the optimization schemes discussed here will still be valid when other techniques are used for discretizing the forward and adjoint RTE.

We also want to mention that, although the results presented in this thesis using minimization in ℓ_1 seem more satisfactory than those using minimization in ℓ_2 in many cases, one should not generalize this without precaution. We refer for a nice discussion of such a comparison to [158].

Finally, we want to list some suggestions for potential future research, in a rather bullet-point fashion. The application of novel regularization techniques to inverse problems governed by the RTE is still quite unexplored, such that this research field is still wide open for research. The following list is only intended to give some rough guidelines, and is not at all exhaustive.

- Our implementation of the LS method can be extended using multiple level set functions and the Tikhonov functional proposed in [48]. Moreover, instead of the classical LS method presented and used in this dissertation, the generalized level set method introduced in [108] can be used. Also, the ENO or WENO and the Runge Kutta schemes [38, 114, 146] for the spatial and time variables respectively may be incorporated for solving the level set equation (4.31).

- The topological derivative can be calculated using the relationship between the shape derivative and the topological derivative given in [26, 32, 90].

- Instead of our finite differences discretization of the radiative transfer equation in $2D$, Born or Rytov approximation techniques for approximating the RTE can be used, or a finite element discretization of the RTE, and extended to $3D$ [1, 30, 75, 153, 156]. The latter is a more accurate discretization than the finite differences scheme. Also, a more efficient scheme for discretizing the RTE as suggested in [75] can be used.

- Instead of using the Landweber-Kaczmarz method, the Steepest-Descent-Kaczmarz scheme [47] can be used, or a Gauss-Newton method [156] or the Newton-type Levenberg-Marquardt-Kaczmarz [16, 27] methods for obtaining a better convergence rate, since the LK method usually shows quite a slow convergence, and the LMK method has shown faster convergence than the LK method [16].

- Instead of using a soft-thresholding operator in the sparsity algorithms presented in chapter 5, the novel accelerated projected steepest descent method as presented in [157] can be used, or alternatively the hard-thresholding operator as defined in [24].

- An analytical convergence analysis for the sparsity algorithms presented in chapter 5 may be performed.

- Instead of using the total variation regularization as presented in chapter 6, the second order total generalized variation method as proposed in [23] may be used.
- The reconstruction of the scattering coefficient and the simultaneous reconstruction of the absorption and scattering coefficients using the total variation regularization can be carried out. Furthermore, the L-curve method can be used for choosing the best regularization parameter in the sparsity and total variation regularization schemes.
- A stopping criterion for the contrast value reconstructions may be implemented.
- It seems possible to recover the clear layer, which simulates the cerebrospinal fluid in our experiments, by using the data of early photons which arrive at the detectors.
- A variable refractive index of the media can be incorporated in our model of the propagation of the photons. Furthermore, the polarization state of the light used to irradiate the examined tissue can be taken into account as suggested in [116].

Bibliography

[1] Ackroyd, R. (1997). *Finite Element Methods for Particle Transport*. John Wiley & Sons Inc.

[2] Agarwal, V. (2003). Total Variation Regularization and L-curve method for the selection of regularization parameter. *ECE*, 599.

[3] Allaire, G., de Gournay, F., Jouve, F., and Toader, A.-M. (2005). Structural optimization using topological and shape sensitivity via a level set method. *Control and Cybernetics*, 34.

[4] Angell, M. (2011). The Epidemic of Mental Illness: Why? *The New York Review of Books*.

[5] Arlotti, L., Banasiak, J., and Lods, B. (2009). A new approach to transport equations associated to a regular field: trace results and well-posedness. *Mediterr. J. Math.*, 6:367–402.

[6] Arridge, S. (1999). Optical tomography in medical imaging. *Inverse Problems*, 15:R41–R93.

[7] Arridge, S., Dorn, O., Kolehmainen, V., Schweiger, M., and Zacharopoulos, A. (2008). Parameter and structure reconstruction in optical tomography. *Journal of Physics: Conference Series*, 135:012001.

[8] Arridge, S., J., K., Kolehmainen, V., and Tarvainen, T. (2011). *Handbook of Mathematical Methods in Medical Imaging*, chapter Optical Imaging, pages 735–780. Springer (New York).

[9] Arridge, S. and Schotland, J. (2009). Optical tomography: forward and inverse problems. *Inverse Problems*, 25(12):59.

[10] Bachmayr, M. (2007). Iterative Total Variation Methods for Nonlinear Inverse Problems. Master's thesis, Johannes Kepler Universität.

[11] Bal, G. (2009). Inverse transport theory and applications. *Inverse Problems*, 25(5):48.

[12] Bal, G. and Jollivet, A. (2010). Stability for time dependent inverse transport. *SIAM Journal on Mathematical Analysis*, 42(2):679–700.

[13] Bal, G. and Monard, F. (2010). An accurate solver for forward and inverse transport. *Journal of Computational Physics*, 229:4952–4979.

[14] Baritaux, J.-C. (2012). *Sparsity-Inducing Reconstruction Methods for Fluorescence Diffuse Optical Tomography*. PhD thesis, École polytechnique fédérale de Lausanne.

[15] Barzilai, J.and Borwein, J. (1988). Two-point step size gradient methods. *IMA Journal of Numerical Analysis*, 8(1):141–148.

[16] Baumeister, J., Kaltenbacher, B., and Leitão, V. (2010). On Levenberg-Marquardt-Kaczmarz iterative methods for solving systems of nonlinear ill-posed equations. *Inverse Problems and Imaging*, 4(3):335–350.

[17] Behrooz, A., Zhou, H.-M., Eftekhar, A., and Adibi, A. (2012). Total variation regulariation for 3D reconstruction in fluorescence tomography: experimental phantom studies. *Applied Optics*, 51(34):8216–8227.

[18] Bluestone, M., Abdoulaev, G., Schmitz, C., Barbour, R., and Hielscher, A. (2001). Three-dimensional optical tomography of hemodynamics in the human head. *Optics Express*, 9(6):272–286.

[19] Boas, D. (1996). *Diffuse photon probes of structural and dynamical properties of turbid media: Theory and biomedical applications*. PhD thesis, Dept. Physics, Univ. Pennsylvania.

[20] Boas, D., Brooks, D., Miller, E., DiMarzio, C., Kilmer, M., Gaudette, R., and Zhang, Q. (2001). Imaging the body with diffuse optical tomography. *IEEE Signal Process Mag.*, 18(6):57–75.

[21] Boas, D., Dale, A., and Franceschini, M. (2004). Diffuse optical imaging of brain activation: approaches to optimizing image sensitivity, resolution, and accuracy. *NeuroImage*, 23:S275–S288.

[22] Bone, B., Pentek, Z., Perbeck, L., and Veress, B. (1997). Diagnostic accuracy of mammography and contrast enhanced MR imaging in 238 histologically verified breast lesions. *Acta Radiol.*, 38:489–496.

[23] Bredies, K., Kunisch, K., and Pock, T. (2010). Total Generalized Variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526.

[24] Bredies, K. and Lorenz, D. (2008). Iterated hard shrinkage for problems with sparsity constraints. *SIAM Journal on Scientific Computing*, 30(2):657–683.

[25] Bredies, K., Lorenz, D., and Maass, P. (2009). A generalized conditional gradient method and its connection to an iterative shrinkage method. *Computational Optimization and Applications*, 42:173–193.

[26] Burger, M., Benjamin, H., and Ring, W. (2004). Incorporating Topological derivatives into Level Set Methods. *Journal of Computational Physics*, 194:344–362.

[27] Burger, M. and Kaltenbacher, B. (2006). Regularizing Newton-Kaczmarz methods for nonlinear ill-posed problems. *SIAM Journal on Numerical Analysis*, 44(1):153–182.

[28] Bushberg, J., Seibert, J., Leidholdt, E., and Boone, J. (2002). *The essential physics of medical imaging*. Lippincott Williams and Wiklkins, second edition edition.

[29] Cai, J.-F., Osher, S., and Shen, Z. (2009). Linearized Bregman iterations for Compressed Sensing. *Mathematics of Computation*, 78:1515–1536.

[30] Cao, N., Nehorai, A., and Jacob, M. (2007). Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm. *Optics Express*, 15(21):13695.

[31] Carpio, A. and Rapún, M. (2008a). *Topological Derivatives for Shape Reconstruction*, volume 1943. Springer.

[32] Carpio, A. and Rapún, M.-L. (2008b). Solving inhomogeneous inverse problems by topological derivative methods. *Inverse Problems*, 24.

[33] Case, K. and Zweifel, P. (1967). *Linear transport theory*. Addison-Wesley.

[34] Céa, J., Garreau, S., Guillaume, P., and Masmoudi, M. (2000). The shape and topological optimizations connection. *Computer Methods in Applied Mechanics and Engineering*, 188(4):713–726.

[35] Cessenat, M. (1984a). Théoremès de trace L^p pour des espaces de fonctions de la neutronique. *C. R. Acad. Sci. Paris Série I*, 299:831–834.

[36] Cessenat, M. (1984b). Théoremès de trace pour des espaces de fonctions de la neutronique. *C. R. Acad. Sci. Paris Série I*, 300:89–92.

[37] Cheung, C., Prahl, S., and Welch, A. (1990). A review of the optical properties of biological tissues. *IEEE Quantum Electron*, 26(12):2166–2185.

[38] Chopp, D. (2005). *Handbook of Biomedical Image Analysis*, volume 1 of *Topics in biomedical engineering international book Series.*, chapter Recent Advances in the Level Set Method, pages 201–256. Kluwer Academic/Plenum Publishers.

[39] Choulli, M. and Stefanov, P. (1996). Inverse scattering and inverse boundary value problems for the linear Boltzman Equation. *Comm. P.D.E*, 21:763–785.

[40] Chung, E., Chang, T., and Tai, X.-C. (2005). Electrical impedance tomography using level set representation and total variational regularization. *Journal of Computational Physics*, 205:357–372.

[41] Cooper, J. (2012). *Sparsity Regularization in Diffuse Optical Tomography*. PhD thesis, Clemson University.

[42] Culver, J., Siegel, A., Stott, J., and Boas, D. (2003). Volumetric diffuse optical tomography of brain activity. *Optics Letters*, 28(21):2061–2063.

[43] Dai, Y.-H., Hager, W., Schittkowski, K., and Zhang, H. (2006). The cyclic Barzilai-Borwein method for unconstrained optimization. *IMA Journal of Numerical Analysis*, 26(3):604–627.

[44] Daubechies, I., Defrise, M., and De Mol, C. (2004). An iterative thersholding algorithm for linear inverse problems with a sparsity constraint. *Communications on Pure and Applied Mathematics*, 57(11):1413–1457.

[45] Dautray, R. and Lions, J.-L. (1993). *Mathematical Analysis and Numerical Methods for Science and Technology*, volume 6. Springer-Verlag.

[46] Davenport, M., Duarte, M., Eldar, Y., and Kutyniok, G. (2012). *Compressed Sensing*, chapter Introduction to compressed sensing, pages 1–68. Cambridge University Press.

[47] De Cezaro, A., Haltmeier, M., Leião, A., and Scherzer, O. (2008). On Steepest-Descent-Kaczmarz methods for regularizing systems of nonlinear ill-posed equations. *Applied Mathematics and Computation*, 202:596–607.

[48] De Cezaro, A., Leião, A., and Tai, X.-C. (2009). On multiple level-set regularization methods for inverse problems. *Inverse Problems*, 25(3):035004.

[49] Delfour, M. and Zolésio, J. (2011). *Shapes and Geometries. Metrics, Analysis, Differential Calculus and Optimization*, volume 16. Society for Industrial and Applied Mathematics, second edition.

[50] Dierkes, T., Dorn, O., Natterer, F., Palamodov, V., and Sielschott, H. (2002). Fréchet Derivatives for some bilinear inverse problems. *SIAM Journal of Applied Mathematics*, 62(6):2092–2113.

[51] Doicu, A., Trautmann, T., and Schreier, F. (2010). *Numerical Regularization for Atmospheric Inverse Problems*. Springer Berlin Heidelberg.

[52] Dorn, O. (1997). *Das inverse transportproblem in der lasertomographie*. PhD thesis, University of Münster.

[53] Dorn, O. (1998). A transport-backtransport method for optical tomography. *Inverse Problems*, 14:1107–1130.

[54] Dorn, O. (2000). Scattering and absorption transport sensitivity functions for optical tomography. *Optics Express*, 7(13):492–506.

[55] Dorn, O. (2002). Shape reconstruction in scattering media with voids using a transport model and level sets. *Canaddian Applied Math Quarterly*, 10(2):239–275.

[56] Dorn, O. (2009). *Numerical Methods in Multidimensional Radiative Transfer*, chapter Shape Reconstruction for an Inverse Radiative Transfer Problem Arising in Medical Imaging, pages 299–309. Springer Berlin Heidelberg.

[57] Dorn, O., Bertete-Aguirre, H., and Papanicolaou, G. (2008). *Inverse Problems and Imaging*, volume 1943 of *Lecture Notes in Mathematics*, chapter Adjoint Fields and Sensitivities for 3D Electromagnetic Imaging in Isotropic and Anisotropic Media, pages 35–66. Springer.

[58] Dorn, O., Miller, E., and Rappaport, C. (2000). A shape reconstruction method for electromagnetic tomography using adjoint fields and level sets. *Inverse Problems*, 16:1119–1156.

[59] Dorn, O. and Prieto, K. (2011). From data to images: A shape based approach for fluorescence tomography. In *Science: Image in Action: Proceedings of the 7th International Workshop on Data Analysis in Astronomy.*, pages 255–266.

[60] Duderstadt, J. and Martin, W. (1979). *Transport theory*. New York: Wiley.

[61] Egger, H. and Schlottbom, M. (2012). A mixed variational framework for the Radiative Transfer Equation. *Mathematical Models and Methods in Applied Sciences*, 22(3).

[62] Eichholz, J. (2011). *Discontinuous Galerkin methods for the radiative transfer equation and its approximations*. PhD thesis, University of Iowa.

[63] Engl, H., Hanke, M., and Neubauer, A. (2000). *Regularization of Inverse Problems*, volume 375 of *Mathematics and Its Applications*. Kluwer Academic Publishers.

[64] Epstein, C. (2008). *Introduction to the Mathematics of Medical Imaging*. Society for Industrial and Applied Mathematics, second edition.

[65] Ernst, M. and Roukema, J. (2002). Diagnosis of non-palpable breast cancer: a review. *The Breast*, 11:13–21.

[66] Fear, E., Meaney, P., and Stuchly, M. (2003). Microwaves for breast detection? *IEEE Potencials*, pages 12–18.

[67] Feijóo, G. (2004). A new method in inverse scattering based on the topological derivative. *Inverse Problems*, 20:1819–1840.

[68] Feijóo, G., Oberai, A., and Pinsky, P. (2004). An application of shape optimization in the solution of inverse acoustic scattering problems. *Inverse Problems*, 20:199–228.

[69] Feng, H. and Bai, J. (2007). Improved perturbation method for absorption map reconstruction in diffusion optical tomography. *Journal of the European Optical Society Rapid Publications*, 2:07015.

[70] Feng, J., Qin, C., Jia, K., Zhu, S., Liu, K., Han, D., Yang, X., Gao, Q., and Tian, T. (2012). Total variation regularization for bioluminescence tomography with the split Bregman method. *Applied Optics*, 51(19):4501–4512.

[71] Fletcher, R. (2005). *Optimization and Control with Applications*, chapter On the Barzilai-Borwein Method, pages 235–256. Springer US.

[72] Fornasier, M. and Rauhut, H. (2011). *Handbook of Mathematical Methods in Imaging*, chapter Compressive Sensing, pages 187–228. Springer.

[73] Fulmanski, P., Laurain, A., Schei, d. J.-F., and Sokołowski, J. (2008). Level set method with topological derivatives in shape optimization. *International Journal of Computer Mathematics*, 85:1491–1514.

[74] Fulmanski, P., Laurain, A., Scheid, J.-F., and Sokołowski, J. (2007). A level set method in Shape and Topology Optimization for Variational Inequalities. *Applied Mathematics and Computer Science*.

[75] Gao, H. and Zhao, H. (2009). A Fast-Forward Solver of Radiative Transfer Equation. *Transport Theory and Statistical Physics*, 38(3):149–192.

[76] Gao, H. and Zhao, H. (2010a). Multilevel bioluminescence tomography based on radiative transfer equation Part 1: regularization. *Optics Express*, 18(3):1854–1871.

[77] Gao, H. and Zhao, H. (2010b). Multilevel bioluminescence tomography based on radiative transfer equation Part 2: total variation and l1 data fidelity. *Optics Express*, 18(3):2894–2912.

[78] Gehre, M., Kluth, T., Lipponen, A., Jin, B., Maass, P., Seppanen, A., and Kaipio, J. (2012). Sparsity Reconstruction in Electrical Impedance Tomography: An Experimental Evaluation. *Journal of Computational and Applied Mathematics*, 236:2126–2136.

[79] Getreuer, P. (2012). Rudin-Osher-Fatemi Total Variation Denoising using Split Bregman. *Image Processing On Line*, 2:74–95.

[80] Giacino, J., Fins, J., Laureys, S., and Schiff, N. (2014). Disorders of consciousness after acquired brain injury: the state of the science. *Nature Reviews Neurology*, 10:99–114.

[81] Gibson, A., Austin, T., Everdell, N., Schweiger, M., Arridge, S., and Meek, J. (2006). Three-dimensional whole-head optical tomography of passive motor evoked responses in the neonate. *NeuroImage*, 30(2):521–528.

[82] Gibson, A., Hebden, J., and Arridge, S. (2005). Recent advances in diffuse optical imaging. *Phys. Med. Biol.*, 50:R1–R43.

[83] Goldstein, T. and Osher, S. (2009). The Split Bregman Method for L1-Regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343.

[84] González-Rodríguez, P., Kindelan, M., Moscoso, M., and Dorn, O. (2005). History matching problem in reservoir engineering using the propagation-backpropagation method. *Inverse Problems*, 21:565–590.

[85] Gratton, G., Fabiani, M., Corballis, P., Hood, D., Goodman-Wood, M., Hirsch, J., Kim, K., Friedman, D., and Gratton, E. (1997). Fast and localized event-related optical signals (EROS) in the human occipital cortex: comparisons with the visual evoked potential and fMRI. *NeuroImage*, 6(3):168–180.

[86] Haltmeier, M., Leião, A., and Scherzer, O. (2007). Kaczmarz methods for regularizing nonlinear ill-posed equations I: Convergence analysis. *Inverse Problems and Imaging*, 1(2):289–298.

[87] Haskell, R., Svaasand, L., Tsay, T., Reng, T., McAdams, M., and Tromberg, B. (1994). Boundary conditions for the diffusion equation in radiative transfer. *J. Opt. Soc. Amer. A*, 11(10):2727–2741.

[88] Haslinger, J. and Mäkinen, R. (2003). *Introduction to Shape Optimization.Theory, Approximation and Computation*. Society for Industrial and Applied Mathematics.

[89] Hata, T., Takahashi, H., Watanabe, K., Takahashi, M., Taguchi, K., Itoh, T., and Todo, S. (2004). Magnetic resonance imaging for preoperative evaluation of breast cancer: a comparative study with mamography and ultrasonography. *J. Am Coll. Surg.*, 198:190–197.

[90] He, L., Kao, C.-Y., and S., O. (2007). Incorporating topological derivatives into shape derivatives based level set methods. *Journal of Computational Physics*, 225:891–909.

[91] Hielscher, A., Klose, A., Catarious, D., and Hanson, K. (1998). Tomographic Imaging of Biological Tissue by Time-Resolved, Model-Based, Iterative Image Reconstruction. In *Advances in Optical Imaging and Photon Migration*, 21, pages 156–161.

[92] Hintermüller, M. and Laurain, A. (2008). Electrical impedance tomography from topology to shape. *Control and Cybernetics*, 37.

[93] Hintz, S., Benaron, D, S.-A., Zourabian, A., Stevenson, D., and D, B. (2001). Bedside functional imaging of the premature infant brain during passive motor activation. *J. Perinat. Med.*, 29:335–343.

[94] Hwang, E., Kinkel, K., Esserman, L., Lu, Y., Weidner, N., and Hylton, N. (2003). Magnetic resonance imaging in patients diagnosed with ductal carcinoma-in-situ value in the diagnosis of residual desease, occult invasion, and multicentricity. *Ann. Surg. Oncol.*, 10:381–388.

[95] Hylton, N. (2005). Magnetic Resonance Imaging of the breast: Opportunites to improve breast cancer management. *J. Clin. Oncol.*, 23:1678–1684.

[96] Irishina, N. (2009). *Microwave medical imaging using level set techniques*. PhD thesis, Universidad III Carlos de Madrid.

[97] Ishimaru, A. (1997). *Wave Propagation and Scattering in Random Media*. New York: IEEE Press.

[98] Jin, B., Kahn, T., and Maass, P. (2012). A reconstruction algorithm for electrical impedance tomography based on sparsity regularization. *International Journal for Numerical Methods in Engineering*, 89:337–353.

[99] Jin, W. and He, Y. (2012). Iterative Reconstruction for Bioluminescence Tomography with Total Variation Regularization. In *Optics in Health Care and Biomedical Optics V*, number 8553 in "SPIE".

[100] Joseph, D., Huppert, T., Franceschini, M., and Boas, D. (2006). Diffuse optical tomography system to image brain activation with improved spatial resolution and validation with functional magnetic resonance imaging. *Applied Optics*, 45(31):8142–51.

[101] Kaipio, J. and Somersalo, E. (2005). *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer.

[102] Kaltenbacher, B., Neubauer, A., and Scherzer, O. (2008). *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*. Radon Series on Computational and Applied Mathematics. de Gruyter.

[103] Kataja, M, editor (2005). *Multiphase Flows in Process Industry ProMini*. Julkaisija-Utgivare. Research notes.

[104] Keith, L., Oleszczuk, J., and Laguens, M. (2002). Are Mammography and Palpation Sufficient for Breast Cancer Screening? A Dissenting Opinion. *Journal of Wmen's Health and Gender-based Medicine*, 11:17–25.

[105] Kerlikowske, K., Grady, D., Rubin, S., Sandrock, C., and Ernster, V. (1995). Efficacy of screening mammography. A meta-analysis. *JAMA-Journal of the American Medical Association*, pages 149–154.

[106] Khairi, M., Ibrahim, S., Yunus, M., and Faramarzi, M. (2012). A review of on applications of Optical Tomography in Industrial Process. *International Journal on Smart Sensing and Intelligent Systems*, 5(4):767–798.

[107] Kolehmainen, V. (2001). *Novel approaches to image reconstruction in diffusion tomography*. PhD thesis, University of Kuopio.

[108] Kolehmainen, V., Lassas, M., and Siltanen, S. (2008). Limited data X-ray tomography using nonlinear evolution equations. *SIAM Journal of Scientific Computation*, 30(3):1413–1429.

[109] Kuzmin, D. (2010). A Guide to Numerical Methods for Transport Equations. note.

[110] Landweber, L. (1951). An iteration formula for Fredholm integral equations of the first kind. *Amer.J. Math.*, 73:615–624.

[111] Larusson, F., Fantini, S., and Miller, E. (2011). Hyperspectral image reconstruction for diffuse optical tomography. *Biomedical Optics Express*, 2(4):946–965.

[112] Leitão, A. and Marques Alves, M. (2012). On Landweber-Kaczmarz methods for regularizing systems of ill-posed equations in Banach spaces. *Inverse Problems*, 28:104008(15pp).

[113] Liu, X. and Hunag, L. (2010). Split Bregman iteration algorithm for total bounde variation regularization based image deblurring. *Journal of Mathematical Analysis and Applications*, 372:486–495.

[114] Macklin, P. and Lowengrub, J. (2005). Evolving interfaces via gradients of geometry-dependent interior Poisson problems: application to tumor growth. *Journal of Computational Physics*, 203:191–220.

[115] Madsen, S., editor (2013). *Optical Methods and Instrumentation in Brain Imaging*. Bioanalysis: Advanced Materials, Methods, and Devices. Springer.

[116] Moscoso, M. (2008). *Inverse Problems and Imaging*, volume 1943 of *Lecture Notes in Mathematics*, chapter Introduction to Image Reconstruction, pages 1–16. Springer.

[117] Muttarak, M., Pojchamarnwiputh, S., and Chaiwun, B. (2003). Preoperative detection by mammography. *Annals Academy of Medicine Singapore*, 32:433–437.

[118] Nass, S., I, H., and Lashof, J. (2001). *Mammography and beyond: developing technologies for the early detection of breast cancer*. National Academy Press: Washington DC.

[119] Nocedal, J. and Wright, J. (1999). *Numerical Optimization*. Series in Operations Research. Springer, second edition.

[120] Novotny, A., Feijóo, R., Taroco, E., and Padra, C. (2003). Topological sensitivity analysis. *Comput. Methods Appl. Mech. Engg.*, 192:803–829.

[121] Novotny, A. and Sokołowski, J. (2013). *Topological Derivatives in Shape Optimization*, volume XII. Springer.

[122] Ogiela, M. and Tadeusiewicz, R. (2008). *Modern Computational Intelligence Methods for the Interpretation of Medical Images*, volume 84 of *Studies in Computational Intelligence*, chapter Sources of medical images and their general characteristics, pages 7–46. Springer.

[123] Olesen, J., Gustavsson, A., Svensson, M., Wittchen, H., and Jönsson, B. (2012). The economic cost of brain disorders in Europe. *European Journal of Neurology*, 19:155–162.

[124] Olesen, J. and Leonardi, M. (2003). The burden of brain diseases in Europe. *European Journal of Neurology*, 10:471–477.

[125] Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. (2005). An iterative regularization method for total variation-based image restoration. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 4(2):460–489.

[126] Osher, S. and Fedkiw, R. (2003). *Level set Method and Dynamic Implicit Surfaces*, volume 153 of *Applied Mathematical Sciences*. Springer.

[127] Osher, S., Mao, Y., Dong, B., and Yin, W. (2010). Fast Linearized Bregman iteration for Compressive sensing and Sparse Denoising. *Comunications in Mathematical Sciences*, 8(1):93–111.

[128] Osher, S. and Sethian, J. (1988). Fronts propagating with curvature -dependent speed: algorithms based on Hamilton-Jacobi formulations. *J. Comput. Phys*, 79:12–49.

[129] Osher, S. and Tsai, Y.-H. R. (2005). Total Variation and Level Set Based Methods in Image Science. *Acta Numerica*, 14:509–573.

[130] Peng, D., Merriman, B., Osher, S., Zhao, H., , and Kang, M. (1999). A PDE-Based Fast Local Level Set Method. *Journal of Computational Physics*, 155:410–438.

[131] Perkins, G. and Middleton, L. (2003). Breast cancen in men. *BMJ*, 327:239–240.

[132] Pisano, E., Gatsonis, C., Hendrick, E., Yaffe, M., Baum, J., Acharyya, S., and et. al. (2005). Diagnostic performance of digital versus film mammography for breast-cancer screening. *Annals Academy of Medicine Singapore*, 353:1773–1783.

[133] Pogue, B., Davis, S., Leblond, F., Mastanduno, M., Dehghani, H., and K, P. (2011). Implicit and explicit prior information in near-infrared spectral imaging: accuracy, quantification and diagnostic value. *Phil. Trans. R. Soc. A*, 369:4531–4557.

[134] Prakash, J. (2014). *Development of next generation image reconstruction algorithms for Diffuse Optical and Photoacoustic Tomography*. PhD thesis, Supercomputer Education and Research Centre (SERC), Indian Institute of Science.

[135] Prakash, J., Shaw, C., Manjappa, R., Kanhirodan, R., and Yalavarthy, P. (2014). Sparse Recovery Methods Hold Promise for Diffuse Optical Tomographic Image Reconstruction. *IEEE Journal of Selected Topics in Quantum Electronics*, 20(2):6800609.

[136] Press, W.and Teukolsky, S., Vatterling, W., and Flannery, B. (1998). *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, second edition.

[137] Ramlau, R. and Teschke, G. (2006). A Tikhonov-Based Projection Iteration for Nonlinear Ill-Posed Problems with Sparsity Constraints. *Numerische Mathematik*, 104:177–203.

[138] Ren, K. Bal, G. and Hielscher, A. H. (2007). Frequency domain optical tomography based on the radiative transfer equation. *SIAM J. Sci. Comput.*, 28(4):1463–1489.

[139] Ren, K., Bal, G., and Hielscher, A. H. (2007). Transport- and diffusion-based optical tomography in small domains: a comparative study. *Applied Optics*, 46(27):6669–6679.

[140] Rossen, L. and Rossen, G. (2015). American Cancer Society, Cancer Facts and Figures 2013, http://www.cancer.org/.

[141] Santosa, F. (1996). A Level-Set Approach for Inverse Problems Involving Obstacles. *ESAIM:Control, Optimization and Calculus of Variations*, 1:17–33.

[142] Saratoon, T., Tarvainen, T., Cox, B., and Arridge, S. (2013). A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation. *Inverse Problems*, 29(7):075006.

[143] Schöpfer, F., Louis, A., and Schuster, T. (2006). Nonlinear iterative methods for linear ill-posed problems in Banach spaces. *Inverse Problems*, 22:311–329.

[144] Schweiger, M., Arridge, S., Dorn, O., Zacharopoulos, A., and Kolehmainen, V. (2006). Reconstruction absorption and diffusion shape profiles in optical tomography using a level set technique. *Optics Letters*, 31(4):471–473.

[145] Schweiger, M., Gibson, A., and Arridge, S. (2003). Computational aspects of diffuse optical tomography. *IEEE Signal Process Mag.*, 5(6):33–41.

[146] Sethian, J. (1996). *Level set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computers Visions and material Sciences*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press.

[147] Sokołowski, J. and Zolésio, J. (1991). *Introduction to Shape Optimization, Shape Sensitivity Analysis*, volume 16. Springer-Verlag, Springer Series in Computational Mathematics.

[148] Spencer, A. (1980). *Continuum Mechanics*. Dover Publication Inc.

[149] Stefan, W. (2008). *Total Variation Regularization for Linear ill-posed Inverse Problems: extensions and applications*. PhD thesis, Arizona State University.

[150] Strangman, G., Boas, D., and Sutton, J. (2002). Non-Invasive Neuroimaging Using Near-Infrared Light. *Biol Psychiatry*, 52(7):679–93.

[151] Suri, J., Liu, K., Singh, S., Laxminarayan, S., Zeng, X., and Reden, L. (2002). Shape recovery algorithms using level sets in 2D/3D medical imagery: a state-of-the-art review. *IEEE Trans. Information Technol. Biomed.*, 6:8–28.

[152] Tarantola, A. (2005). *Inverse Problem Theory and Methods for Model Parameter Estimation*. Society for Industrial and Applied Mathematics.

[153] Tarvainen, T. (2006). *Computational Methods for Light Transport in Optical Tomography*. PhD thesis, University of Kuopio, Finland.

[154] Tarvainen, T., Cox, B., Kaipio, J., and Arridge, S. (2012). Reconstructing absorption and scattering disctributions in quantitative photoacoustic tomography. *Inverse Problems*, 28(8):084009, 17pp.

[155] Tarvainen, T., Vauhkonen, M., Kolehmainen, V., Arridge, S., and Kaipio, J. (2005). Coupled radiative transfer equation and diffusion approximation model for photon migration in turbid medium with low-scattering and non-scattering regions. *Physics in medicine and bioloby*, 50(20):4913–4930.

[156] Tarvainen, T., Vauhkonen, M., and S.R., A. (2008). Gauss-Newton reconstruction method for optical tomography using the ginete element solution of the radiative transfer equation. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 109:2767–2778.

[157] Teschke, G. and Borries, C. (2010). Accelerated projected steepest descent method for nonlinear inverse problems with sparsity constraints. *Inverse Problems*, 26:025007.

[158] van den Doel, K., Ascher, U., and Haber, E. (2013). *The lost honour of ℓ_2-based regularization*, volume 13 of *Radon Series on Computational and Applied Mathematics*, pages 1–25. Large Scale Inverse Problems. Computational Methods and Applications in the Earth Sciences.

[159] Venugopal, V. and Intes, X. (2012). Recent Advances in Optical Mammography. *Current Medical Imaging Reviews*, 8(3):244–259.

[160] Vidal-Rosas, E., Billings, S., Zheng, Y., Mayhew, J., Johnston, D., Kennerley, A., and Coca, D. (2014). Reduced-order modeling of light transport in tissue for real-

time monitoring of brain hemodynamics using diffuse optical tomography. *Journal of Biomedical Optics*, 19(2):026008.

[161] Villringer, A. and Britton, C. (1997). Non-invasive optical spectroscopy and imaging of human brain function. *Trends Neurosci.*, 20(10):435–442.

[162] Vogel, C. (2002). *Computational Methods for Inverse Problems*. Frontier in Applied Mathematics. Society for Industrial and Applied Mathematics.

[163] Vogel, C. and Oman, M. (1998). Fast, robust total variation-based reconstruction of noisy blurred images. *IEEE Transactions on Image Processing*, 7(6):813–824.

[164] Wang, L. e. (2009). *Photoacoustic Imaging and Scpectroscopy*. CRC Press, Boca Raton, FL.

[165] Wang, Y., Yang, J., Yin, W., and Zhang, Y. (2008). A new alternating minimization algorithm for total Variation Image Reconstruction. *SIAM Journal on Imaging Sciences*, 1(3):248–272.

[166] Wang, Y., Ying, W., and Zhang, Y. (2007). A Fast Algorithm for Image Deblurring with Total Variation Regularization. Technical report, Department of Computational and Applied Mathematics, USA.

[167] Wang, Y. and Yagola, A. and Yang, C. (2011). *Optimization and Regularization for Computational Inverse Problems and Applications*. Springer Berling Heidelberg.

[168] Widmer, G. (2009). An Efficient Sparse Finite Element Solver for the Radiative Transfer Equation. *J. Heat Transfer*, 132(2):023403.

[169] Williams, R. and Beck, M., editors (1995). *Process Tomography: Principles, Techniques and Applications*. Butterworth-Heinemann Ltd.

[170] Wittchen, H., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., and et. al. (2011). The size and burden of mental disorders and other disorders of the brain in Europe. *European Neuropsychopharmacology*, 21(9):655–679.

[171] Wright, S., Nowak, R., and Figueiredo, M. (2009). Sparse Reconstruction by Separable Approximation. *IEEE Transactions on Signal Processing*, 57(7):2479–2493.

[172] Xu, Y., Huang, T.-Z., Liu, J., and Lv, X.-G. (2013). Split Bregman Iteration Algorithm for Image Deblurring Using Fourth-Order Total Bounded Variation Regularization Model. *Journal of Applied Mathematics*, 2013:1–11.

[173] Yates, T. (2005). *Time-resolved Optical Tomography for the Detection and Specification of Breast Disease*. PhD thesis, University College London,UK.

[174] Yin, W., Osher, S., Goldfarb, D., and Darbon, J. (2008). Bregman Iterative Algorithms for ℓ_1 Minimization with Applications to Compressed Sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168.

[175] Zhang, X., Burger, M., Bresson, X., and Osher, S. (2010). Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction. *SIAM Journal on Imaging Sciences*, 3(3):253–276.

[176] Zhu, Y. and Yang, X. (2011). TV Sparsifying MR Image REconstruction in Compressive Sensing. *Journal of Signal and Information Processing*, 2:44–51.

[177] Zijlstra, W. and Buursma, A. (1997). Spectrophotometry of Hemoglobin: Absorption Spectra of Bovine Oxyhemoglobin, Deoxyhemoglobin, Carboxyhemoglobin, and Methemoglobin. *Comp. Biochem. Physiol.*, 118B(4):743–749.