

**Inverse Problems in Transport and Diffusion
Theory with Applications in Optical Tomography**

Kui Ren

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2006

© 2006

Kui Ren

All Rights Reserved

ABSTRACT

Inverse Problems in Transport and Diffusion Theory with Applications in Optical Tomography

Kui Ren

The work in this thesis mainly concerns inverse problems in transport and diffusion theory with an emphasis on applications in imaging techniques such as optical tomography and atmospheric remote sensing. Mathematically, inverse problems here involve the reconstruction of coefficients in partial differential (and integro-differential) equations from boundary measurements.

The first half of the thesis are devoted to the analysis and numerical solutions of inverse transport problems in optical tomography and atmospheric remote sensing. We developed two reconstruction algorithms for optical tomography in which we use the frequency domain transport equation as the forward model of light propagation in tissues. We show by numerical examples that the usage of the frequency domain information allows us to reduce the crosstalk between absorption and scattering coefficients in transport reconstructions from boundary current measurements. The crosstalk is much severe when steady-state data are used in the reconstruction. We have also analyzed an inverse problem related to the scattering-free atmospheric radiative transport equation. The inverse problem aims at reconstructing the concentration profiles of atmospheric gases (parameterized as functions of altitude in both the coefficient and the source term of the transport equation) from wavenumber-dependent boundary radiation measurement taken by space-borne infrared spectrometer. We showed in simplified situations that although the problem does admit a unique solution, it is severely ill-posed. We proposed an explicit procedure based on asymptotic analysis to reconstruct localized structures in the profile.

Modeling microscopic transport processes by macroscopic diffusion equations has its advantage many applications. Mathematically the modeling problem corresponds

to the derivation of diffusion equations from transport equations. The second half of the thesis is devoted to such modeling problems and inverse problems related to them. We first compared in detail numerical reconstructions based the transport and diffusion equations in highly scattering and low absorbing media of small size. We characterized quantitatively the effect of inaccuracy in the diffusion approximation on the quality of the reconstructions. We then derived a generalized diffusion approximation for light propagation in highly diffusive media with extended thin non-scattering regions based on several previously reported results. We modeled those non-scattering extended regions by co-dimension one surfaces and used localized surface conditions to account for the effects of those non-scattering regions. Numerical simulations confirmed the accuracy of the new diffusion approximation. An inverse problem related to this generalized diffusion equation was then analyzed. The aim of this inverse problem is to reconstruct the locations of those extended non-scattering regions. We showed by numerical simulation that those regions be reconstructed from over-determined boundary measurements. The reconstruction method is based on shape sensitivity analysis and the level set method.

Contents

List of Figures	viii
List of Tables	x
Acknowledgments	xi
1 Introduction	1
1.1 Optical tomography	1
1.2 The inverse transport problems	3
1.3 Diffusion approximations and inversions	4
1.4 Outline of the thesis	5
2 Inverse transport problem in frequency domain optical tomography	7
2.1 Problem formulation	7
2.1.1 Forward problem	10
2.1.2 Least square formulation	11
2.2 Discretization methods	15
2.2.1 The discrete ordinates formulation	15
2.2.2 Spatial discretization	16
2.2.3 Discrete adjoint problem	19
2.3 Numerical implementation	22
2.3.1 Numerical optimization	23
2.3.2 Solving algebraic systems	25
2.3.3 Selecting regularization parameter	26

2.3.4	Cost of the numerical method	27
2.4	Numerical examples	28
2.4.1	Forward simulations	28
2.4.2	Setup for the reconstructions	31
2.4.3	Generating synthetic data	32
2.4.4	Single parameter reconstructions	33
2.4.5	Frequency-domain versus steady-state	37
2.5	Conclusions and remarks	41
3	Inverse transport as a PDE-constrained optimization problem	44
3.1	Problem statement	44
3.2	The augmented Lagrangian method for inverse transport	48
3.2.1	The augmented-Lagrangian algorithm	49
3.2.2	Interpretation and discussion	52
3.3	Numerical reconstructions	54
3.3.1	The test problem setup	54
3.3.2	Reconstruction of absorption coefficients	57
3.3.3	Reconstruction of scattering coefficients	64
3.3.4	Simultaneous reconstruction of two coefficients	69
3.4	Conclusions and remarks	70
4	Inverse transport problem in atmospheric remote sensing	72
4.1	Problem statement	72
4.2	The mathematical model	74
4.3	Uniqueness and ill-posedness of a simplified model	76
4.3.1	The case of a single gas	77
4.3.2	The case of multiple gases	82
4.4	Small inclusions	84
4.4.1	The case of a single gas	84
4.4.2	The case of multiple gases	87
4.5	Numerical reconstructions	89

4.5.1	The case of a single gas	89
4.5.2	The case of two gases	93
4.6	Conclusions and remarks	94
5	Comparison of transport and diffusion reconstructions in small domains	96
5.1	Problem statement	96
5.1.1	Transport and diffusion approximations	97
5.2	Reconstruction methods	100
5.2.1	Reconstruction algorithms	100
5.2.2	Discretization of forward models	101
5.3	Numerical results	103
5.3.1	Setup for the reconstructions	103
5.3.2	Diffusive media of small size	105
5.3.3	Effects of modulation frequency	108
5.3.4	The impact of the extrapolation length	110
5.3.5	Diffusive media with void regions	112
5.4	Conclusions and remarks	114
6	Generalized diffusion approximation and its validations	116
6.1	Problem statement	116
6.2	Generalized diffusion model	120
6.2.1	Notation and Geometry.	120
6.2.2	Generalized diffusion equation with non-local interface conditions.	122
6.2.3	Localization of the interface conditions.	123
6.2.4	Tangential diffusion coefficient for circular layers.	125
6.2.5	Generalized diffusion model with local interface conditions.	126
6.2.6	Remarks on the mathematical model.	127
6.3	Validation of the model with forward simulations	129
6.3.1	Two dimensional numerical simulations.	130
6.3.2	Interpretation of results.	132

6.3.3	Three dimensional numerical simulations.	133
6.4	Conclusions and remarks	134
7	Surface identifications by shape sensitivity analysis and the level set method	138
7.1	The singular surface problem	138
7.1.1	Forward model	139
7.1.2	Inverse surface problem	142
7.1.3	Comparison with the reconstruction of inclusions	144
7.2	Shape sensitivity analysis	145
7.2.1	The material derivatives	147
7.2.2	The shape derivative	151
7.3	Choosing the direction of descent	156
7.4	Level set implementation	158
7.4.1	Representing and moving interfaces	159
7.4.2	Implementation of the level set method	159
7.5	Numerical simulations	162
7.5.1	Reconstructions of ellipses	163
7.5.2	Reconstruction of more complicated surfaces	165
7.6	Conclusions and remarks	169
8	Summary	170
	Bibliography	186

List of Figures

1-1	Schematic illustration of optical tomographic problem. Near infra-red lights are sent into the tissue Ω from point sources located at the surface and the outgoing currents of photons are measured by some detectors (\square). Optical properties, absorption $\sigma_a(\mathbf{x})$ and scattering $\sigma_s(\mathbf{x})$, of the tissue are objects that are sought.	2
2-1	Geometrical settings of the computational domains. Diamond (\diamond) and circle (\circ) denote source and detectors, respectively.	29
2-2	AC amplitude (a) phase delay (b) computed at the detectors for different optical parameters	30
2-3	The difference of (a) AC amplitude ($I^v - I^h$) and (b) phase delay ($\theta^v - \theta^h$) calculated at the detectors for various modulation frequencies in domain with a void inclusion	31
2-4	Reconstructed absorption coefficients using data with different noise levels.	34
2-5	Evolution of normalized objective function with respect to the number of iterations and the L -curve used to choose optimal regularization parameter.	35
2-6	Reconstructed reduced scattering coefficients using data with different noise levels.	36
2-7	Reconstructed absorption and reduced scattering coefficients at different iterations using frequency domain data.	37
2-8	Reconstructed absorption and reduced scattering coefficients at different iterations using steady state data.	39
2-9	Evolution of normalized objective function with respect to the number of iterations and the L -curve used to choose optimal regularization parameter in Example 3.	40
2-10	Cross sections of reconstructed absorption and reduced scattering coefficients in a cylinder using frequency domain data.	41
2-11	Cross sections of reconstructed absorption and reduced scattering coefficients in a cylinder using steady state data	42
3-1	A simple illustration of the iteration process of unconstrained (\square) and constrained (\circ) optimization approaches to optical tomography. The subscript u and c denotes quantities in unconstrained and constrained minimization process, respectively.	53

3-2	Test problems setup. Cylinder height: $H = 2$ cm, radius $r = 1$ cm; radius of the embedded small cylinder $r = 0.25$ cm. (a) source-detector layout with 8 sources (\square), 64 detectors (\circ); (b) finite-volume mesh with 6727 tetrahedrons.	54
3-3	Convergence history of $E(\hat{\mathbf{U}}^k)/E(\hat{\mathbf{U}}^0)$ for σ_a reconstruction (in \log_{10} scale). (a) The lm-BFGS unconstrained optimization method with no noise. (b) The augmented Lagrangian method, $\chi = \infty$ (no noise), and $\chi = 15$ dB; (c) The augmented Lagrangian method, $\chi = 20$ dB and different regularization parameters. All the values of β are given in units of $[10^{-10}]$	58
3-4	Cross sections of the reconstructed absorption coefficient in the plane $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the quasi-Newton lm-BFGS method for the unconstrained optimization and the ALM for problem 1 with different noise levels. The target optical properties are $\sigma_a = 0.2 \text{ cm}^{-1}$ in the inclusion and $\sigma_a = 0.1 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to the reconstruction with unconstrained minimization approach; (c) and (d) correspond to the ALM reconstruction with noise free data; (e) and (f) correspond to the ALM reconstruction with 15 dB added noise.	61
3-5	Cross sections of the reconstructed absorption coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the ALM for problem 1 with different regularization parameters. The target optical properties are $\sigma_a = 0.2 \text{ cm}^{-1}$ in the inclusion and $\sigma_a = 0.1 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to the ALM reconstruction with $\beta = 10 \times 10^{-10}$; (c) and (d) correspond to the ALM reconstruction with $\beta = 200 \times 10^{-10}$; (e) and (f) correspond to the ALM reconstruction with $\beta = 500 \times 10^{-10}$	63
3-6	Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the augmented Lagrangian method for problem 2. The target optical properties are $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to the reconstruction after 50 iterations of the ALM; (c) and (d) correspond to the reconstruction after 200 iterations of the ALM; (e) and (f) correspond to the reconstruction at convergence (498 iterations).	64
3-7	Convergence history of $E(\hat{\mathbf{U}}^k)/E(\hat{\mathbf{U}}^0)$ for σ_s reconstruction (in \log_{10} scale). (a) The augmented Lagrangian method, $\chi = 20$ dB and $\beta = 500 \times 10^{-10}$ dB; (b) The augmented Lagrangian method with different initial guesses, $\chi = \infty$ dB and $\beta = 300 \times 10^{-10}$; (c) the lm-BFGS unconstrained optimization method with no noise.	65

3-8	Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the augmented Lagrangian method for problem 2 with different initial guesses. The target optical properties are $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to initial guess $\sigma_s^0 = 10 \text{ cm}^{-1}$; (c) and (d) correspond to initial guess $\sigma_s^0 = 11 \text{ cm}^{-1}$; (e) and (f) correspond to initial guess $\sigma_s^0 = 12 \text{ cm}^{-1}$	66
3-9	Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the augmented Lagrangian method for problem 2 with different meshes. The target optical properties are $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to mesh with 10062 tetrahedrons; (c) and (d) correspond to mesh with 15612 tetrahedrons; (e) and (f) correspond to mesh with 19489 tetrahedrons.	68
3-10	Test problem 3 setup. Cylinder height: $H = 5 \text{ cm}$, radius $r = 1.5 \text{ cm}$; radius of the embedded small cylinder $r = 0.5 \text{ cm}$. (a) source-detector layout with 24 sources (\square), 24 detectors (\circ); (b) finite-volume mesh with 13867 tetrahedrons.	69
3-11	Cross sections of the reconstructed absorption and scattering coefficients in the planes $y = 0$, $z = 2.2$ and $z = 3.5$ with the augmented-Lagrangian method for problem 3. The target optical properties are $\sigma_a = 1.0 \text{ cm}^{-1}$, $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_a = 0.5 \text{ cm}^{-1}$, $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) Reconstruction of σ_a at convergence (712 iterations), left-top: cross section $z = 3.5$, left-bottom: cross section $z = 2.2$, right: cross section $y = 0$. (b) Reconstruction of σ_s at convergence (712 iterations).	70
4-1	Profiles used in the calculation. (a) Temperature profile as a function of z . (b) Rescaled absorption as a function of wavelength. (c) Ozone concentration as a function of z . (d) Data $D(Z, \mu(\nu))$ as a function of wavenumber ν	89
4-2	Cross section of the error functional in the parameter space. (a) functional at $z_0 = 0.3$; (b) functional at $\delta z = 0.06$ and $\delta c = 1.0$	91
5-1	XZ ($y = 0$) and XY ($z = 1$) cross-sections of the computational domain.	104
5-2	Cross sections of reconstructed absorption coefficients in small media	106
5-3	Quality in transport and diffusion reconstructions using data with different noise levels.	107
5-4	Cross sections of reconstructed absorption coefficients with intensity modulated source	108
5-5	Quality of reconstructions as functions of modulation frequencies (in unit of GHz). Left: reconstructions with noise-free data; Right: reconstructions with 12% noise in the data.	109
5-6	Reconstructed absorption coefficients with transport and diffusion models	110

5-7	Quality of reconstructions as functions of extrapolation length. Left: reconstructions with noise-free data; Right: reconstructions with 12% noise in the data. Transport reconstructions are shown here just as a reference.	111
5-8	XZ ($y = 0$) and XY ($z = 1$) cross-sections of the computational domain with a void inclusion.	112
5-9	Cross sections of reconstructed absorption coefficients in media with void regions	113
5-10	Quality in transport and diffusion reconstructions using data with different noise levels.	114
6-1	Local geometry of the clear layer.	121
6-2	Geometrical settings of numerical simulations	130
6-3	Plots of boundary current calculated with different 2-dimensional models	136
6-4	Plots of boundary current calculated with different 3-dimensional models	137
7-1	Geometric setting of the problem in the two-dimensional setting with $\Omega = \Omega_I \cup \Omega_E \cup \Sigma$	141
7-2	Reconstruction of an elliptic interface with synthetic data at different noise levels for full and local Neumann to Dirichlet maps	164
7-3	Reconstruction of a star-shaped interface with synthetic data at different noise levels for full and local Neumann to Dirichlet maps	166
7-4	Same as in Fig. 7-3 except that $N = 5$	167
7-5	Errors in the reconstructions of star-shaped interface for different noise levels	168

List of Tables

2.1	Optimal regularization parameters β and errors in reconstructions for different cases in Example 1 and Example 2.	36
2.2	Error estimates for the reconstructions of Example 3 (<i>E3</i>) and Example 4 (<i>E4</i>) for several iteration steps (k) in the optimization process. Here, “f” refers to frequency-domain calculations and “s” to steady state calculations.	43
3.1	Parameters used in three different problems	55
3.2	Quality of reconstruction of the absorption coefficient for different reconstruction methods, different noise levels and different regularization parameters. The parameter β is given in unit of $[10^{-10}]$	59
3.3	Quality of reconstruction of the scattering coefficient as a function of ALM iteration step.	66
3.4	Quality of reconstruction of the scattering coefficient as a function of the initial guess.	66
3.5	Quality of reconstruction of the scattering coefficient as a function of the mesh size.	68
3.6	Quality of reconstruction of the absorption and scattering coefficients as a function of the ALM iteration step.	69
4.1	Characteristics of the inclusion reconstructed by a full search algorithm. The true values are $z_0 = 0.3$, $\delta z = 0.06$, $\delta c = 1.0$, hence $\delta z \delta c = 0.06$. The numbers in parentheses denote the relative error in percentage between the reconstructed parameters and their true values.	92
4.2	Same as Tab. 4.1 (with the <i>same</i> noisy measurements) except that the Conjugate Gradient algorithm is used in the optimization process.	92
4.3	Characteristics of the inclusion reconstructed by the Conjugate Gradient algorithm when the inclusion is placed in a region with vanishing temperature gradient. The real values for those variables are $z_0 = 0.25$, $\delta z = 0.08$, $\delta c = 1.20$ and $\delta z \delta c = 0.096$. The numbers in parentheses denote the relative error in percentage between the reconstructed parameters and their true values.	93

4.4	Characteristics of the inclusions in the two-particle model reconstructed from noise free data. The initial guess is $z_1 = 0.32$, $\delta z_1 = 0.05$, $\delta c_1 = 0.8$ and $z_2 = 0.28$, $\delta z_2 = 0.10$, $\delta c_1 = 1.0$. The numbers in parentheses denote the relative error in percentage between the reconstructed parameters and their true values.	93
4.5	Same as Tab.4.4 with 0.10% noise.	94
4.6	Same as Tab. 4.4 with 1% noise.	94
6.1	Tangential diffusion coefficients for clear layers of different thickness .	131
7.1	Errors in the reconstructions of ellipses (7.69) with different values of (a, b) using model (7.1) with full measurements. The center of original interfaces $(x_0, y_0) = (0, 0)$	165
7.2	Same as Tab. 7.1 except that the reconstructions are obtained from partial measurements.	165
7.3	Reconstructed centers for the cases presented in Figs. 7-3 ($N = 3$) and 7-4 ($N = 5$).	167

Acknowledgments

I would like to first thank my PhD advisor, Professor Guillaume Bal, for his continuous guidance, advice and support during the past years. To him, I own much more than I can say. I am very fortunate and pleased to work with him. In fact, this thesis will never be possible without the tremendous help he has provided.

I would also like to thank Professor Andreas H. Hielscher for his guidance, advice and support. He showed me many practical problems in optical tomography and provided many insightful discussions. I enjoyed working with him during my PhD study and I am looking forward to future collaborations with him.

I thank Professor C. K. Chu for his continuous encouragement and help during my years at Columbia. I enjoyed every talk with him.

My gratitude also goes to Professor Michael I. Weinstein. I thank him for taking the time to serve on both my oral exam committee and thesis defense committee. I also thank Professor Donald Goldfarb and Professor Harish S. Bhat for serving on my thesis defense committee.

Professor David E. Keyes introduced me the whole idea of domain decomposition methods for partial differential equations which, although are not directly applied in thesis, are very useful for my research. I own him everything in this aspect.

I thank members in Biophotonics and Optical Radiology Lab in Biomedical Engineering Department for many useful discussions on optical tomography problems.

The wonderful administrative staff in the Department of Applied Physics and Applied Mathematics, Marlene Arbo, Lydia Argote, Christina Rohm and Shoshana Filene, have provided me numerous help on various issues during my years at Columbia. I thank them for everything they have done for me and for the department.

It has to be pointed out that many results presented in the thesis are obtained collaboratively with Professor Guillaume Bal and Professor Andreas H. Hielscher. The results in chapter 3 is based on a joint paper with Dr. Gassan Abdoulaev. I thank all of them for allowing me to present those results in the thesis.

Finally I thank my wife Yuan, who was always with me for the last five years. I own her more than what I can express in English. I also thank my family in China

for their continuous love and support.

The research in this work was partially supported by funding provided by the Department of Applied Physics and Applied Mathematics, Academic Quality Fund of Columbia University, Institute of Pure and Applied Mathematics at University of California, Los Angeles (UCLA), National Institute of Health (NIH) (through Professor Andreas H. Hielscher) and National Science Foundation (NSF) (through Professor Guillaume Bal). Their support is greatly acknowledged.

Chapter 1

Introduction

Inverse problems related to the transport and diffusion equations have long been of interest in many fields of applied sciences. In those problems, one attempts to determine the spatial distribution of constitutive parameters in the equations. Applications include for instance astronomy [182], nuclear science [109, 110], atmospheric science [45, 46] as well as many other fields [3, 11, 51, 127]. Recent advances in inverse transport and diffusion theory have been fueled by an increased interest in optical tomography where one attempts to reconstruct the absorption and scattering coefficients inside the body using boundary measurements of near infra-red light [11, 13, 36, 48, 86, 93, 99, 114, 122, 123, 130, 142, 151, 162, 185].

1.1 Optical tomography

In optical tomography experiment, one sends near infra-red light into biological tissues and measures the outgoing photon current at the surfaces of the tissues. One then wants to infer the optical properties of the tissues from knowledge of those measurements. These optical properties can be used for diagnostic purposes [48]. Applications of optical tomography include, for example, brain [30, 37, 93], breast [114, 143, 144, 177, 174] and joint imaging [92, 108, 131, 145, 184]. We refer to Fig. 1-1 for a schematic illustration of optical tomography problem and to [11, 13, 19, 34, 62, 122, 123, 130, 170, 178] for discussions on practical, theoretical, and computational aspects of optical tomography.

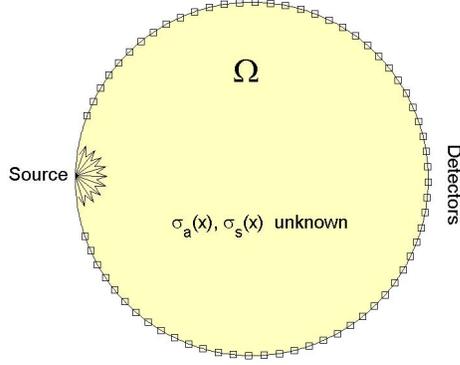


Figure 1-1: Schematic illustration of optical tomographic problem. Near infra-red lights are sent into the tissue Ω from point sources located at the surface and the outgoing currents of photons are measured by some detectors (\square). Optical properties, absorption $\sigma_a(\mathbf{x})$ and scattering $\sigma_s(\mathbf{x})$, of the tissue are objects that are sought.

Since the propagation of near infra-red light in tissues is best modeled by the radiative transport equation, mathematically optical tomography reduces to an inverse problem in transport theory. The term “inverse problem” here refers to the case where instead of solving the transport equation with given coefficients, one try to reconstruct those coefficient from partial information (typically boundary information) about a family of solutions.

An important issue in optical tomography is the so-called crosstalk phenomenon. What we mean by cross-talk is that purely scattering (or purely absorbing) inclusions are often reconstructed with unphysical absorption (or scattering) properties. This behavior is well-understood from the theoretical viewpoint: different optical distributions inside the medium can lead to the same measurements collected at the surface of the medium [14, 96]. Additional information needs to be obtained to improve the reconstructions, and multiple frequency data is a way to do so.

In frequency domain optical tomography experiments, the light source intensity is modulated, typically at 100 – 500 MHz, which leads to the propagation of so-called photon density waves in scattering media. In chapter 2 and chapter 3 of this thesis, we have developed numerical reconstruction algorithms that are based on the frequency-domain transport equation. We show by numerical examples that the usage of the frequency domain information allows us to reconstruct simultaneously the two

coefficients in the transport equation from boundary current measurements.

1.2 The inverse transport problems

In many recent applications of transport theory, one aims at reconstructing the constitutive coefficients in the transport equation from measured data around domain boundary [15, 16, 126, 127, 169, 52, 170]. As we have seen, optical tomography is one of such applications. In fact, there exist a relatively large body of inverse problems related to the transport equation in mathematical and engineering sciences [10, 66, 74, 109, 110, 182]; see for example also the reviews in [126, 127]. These inverse problems pose difficult analytical and computational questions. Most of the inverse transport problems are ill-posed [67, 96, 103], meaning that either a) there exists no solution to the reconstruction of the coefficients from available data, or b) there are more than one solutions, or c) the solution of the problem does not depend continuously on the available data; see above references for more precise definitions of ill-posedness. Among ill-posed problems, some are called mildly ill-posed, the others are called severely ill-posed [67]. Essentially mildly ill-posedness means that assuming the uniqueness of reconstruct holds, when no regularization is applied, noise contained in the data is amplified during the inversion procedure comparable to what would result from a finite number of differentiations. When noise is more amplified than what would result from an arbitrary number of differentiations, we say the problem is severe ill-posed [67].

Atmospheric remote sensing is one of such severely ill-posed inverse transport problems. In atmospheric remote sensing, one aims at reconstructing the concentration of atmospheric gases (parameterized as functions in the source term of the transport equation) from wavenumber-dependent boundary radiation measurement taken by space-borne infrared spectrometer. The problem is severely ill-posed because the reconstruction invokes the inversion of Laplace transform which is a notoriously unstable process. We demonstrate in chapter 4 that although the problem does admit a unique solution, it is severely ill-posed. Instead of attempting to reconstruct the

whole concentration profile, one should really focus on feature reconstruction. We propose an explicit procedure based on asymptotic analysis to reconstruct localized structures in the profile.

1.3 Diffusion approximations and inversions

Derivation of macroscopic diffusion models for microscopic transport processes is important in many applications. Diffusion equation models the spatial (and not the phase space) particle density. It is both analytically more tractable and computationally less expensive than the transport equation.

Inverse problems in diffusion theory have been extensively studied. Many theoretical [14, 19, 51, 122, 123, 120, 121, 172, 178] and computational [11, 13, 47, 141, 152, 153, 162] analysis on inverse diffusion problem have been done in the past years.

The derivation of diffusion to model photon propagation is quite classical [11, 58, 65, 111]. Essentially diffusion can be used when scattering is high and absorption small. Such assumptions are verified by almost all human tissues in the head but for a thin layer filled with cerebrospinal fluid. This layer is almost collision-less and absorption-less. Several studies show that diffusion models perform very poorly in such layers [13, 59, 90, 149].

There exists a large literature on numerical techniques that allow us to use coarser schemes (modeling transport or diffusion equations) in the regions where multiple scattering makes the simulation relatively straightforward and finer schemes in the vicinity of the clear layer where transport effects must be calculated accurately [22, 26, 77, 112, 175].

Because clear layers are thin in practice, an alternative solution exists to solving transport equations. Hybrid models that would solve a diffusion equation where the tissues are highly scattering and model the transport behavior in the clear layer have been developed [13, 59, 149]. Similar models were developed using an approach based on the asymptotic expansion of transport equations [17, 18].

Based on the work in [17, 18], we develop in chapter 6 of this thesis a generalized

diffusion model that can correctly model the effect of clear layers while keeping the computational cost on the same level as classical diffusion approximations. In the new diffusion equation, the clear layers are replaced by co-dimension one surfaces and their effects are modeled by tangential diffusion process supported on the surfaces. We present numerical simulations that confirm the accuracy of the new diffusion approximation. An inverse problem related to this generalized diffusion equation is then analyzed in chapter 7. The aim of this inverse problem is to reconstruct the location of those extended non-scattering regions. We show by numerical simulations that those regions be reconstructed from over-determined boundary measurements. The reconstruction method is based on shape sensitivity analysis and the level set method.

1.4 Outline of the thesis

The thesis is organized as follows. We develop in chapter 2 and chapter 3 two reconstruction algorithms for optical tomography that are based on the frequency domain transport equation. We introduce a numerical procedure that combines a spatial finite volume discretization, an angular discrete ordinate method, with a GMRES algorithm to solve forward problems of the transport equation. The inversion methods are based on numerical optimization techniques. A quasi-Newton type of method is tested in chapter 2 and a PDE-constrained optimization method is presented in chapter 3. The comparison between those methods are also considered. We then consider an inverse transport problem in atmospheric remote sensing in chapter 4 where we show in simplified situations that although the problem admits a unique solution, it is severely ill-posed. We also propose an explicit procedure based on asymptotic analysis to reconstruct localized structures in the profile. Chapter 5 is devoted to a detailed comparison between transport-based and diffusion-based reconstructions in small domains. In chapter 6 we derive a generalized diffusion equation for photon propagation in diffusive media with clear layers. We also present numerical simulations of the new diffusion equations. Chapter 7 is devoted to a study of an inverse

problem of the generalized diffusion model where we reconstruct clear layers from boundary measurement. A summary of the thesis is offered in chapter 8.

The chapters in this thesis are based on published [2, 23, 24, 25, 146] and submitted [147, 148] research papers. We have tried to keep them relatively self-contained, which causes some repetition in the presentation. Chapter 2 is based on [146, 147]; chapter 3 on [2]; chapter 4 on [24]; chapter 5 on [148]; chapter 6 on [23]; and chapter 7 on [25].

Chapter 2

Inverse transport problem in frequency domain optical tomography

Optical tomography is increasingly being used as a medical imaging tool to assess the scattering and absorbing properties of human tissues probed by near-infra-red photons. Mathematically, optical tomography reduces to parameter identification problems (inverse problems) for the ERT, also referred to as the linear Boltzmann equation or the transport equation. The aim of this chapter is to design a reconstruction algorithm that solves this inverse transport problem and can be used in practical optical tomography applications. The presentation of this chapter is based on reference [146, 147].

2.1 Problem formulation

Transport-based reconstruction codes for use in biomedical optical tomography have recently been developed in several situations. First, a transport back-transport method, a nonlinear inversion method, applied to the two-dimensional time-dependent equation of radiative transfer was reported in [62]. New algorithms were developed and experimentally tested for two- and three- dimensional cases using a time-independent ERT in [104, 105, 106, 107]. While these works, which address real-life three-dimensional problems, are an important step towards practical applications,

they still suffer from considerable cross-talk between absorption and scattering reconstructions. What we mean by cross-talk is that purely scattering (or purely absorbing) inclusions are often reconstructed with unphysical absorption (or scattering) properties. This behavior is well-understood from the theoretical viewpoint: Different optical distributions inside the medium can lead to the same measurements collected at the surface of the medium [14, 96]. To avoid such cross-talks, which may lead to wrong diagnostics, we need different data. An experimental technique increasingly employed in recent years to obtain additional information is to use frequency domain measurements. In this case the source intensity is modulated (typically between 100-1000 MHz), leading to the propagation of so-called photon density waves. Since frequency domain measurements provide information about both the phase and the intensity of the waves (and not only the intensity as in steady-state measurements), it is expected that frequency-domain techniques will allow for better separation of absorption and scattering effects [14, 124]. Numerical reconstructions based on frequency-domain ERT, however, have not yet been developed in the literature. This is the major motivation for the present work.

We now formulate the optical tomography problem. Let $\Omega \subset \mathbb{R}^n$ be our domain of interest, with sufficiently regular boundary $\partial\Omega$. Then the frequency-domain equation of radiative transfer that describes the photon density in the phase space, i.e., as a function of position $\mathbf{x} \in \Omega$ and direction $\boldsymbol{\theta} \in S^{n-1}$ (unit sphere of \mathbb{R}^n) is given by [11]

$$\begin{aligned} \mathcal{T}u \equiv \left(\frac{i\omega}{v} + \boldsymbol{\theta} \cdot \nabla + \sigma_a(\mathbf{x}) \right) u(\mathbf{x}, \boldsymbol{\theta}) + Q(u)(\mathbf{x}, \boldsymbol{\theta}) &= 0 && \text{in } \Omega \times S^{n-1} \\ u(\mathbf{x}, \boldsymbol{\theta}) &= f(\mathbf{x}, \boldsymbol{\theta}) && \text{on } \Gamma_-, \end{aligned} \quad (2.1)$$

where $i = \sqrt{-1}$, $n = 2, 3$ is the space dimension, $v \in \mathbb{R}^+$ is the speed of light in the medium, and ω is the modulation frequency of the boundary source $f(\mathbf{x}, \boldsymbol{\theta})$. The non-negative function $\sigma_a(\mathbf{x}) \in L^\infty(\Omega)$ is the absorption coefficient. The unknown quantity, $u(\mathbf{x}, \boldsymbol{\theta})$, is the radiant power per unit solid angle per unit area perpendicular to the direction of propagation at \mathbf{x} in the direction $\boldsymbol{\theta}$. Note that $u(\mathbf{x}, \boldsymbol{\theta})$ depends also on ω although, for simplicity, we do not write this dependency explicitly. The boundary

sets Γ_{\pm} are defined as

$$\Gamma_{\pm} = \{(\mathbf{x}, \boldsymbol{\theta}) \in \partial\Omega \times S^{n-1} \text{ s.t. } \pm \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) > 0\},$$

with $\boldsymbol{\nu}(\mathbf{x})$ the outward unit normal to the domain at $\mathbf{x} \in \partial\Omega$. The scattering operator Q is defined as

$$Q(u)(\mathbf{x}, \boldsymbol{\theta}) = \sigma_s(\mathbf{x}) \left(u(\mathbf{x}, \boldsymbol{\theta}) - \int_{S^{n-1}} k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') u(\mathbf{x}, \boldsymbol{\theta}') d\mu(\boldsymbol{\theta}') \right). \quad (2.2)$$

Here, $\sigma_s(\mathbf{x}) \in L^\infty(\Omega)$ is the scattering coefficient and $d\mu$ is the surface measure on S^{n-1} normalized so that $\int_{S^{n-1}} d\mu(\boldsymbol{\theta}) = 1$. The ‘‘collision’’ kernel $k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}')$, which describes the probability that photons traveling in direction $\boldsymbol{\theta}'$ scatter into direction $\boldsymbol{\theta}$, is a positive function independent of \mathbf{x} and satisfies the normalization condition:

$$\int_{S^{n-1}} k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') d\mu(\boldsymbol{\theta}') = 1. \quad (2.3)$$

The scattering kernel for light propagation in tissues is highly peaked forward and is chosen as the Henyey-Greenstein phase function [87, 183]

$$k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}') = C \frac{1 - g^2}{(1 + g^2 - 2g \cos \phi)^{3/2}}, \quad (2.4)$$

where ϕ is the angle between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}'$, i.e., $\boldsymbol{\theta} \cdot \boldsymbol{\theta}' = \cos \phi$ and where $g \in [0, 1]$ is the anisotropy factor, which measures how peaked forward the phase function is. The larger g is, the more forward the scattering. The anisotropy factor is often used to define the so-called effective scattering coefficient through $\sigma'_s = (1 - g)\sigma_s$. C is a normalization constant such that (2.3) hold. We mention that scattering kernels other than (2.4) have also been used in some situations [101] and that simplified (Fokker-Planck) models can also be used to analyze highly peaked scattering in biological tissues [102].

The optical tomography problem thus consists of reconstructing $\sigma_a(\mathbf{x})$ and $\sigma_s(\mathbf{x})$ in (2.1) from boundary current measurements; see (2.6) below. Our objective in this

work is to present a numerical scheme that performs the reconstruction.

2.1.1 Forward problem

The absorption and scattering coefficients σ_a and σ_s cannot take negative values and have to be bounded. We thus introduce the following parameter space \mathcal{Q} :

$$\mathcal{Q} := \{(\sigma_a, \sigma_s) : \sigma_a \geq 0, \sigma_s \geq 0, \text{ and } (\sigma_a, \sigma_s) \in L^\infty(\Omega) \times L^\infty(\Omega)\}.$$

We also introduce the functional spaces [6, 58]:

$$\begin{aligned} L^2_{\boldsymbol{\theta}, \nu}(\Gamma_\pm) &:= \left\{ u(\mathbf{x}, \boldsymbol{\theta}) : \int_{\Gamma_\pm} |u(\mathbf{x}, \boldsymbol{\theta})|^2 |\boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x})| d\sigma(\mathbf{x}) d\mu(\boldsymbol{\theta}) < +\infty \right\} \\ \mathcal{W}^2(\Omega \times S^{n-1}) &:= \{u(\mathbf{x}, \boldsymbol{\theta}) : u \in L^2(\Omega \times S^{n-1}) \text{ and } \boldsymbol{\theta} \cdot \nabla u \in L^2(\Omega \times S^{n-1})\}. \end{aligned}$$

Adapting well-known results [6, 58] with complex-valued absorption coefficient $\sigma_a + \frac{i\omega}{v}$ in $L^\infty(\Omega)$, we have the following statement about the forward problem

Proposition 2.1.1. *Assume that $(\sigma_a, \sigma_s) \in \mathcal{Q}$, the modulation frequency is finite $\omega < +\infty$, and $f \in L^2_{\boldsymbol{\theta}, \nu}(\Gamma_-)$. Then the forward problem (2.1) is well-posed and admits a unique solution $u(\mathbf{x}, \boldsymbol{\theta}) \in \mathcal{W}^2(\Omega \times S^{n-1})$.*

We can then define the following *albedo* operator (as well as its adjoint) [52, 129]:

$$\Lambda : \begin{aligned} f &\longmapsto u|_{\Gamma_+} \\ L^2_{\boldsymbol{\theta}, \nu}(\Gamma_-) &\longmapsto L^2_{\boldsymbol{\theta}, \nu}(\Gamma_+). \end{aligned} \tag{2.5}$$

The albedo operator Λ maps the incoming flux on the boundary into the outgoing flux and is a functional of the optical parameters σ_a and σ_s .

A major difficulty in optical tomography comes from the fact that in practice, only outgoing currents, which are angular averages of the outgoing flux and are similar to diffusion-type measurements, are available. This prevents us from using classical uniqueness and stability results in inverse transport theory [52]. In fact, the inverse problem we solve in this paper is very similar to the diffusion-based inverse problem [11], on which many more theoretical results exist. To date, we do not know

of any theoretical result on the reconstruction of optical properties from outgoing currents for arbitrary geometries. This makes the development of numerical tools all the more important.

To be consistent with existing measurement technologies, we define the following “measurement operator”:

$$\begin{aligned} \mathcal{G}u|_{\Gamma_+} &:= \int_{S_+^{n-1}} \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) u|_{\Gamma_+} d\mu(\boldsymbol{\theta}) \equiv z(\mathbf{x}) \\ \mathcal{G} &: L_{\boldsymbol{\theta}, \boldsymbol{\nu}}^2(\Gamma_+) \mapsto L^2(\partial\Omega) \equiv \mathcal{Z} \end{aligned} \quad (2.6)$$

with $S_+^{n-1} := \{\boldsymbol{\theta} : \boldsymbol{\theta} \in S^{n-1} \text{ s.t. } \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) > 0\}$. We will call \mathcal{Z} the “measurement space”. Now the composite operator $\mathcal{G}\Lambda : f \mapsto z$ maps the incoming flux into the tomographic measurements. The adjoint operator \mathcal{G}^* of \mathcal{G} is defined via the identity

$$\langle \overline{\mathcal{G}^* g_1}, g_2 \rangle_{L_{\boldsymbol{\theta}, \boldsymbol{\nu}}^2(\Gamma_+)} = \langle \overline{g_1}, \mathcal{G} g_2 \rangle_{\mathcal{Z}}, \quad (2.7)$$

for all $g_1 \in \mathcal{Z}$ and $g_2 \in L_{\boldsymbol{\theta}, \boldsymbol{\nu}}^2(\Gamma_+)$, where the symbol $\overline{\mathcal{Y}_1}$ denotes the complex conjugate of \mathcal{Y}_1 , and $\langle \cdot, \cdot \rangle_X$ is the usual inner product in a Hilbert space X . One observe that \mathcal{G}^* is nothing but the operation of multiplication by $\boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x})$.

2.1.2 Least square formulation

The inverse problem of optical tomography can be formulated as follows: determine $(\sigma_a, \sigma_s) \in \mathcal{Q}$ such that

$$\mathcal{G}\Lambda f = z \quad (2.8)$$

holds for all possible source-measurement pairs (f, z) . Here $z \in \mathcal{Z} \equiv L^2(\partial\Omega)$ is the measured data corresponding to source f . This problem is in general severely ill-posed (assuming that uniqueness of reconstruction holds as in diffusion theory [11]) in the sense that when no regularization is applied, noise contained in the data z is more amplified during the inversion procedure than what would results from an arbitrary number of differentiations [67]. Another practical difficulty in solving (2.8) lies in the fact that the amount of available data may be quite limited [119]. For

example, one may only be able to use a limited number (say, N_q) of light sources. After discretizing (2.8) on a reasonable mesh, we will end up with a very under-determined nonlinear system. A classical way to resolve the lack of measurements is to turn to the following least square formulation: find (σ_a, σ_s) solving:

$$\mathcal{F}(\sigma_a, \sigma_s) =: \frac{1}{2} \sum_{q=1}^{N_q} \|\mathcal{G}\Lambda f_q - z_q\|_{\mathcal{Z}}^2 \rightarrow \min. \quad (2.9)$$

Here, $1 \leq q \leq N_q$ denotes the light source number. For reasons we have mentioned earlier, the least square problem (2.9) is usually not stable [11]. To stabilize the problem, we impose additional smoothness restrictions on the coefficients we wish to reconstruct. In other words, we look for optical properties in a space that is much smaller than \mathcal{Q} . We call this space the space of *admissible parameters*:

$$\mathcal{Q}_{ad} := \{(\sigma_a, \sigma_s) : (\sigma_a, \sigma_s) \in [\sigma_a^l, \sigma_a^u] \times [\sigma_s^l, \sigma_s^u], \text{ and } (\sigma_a, \sigma_s) \in \mathcal{H}^1(\Omega) \times \mathcal{H}^1(\Omega)\},$$

where σ_a^l (resp. σ_a^u) and σ_s^l (resp. σ_s^u) are lower (resp. upper) bounds of σ_a and σ_s , respectively, with $\sigma_a^l > 0$ and $\sigma_s^l > 0$. $\mathcal{H}^1(\Omega)$ is the usual Hilbert space of $L^2(\Omega)$ functions with first-order partial derivatives in $L^2(\Omega)$:

$$\|\mathcal{Y}\|_{\mathcal{H}^1(\Omega)}^2 := \|\mathcal{Y}\|_{L^2(\Omega)}^2 + \|\nabla \mathcal{Y}\|_{L^2(\Omega)}^2, \quad \text{for } \mathcal{Y} \in \mathcal{H}^1(\Omega). \quad (2.10)$$

It is known that \mathcal{Q}_{ad} is a closed and convex subset of $\mathcal{H}^1(\Omega) \times \mathcal{H}^1(\Omega)$. We can thus introduce the following regularized least square functional:

$$\mathcal{F}_\beta(\sigma_a, \sigma_s) := \mathcal{F}(\sigma_a, \sigma_s) + \frac{\beta}{2} \mathcal{J}(\sigma_a, \sigma_s), \quad (2.11)$$

where the last term is a regularization term and β is the regularization parameter [67]. The method for choosing β will be described in section 2.3.3. We use the Tikhonov regularization functional in our problem:

$$\mathcal{J}(\sigma_a, \sigma_s) = \|\sigma_a - \sigma_a^0\|_{\mathcal{H}^1(\Omega)}^2 + \epsilon \|\sigma_s - \sigma_s^0\|_{\mathcal{H}^1(\Omega)}^2, \quad (2.12)$$

where σ_a^0 and σ_s^0 are initial guesses for the σ_a and σ_s profiles, and ϵ is a small constant. The choice of ϵ is addressed in section 2.3.3. We thus formulate the optical tomography problem as the following regularized least square problem:

$$\begin{aligned}
 \text{(RLS)} \quad & \min_{(\sigma_a, \sigma_s)} \mathcal{F}_\beta \\
 & \sigma_a^l \leq \sigma_a \leq \sigma_a^u \\
 & \sigma_s^l \leq \sigma_s \leq \sigma_s^u
 \end{aligned}$$

We first observe that problem **(RLS)** has at least one solution in the sense that the functional $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ admits at least one minimizer. This existence result is classical and follows from the weak lower semicontinuity and coercivity of $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ [115, 180]. However, we cannot show that $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ is strictly convex and cannot conclude that the minimizer is unique [180].

Our implementation of the inverse problem of optical tomography is a gradient-based minimization approach. We thus need to compute the Fréchet derivative of the least square functional $\mathcal{F}_\beta(\sigma_a, \sigma_s)$. Direct estimates of the Fréchet derivatives being quite costly because the optical parameters are (at least at the continuous level) infinite dimensional objects, we adopt the adjoint state (or co-state) approach [180] to estimate the derivatives. We have the following result:

Theorem 2.1.2 (Fréchet derivatives). *The functional $\mathcal{F}_\beta(\sigma_a, \sigma_s)$ is Fréchet differentiable with respect to σ_a and σ_s . The derivative at (σ_a, σ_s) in the direction (h_a, h_s) is given by*

$$\begin{pmatrix} \mathcal{F}'_\beta h_a \\ \mathcal{F}'_\beta h_s \end{pmatrix} = \begin{pmatrix} \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a \right) u_q \right\rangle_{L^2(\Omega \times S^{n-1})} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\Omega)} \\ \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_s} h_s \right) u_q \right\rangle_{L^2(\Omega \times S^{n-1})} + \beta \epsilon \langle \sigma_s - \sigma_s^0, h_s \rangle_{\mathcal{H}^1(\Omega)} \end{pmatrix}, \quad (2.13)$$

where \mathcal{T} is the transport operator defined in (2.1); u_q and φ_q are the solutions of the forward problem (2.1) with source f_q and its adjoint problem (2.16) (defined below), respectively. *Re* means taking the real part.

Proof. Let us denote by r_q the residual $\mathcal{G}\Lambda f_q - z_q = \mathcal{G}u_q|_{\Gamma_+} - z_q$. According to [62, 63],

r_q is Fréchet differentiable with respect to both σ_a and σ_s . The L^2 -norm is Fréchet differentiable as shown in [115]. By the chain rule, $\|r_q\|_{\mathcal{Z}}^2$ is Fréchet differentiable. Since the summation is finite, we deduce that \mathcal{F} is differentiable. Together with the fact that \mathcal{J} is differentiable, we conclude that \mathcal{F}_β is Fréchet differentiable with respect to σ_a and σ_s .

We now compute these Fréchet derivatives. Let us compute the derivative with respect to σ_a :

$$\begin{aligned} \mathcal{F}'_\beta(\sigma_a, \sigma_s)h_a &= \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \bar{r}_q, \mathcal{G} \left(\frac{\partial u_q|_{\Gamma_+}}{\partial \sigma_a} h_a \right) \right\rangle_{\mathcal{Z}} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\Omega)} \\ &= \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \overline{\mathcal{G}^* r_q}, \frac{\partial u_q|_{\Gamma_+}}{\partial \sigma_a} h_a \right\rangle_{L^2_{\theta, \nu}(\Gamma_+)} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\Omega)} \end{aligned} \quad (2.14)$$

where we have used the properties of the adjoint operator (2.7). On the other hand, differentiating the transport equation (2.1) for source f_q gives:

$$\begin{aligned} \mathcal{T} \phi_q + \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a \right) u_q &= 0 \quad \text{in } \Omega \times S^{n-1} \\ \phi_q &= 0 \quad \text{on } \Gamma_-, \end{aligned} \quad (2.15)$$

where $\phi_q \equiv \frac{\partial u_q}{\partial \sigma_a} h_a$, and \mathcal{T} is the transport operator defined in (2.1). We need also to introduce an adjoint variable φ_q of u_q which is the solution of the following adjoint transport equation:

$$\begin{aligned} \mathcal{T}^* \varphi_q \equiv \left(\frac{i\omega}{v} - \boldsymbol{\theta} \cdot \nabla + \sigma_a(\mathbf{x}) \right) \varphi_q(\mathbf{x}, \boldsymbol{\theta}) + Q(\varphi_q)(\mathbf{x}, \boldsymbol{\theta}) &= 0 \quad \text{in } \Omega \times S^{n-1} \\ \varphi_q(\mathbf{x}, \boldsymbol{\theta}) &= -\overline{\mathcal{G}^* r_q} \quad \text{on } \Gamma_+. \end{aligned} \quad (2.16)$$

Here we have used that $Q^* = Q$, which follows from the definition (2.2). Multiplying (2.15) by φ_q and (2.16) by ϕ_q , then integrating over $\Omega \times S^{n-1}$, we obtain

$$\left\langle \overline{\mathcal{G}^* r_q}, \phi_q \right\rangle_{L^2_{\theta, \nu}(\Gamma_+)} = \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a \right) u_q \right\rangle_{L^2(\Omega \times S^{n-1})}, \quad (2.17)$$

which leads to

$$\mathcal{F}'_{\beta}(\sigma_a, \sigma_s)h_a = \operatorname{Re} \sum_{q=1}^{N_q} \left\langle \varphi_q, \left(\frac{\partial \mathcal{T}}{\partial \sigma_a} h_a \right) u_q \right\rangle_{L^2(\Omega \times S^{n-1})} + \beta \langle \sigma_a - \sigma_a^0, h_a \rangle_{\mathcal{H}^1(\Omega)}. \quad (2.18)$$

The derivative with respect to σ_s can be computed similarly. \square

This result shows that in order to compute the Fréchet derivative of the objective functional $\mathcal{F}_{\beta}(\sigma_a, \sigma_s)$, we need to solve one forward transport problem (2.1) and one adjoint transport problem (2.16).

2.2 Discretization methods

There is a vast literature on the discretization of radiative transfer equations; see for instance [4, 77, 113]. In this paper, we have chosen to use the discrete ordinates method to discretize the directional variables and the finite volume method [68] to discretize the spatial variables.

2.2.1 The discrete ordinates formulation

In the discrete ordinates method [4, 113], we approximate the total scalar flux, defined as the integral of $u(\mathbf{x}, \boldsymbol{\theta})$ over S^{n-1} , by the following quadrature rule

$$\int_{S^{n-1}} u(\mathbf{x}, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) \approx \sum_{j=1}^J \eta_j u(\mathbf{x}, \boldsymbol{\theta}_j), \quad (2.19)$$

where $\boldsymbol{\theta}_j$ is the j th direction and η_j the associated weight, for $1 \leq j \leq J$. Details on how to choose the set of directions $\{\boldsymbol{\theta}_j\}_{j=1}^J$ and the corresponding weights $\{\eta_j\}_{j=1}^J$ can be found in [113]. To ensure particle conservation, we impose that

$$\sum_{j=1}^J \eta_j = 1. \quad (2.20)$$

The equation of radiative transfer is now decomposed as a discrete set of J coupled

differential equations that describe the photon flux field along J directions:

$$\nabla \cdot (\boldsymbol{\theta}_j u) + \left(\sigma_t + \frac{i\omega}{v}\right)u(\mathbf{x}, \boldsymbol{\theta}_j) = \sigma_s(\mathbf{x}) \sum_{j'=1}^J \eta_{j'} k_{jj'} u(\mathbf{x}, \boldsymbol{\theta}_{j'}), \quad (2.21)$$

for $j = 1, 2, \dots, J$, where $k_{jj'} = k(\boldsymbol{\theta}_j \cdot \boldsymbol{\theta}_{j'})$, and where $\sigma_t = \sigma_a + \sigma_s$. We impose

$$\sum_{j=1}^J \eta_j k_{jj'} = 1, \quad 1 \leq j' \leq J, \quad (2.22)$$

so that the number of photons in the system is preserved by the scattering process.

2.2.2 Spatial discretization

We use a finite volume method to perform the spatial discretization. Finite volume methods [68] ensure the conservation of mass (or momentum, energy) in a discrete sense, which is important in transport calculations. They also have the advantage of easily handling complicated geometries by arbitrary triangulations, which we need in tomographic applications.

We implement a cell-centered version of the finite volume methods. Consider a mesh of \mathbb{R}^n , \mathcal{M} , consisting of polyhedral bounded convex subsets of \mathbb{R}^n which covers our computational domain Ω . Let $\mathcal{C} \in \mathcal{M}$ be a *control cell*, that is an element of the mesh \mathcal{M} , $\partial\mathcal{C}$ its boundary, and $V_{\mathcal{C}}$ its Lebesgue measure. We assume that the unknown quantity, for example $u(\mathbf{x}, \boldsymbol{\theta}_j)$, takes its averaged value in \mathcal{C} (thus is constant). We denote this value by $u_j^{\mathcal{C}}$:

$$u_j^{\mathcal{C}} \equiv \frac{1}{V_{\mathcal{C}}} \int_{V_{\mathcal{C}}} u(\mathbf{x}, \boldsymbol{\theta}_j) d\mathbf{x}. \quad (2.23)$$

Integrating the above discrete ordinates equations (2.21) over cell \mathcal{C} and using the divergence theorem on the first term, we obtain the following equations

$$\int_{\partial\mathcal{C}} \boldsymbol{\theta}_j \cdot \mathbf{n}_{\mathcal{C}}(\mathbf{x}) u_j d\gamma(\mathbf{x}) + \left(\sigma_t^{\mathcal{C}} + \frac{i\omega}{v}\right) V_{\mathcal{C}} u_j^{\mathcal{C}} = V_{\mathcal{C}} \sigma_s^{\mathcal{C}} \sum_{j'=1}^J \eta_{j'} k_{jj'} u_{j'}^{\mathcal{C}}, \quad (2.24)$$

for $1 \leq j \leq J$, where, $\mathbf{n}_C(\mathbf{x})$ denotes the outward normal to ∂C at point $\mathbf{x} \in \partial C$, $d\gamma(\mathbf{x})$ denotes the surface Lebesgue measure on ∂C and σ_s^C (σ_t^C) is the value of σ_s (σ_t) on cell C .

Now we have to approximate the flux through the boundary of C , i.e., the first integral term in equation (2.24). Let $\{C_i\}_{i=1}^I$ be the set of neighboring cells of C . We denote by $S_{C,i}$ the common edge of cell C and C_i , i.e., $S_{C,i} = \partial C \cap \partial C_i$. We then have

$$\int_{\partial C} \boldsymbol{\theta}_j \cdot \mathbf{n}_C(\mathbf{x}) u_j d\gamma(\mathbf{x}) = \sum_i \int_{S_{C,i}} \boldsymbol{\theta}_j \cdot \mathbf{n}_C(\mathbf{x}) u_j d\gamma(\mathbf{x}). \quad (2.25)$$

The flux $\int_{S_{C,i}} \boldsymbol{\theta}_j \cdot \mathbf{n}_C(\mathbf{x}) u_j d\gamma(\mathbf{x})$ can be approximated by various numerical schemes. In this work, we take a first-order upwind scheme:

$$F_{j,i}^C := \int_{S_{C,i}} \boldsymbol{\theta}_j \cdot \mathbf{n}_C(\mathbf{x}) u_j d\gamma(\mathbf{x}) = \begin{cases} \boldsymbol{\theta}_j \cdot \mathbf{n}_C |S_{C,i}| u_j^C & \text{if } \boldsymbol{\theta}_j \cdot \mathbf{n}_C \geq 0 \\ \boldsymbol{\theta}_j \cdot \mathbf{n}_C |S_{C,i}| u_j^{C_i} & \text{if } \boldsymbol{\theta}_j \cdot \mathbf{n}_C < 0, \end{cases} \quad (2.26)$$

where $|S_{C,i}|$ is the Lebesgue measure of $S_{C,i}$. We then obtain a full discretization of the discrete ordinates equations

$$\sum_i F_{j,i}^C + (\sigma_t^C + \frac{i\omega}{v}) V_C u_j^C = V_C \sigma_s^C \sum_{j'=1}^J \eta_{j'} k_{jj'} u_{j'}^C, \quad (2.27)$$

for $j = 1, 2, \dots, J$. Let N denote the total number of control cells. After collecting the discretized transport equation (2.27) on all control cells, we arrive at the following system of complex-valued algebraic equations

$$\mathbf{A}\mathbf{U} = \mathbf{S}\mathbf{U} + \mathbf{G} \quad (2.28)$$

where $\mathbf{A} \in \mathbb{C}^{NJ \times NJ}$ and $\mathbf{S} \in \mathbb{C}^{NJ \times NJ}$ are the discretized streaming-collision and scattering operators, respectively. The boundary source $f(\mathbf{x}, \boldsymbol{\theta})$, which comes into the discretized system via the flux approximation (2.26) is denoted by \mathbf{G} . The vector $\mathbf{U} \in \mathbb{C}^{NJ \times 1}$, which contains the values of $u(\mathbf{x}, \boldsymbol{\theta})$ on the cell C in the direction $\boldsymbol{\theta}_j$ is

organized as

$$\mathbf{U} = \begin{pmatrix} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_J \end{pmatrix}, \text{ with } \mathbf{U}_j = \begin{pmatrix} u_j^1 \\ \vdots \\ u_j^N \end{pmatrix} \in \mathbb{C}^N. \quad (2.29)$$

The matrices \mathbf{A} and \mathbf{S} have sparse structures. In fact, they are sparse block matrices. \mathbf{A} is a block diagonal matrix that can be written as:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & & \\ & \ddots & \\ & & \mathbf{A}_J \end{pmatrix} + \begin{pmatrix} \mathbf{C}_0 & & \\ & \ddots & \\ & & \mathbf{C}_0 \end{pmatrix}, \quad (2.30)$$

where $\mathbf{A}_j \in \mathbb{C}^{N \times N}$ is the discretization of the advection operator \mathcal{A} defined by $\mathcal{A}u := \boldsymbol{\theta}_j \cdot \nabla u$. From (2.26) we can deduce that \mathbf{A}_j has no more than $N \times N_E$ non-zero elements, where N_E is the total number of edges (surfaces in 3-dimension) each control cell has.

Matrix $\mathbf{C}_0 \in \mathbb{C}^{N \times N}$ is diagonal:

$$\mathbf{C}_0 = \begin{pmatrix} V_1(\sigma_t^1 + \frac{i\omega}{v}) & & \\ & \ddots & \\ & & V_N(\sigma_t^N + \frac{i\omega}{v}) \end{pmatrix},$$

where we recall $\sigma_t^i \equiv \sigma_a^i + \sigma_s^i$ ($i = 1, \dots, N$).

The matrix \mathbf{S} can be expressed as the direct product of two smaller matrices:

$$\mathbf{S} = \mathbf{K} \otimes \mathbf{D}_0, \quad (2.31)$$

with $\mathbf{D}_0 \in \mathbb{C}^{N \times N}$ a diagonal matrix given by

$$\mathbf{D}_0 = \begin{pmatrix} V_1\sigma_s^1 & & \\ & \ddots & \\ & & V_N\sigma_s^N \end{pmatrix},$$

and $\mathbf{K} \in \mathbb{C}^{J \times J}$ a dense matrix with component $(\mathbf{K})_{jj'} = \eta_{j'} k_{jj'}$. In practical applications, the number of directions is much smaller than the number of spatial mesh elements ($J \ll N$). So although \mathbf{K} is dense, the scattering matrix \mathbf{S} is sparse. However, in general the matrix \mathbf{K} is not symmetric unless we choose η_j to be constant. The matrix $\mathbf{A} - \mathbf{S}$ is thus neither symmetric nor positive definite, which is the reason for us to choose a GMRES solver in section 2.3.2.

Let us remark here that our finite volume discretization reduces to an upwind finite difference scheme on usual finite difference grids. We refer to our earlier work [146] for some numerical tests on the finite volume discretization of the transport equation.

2.2.3 Discrete adjoint problem

We present in this section the numerical method we have employed to compute the gradient of discrete objective function with respect to the optical properties on each cell.

To simplify the notation, we denote from now on by $\sigma_a \in \mathbb{R}^{N \times 1}$ the absorption coefficient vector $(\sigma_a^1, \dots, \sigma_a^C, \dots, \sigma_a^N)^T$ and $\sigma_s \in \mathbb{R}^{N \times 1}$ the scattering coefficient vector $(\sigma_s^1, \dots, \sigma_s^C, \dots, \sigma_s^N)^T$.

We want to minimize the discrepancy between model predictions and measurements over a set of source and detector pairs. Let N_q denote the number of sources used an experiment, and N_d denote the number of detectors used for each source. Then the following objective function we employed takes the following form

$$\mathcal{F}_\beta(\sigma_a, \sigma_s) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} |\mathcal{P}_d \mathbf{U}^q - z_{q,d}^\delta|^2 + \frac{\beta}{2} \mathcal{J}(\sigma_a, \sigma_s) \quad (2.32)$$

where $z_{q,d}^\delta$ denote the d -th measurement of the q -th source. The superscript δ is used to denote the level of noise contained in the measurements. \mathbf{U}^q is solution of the transport equation for the q -th source. $\mathcal{P}_d \in \mathbb{R}^{1 \times N}$ is a discretized version of the measurement operator. It takes the outgoing flux at detector d and averages over

S_+^{n-1} . The discretized regularization term is given by

$$\begin{aligned} \mathcal{J}(\sigma_a, \sigma_s) = & \sum_{\mathcal{C}=1}^N \left(\sum_{\kappa=\{x,y,z\}} [\Omega_{\kappa}^{\mathcal{C}}(\sigma_a - \sigma_a^0)]^2 + (\sigma_a^{\mathcal{C}} - \sigma_a^{0,\mathcal{C}})^2 \right) \\ & + \epsilon \sum_{\mathcal{C}=1}^N \left(\sum_{\kappa=\{x,y,z\}} [\Omega_{\kappa}^{\mathcal{C}}(\sigma_s - \sigma_s^0)]^2 + (\sigma_s^{\mathcal{C}} - \sigma_s^{0,\mathcal{C}})^2 \right) \end{aligned} \quad (2.33)$$

where $\Omega_{\kappa}^{\mathcal{C}} \in \mathbb{R}^{1 \times N}$ denotes the discretized partial differential operator at cell \mathcal{C} in the κ ($= x, y, z$) direction.

We now start to compute the gradient of objective function (2.32) with respect to optical properties on each mesh element. It is straightforward to check that

$$\frac{\partial \mathcal{F}_{\beta}}{\partial \sigma_a^{\mathcal{C}}} = \left[\sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \mathbf{r}_d^q \mathcal{P}_d \frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}} \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_a^{\mathcal{C}}}, \quad (2.34)$$

with $\mathbf{r}_d^q = \mathcal{P}_d \mathbf{U}^q - z_{q,d}^{\delta}$, and $[\cdot]_{\text{Re}}$ denotes the real part of $[\cdot]$.

At the same time, we notice from (2.28) that:

$$\frac{\partial \mathbf{A}}{\partial \sigma_a^{\mathcal{C}}} \mathbf{U}^q + \mathbf{A} \frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}} = \frac{\partial \mathbf{S}}{\partial \sigma_a^{\mathcal{C}}} \mathbf{U}^q + \mathbf{S} \frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}}, \quad (2.35)$$

for source $q = 1, \dots, N_q$, which is equivalent to saying that

$$\frac{\partial \mathbf{U}^q}{\partial \sigma_a^{\mathcal{C}}} = -(\mathbf{A} - \mathbf{S})^{-1} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^{\mathcal{C}}} \mathbf{U}^q, \quad (2.36)$$

since $\mathbf{A} - \mathbf{S}$ is invertible. It is very important to note that the matrices \mathbf{A} and \mathbf{S} are independent of the source used. Thus, there are no superscripts q associated with them. We thus have

$$\frac{\partial \mathcal{F}_{\beta}}{\partial \sigma_a^{\mathcal{C}}} = - \left[\sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \mathbf{r}_d^q \mathcal{P}_d (\mathbf{A} - \mathbf{S})^{-1} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^{\mathcal{C}}} \mathbf{U}^q \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_a^{\mathcal{C}}}. \quad (2.37)$$

We now introduce a new state variable $\Psi^q \in \mathbb{C}^{N \times 1}$ (called adjoint variable of \mathbf{U}^q)

given by

$$-\sum_{d=1}^{N_d} \bar{\mathbf{r}}_d^q \mathcal{P}_d (\mathbf{A} - \mathbf{S})^{-1} = \mathbf{\Psi}^{qT}. \quad (2.38)$$

where $\mathbf{\Psi}^{qT}$ denotes the transpose of $\mathbf{\Psi}^q$. We then say that $\mathbf{\Psi}^q$ is the solution of the following *adjoint* equation of (2.28):

$$(\mathbf{A} - \mathbf{S})^T \mathbf{\Psi}^q = -\sum_{d=1}^{N_d} \bar{\mathbf{r}}_d^q \mathcal{P}_d^T. \quad (2.39)$$

One then arrives at

$$\frac{\partial \mathcal{F}_\beta}{\partial \sigma_a^c} = \left[\sum_{q=1}^{N_q} \mathbf{\Psi}^{qT} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^c} \mathbf{U}^q \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_a^c}, \quad (2.40)$$

with

$$\frac{\partial \mathcal{J}}{\partial \sigma_a^c} = 2 \left(\sum_{\kappa=\{x,y,z\}} \Omega_\kappa^c (\sigma_a - \sigma_a^0) (\Omega_\kappa^c \mathcal{I}_c) + (\sigma_a^c - \sigma_a^{0,c}) \right),$$

where the unit direction vector $\mathcal{I}_c \in \mathbb{R}^{N \times 1}$ is a vector whose c -th element is 1 and all other components are zero.

Very similar computation leads to the fact that the derivatives of the objective functional with respect to σ_s^c are given by

$$\frac{\partial \mathcal{F}_\beta}{\partial \sigma_s^c} = \left[\sum_{q=1}^{N_q} \mathbf{\Psi}^{qT} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_s^c} \mathbf{U}^q \right]_{\text{Re}} + \frac{\beta}{2} \frac{\partial \mathcal{J}}{\partial \sigma_s^c}, \quad (2.41)$$

with

$$\frac{\partial \mathcal{J}}{\partial \sigma_s^c} = 2\epsilon \left(\sum_{\kappa=\{x,y,z\}} \Omega_\kappa^c (\sigma_s - \sigma_s^0) (\Omega_\kappa^c \mathcal{I}_c) + (\sigma_s^c - \sigma_s^{0,c}) \right).$$

Formulas (2.40) and (2.41) are what we used to compute the derivatives of objective function with respect to optical properties on each element. Note that we did not form explicitly the matrix $\frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^c}$ (resp. $\frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_s^c}$) in the evaluations of $\mathbf{\Psi}^{qT} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_a^c} \mathbf{U}^q$ (resp. $\mathbf{\Psi}^{qT} \frac{\partial (\mathbf{A} - \mathbf{S})}{\partial \sigma_s^c} \mathbf{U}^q$) because this matrix has a very simple sparse structure according

to (2.30) and (2.31). Instead, a matrix-free method was adopted. In fact, since

$$\left[\frac{\partial(\mathbf{A} - \mathbf{S})}{\partial\sigma_a^c}\right]_{ij} = \begin{cases} V_c, & i = j \text{ and } \text{mod}(i, N) = \mathcal{C} \\ 0, & \text{otherwise,} \end{cases} \quad (2.42)$$

where recall that N is the total number of volume cells that cover our computational domain, we have

$$\Psi^{qT} \frac{\partial(\mathbf{A} - \mathbf{S})}{\partial\sigma_a^c} \mathbf{U}^q = \sum_{j=1}^J \Psi_{(j-1) \times N + \mathcal{C}}^q V_c \mathbf{U}_{(j-1) \times N + \mathcal{C}}^q. \quad (2.43)$$

Note that here $\Psi_{(j-1) \times N + \mathcal{C}}^q$ (resp. $\mathbf{U}_{(j-1) \times N + \mathcal{C}}^q$) denotes the $[(j-1) \times N + \mathcal{C}]$ -th element of Ψ^q (resp. \mathbf{U}^q).

The same observation on $\frac{\partial(\mathbf{A} - \mathbf{S})}{\partial\sigma_s^c}$ leads to

$$\begin{aligned} \Psi^{qT} \frac{\partial(\mathbf{A} - \mathbf{S})}{\partial\sigma_s^c} \mathbf{U}^q &= \sum_{j=1}^J \Psi_{(j-1) \times N + \mathcal{C}}^q V_c \mathbf{U}_{(j-1) \times N + \mathcal{C}}^q \\ &\quad - \sum_{j'=1}^J \sum_{j=1}^J (\mathbf{K})_{j'j} \Psi_{(j'-1) \times N + \mathcal{C}}^q V_c \mathbf{U}_{(j-1) \times N + \mathcal{C}}^q. \end{aligned} \quad (2.44)$$

where the (j', j) th component of matrix \mathbf{K} , $(\mathbf{K})_{j'j} = \eta_{j'} k_{jj'}$, as given in (2.31). We can thus evaluate (2.40) and (2.41) without forming any intermediate matrices.

2.3 Numerical implementation

We have implemented the quasi-Newton optimization algorithm to solve the regularized least-square problem (**RLS**) introduced in section 2.1.2. We have found in practice that this method converged much faster (in terms of function evaluations) than the nonlinear conjugate gradient method with either the Fletcher-Reeves or the Polak-Ribière updating formula [132]. This is expected from theory [132] and is consistent with practical applications tested in [107]. We have also implemented a Gauss-Newton method [132] to solve the least square problem (without the bounds

constraints), and found that the method converges extremely slow in our case. This is probably due to the fact that our problem is highly nonlinear and Gauss-Newton method usually does not work well in this kind of situations [81, 132]. Detail comparison between various method of solving the least-square reconstruction problem is an ongoing project.

In this work, we employ the BFGS update rule [132] of inverse Hessian matrix for our quasi-Newton method. The usual BFGS method, however, requires the explicit construction of the Hessian matrix, which is unrealistic for large problems. The memory size required to store the Hessian matrix is roughly proportional to the square of the memory used for the unknown parameters. We have thus resorted to a limited-memory version of BFGS method which avoids the explicit construction of the inverse Hessian matrix.

2.3.1 Numerical optimization

The BFGS algorithm can be viewed as a special case of quasi-Newton method [132]. With $\boldsymbol{\sigma}$ denoting the vector of discretized optical properties, the quasi-Newton methods can be characterized by the following iterative process:

$$\boldsymbol{\sigma}_{k+1} = \boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k, \quad k \in \mathbb{N}^+ \quad (2.45)$$

where \mathbf{p}_k is a descent direction vector and α_k is the step length. The BFGS algorithm chooses \mathbf{p}_k to be the solution of an approximated solution of Newton-type optimality equation, i.e.,

$$\mathbf{p}_k = H_k \mathbf{g}_k, \quad (2.46)$$

where \mathbf{g}_k is the gradient of the least-square functional, $\mathbf{g}_k = -\nabla_{\boldsymbol{\sigma}} \mathcal{F}_\beta(\boldsymbol{\sigma}_k)$. H_k is the inverse Hessian matrix of \mathcal{F}_β at step k . Instead of computing real inverse Hessian matrices, which is very time-consuming, the BFGS algorithm chooses to approximate H_k by the following updating rule

$$H_{k+1} = W_k^T H_k W_k + \rho_k \mathbf{s}_k \mathbf{s}_k^T \quad (2.47)$$

with $W_k = \mathbf{I} - \rho_k \mathbf{y}_k \mathbf{s}_k^T$, $\mathbf{s}_k = \boldsymbol{\sigma}_{k+1} - \boldsymbol{\sigma}_k$, $\mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$, and $\rho_k = \frac{1}{\mathbf{y}_k^T \mathbf{s}_k}$. \mathbf{I} is the identity matrix. As we mentioned above, forming (2.47) takes tremendous computer memory for large problems. To overcome this shortcoming, the limited-memory version of BFGS only stores the vector \mathbf{y}_k and \mathbf{s}_k obtained in the last m ($3 \leq m \leq 7$ usually) iterations [100] and discards the rest. Thus after first m iterations, (2.47) can be expressed as:

$$\begin{aligned}
H_{k+1} &= (W_k^T \cdots W_{k-m}^T) H_{k+1}^0 (W_{k-m} \cdots W_k) \\
&+ \rho_{k-m} (W_k^T \cdots W_{k-m+1}^T) \mathbf{s}_{k-m} \mathbf{s}_{k-m}^T \times (W_{k-m+1} \cdots W_k) \\
&+ \rho_{k-m+1} (W_k^T \cdots W_{k-m+2}^T) \mathbf{s}_{k-m+1} \mathbf{s}_{k-m+1}^T \times (W_{k-m+2} \cdots W_k) \\
&\vdots \\
&+ \rho_k \mathbf{s}_k \mathbf{s}_k^T
\end{aligned} \tag{2.48}$$

with the sparse initial guess H_{k+1}^0 given by $H_{k+1}^0 = \frac{\mathbf{y}_{k+1}^T \mathbf{s}_{k+1}}{\mathbf{y}_{k+1}^T \mathbf{y}_{k+1}} \mathbf{I}$.

We refer interested readers to [100, 132] for more details on the limited-memory BFGS algorithms, and to reference [107] for applications of those algorithms to optical tomographic problems. Convergence of BFGS algorithms has been proved under certain conditions and has been tested on many applications [41, 132].

To impose bounds on optical parameters, we have to modify the relation (2.46) slightly. We adopt a gradient projection method [41, 100, 132] to do this. At the beginning of each iteration, we use the gradient projection method to find a set of active bounds. We then solve a sub-minimization problem

$$\min_{\boldsymbol{\sigma}} Q_k(\boldsymbol{\sigma}) \equiv \mathcal{F}_\beta(\boldsymbol{\sigma}_k) + \mathbf{g}_k^T (\boldsymbol{\sigma} - \boldsymbol{\sigma}_k) + \frac{1}{2} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_k)^T H_k^{-1} (\boldsymbol{\sigma} - \boldsymbol{\sigma}_k), \tag{2.49}$$

on the set of free variables to find an approximation solution $\bar{\boldsymbol{\sigma}}_{k+1}$, treating the active bounds as equality constraints by Lagrange multipliers. After we find an approximation solution $\bar{\boldsymbol{\sigma}}_{k+1}$, a line search along $\mathbf{p}_k = \bar{\boldsymbol{\sigma}}_{k+1} - \boldsymbol{\sigma}_k$ is done to find the step length α_k in (2.45). We use a line search method that enforces the Wolfe conditions [132].

That is, we look for an α_k that solves:

$$\min_{\alpha_k > 0} \mathcal{F}_\beta(\boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k), \quad (2.50)$$

and satisfies:

$$\mathcal{F}_\beta(\boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k) \leq \mathcal{F}_\beta(\boldsymbol{\sigma}_k) + c_1 \alpha_k \nabla \mathcal{F}_\beta^T(\boldsymbol{\sigma}_k) \mathbf{p}_k \quad (2.51)$$

$$\nabla \mathcal{F}_\beta^T(\boldsymbol{\sigma}_k + \alpha_k \mathbf{p}_k) \mathbf{p}_k \geq c_2 \nabla \mathcal{F}_\beta^T(\boldsymbol{\sigma}_k) \mathbf{p}_k \quad (2.52)$$

where $c_1 = 10^{-4}$, $c_2 = 0.1$ in our case. More details on how to impose bound constraints in BFGS algorithms can be found in [100, sec. 5.5] and [41, 186].

2.3.2 Solving algebraic systems

As we have mentioned before, at each step of the minimization process, we have to solve both a discretized transport equation (2.28) and its adjoint problem (2.39) to compute the Fréchet derivatives (2.40) and (2.41) of the objective functional, forming the gradient vector \mathbf{g}_k in (2.46). In fact, almost all of the computational time in the reconstruction process is devoted to the solution of these transport equations. In this work, instead of using the popular source iteration (SI) method, which converges very slowly in diffusive regimes unless it is properly accelerated [4], we choose to solve the forward problems by a preconditioned GMRES(n) algorithm [157, 158], where n denotes the number of iterative steps after which GMRES is restarted. Our general principle is to choose n large when the problem size is small and n small when the problem size is large. The implementation of the algorithm is based on the template provided in [27]. The preconditioner we employ is the zero fill-in incomplete LU factorization (ILU(0)) [139, 157] that has been proved to be efficient in transport calculations [139]. Details about this factorization can be found in reference [157]. In all of the numerical examples in section 2.4, we pick $n = 7$, and the GMRES algorithm is stopped if the relative residual is small enough. For example, the stopping criteria $\|\mathbf{G} - (\mathbf{A} - \mathbf{S})\mathbf{U}_k\|_{l^2} / \|\mathbf{G} - (\mathbf{A} - \mathbf{S})\mathbf{U}_0\|_{l^2} \leq 10^{-10}$, is used to solve (2.28). Here \mathbf{U}_0 is

the initial guess and \mathbf{U}_k is the \mathbf{U} value at the k -th GMRES iteration.

2.3.3 Selecting regularization parameter

To choose the optimal regularization parameter β in (2.11), we adopt the L -curve method in this study. Although there exist proofs that the L -curve method fails to convergence for certain classes of inverse problems [179], we have observed satisfactory results in our applications. We plot the log of the regularization functional against the squared norm of the regularized residual, say, r_β , for a range of values of the regularization parameter. The right parameter β is the one at which the L -curve reaches the maximum of its curvature [82, 180]. One can show that the right β maximizes the following curvature function [180]

$$\kappa(\beta) = -\frac{R(\beta)S(\beta)[\beta R(\beta) + \beta^2 S(\beta)] + [R(\beta)S(\beta)]/S'(\beta)}{[R^2(\beta) + \beta^2 S^2(\beta)]^{3/2}}, \quad (2.53)$$

where $R(\beta)$ and $S(\beta)$ are defined by

$$R(\beta) := \|r_\beta\|_{L^2}^2, \quad S(\beta) := \mathcal{J}(\sigma_a, \sigma_s).$$

We recall that β is not included in $\mathcal{J}(\sigma_a, \sigma_s)$. One notices immediately that the L -curve method requires several reconstructions for any single problem, and thus is very time-consuming. A simple *continuation* method is suggested in [81] to reduce the computational cost of regularization parameter selection. In this method, one start the first reconstruction with a relatively large β . The result of this reconstruction is then taken to be the initial guess of next reconstruction with a smaller β . If the two β are not dramatically different from each other, then the two reconstructions should converge to similar results. Thus, the reconstruction with smaller β is supposed to converge fast since its initial guess is chosen to be close enough to its real solution. The process can be repeated to perform reconstructions with several values of β . We adopt this continuation method in our three-dimensional numerical example (Example 4) in the next section. We present in Fig. 2-5 (B) and Fig. 2-9 (B) the L -curve we have

used in Example 1 and Example 3, respectively, to choose the optimal regularization parameter β . Note that in the Fig. 2-5 (B), $\mathcal{J}(\sigma_a, \sigma_s)$ simplifies to $\|\sigma_a\|_{\mathcal{H}^1}^2$ since we reconstruct only σ_a and we have chosen $\sigma_a^0 = 0$ in that case.

Another important issue is to choose an appropriate weight ϵ in the regularization functional defined in (2.12). This weight is necessary because, in practice, σ_s takes values that are about two order of magnitude larger than σ_a . The weight is used to bring the two terms in the regularization functional to the same level so that the regularization term has an effect on both σ_a and σ_s . In all our numerical simulations in section 5, we choose ϵ to be the ratio $(\sigma_a^b/\sigma_s^b)^2$, where σ_a^b and σ_s^b are optical properties of background media.

We remark finally that the \mathcal{H}^1 norm we use in the regularization functional can be replaced by other norms or semi-norms. For example, we have performed reconstructions on numerical Example 1 in section 5.3 with stricter bounds on σ_a and $\|\nabla\sigma_a\|_{L^2}$ (instead of $\|\sigma_a\|_{\mathcal{H}^1}$) as the regularization functional. We have obtained very similar results (although the optimal regularization parameter changes). The main reason for us to use the \mathcal{H}^1 norm is that in many practical applications, we want to find solutions near some reference (σ_a^0, σ_s^0) , for example, some known background.

2.3.4 Cost of the numerical method

The computational cost of our method consists of two main parts. The first part is the evaluation of the objective function and its gradient in the optimization process. The second part is the updating of the BFGS matrices and vectors.

The costs of the function evaluation and of its gradient scale linearly with the number of optical sources N_q . Since each forward problem and its corresponding adjoint problem cost about the same, each gradient calculation (about $2N_q$ forward solves) is approximately twice as expensive as a function evaluation (about N_q forward solves). The cost in updating BFGS matrices and vectors can be neglected compared to function and gradient evaluations. The reason is that BFGS vectors (in \mathbb{R}^{2N}) are dramatically smaller than the vectors appearing in the forward and adjoint problems (in \mathbb{C}^{JN}).

In our computations, we store the non-zero elements of the matrix $\mathbf{A} - \mathbf{S}$ by using the compressed row storage scheme [27] whenever it is possible to do so. When it is not possible to store $\mathbf{A} - \mathbf{S}$, we store \mathbf{A} , \mathbf{C}_0 , \mathbf{K} and \mathbf{D}_0 defined in (2.30) and (2.31). This requires much less memory with the price that extra efforts has to be paid to evaluate matrix vector products in GMRES. We use the following procedure to compute $\mathbf{Y} \equiv (\mathbf{A} - \mathbf{S})\mathbf{X}$ for any vector \mathbf{X} with the same structure as \mathbf{U} in (2.29):

1. For $j = 1, \dots, J$,
 - Compute $\mathbf{X}'_j = \mathbf{D}_0\mathbf{X}_j$;
 - Compute $\mathbf{X}''_j = \mathbf{C}_0\mathbf{X}_j$;
 - Compute $\mathbf{X}''_j = \mathbf{X}''_j + \mathbf{A}_j\mathbf{X}_j$;
2. For $j = 1, \dots, J$, $\mathbf{Y}_j = \mathbf{X}''_j - \sum_{j'} \mathbf{K}_{jj'}\mathbf{X}'_{j'}$.

We prefer to store the matrix $\mathbf{A} - \mathbf{S}$ because it saves computational time when matrix-vector products are calculated. In all the numerical examples shown in the following section, we were able to store $\mathbf{A} - \mathbf{S}$. Note that the storage requirement does not increase with the number of sources (N_q) because we solve the transport equation (and its adjoint) with different sources sequentially. The storage cost of BFGS vectors can be neglected compared with the storage of the forward and adjoint matrices and vectors.

2.4 Numerical examples

We provide in this section several numerical examples that illustrate the performance of our numerical method. We will first show some forward simulations and then show some reconstructions.

2.4.1 Forward simulations

To initially test the performance of the transport solver, we chose two examples. In the first example, we consider a 2-dimensional homogeneous medium of size $5 \text{ cm} \times 5$

cm, defined as $\Omega := \{(x, y)^T | 0 < x, y < 5\}$. A point source is placed at $\mathbf{x}_s = (0, 2.5)^T$ and 49 detectors are uniformly distributed on the right boundary of the domain, i.e, positions for the detector d ($1 \leq d \leq 49$) is $\mathbf{x}_d = (5, 0.1d)^T$; see Fig. 2-1 (a). The computational domain is discretized into 100×100 square cells. 128 directions (uniformly distributed on unit circle) with equal weights are used. The scattering kernel we employed is the Henyey-Greenstein phase function [4] with an anisotropic factor $g = 0.9$. In all computations, we set the refractive index of the medium to be constant and equal to 1.37.

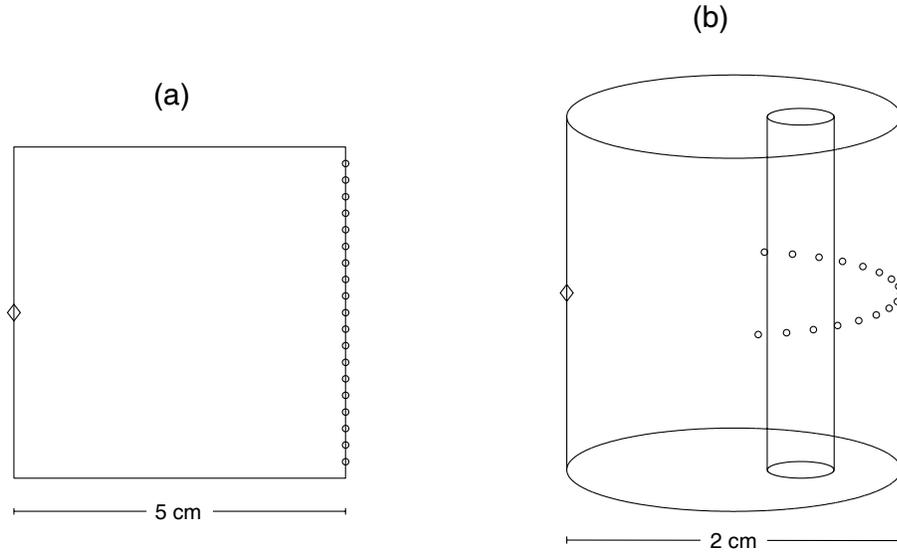


Figure 2-1: Geometrical settings of the computational domains. Diamond (\diamond) and circle (\circ) denote source and detectors, respectively.

Fig. 2-2 (a) and (b) show AC amplitude and phase delay for the first example calculated at detector positions assuming different optical properties. The modulation frequency for the source is taken to be 200 MHz. We observe that at a fixed modulation frequency, an increase in either absorption or scattering will cause a decrease of the AC amplitude computed at the detectors, see Fig. 2-2 (a). Phase delays obtained at the detectors (see Fig. 2-2 (b)) increase with scattering effects but decrease with increasing absorption. These observations agree with the underlying physics of the transport processes [43].

In the second example, we compare the results obtained for a cylindrical domain

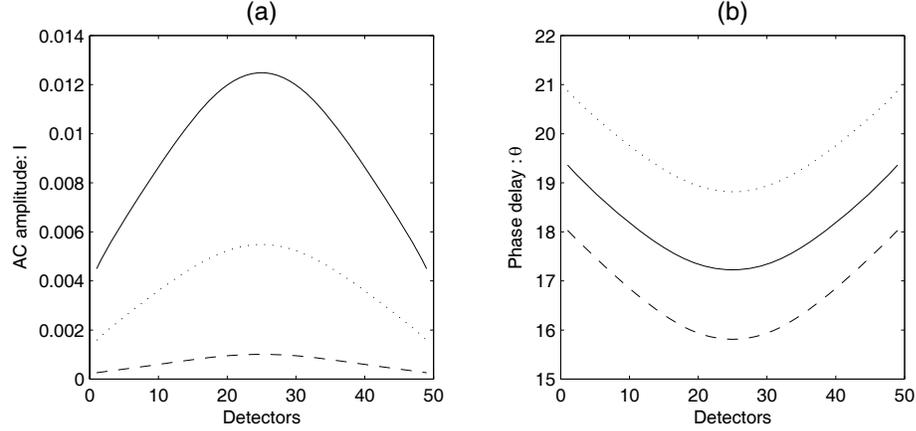


Figure 2-2: AC amplitude (a) phase delay (b) computed at the detectors for different optical parameters. $g = 0.9$ and 128 directions are used. Solid line: $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 120 \text{ cm}^{-1}$; dash line: $\sigma_a = 0.2 \text{ cm}^{-1}$, $\sigma_s = 120 \text{ cm}^{-1}$; dotted line: $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 240 \text{ cm}^{-1}$. The modulation frequency of the source is 200 MHz.

with and without a void-like inclusion. By void-like inclusion we mean a region in which both optical parameters are very small ($\sigma_a=0.001 \text{ cm}^{-1}$ and $\sigma_s=0.01 \text{ cm}^{-1}$). The domain is defined by $\Omega := \{(x, y, z)^T | x^2 + y^2 < 1; 0 < z < 2\}$ and the void by $\Omega_v = \{(x, y, z) | (x - 0.4)^2 + y^2 < 0.2^2, 0 < z < 2\}$. A point source is placed on $\mathbf{x}_s = (-1, 0, 1)^T$ and the detectors are uniformly distributed on the half circle $\Gamma = \{(x, y, z) | x^2 + y^2 = 1, x \geq 0, z = 1\}$, see Fig. 7-1(b). The domain is discretized into 11836 tetrahedral elements, and 120 directions (S10) with full level symmetry [113] are used. For the convenience of notation, we assign a superscript h to those quantities associated with the homogeneous medium and a superscript i to those with the inhomogeneities.

Fig. 2-3 shows results for the second example. We plot here the difference between the quantities calculated with and without the void inclusion as a function of detector positions. We assign the superscript v to those quantities computed in the former case and the superscript h to those in the latter case. We show the comparison at several modulation frequencies. It can be seen from Fig. 2-3 that the AC amplitude increases at the detectors right behind the void inclusion. This well-known effect is due to the non-scattering and non-absorbing nature of void regions. A change of phase of a few degrees is observed, as well as phase change increase with the modulation frequency

of the source.

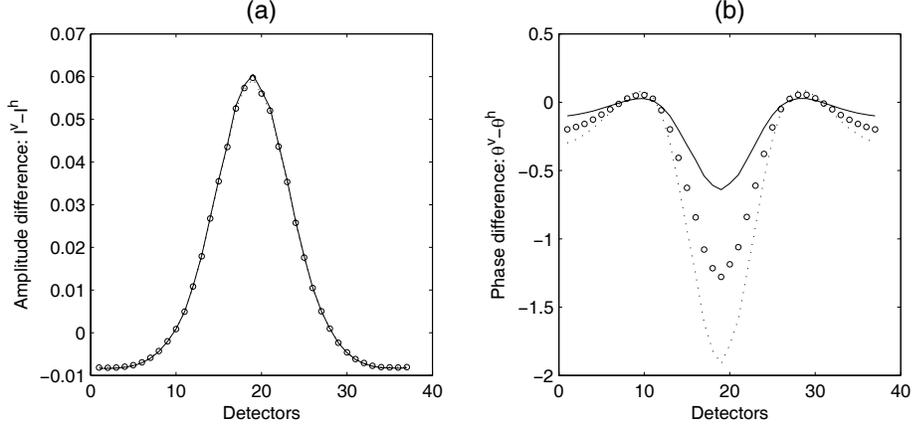


Figure 2-3: The difference of (a) AC amplitude ($I^v - I^h$) and (b) phase delay ($\theta^v - \theta^h$) calculated at the detectors for various modulation frequencies in domain with a void inclusion. $g = 0.9$ and 120 fully level-symmetric directions [113] are used. The optical parameters are: $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 120 \text{ cm}^{-1}$. Solid line: $\omega = 200$ MHz; empty circle: $\omega = 400$ MHz; dotted line: $\omega = 600$ MHz.

2.4.2 Setup for the reconstructions

We now provide four numerical examples that illustrate the performance of our numerical reconstruction method. In the first example, we reconstruct the spatial distribution of the absorption coefficient while keeping the scattering coefficient fixed. In the second example the spatially varying scattering coefficient is reconstructed, while the absorption coefficient is fixed. We then show an example in which both optical properties are reconstructed simultaneously. All the first three examples are done in two-dimensional settings. In the fourth example, we show three-dimensional simultaneous reconstructions of both optical properties. The setting for our calculations in the following.

For our two-dimensional simulations, we consider a computational domain of size $2 \times 2 \text{ cm}^2$, denoted by $\bar{\Omega} \equiv \Omega \cup \partial\Omega = [0, 2] \times [0, 2]$. We cover the domain by 80×80 cells of uniform size whose nodes are given by

$$\Omega_h = \{\mathbf{x}_{i,j} = (x_i, y_j), x_i = i\Delta x, y_j = j\Delta y, i, j = 0, 1, \dots, 80\},$$

with $\Delta x = \Delta y = 0.025$. The direction space is discretized into 128 uniformly distributed (over $[0, 2\pi)$) directions with identical quadrature weight:

$$S_{\Delta\theta}^1 = \{\boldsymbol{\theta}_i : \boldsymbol{\theta}_i = (i - 1) * \Delta\theta, i = 1, \dots, 128\}.$$

where $\Delta\theta = 2\pi/128$. The above discretizations yield a total number of 819200 unknowns for one forward problem (solving for \mathbf{U} in (2.28)), which is also true for the corresponding adjoint problem (Ψ in (2.39)). In all two-dimensional simulations, four sources ($N_q = 4$) are used. They are located at $(0, 1)$, $(1, 0)$, $(2, 1)$ and $(1, 2)$, respectively. For each sources, 20 detectors ($N_d = 20$) are used. The detectors are uniformly distributed along the sides of the square.

For the three-dimensional simulation, we consider a cylindrical domain given by

$$\bar{\Omega} := \{(x, y, z) : x^2 + y^2 \leq 1; 0 \leq z \leq 2\}.$$

We cover the domain by 22022 tetrahedral elements. For the integration over S^2 , we employ the full level symmetric S_8 discrete ordinate set of [113]. This set consists of 80 directions. A total number of 16 sources ($N_q = 16$) are used in the simulation. The sources are uniformly distributed on the two circles defined by $\Gamma_i^q = \{(x, y, z) : x^2 + y^2 = 1; z = z_i\}$ ($i = 1, 2$), where $z_1 = 0.5$, $z_2 = 1.5$. We arrange 8 layers of totally 128 detectors on the boundary of the domain. Those detectors are located at $\Gamma_j^d = \{(x, y, z) : x^2 + y^2 = 1; z = z_j\}$ ($j = 1, \dots, 8$), with $z_j = 0.3 + (j - 1) * 0.2$. Note that some detectors are placed on the same xy-plane as sources (but they do not overlap with each other).

2.4.3 Generating synthetic data

In all the numerical reconstructions presented in section 2.4, the “measurements” are synthetic rather than coming from real experimental data. A straightforward way of generating synthetic data is to use the same discretization for the forward model and for the inversion procedure. This may lead to somewhat simplified inversions of

the finite dimensional problem, which is often referred to as *inverse crimes* in the inverse problems community [54, p.133]. To avoid committing these “crimes”, the data are generated with a discretization about twice as fine (in all variables) as the discretization used in the inversions.

In the following sections, our “exact data” will thus be those obtained from the fine mesh calculations. By noisy data, we mean the “exact data” polluted by additional multiplicative noise. Let z_d be the exact data, the “noisy” data are simulated according to the rule: $z_d^\delta = (1 + \delta * \text{random})z_d$, where “random” is a uniformly distributed random variable in $[-1, 1]$ and $\delta \geq 0$ will vary in our numerical simulations.

It should be noted that the “exact data” seen on the coarse grid used in the inversion actually already contain some “noise” because they were generated by the fine mesh calculations.

2.4.4 Single parameter reconstructions

We start with a simpler case where only one optical parameter needs to be reconstructed. Such reconstructions are often useful in practical applications. For example, it is generally believed that changes in the oxygenation of tissues correspond mainly to changes in the absorption property of tissues [177].

Example 1: Reconstructing σ_a We first reconstruct a small absorbing disc centered at (1.15 cm, 1.15 cm) of radius 0.2 cm embedded in the two-dimensional computational domain. The optical parameters for the background and the disc are $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$ and $\sigma_a = 0.2 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$, respectively. The anisotropy factor $g = 0.90$. Those are typical parameter values as they are encountered in biomedical applications. The modulation frequency of the source is $\omega = 600\text{MHz}$. Each reconstruction here takes approximately 5 hours on a 2.4GHz Pentium XEON processor.

Fig. 2-4 shows the real absorption map and the reconstructed maps for different noise levels. We also provide in Fig. 2-5 (A) (solid line) the evolution of the normalized objective function versus the iteration step in the case of reconstructions using noise

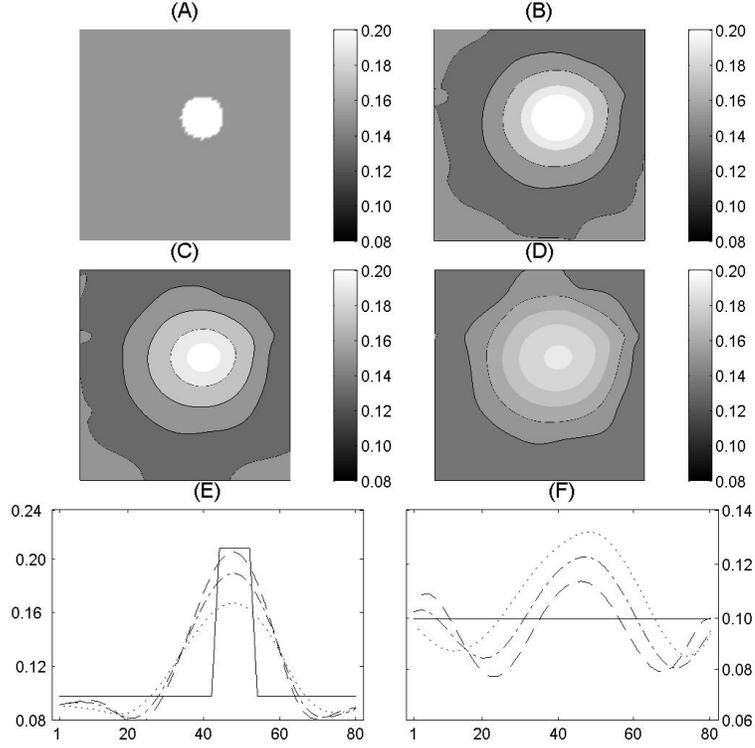


Figure 2-4: Maps of the reconstructed absorption coefficient σ_a [cm^{-1}] in Example 1. (A): real absorption map; (B): reconstructed absorption map with exact synthetic data; (C): reconstruction with 10% random noise; (D): reconstruction with 20% random noise; (E): cross sections of map (A) (solid line), (B) (dashed line), (C) (dash-dotted line) and (D) (dotted line) along the bottom left to top right diagonal. (F): same as (E) except that the cross section is along $y = 0.4$.

free synthetic data. Note that although the total number of BFGS iterations can be reduced by stricter line search scheme in the optimization algorithm, the total computational time remains almost constant. We have tested many classical line search algorithms. The results shown here use the one that works best for our application.

The quality of the reconstructions is measured as follows. Denote by $M^o \in \mathbb{R}^N$ ($M^r \in \mathbb{R}^N$) an exact (reconstructed) quantity, which can be either the absorption or the scattering map. We then define the relative l^2 error between M^o and M^r by:

$$\mathcal{E}_{l^2} = \frac{\|M^r - M^o\|_{l^2}}{\|M^o\|_{l^2}} := \frac{\sqrt{\sum_{i=1}^N (M_i^r - M_i^o)^2}}{\sqrt{\sum_{i=1}^N (M_i^o)^2}}. \quad (2.54)$$

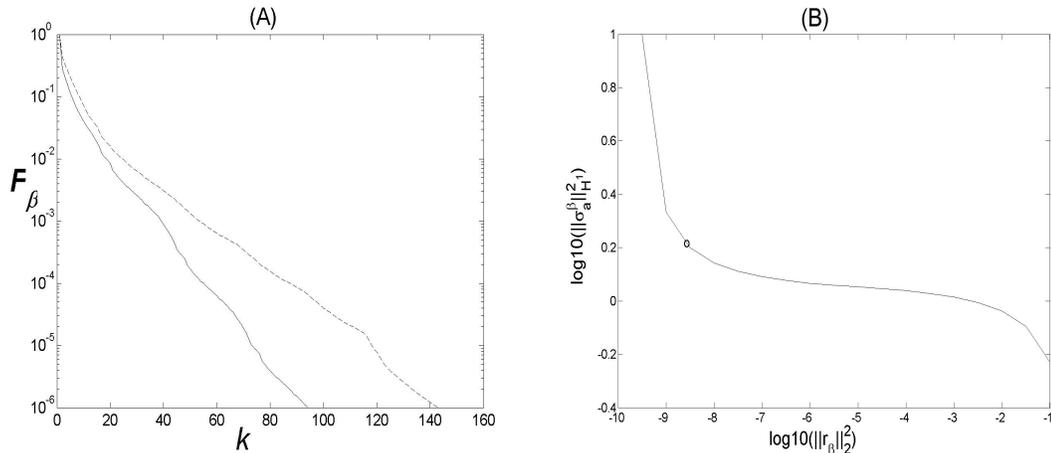


Figure 2-5: (A). Evolution of normalized objective functional \mathcal{F}_β with respect to the number of iteration steps k for Example 1. Solid line: reconstruction of an absorbing disc; dashed line: reconstruction of a scattering disc. (B). L -curve used to choose optimal regularization parameter β for reconstruction with noise-free data in the reconstruction of an absorbing disc. The circle (\circ) denotes the place where β is chosen. Note that here $\mathcal{J} = \|\sigma_a\|_{\mathcal{H}^1}^2$ because we reconstruct only σ_a and we have set $\sigma_a^0 = 0$.

The quality of the reconstruction in the case of an absorbing disc is given by the parameters shown in the second column of Tab. 2.4.4. In Fig. 2-5 (B), we display the L -curve we use to choose the optimal parameter β in the above reconstructions. We mention again that due to the acquisition of our data presented in section 2.4.3, the noise-free data actually contain “noise” on the coarse grid.

The reconstructions are classical examples of what we may expect in the field. Because of the severe ill-posedness of the inverse problem, it is difficult to reconstruct localized objects unless additional information is included in the reconstruction (which we do not want to do here). Note that the center and the integral of the absorption inclusion are more or less well reconstructed, even if the localized absorption map is somewhat smeared over a relatively large domain (whose size increases as the noise level, whence the regularization parameter β , increases). This is consistent with results obtained from asymptotic theories; see for instance [19].

Example 2: Reconstructing σ_s We now replace the highly absorbing disc in the previous case by a highly scattering disc at the same position and with the same

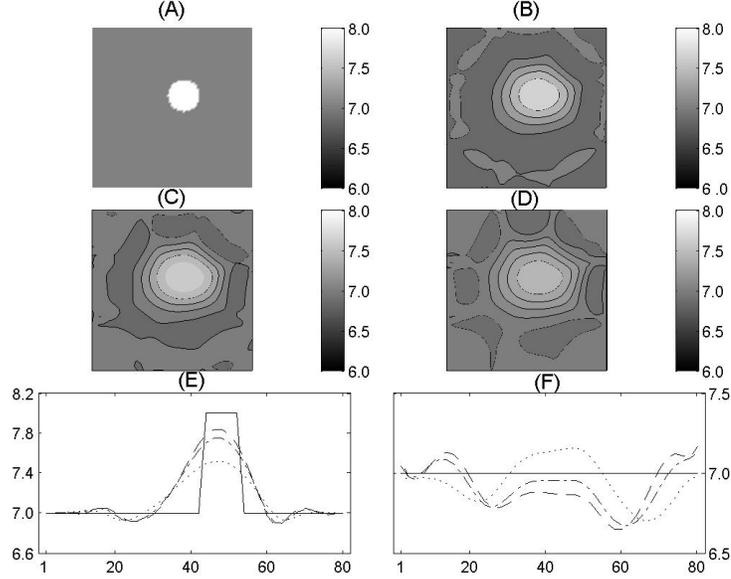


Figure 2-6: Maps of reconstructed reduced scattering coefficients $\sigma'_s = (1 - g)\sigma_s$ [cm^{-1}] for Example 1. (A): real scattering map. (B): reconstructed scattering map with noise free synthetic data. (C): reconstruction with 10% random noise. (D): reconstruction with 20% random noise. (E): cross sections of map (A) (solid line), (B) (dashed line), (C) (dash-dotted line) and (D) (dotted line) along the bottom left to top right diagonal. (F): same as (E) except that the cross section is along $y = 0.4$.

size. The optical parameters for the background and the disc are $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 70 \text{ cm}^{-1}$ and $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$, respectively. Again, the anisotropy factor $g = 0.9$, the modulation frequency $\omega = 600\text{MHz}$. Each reconstruction here takes approximately 6 hours on a 2.4GHz Pentium XEON processor. Fig. 2-6 shows the exact scattering map and the reconstructed maps for different noise levels. Error estimates are presented the third column of Tab. 2.4.4.

Cases	absorbing disc			scattering disc		
Noise level	0%	10%	20%	0%	10%	20%
$\beta \times 10^8$	1.0	2.8	4.6	1.3	2.9	5.7
\mathcal{E}_{l_2}	0.064	0.069	0.072	0.065	0.074	0.080

Table 2.1: Optimal regularization parameters β and errors in reconstructions for different cases in Example 1 and Example 2.

The quality of the reconstructions is very similar in the above two examples and is quite satisfactory. In spite of the fact that stronger regularizations have to be

imposed as the noise level increases, the localization and the estimate for the optical parameters in the presence of moderate noise indeed allow us to obtain reasonably accurate information toward diagnostic purposes.

2.4.5 Frequency-domain versus steady-state

One of the main reasons for introducing frequency-domain reconstructions is that they allow for a better separation between the scattering and absorption properties of the inclusions [124]. Diffusion-based theories show that both coefficients cannot be reconstructed simultaneously without additional geometrical hypotheses [14, 96]. We now show on two numerical examples that frequency-domain data indeed substantially improve the reconstruction of both coefficients.

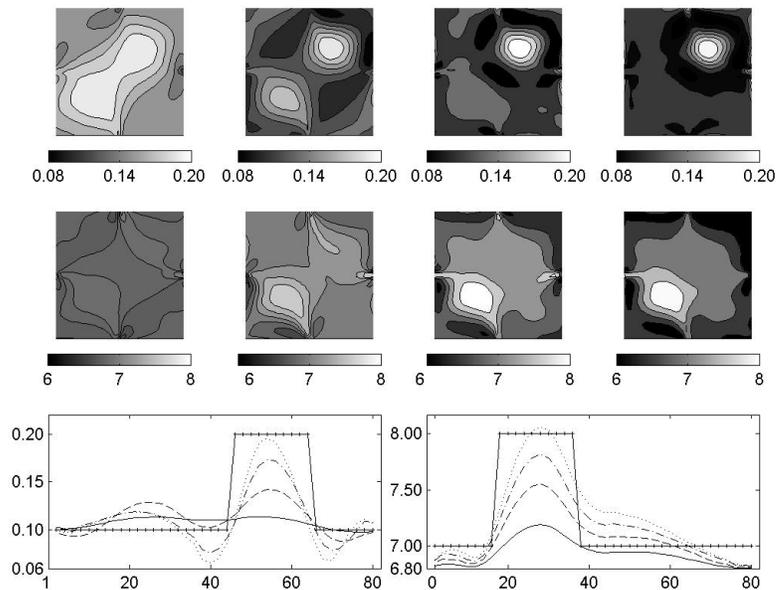


Figure 2-7: Top row: Maps of reconstructed absorption coefficients σ_a [cm^{-1}] at BFGS iteration $k = 40, 80, 120$ and 156 (final), respectively, for the frequency domain reconstruction in Example 3. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1 - g)\sigma_s$ [cm^{-1}]. Bottom row: cross section of real (solid line with +) and reconstructed absorption maps (left), reduced scattering maps (right) along the diagonal at iterations $k = 40$ (solid line), 80 (dashed line), 120 (dash-dotted line) and 156 (dotted line). The reconstructions are done with noise-free synthetic data.

Example 3: Two-dimensional simultaneous reconstruction We reconstruct here in the square domain two small discs of radius 0.2 cm and centered at (1.35 cm, 1.35 cm) and (0.65 cm, 0.65 cm), respectively. The first disc is highly absorbing and the second one is highly scattering. Optical properties for the two discs are $\sigma_a = 0.2 \text{ cm}^{-1}$, $\sigma_s = 70 \text{ cm}^{-1}$ and $\sigma_a = 0.1 \text{ cm}^{-1}$, $\sigma_s = 80 \text{ cm}^{-1}$, respectively. The background parameters are $\sigma_a = 0.1 \text{ cm}^{-1}$ and $\sigma_s = 70 \text{ cm}^{-1}$. As before, $g = 0.9$, $\omega = 600\text{MHz}$. Each reconstruction takes approximately 8 hours on a 2.4GHz Pentium XEON processor.

We compare the reconstructions based on frequency-domain ERT with those based on steady-state ERT. The latter is obtained by setting the frequency $\omega = 0$ in our formulation and keeping everything else the same. We present in Fig. 2-7 and Fig. 2-8 the reconstructions obtained by the frequency-domain method and the steady-state method, respectively. We also list the parameters which measure the quality of the reconstructions at different iteration steps in Tab. 2.2.

Example 4: Three-dimensional simultaneous reconstruction In the last numerical test, we show simultaneous reconstructions of two optical properties in a three-dimensional setting. We try to reconstruct a small cylinder $\bar{\Omega}_s = \{(x, y, z) : (x - 0.5)^2 + y^2 \leq 0.2^2, 0 \leq z \leq 2\}$ embedded in the cylindrical domain defined in section 2.4.2. Both the absorption and the scattering coefficients of the small cylinder are different from those of the background. This is different from the case in Example 3 where absorption and scattering anomalies are located at different places. Optical properties for the small cylinder are $\sigma_a = 0.2 \text{ cm}^{-1}$ and $\sigma_s = 80 \text{ cm}^{-1}$, while those for the background are $\sigma_a = 0.1 \text{ cm}^{-1}$ and $\sigma_s = 70 \text{ cm}^{-1}$. The anisotropy factor $g = 0.9$ and the modulation frequency $\omega = 600\text{MHz}$. Each reconstruction takes approximately 22 hours on a 3GHz Pentium XEON processor. As in Example 3, we compare the frequency domain reconstructions with the steady state reconstructions. Cross sections of reconstructions obtained by frequency-domain method and steady-state method are presented in Fig. 2-10 and Fig. 2-11, respectively. Qualities of the reconstructions at different iteration steps are again listed in Tab. 2.2.

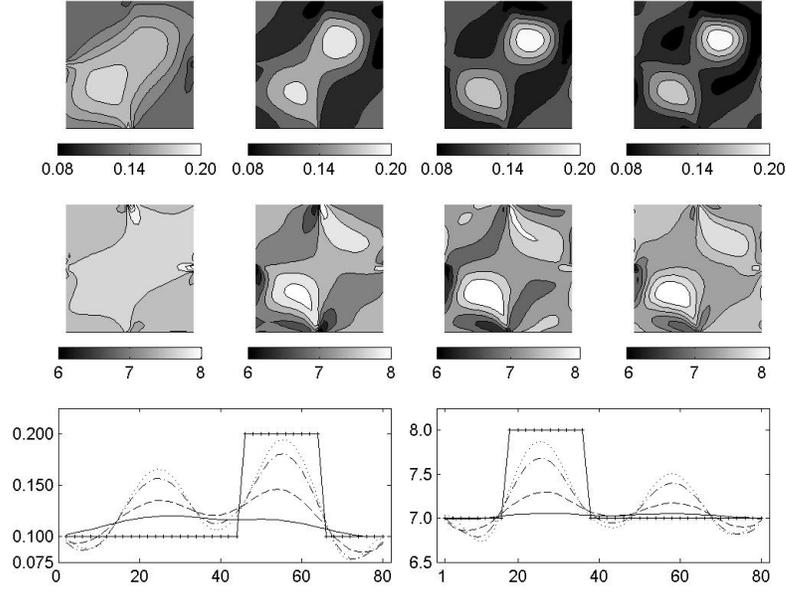


Figure 2-8: Top row: Maps of reconstructed absorption coefficients σ_a [cm^{-1}] at BFGS iteration $k = 40, 80, 120$ and 354 (final), respectively, for the steady state reconstruction in Example 3. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1-g)\sigma_s$ [cm^{-1}]. Bottom row: cross section of real (solid line with +) and reconstructed absorption maps (left), reduced scattering maps (right) along the diagonal at iterations $k = 40$ (solid line), 80 (dashed line), 120 (dash-dotted line) and 354 (dotted line). The reconstructions are done with noise-free synthetic data.

We first observe that in both two-dimensional (Example 3) and three-dimensional (Example 4) reconstructions, the frequency domain reconstruction converges faster (in terms of BFGS iterations) than the steady state reconstruction; see for example the results on Fig. 2-9 (A) and in Tab. 2.2. This has been confirmed in many other geometrical settings we have tested: the speed of convergence of the steady-state reconstruction presented here is one of the most favorable we have obtained, whereas the speed of convergence of the frequency-domain reconstructions was very often similar to what we have presented here.

As far as quality of the reconstruction is concerned, we observe a significant improvement in the frequency domain reconstructions compared to the steady-state reconstructions. In all simulations, the stopping criteria is the same: $\frac{\mathcal{F}_\beta^k(\sigma_a, \sigma_s)}{\mathcal{F}_\beta^0(\sigma_a, \sigma_s)} \leq 10^{-5}$.

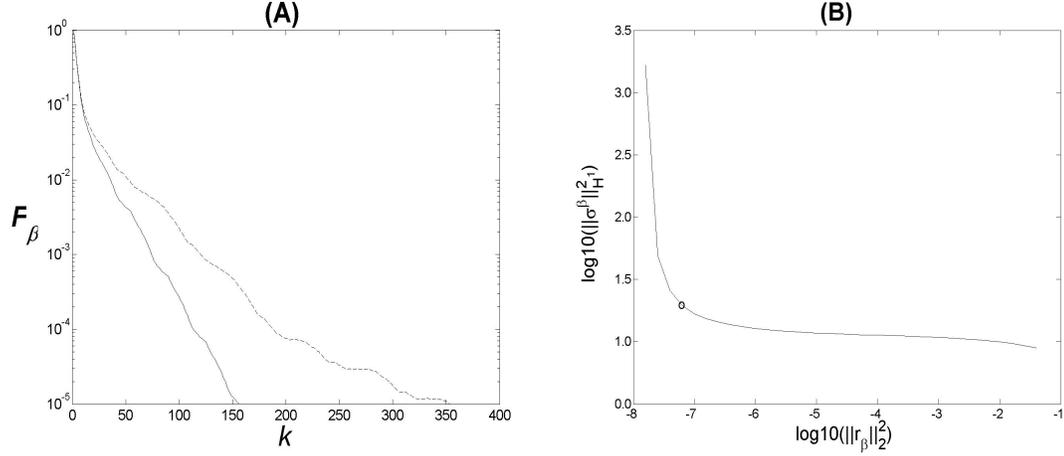


Figure 2-9: (A). Evolution of the normalized objective functional \mathcal{F}_β with respect to the number of iteration steps k for Example 3. Solid line: frequency domain reconstruction of both coefficient simultaneously; dashed line: steady state reconstruction of both coefficients simultaneously. (B.) L -curve used to choose optimal regularization parameter β for reconstruction with noise-free data in the frequency domain simultaneous reconstruction of absorbing and scattering coefficients. The circle (o) denotes the place where β is chosen. Note that $\mathcal{J} \equiv \|\sigma^\beta\|_{\mathcal{H}^1}^2 := \|\sigma_a^\beta\|_{\mathcal{H}^1}^2 + \epsilon\|\sigma_s^\beta\|_{\mathcal{H}^1}^2$ because we have set $\sigma_a^0 = \sigma_s^0 = 0$.

Although the L^2 errors may not enjoy a dramatic improvement (see Tab. 2.2), they are still significantly reduced. More importantly, the last rows in Fig. 2-7 and Fig. 2-8 show dramatic reductions (at least by a factor 2) of the cross-talk between the absorption and scattering reconstructions: the spurious bumps (left of the left picture on the bottom row and right of the right picture on the bottom row in Fig. 2-7 and Fig. 2-8) are clearly much stronger in the steady-state calculations than in the frequency-domain calculations. This is the major advantage of the frequency-domain calculations. Very similar phenomena are observed in the three-dimensional simulations; see Fig. 2-10 and Fig. 2-11. Although the absorption coefficient is still over-estimated in the frequency domain case (last row of Fig. 2-10), it is better than the situation in the steady-state case (last row of Fig. 2-11) where the absorption coefficient σ_a is severely over-estimated while the reduced scattering coefficient σ'_s is severely under-estimated. As predicted by theory, we have observed that an increase in ω led to reduced cross-talks. How much this effect depends on the choice of the frequency (as well as on possible combinations of different frequencies) and on the

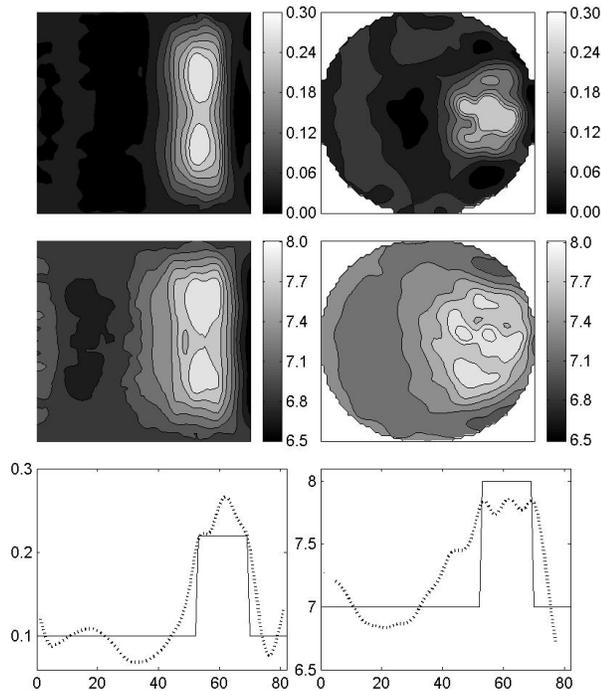


Figure 2-10: Top row: XY cross-section (at $z = 1$) and XZ cross-section (at $y = 0$) of the reconstructed absorption coefficient σ_a [cm^{-1}] for the frequency domain reconstruction in Example 4. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1 - g)\sigma_s$ [cm^{-1}]. Bottom left: real (solid) and reconstructed (dotted) absorption coefficient along line $y = 0, z = 1$. Bottom right: same as bottom left but for reconstructed reduced scattering coefficient.

geometrical setting will be explored in future works.

2.5 Conclusions and remarks

We have formulated an inverse problem in optical tomography as a regularized least square problem based on the frequency-domain equation of radiative transfer to model light propagation in biological tissues. In the inversion procedure, the forward model is discretized by using a finite volume method and a discrete ordinates method. We solve the regularized least square problem by using a limited-memory Quasi-Newton method with BFGS type updating rule for the Hessian matrix, and have incorporated positivity constraints and L^∞ bounds on the optical parameters. Numerical recon-

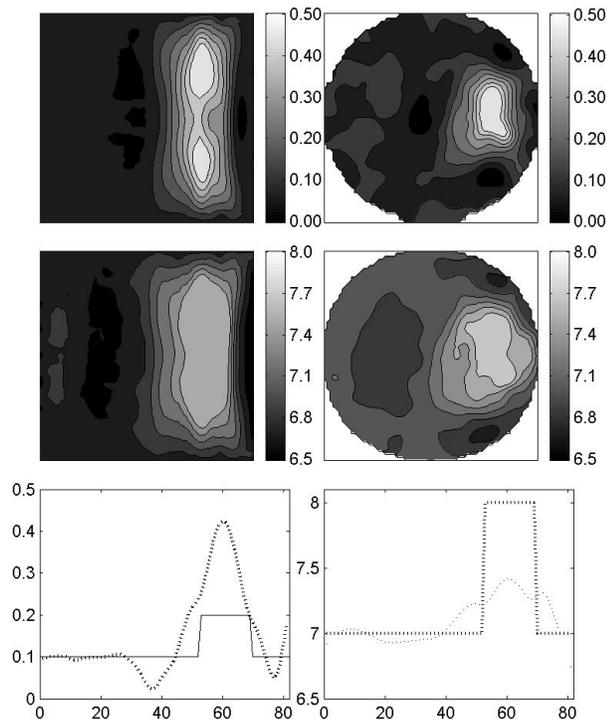


Figure 2-11: Top row: XY cross-section (at $z = 1$) and XZ cross-section (at $y = 0$) of the reconstructed absorption coefficient σ_a [cm^{-1}] for the steady-state reconstruction in Example 4. Middle row: same as the top row but for the reduced reconstructed scattering coefficients $\sigma'_s = (1 - g)\sigma_s$ [cm^{-1}]. Bottom left: real (solid) and reconstructed (dotted) absorption coefficient along line $y = 0, z = 1$. Bottom right: same as bottom left but for reconstructed reduced scattering coefficient.

structions based on synthetic data provide results that are in agreement with the expected reconstructions. Notably, the crosstalk between the two optical parameters is significantly reduced in frequency-domain reconstructions.

The method presented here also overcomes several of the shortcomings of diffusion-equation based optical tomography [11, 58], which provides a very useful tool in many problems but fails to adequately model strongly absorbing regions (e.g. large blood-filled spaces such as brain hematoma), low-scattering void-like inclusions (e.g. spaces filled with cerebrospinal fluid, amniotic fluid, or synovial fluid) and optically relatively thin media such as fingers and small animals. Main domains of application of the proposed method will be functional imaging of rheumatoid arthritis (RA) of human

		$k = 40$		$k = 80$		$k = 120$		Final	
		σ_a	σ_s	σ_a	σ_s	σ_a	σ_s	σ_a	σ_s
$E3$	$\mathcal{E}_{l^2}(f)$	0.121	0.144	0.092	0.112	0.080	0.092	0.063	0.076
	$\mathcal{E}_{l^2}(s)$	0.181	0.224	0.127	0.132	0.113	0.112	0.094	0.106
$E4$	$\mathcal{E}_{l^2}(f)$	0.194	0.252	0.173	0.208	0.147	0.189	0.131	0.171
	$\mathcal{E}_{l^2}(s)$	0.342	0.422	0.287	0.366	0.245	0.322	0.210	0.292

Table 2.2: Error estimates for the reconstructions of Example 3 ($E3$) and Example 4 ($E4$) for several iteration steps (k) in the optimization process. Here, “f” refers to frequency-domain calculations and “s” to steady state calculations.

finger joints, and small animal imaging [88, 92], where the diffusion equation typically fails to generate accurate forward predictions.

One problem with the methods presented in the chapter is that the speed of the inversion depends strongly on how fast and accurately the forward problem is solved. Codes that use the radiative transport equation as a forward model are especially slow, since numerical solutions of ERT require in general the discretization of an angular variable in addition to spatial and time variables (in time-dependent case). In next chapter, we will introduce a new kind of method that can accelerate the reconstruction process considerably. We will compare numerically the performance of the two methods in various settings.

Chapter 3

Inverse transport as a PDE-constrained optimization problem

As we have seen in the previous chapter, numerical optimization techniques play an important role in the solution of the inverse transport problems. We have developed an algorithm that is based on quasi-Newton type of minimization technique. This method requires solving the forward problem for some approximations of absorption and scattering coefficients over and over again in each reconstruction. The speed of the algorithm depends strongly on how fast and accurately the forward problems are solved. Since numerical solutions of the ERT are very expensive, the inversion is usually very slow. In the chapter, we introduce a new type of minimization technique that can potentially speed up the reconstruction process. We will compare the advantages and disadvantages of the new method with the quasi-Newton method in the previous chapter. The presentation of this chapter is based on reference [2].

3.1 Problem statement

Let us recall first the problem of optical tomography. In general, the propagation of the near-infrared light in tissue can be modeled by a set of differential equations and boundary conditions that can be written abstractly into the form

$$\mathcal{O}(\sigma, u) = 0, \quad (\sigma, u) \in \boldsymbol{\sigma} \times \mathcal{U}, \quad (3.1)$$

where $\sigma \equiv (\sigma_a, \sigma_s)$ denotes the optical property of the tissue while u essentially measures the distribution of light in the tissue. The σ denotes the function space where the optical property σ takes value in, and the \mathcal{U} denotes the space that solution u belongs to. $\mathcal{O}(\sigma, u)$ should be understood as an operator from $\sigma \times \mathcal{U}$ to \mathcal{U} .

Model (3.1) can be either the radiative transfer equation [11, 43, 58], which is believed to be very accurate but hard to solve as we have seen in chapter 2, or the less accurate diffusion equation [11], which is an approximation of the ERT in the cases where scattering is sufficiently high ($\sigma_s \gg 1$) and absorption is sufficiently small ($\sigma_a \ll 1$).

The measured data (in data space \mathcal{Z}) in optical tomography is usually a bounded linear functional of u , $\mathcal{G}u$, where $\mathcal{G} : \mathcal{U} \mapsto \mathcal{Z}$ is called the measurement operator. Note that \mathcal{G} has to be introduced because in practice only partial information about the forward solution, u , can be measured [11]. An example of the measurement operator is the one defined in (2.6) in the previous chapter. We remark here that when u can be measured directly, \mathcal{G} is nothing but an identity operator.

The objective of optical tomography is to transform the measured data into accurate approximation of the spatial distribution of optical properties σ inside the tissue. Due to the lack of analytical inversion formulas, this transformation is usually done through numerical optimization tools [11, 13, 147]. The guiding principle of optimization approaches to the inverse problems in optical tomography is to update iteratively model parameter σ such that the forward model (3.1) generates a set of data (predictions) that best match measurements. To be more specific, let us first denote by $z \in \mathcal{Z}$ our experimental measurements. We then introduce a real-valued non-negative function, called objective function, to be minimized to generate the inverse solution, $\phi : \sigma \times \mathcal{U} \mapsto \mathbb{R}^+$:

$$\phi(\sigma, u) = \frac{1}{2} \|\mathcal{G}u - z\|_{\mathcal{Z}}^2 + \mathcal{R}(\sigma). \quad (3.2)$$

Again, the first term is used to quantify the difference between the measurements and model predictions, while the second term, $\mathcal{R}(\sigma)$, is a regularization term. The parameter σ that minimizes ϕ is the one we want.

The diffuse optical tomographic imaging can now be formulated as the following equation-constrained optimization problem

$$\begin{aligned} & \text{minimize} && \phi(\sigma, u) \\ & \text{subject to} && (3.1). \end{aligned} \tag{3.3}$$

Note that we do not consider here additional constraints on the optical property σ , which might be added to (3.3) as we have done in the previous chapter; see also references [91, 141, 154, 155].

There are several techniques to solve (3.3) in optimization theory [132]. All the existing algorithms in optical tomography community convert (3.3) into an unconstrained optimization problem, just as what we did in chapter 2. Essentially one first solves the equation (3.1) to obtain u as a function of σ , $u = u(\sigma)$ and then uses this fact to eliminate the explicit dependency of the objective function (3.2) on function u , giving rise to another function that should be minimized only with respect to the optical property σ , $\tilde{\phi} : \boldsymbol{\sigma} \mapsto \mathbb{R}^+$:

$$\tilde{\phi}(\sigma) \equiv \phi(\sigma, u(\sigma)). \tag{3.4}$$

To minimize $\tilde{\phi}(\sigma)$, one can use methods such as the conjugate gradient and quasi-Newton methods we adopted in chapter 2; see also references [1, 79, 107, 133, 147, 162] for details in implementation of algorithms based on this unconstrained approach.

One drawback of the above-mentioned method is that for each evaluation of the function $\tilde{\phi}(\sigma)$ during the minimization process, one has to solve the forward problem (3.1) once to obtain $u(\sigma)$. So each inversion procedure will require hundreds of forward-problem-solving, which usually takes extremely long computational time.

The approach that is considered in this chapter does not require solving the forward problem many times. Instead of working with (3.4), we start directly from problem (3.3). To do that, we introduce the following Lagrangian functional for

problem (3.3), $\mathcal{L} : \boldsymbol{\sigma} \times \mathcal{U} \times \mathcal{V}^* \mapsto \mathbb{R}$,

$$\mathcal{L}(\boldsymbol{\sigma}, u; \lambda) = \phi(\boldsymbol{\sigma}, u) - \langle \lambda, \mathcal{O}(\boldsymbol{\sigma}, u) \rangle, \quad (3.5)$$

where $\lambda \in \mathcal{U}^*$ is the Lagrange multiplier, \mathcal{U}^* is the dual space of \mathcal{U} , and $\langle \cdot, \cdot \rangle$ denotes the duality pair between \mathcal{U} and \mathcal{U}^* . From the theory of constrained optimization theory [32, 132], it is known that the solution to (3.3) satisfies the following optimality condition for \mathcal{L} [32, 132]:

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\sigma}}(\bar{\boldsymbol{\sigma}}, \bar{u}; \bar{\lambda}) = 0, \quad \frac{\partial \mathcal{L}}{\partial u}(\bar{\boldsymbol{\sigma}}, \bar{u}; \bar{\lambda}) = 0, \quad \mathcal{O}(\bar{\boldsymbol{\sigma}}, \bar{u}; \bar{\lambda}) = 0, \quad (3.6)$$

The system (3.6) is also known as the Karush-Kuhn-Tacker (KKT) condition [132]. Now the optimal point of (3.3) can be found by solving this KKT system. This approach is often referred to as *all-at-once* method because it solves the forward and minimization problems simultaneously as compared to the unconstrained optimization approach.

This kind of constrained optimization technique promises a much faster solution of the inverse problem and has already proved to be very useful in other applications such as shape design in aerodynamics [134, 164] and optimal control of incompressible flows [75]; see also [33] for a general overview. In all those problems, there are forward models that can be written in the form of (3.1). This kind of problem is often called simulation-based optimization in the literature. In the case when the forward model is a partial differential equation, the term PDE-constrained optimization is used instead [33]. For the optical tomography problem, the integro-differential radiative transport equation can be written as a system of coupled first-order partial differential equations after passing to the discrete ordinate formulation (see the previous chapter), we can thus put our inverse problem in a general PDE-constrained optimization framework. However, implementation of this PDE-constrained optimization approach for optical tomography has never been reported to the best of our knowledge. In the rest of the chapter, we will develop a constrained optimization approach for optical tomographic imaging and compare its performance with the un-

constrained optimization approach. Throughout this paper, we will not worry about the existence and uniqueness problems of (3.3), but rather focus on the numerical aspects.

We remarks finally that in general, optimization with constraints is not completely new in optical tomography. Many authors have exploited simple-bound constraints, which specifies upper and lower bounds for the optical properties such as those in chapter 2; see also [91, 141, 154, 155]. However, in all these case the forward problem still needs to be solved many times, and one does not make use of the advantages of novel PDF-constrained concepts in which the governing equation itself is considered as a constraint.

3.2 The augmented Lagrangian method for inverse transport

As before, let us denote by N_q the total number of light sources and denote by N_d the number of detectors readings corresponding to each source. The solution of the forward problem (2.28) corresponding to \mathbf{G}_q is denoted by \mathbf{U}_q and the set of solutions for all sources is denoted by $\widehat{\mathbf{U}} \equiv (\mathbf{U}_1, \dots, \mathbf{U}_{N_q})$. The set of measurements is denoted by $\mathbf{z} \equiv \{z_{q,d}, d = 1, \dots, N_d, q = 1, \dots, N_q\}$. Finally, for the detector located at a mesh node $\mathbf{x}_d \in \partial\Omega$, we denote by $P_d : \mathbb{R}^{2LN} \mapsto \mathbb{C}$ the discretized version of the measurement operator that measures the outgoing current at \mathbf{x}_d . Here L is the number of discrete ordinates used, N is the number of volumes used in the finite volume discretization of the domain.

With all these notations, we are ready to introduce the objective function to be minimized in optical tomography

$$\Phi(\Sigma, \widehat{\mathbf{U}}) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N_d} \frac{|P_d \mathbf{U}_q - z_{q,d}|^2}{|z_{q,d}|^2} + \frac{\beta}{2} R(\Sigma), \quad (3.7)$$

where β is a the regularization parameter. Σ is the discretization of σ . It is important

to note that the objective function Φ takes into account solutions \mathbf{U}_q of the forward problem for all N_q sources simultaneously.

Now on the discretized level, optical tomography can be formulated as a minimization problem subject to the constraints (2.28) for N_q different sources:

$$\begin{aligned} \min_{(\Sigma, \hat{\mathbf{U}}) \in \mathbb{R}^{2N} \times \mathbb{R}^{2LN N_q}} \quad & \Phi(\Sigma, \hat{\mathbf{U}}), \\ \text{subject to} \quad & (\mathbf{A} - \mathbf{S})\mathbf{U}_q - \mathbf{G}_q = 0, \quad 1 \leq q \leq N_q. \end{aligned} \quad (3.8)$$

Note that each discrete constraint $(\mathbf{A} - \mathbf{S})\mathbf{U}_q - \mathbf{G}_q = 0$ binds the vector of optical properties Σ and the vector of fluences \mathbf{U}_q for a particular source \mathbf{G}_q . The overall number of discretized constraints is (number of sources) \times (number of ordinates) \times (number of finite volumes).

3.2.1 The augmented-Lagrangian algorithm

Let us write $\mathbf{O}_q(\Sigma, \mathbf{U}_q) \equiv (\mathbf{A} - \mathbf{S})\mathbf{U}_q - \mathbf{G}_q = 0$ from now on. We introduce a Lagrangian function $\mathcal{L} : \mathbb{R}^{2N} \times \mathbb{R}^{2LN N_q} \times \mathbb{R}^{2LN N_q} \mapsto \mathbb{R}$ defined by

$$\mathcal{L}(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}}) = \Phi(\Sigma, \hat{\mathbf{U}}) - \sum_{q=1}^{N_q} \boldsymbol{\lambda}_q^T \mathbf{O}_q(\Sigma, \mathbf{U}_q), \quad \boldsymbol{\lambda}_q \in \mathbb{R}^{2LN}. \quad (3.9)$$

where $\hat{\boldsymbol{\lambda}} \equiv (\boldsymbol{\lambda}_1, \dots, \boldsymbol{\lambda}_{N_q})$ is introduced just to simplify the notation. The solution to the optimization problem (3.8) satisfies the optimality condition of (3.9), which is given by the following system:

$$\begin{aligned} \nabla_{\Sigma} \mathcal{L}(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}}) &= \nabla_{\Sigma} \Phi(\Sigma, \hat{\mathbf{U}}) - \sum_{q=1}^{N_q} \boldsymbol{\lambda}_q^T \nabla_{\Sigma} \mathbf{O}_q(\Sigma, \mathbf{U}_q) = 0, \\ \nabla_{\hat{\mathbf{U}}} \mathcal{L}(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}}) &= \nabla_{\hat{\mathbf{U}}} \Phi(\Sigma, \hat{\mathbf{U}}) - \sum_{q=1}^{N_q} \boldsymbol{\lambda}_q^T \nabla_{\mathbf{U}_q} \mathbf{O}_q(\Sigma, \mathbf{U}_q) = 0, \\ \nabla_{\boldsymbol{\lambda}_q} \mathcal{L}(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}}) &= \mathbf{O}_q(\Sigma, \mathbf{U}_q) = 0, \quad q = 1, \dots, N_q. \end{aligned}$$

In other words, if $(\Sigma^*, \widehat{\mathbf{U}}^*)$ provides an optimal solution of (3.8), then there exist $\boldsymbol{\lambda}_q^* \in \mathbb{R}^{2LN}$, $1 \leq q \leq N_q$, such that $(\Sigma^*, \mathbf{U}^*, \widehat{\boldsymbol{\lambda}}^*)$ is a stationary point of the Lagrangian function (3.9).

There exists several methods to find the stationary point of the Lagrangian function. For example, one can solve directly system (3.10) if an efficient nonlinear algebraic solver is available as in [33]. Here, we propose to use an augmented Lagrangian method, which is an iterative method easy to implement. The method defines an augmented Lagrangian function by

$$\mathcal{L}_K(\Sigma, \widehat{\mathbf{U}}; \widehat{\boldsymbol{\lambda}}) = \mathcal{L}(\Sigma, \widehat{\mathbf{U}}; \widehat{\boldsymbol{\lambda}}) + \frac{1}{2K} \sum_{q=1}^{N_q} \|\mathbf{O}_q(\Sigma, \mathbf{U}_q)\|_{l^2}^2, \quad (3.10)$$

where the term $(1/2K) \sum_{q=1}^{N_q} \|\mathbf{O}_q(\Sigma, \mathbf{U}_q)\|_{l^2}^2$ is a penalty for violating the constraints $\mathbf{O}_q(\Sigma, \mathbf{U}_q) = 0$, $1 \leq q \leq N_q$. The augmented Lagrangian method will look for a stationary point of the function \mathcal{L}_K instead of \mathcal{L} . Note that in the limit that the constraints are satisfied exactly, $\mathcal{L}_K = \mathcal{L}$. We send interested readers to reference [132] for more theoretical issues.

Assume that at the k -th iteration of the algorithm we have an approximation $(\Sigma^k, \widehat{\mathbf{U}}^k, \widehat{\boldsymbol{\lambda}}^k)$ to the stationary point $(\Sigma^*, \widehat{\mathbf{U}}^*, \widehat{\boldsymbol{\lambda}}^*)$ of the Lagrangian function. Let us fix the current estimates of the Lagrangian multipliers $\widehat{\boldsymbol{\lambda}}^k$ and a penalty parameter K_k . Minimization of $\mathcal{L}_{K_k}(\Sigma, \widehat{\mathbf{U}}; \widehat{\boldsymbol{\lambda}}^k)$ with respect to Σ and $\widehat{\mathbf{U}}$ yields the following system:

$$\begin{aligned} \nabla_{\Sigma} \Phi(\Sigma, \widehat{\mathbf{U}}) - \sum_{q=1}^{N_q} \left[(\boldsymbol{\lambda}_q^k)^T - \frac{1}{K_k} \mathbf{O}_q^T(\Sigma, \mathbf{U}_q) \right] \nabla_{\Sigma} \mathbf{O}_q(\Sigma, \mathbf{U}_q) &= 0, \\ \nabla_{\widehat{\mathbf{U}}} \Phi(\Sigma, \widehat{\mathbf{U}}) - \sum_{q=1}^{N_q} \left[(\boldsymbol{\lambda}_q^k)^T - \frac{1}{K_k} \mathbf{O}_q^T(\Sigma, \mathbf{U}_q) \right] \nabla_{\mathbf{U}_q} \mathbf{O}_q(\Sigma, \mathbf{U}_q) &= 0. \end{aligned} \quad (3.11)$$

Suppose that $(\Sigma^k, \widehat{\mathbf{U}}^k)$ is an approximate minimizer of $\mathcal{L}_{K_k}(\Sigma, \widehat{\mathbf{U}}; \widehat{\boldsymbol{\lambda}}^k)$, i.e. an approximate solution of system (3.11). Comparing this system with the optimality

condition (3.10), we conclude that $\boldsymbol{\lambda}_q^k - (1/K_k)\mathbf{O}_q(\Sigma^k, \mathbf{U}_q^k)$ approximates $\boldsymbol{\lambda}_q^*$:

$$\boldsymbol{\lambda}_q^* \approx \boldsymbol{\lambda}_q^k - (1/K_k)\mathbf{O}_q(\Sigma^k, \mathbf{U}_q^k), \quad 1 \leq q \leq N_q. \quad (3.12)$$

This formula can be rearranged to produce an estimate of $\mathbf{O}_q(\Sigma, \mathbf{U}_q)$:

$$\mathbf{O}_q(\Sigma^k, \mathbf{U}_q^k) \approx K_k(\boldsymbol{\lambda}_q^k - \boldsymbol{\lambda}_q^*). \quad (3.13)$$

Hence, we deduce that if $\boldsymbol{\lambda}_q^k$ is close to the optimal Lagrangian multiplier $\boldsymbol{\lambda}_q^*$, and K_k is small enough, then the pair $(\Sigma^k, \widehat{\mathbf{U}}^k)$ satisfies the corresponding constraint with a high accuracy. Formula (3.12) prompts a rule for iterative updating of the Lagrangian multipliers:

$$\boldsymbol{\lambda}_q^{k+1} = \boldsymbol{\lambda}_q^k - (1/K_k)\mathbf{O}_q(\Sigma^k, \mathbf{U}_q^k), \quad 1 \leq q \leq N_q. \quad (3.14)$$

We thus arrive to the following augmented Lagrangian algorithm.

Augmented Lagrangian Algorithm:

Initially choose $K_0, \tau_0 > 0$ and maximum iteration step *MAXIT* and Lagrangian multiplier $\widehat{\boldsymbol{\lambda}}^0$; choose also the initial guess $(\widetilde{\Sigma}^0, \widetilde{\mathbf{U}}^0)$ for subproblem (1) in the next loop;

- **FOR** $k = 0, 1, 2, \dots, \text{MAXIT}$

1. Find the minimizer $(\Sigma^k, \widehat{\mathbf{U}}^k)$ of $\mathcal{L}_{K_k}(\Sigma, \widehat{\mathbf{U}}; \widehat{\boldsymbol{\lambda}}^k)$ by an iterative method that
 - starts from initial value $(\widetilde{\Sigma}^k, \widetilde{\mathbf{U}}^k)$;
 - terminates when $\|\nabla_{(\Sigma, \widehat{\mathbf{U}})} \mathcal{L}_{K_k}(\Sigma^k, \widehat{\mathbf{U}}^k; \widehat{\boldsymbol{\lambda}}^k)\|_{l^2} \leq \tau_k$ is satisfied;
2. **IF** (final convergence criteria reached)
 - STOP**, $(\Sigma^k, \widehat{\mathbf{U}}^k)$ is the final solution;
3. Update Lagrangian multipliers

$$\boldsymbol{\lambda}_q^{k+1} = \boldsymbol{\lambda}_q^k - (1/K_k)\mathbf{O}_q(\Sigma^k, \mathbf{U}_q^k), \quad 1 \leq q \leq N_q; \quad (3.15)$$

4. Choose new penalty parameter $K_{k+1} \in (0, K_k)$ and new subproblem-stopping parameter τ_k ;
5. Set starting point for the next iteration:

$$(\tilde{\Sigma}^{k+1}, \tilde{\mathbf{U}}^{k+1}) = (\Sigma^k, \hat{\mathbf{U}}^k);$$

• **END**

We refer to references [32, 97, 132] for more detailed discussion of the augmented Lagrangian method, including the choice of parameters K_k and τ_k .

To solve the sub-optimization problem in step 1 of the above algorithm, we can use the limited-memory version of BFGS algorithm as we have introduced before in chapter 2. The BFGS algorithm requires computing of the gradient of objective function with respect to Σ and $\hat{\mathbf{U}}$. Those gradients can be analytical computed through

$$\nabla_{\Sigma} \mathcal{L}_K(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}}) = \nabla_{\Sigma} R(\Sigma) - \sum_{q=1}^{N_q} \boldsymbol{\lambda}_q^T \nabla_{\Sigma} \mathbf{O}_q(\Sigma, \mathbf{U}_q) + \frac{1}{K} \sum_{q=1}^{N_q} (\mathbf{O}_q(\Sigma, \mathbf{U}_q))^T \nabla_{\Sigma} \mathbf{O}_q(\Sigma, \mathbf{U}_q), \quad (3.16)$$

where $\nabla_{\Sigma} \mathbf{O}_q = (\nabla_{\Sigma_a} \mathbf{O}_q, \nabla_{\Sigma_s} \mathbf{O}_q)$, and partial derivatives

$$\frac{\partial \mathbf{O}_q}{\partial [\Sigma_a]_i} \equiv \begin{pmatrix} \frac{\partial [\mathbf{O}_q]_1}{\partial [\Sigma_a]_i} \\ \frac{\partial [\mathbf{O}_q]_2}{\partial [\Sigma_a]_i} \\ \vdots \end{pmatrix}$$

can be easily computed explicitly using the matrix representation (2.30) and (2.31). Gradient $\nabla_{\hat{\mathbf{U}}} \mathcal{L}_K(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}})$ can similarly be found in a closed form.

3.2.2 Interpretation and discussion

As follows from the estimate (3.13), the constraints $\mathbf{O}_q(\Sigma, \mathbf{U}_q) = 0$, $q = 1, \dots, N_q$ are in general not satisfied with $(\Sigma^k, \hat{\mathbf{U}}^k)$ on every iteration of the ALM. This means that

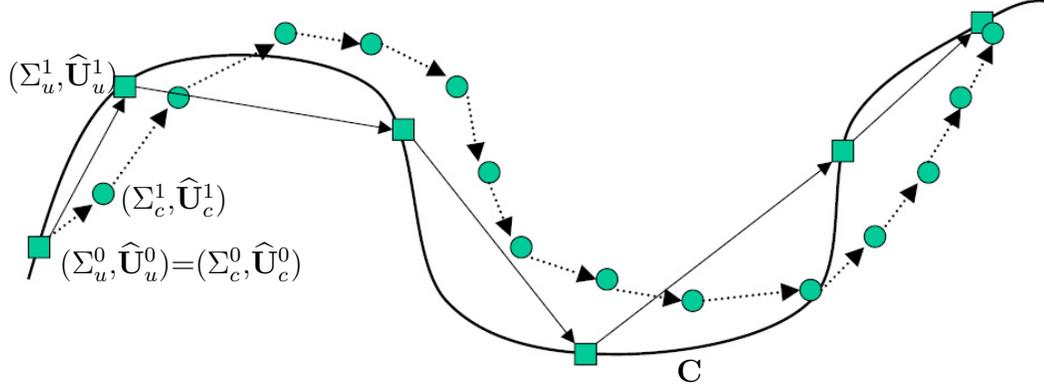


Figure 3-1: A simple illustration of the iteration process of unconstrained (\square) and constrained (\circ) optimization approaches to optical tomography. The subscript u and c denotes quantities in unconstrained and constrained minimization process, respectively.

\mathbf{U}_q^k is not the exact solution of the radiative transport equation with the optical properties $\Sigma^k \equiv (\Sigma_a^k, \Sigma_s^k)$ and source \mathbf{G}_q , which is quite different from the unconstrained optimization approach in chapter 2, in which \mathbf{U}_q^k is the solution of the ERT with optical properties (Σ_a^k, Σ_s^k) on every step k of an iterative minimization algorithm such as the quasi-Newton method. To outline the difference between the two methods we give a simple geometrical interpretation of the unconstrained and constrained optimization approaches. The equations $\mathbf{O}_q(\Sigma, \mathbf{U}_q) = 0$, $q = 1, \dots, N_q$ define a set \mathbf{C} of the space $\mathbb{R}^N \times \mathbb{R}^N \times (\mathbb{R}^{2LN})^{N_q}$. On every iteration of the unconstrained optimization method vectors $(\Sigma^k, \hat{\mathbf{U}}^k)$ are forced to belong to the set \mathbf{C} by construction. Satisfying this condition exactly requires solving forward problems for different right-hand sides on every iteration of the algorithm. In the constrained optimization approach, however, conditions $\mathbf{O}_q(\Sigma^k, \mathbf{U}_q^k) = 0$, $q = 1, \dots, N_q$ will be satisfied only in the limit when the algorithm converges. At a specific iteration before this convergence, $(\Sigma^k, \hat{\mathbf{U}}^k)$ does not have to belong to the set \mathbf{C} ; see Fig. 3-1. This property provides a higher degree of freedom in choosing a new iteration, and thus can lead to computationally less intensive algorithms.

An advantageous property of the augmented Lagrangian method for optical tomography is that it does not have significantly larger memory requirements than unconstrained optimization methods. Since the discretized transport operator $\mathbf{T} \equiv$

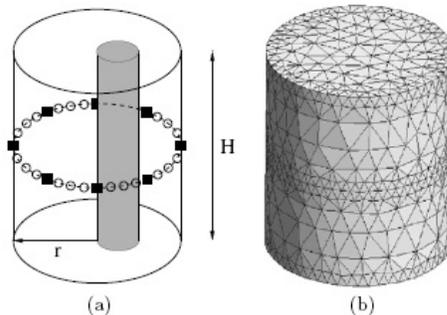


Figure 3-2: Test problems setup. Cylinder height: $H = 2$ cm, radius $r = 1$ cm; radius of the embedded small cylinder $r = 0.25$ cm. (a) source-detector layout with 8 sources (\square), 64 detectors (\circ); (b) finite-volume mesh with 6727 tetrahedrons.

$\mathbf{A} - \mathbf{S}$ is the same for all sources \mathbf{G}_q , one needs to store only one \mathbf{T} as in the unconstrained optimization cases. Actually, the only extra storage needed are the several source, radiance vectors and Lagrangian multipliers. This storage is much smaller than the storage of \mathbf{T} .

We also remark that the augmented Lagrangian method proposed above can be easily parallelized. First, in the sub-optimization problem, step 1, the gradient of the augmented Lagrangian function has an analytical form (3.16), which involves only the summation of local matrix-vector and vector-vector products. The computation of this gradient can thus be done on separate processors and then collected. A trivial example is to send those terms with different subscript q to different processors. Second, the update of Lagrangian multipliers in step 2 of the method can also be sent to parallel processors.

3.3 Numerical reconstructions

3.3.1 The test problem setup

To illustrate the performance of the ALM, we consider three types of media and measurement geometries. In the first problem, we reconstruct the absorption coefficient in a cylinder with a smaller cylindrical inhomogeneity, in which the absorption coefficient is twice as high as in the background (see Fig. 3-2 for the geometrical

	Problem 1	Problem 2	Problem 3
Anisotropy factor g	0.0	0.5	0.9
Background σ_a (cm^{-1})	0.1	0.1	0.5
Inhomogeneity σ_a (cm^{-1})	0.2	0.1	1.0
Background σ'_s (cm^{-1})	10.0	10.0	10.0
Inhomogeneity σ'_s (cm^{-1})	10.0	15.0	15.0
Number of sources	8	8	24
Number of detectors	64	64	24
Modulation frequency ω (MHz)	400	400	400
Number of finite volumes	6747	6747	13867
Number of ordinates	8	48	80

Table 3.1: Parameters used in three different problems

set-up). In the second problem, which uses the same overall geometry, we reconstruct the spatial distribution of the scattering coefficient in a moderately anisotropic medium. In these two problems all sources and detectors are located on the circle defined by $\Gamma := \{(x, y, z) : x^2 + y^2 = 1, z = 1\}$. Finally, we show an example in which we simultaneously reconstruct the absorption and scattering coefficients in a highly anisotropic medium (Fig. 3-10 and Tab. 3.1). In this case, all sources and detectors are located on two circles defined by $\Gamma_1 = \{(x, y, z) : x^2 + y^2 = 1.5^2, z = 2.2\}$ and $\Gamma_2 = \{(x, y, z) : x^2 + y^2 = 1.5^2, z = 3.5\}$. Similar measurement configurations are commonly used [184], for example, for optical measurements in finger joints, or small animal studies. For the discrete ordinate method, we use the level symmetric discrete ordinate arrangement and the corresponding weight set from [113]. All reconstructions were performed on a Linux workstation with a 700 MHz Pentium XEON processor.

Synthetic measurements

As measurements we use synthetic data, which are obtained by implementing a forward solver to the problem with exact optical properties. An algorithm for the forward problem solution in frequency domain is presented in [146]; see also chapter 2. To avoid a cancellation of numerical errors when the same model is used for the inverse problem and for generation of the synthetic data, which is often referred to as an

“inverse crime” [54], these data were generated using a much finer mesh (64280 finite volumes) and 48 discrete ordinates. Along with the “exact” synthetic measurements we consider synthetic measurements with added interval Gaussian noise. If $z_{q,d}$ is a synthetic measurement corresponding to the source q and detector d , then instead of $z_{q,d}$ we use $z_{q,d}^\nu = z_{q,d} + \nu_{q,d}\mathcal{N}(0, 1)I_{[-1,1]}$, where $\mathcal{N}(0, 1)$ is a standard normal distribution, and $I_{[-1,1]}$ is an interval indicator function that is equal to 1 on the interval $[-1, 1]$ and zero otherwise. Parameter $\nu_{q,d}$ is the standard deviation of the added noise. The signal-to-noise ratio χ is then defined as

$$\chi = 10 \log_{10} \frac{z_{q,d}}{\nu_{q,d}} \quad (3.17)$$

In our experiments, the signal-to-noise ratio χ is the same for all sourcedetector pairs and is equal to 20 dB or 15 dB, which are typical values for optical tomography systems [107].

The initial guess

In all three problems considered in the paper we use homogeneous distributions as initial guesses of σ_a and σ_s for the absorption and scattering coefficients. Solution of the forward problem with these optical properties provides us with the initial guess for the radiance u . We solve the forward problems with different sources only once to ensure that the ALM starts from the point in the space of “optical properties–radiance” pairs that belongs to the set \mathbf{C} (see Fig. 3-1).

The stopping criterion

The stopping criterion for the algorithm is given by

$$|E(\hat{\mathbf{U}}^{k+1}) - E(\hat{\mathbf{U}}^k)| < \varepsilon,$$

with $\varepsilon = 10^{-6}$ and the error function $E(\widehat{\mathbf{U}})$ being the first term in (3.7), in other words,

$$E(\widehat{\mathbf{U}}) = \frac{1}{2} \sum_{q=1}^{N_q} \sum_{d=1}^{N^d} \frac{|P_d \mathbf{U}_q - z_{q,d}|^2}{|z_{q,d}|^2}.$$

Therefore the algorithm is stopped when the error function $E(\widehat{\mathbf{U}})$ does not decrease anymore.

Reconstruction error measurement

To evaluate the quality of reconstruction we use the correlation coefficient $\rho(x, y)$ and the deviation factor $\delta(x, y)$ between two vectors that are defined as

$$\rho(x, y) = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{(N-1)Var(x)Var(y)}, \quad \delta(x, y) = \frac{\sqrt{\sum_{i=1}^N (x_i - y_i)^2 / N}}{Var(y)}, \quad (3.18)$$

where \bar{x} (\bar{y}) denotes mean value of x (y) and $Var(x)$ ($Var(y)$) denotes variance of x (y).

We will compute the correlation coefficient and deviation factor between reconstructed (with a superscript r) and exact (with a superscript e) optical properties. The larger the correlation coefficient the closer the reconstructed quantities resemble the exact ones. If the reconstructed quantities are identical to the exact ones, then their correlation coefficient is 1. The smaller δ the smaller the absolute difference between reconstructed and exact quantities are. In the ideal case $\delta = 0$. Note that, since it is possible to obtain reasonably good reconstructions only in the vicinity of the planes where sources and detectors are located (Fig. 3-2(a)), we compute ρ and δ only in this plane.

3.3.2 Reconstruction of absorption coefficients

The constrained and unconstrained algorithms were used to perform reconstructions for various combinations of different noise levels and regularization parameters. The results with respect to the correlation coefficient, deviation factor, CPU time, num-

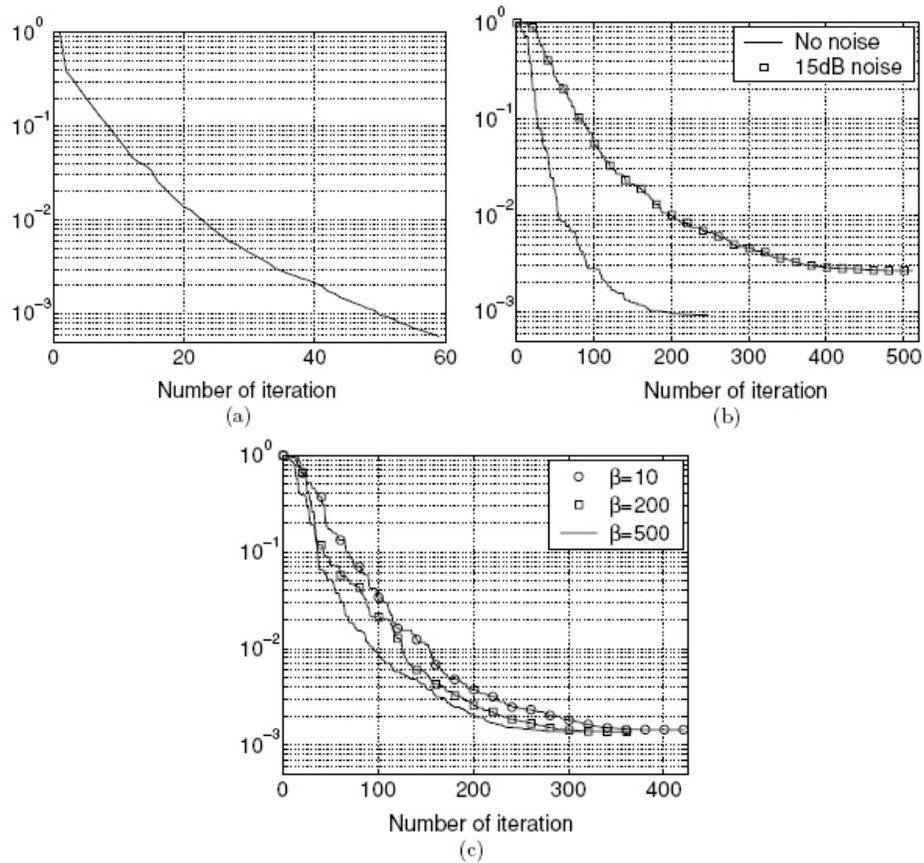


Figure 3-3: Convergence history of $E(\hat{\mathbf{U}}^k)/E(\hat{\mathbf{U}}^0)$ for σ_a reconstruction (in \log_{10} scale). (a) The lm-BFGS unconstrained optimization method with no noise. (b) The augmented Lagrangian method, $\chi = \infty$ (no noise), and $\chi = 15$ dB; (c) The augmented Lagrangian method, $\chi = 20$ dB and different regularization parameters. All the values of β are given in units of $[10^{-10}]$.

Method	ALM	ALM	ALM	ALM	ALM	lm-BFGS
Signal to noise ration χ (dB)	15	20	20	20	∞	∞
Regularization parameter β	200	10	200	500	200	-
Iteration to convergence	512	422	367	315	248	60
CPU time (h)	9.1	7.9	6.4	5.5	3.9	103.2
Correlation $\rho(\sigma_a^e, \sigma_a^r)$	0.63	0.53	0.68	0.71	0.76	0.79
Deviation $\delta(\sigma_a^e, \sigma_a^r)$	0.85	0.97	0.79	0.81	0.69	0.64

Table 3.2: Quality of reconstruction of the absorption coefficient for different reconstruction methods, different noise levels and different regularization parameters. The parameter β is given in unit of $[10^{-10}]$.

ber of iterations to convergence are summarized in Tab. 3.2, and the corresponding convergence history (the error function $E(\widehat{\mathbf{U}}^k)$ versus the iteration number k) of the iterative methods is shown in Fig. 3-3.

Comparison of constrained and unconstrained optimization

We first compare the performance of our newly developed constrained optimization code with the unconstrained optimization code developed in chapter 2. Fig. 3-4 (a) and (b) show the reconstruction of the absorber obtained with that unconstrained optimization code using noise-free data ($\chi = \infty$). It took approximately 103.2 h of CPU time to complete 60 lm-BFGS iterations (Tab. 3.2, Fig. 3-3(a)). The correlation coefficient is $\rho(\sigma_a^e, \sigma_a^r) = 0.79$ and the deviation factor is $\delta(\sigma_a^e, \sigma_a^r) = 0.64$. Fig. 3-4 (c) and (d) display the images obtained with the augmented Lagrangian method, using the same spatial and angular discretization and the same finite-volume mesh. A comparable image quality is achieved in only 3.9 h (248 ALM iterations), which is almost 27 times faster. This acceleration factor is essentially determined by the number of matrix-vector multiplications with the matrix $\mathbf{T} \equiv \mathbf{A} - \mathbf{S}$, required to complete the reconstruction. For example, for the absorption reconstruction problem considered in this section, on every ALM iteration one has to make on average 1.5×10^3 multiplications with the matrix \mathbf{T} for each source. The unconstrained optimization code makes approximately 1.5×10^5 matrix-vector multiplications for each source on every iteration of the lm-BFGS method. Since 60 lm-BFGS iterations are needed to get Fig. 3-4 (a) and (b), while 248 ALM iterations are required for Fig. 3-4 (c) and (d), the overall

number of matrix-vector multiplications used by the ALM code is approximately 24 times less than in the unconstrained minimization code, which accords well with the observed CPU time differences. We also explored if relaxing the accuracy requirements for the forward and adjoint solutions in the unconstrained optimization code can yield similar acceleration. We observed a moderate (up to 25%) acceleration of the reconstruction in the cases where appropriate accuracy requirements were chosen. Here an appropriate accuracy requirement means a stopping criterion for the forward GMRES solver such that it still ensures convergence of the reconstruction algorithm. For example, we usually stop the GMRES solver when the relative residual is smaller than 10^{-13} . If we stop the GMRES when the relative residual is smaller than 10^{-7} in the first several quasi-Newton steps, the reconstruction still converges to very similar results, and the computational time is reduced by approximately 25%. However, if we further relax the accuracy requirements our minimization algorithm no longer converges to the right solution. The reason for this, in our opinion, is because we use the adjoint differentiation method to compute the gradient of the objective function with respect to optical parameters. This gradient is calculated using the solution of the forward and adjoint problems. In the cases where the forward and adjoint solutions are not accurate enough, the approximate gradient can be computed only approximately which results in poor convergence of the minimization algorithm, or no convergence at all.

Note that the reconstructions of the absorption coefficient obtained by different methods (ALM versus lm-BFGS) do not coincide completely. But this fact is expected and can be easily explained. Indeed, ALM and lm-BFGS generate iterations that do not have to be the same (see Fig. 3-1), but converge to each other only in the limit as $k \rightarrow \infty$. In the lm-BFGS method for the unconstrained optimization the intensity \mathbf{U}_q^k is a solution of the forward problem (the radiative transfer equation) with $\sigma_a = \sigma_a^k$, on each iteration k , by construction, whereas this property does not have to hold for the augmented Lagrangian method.

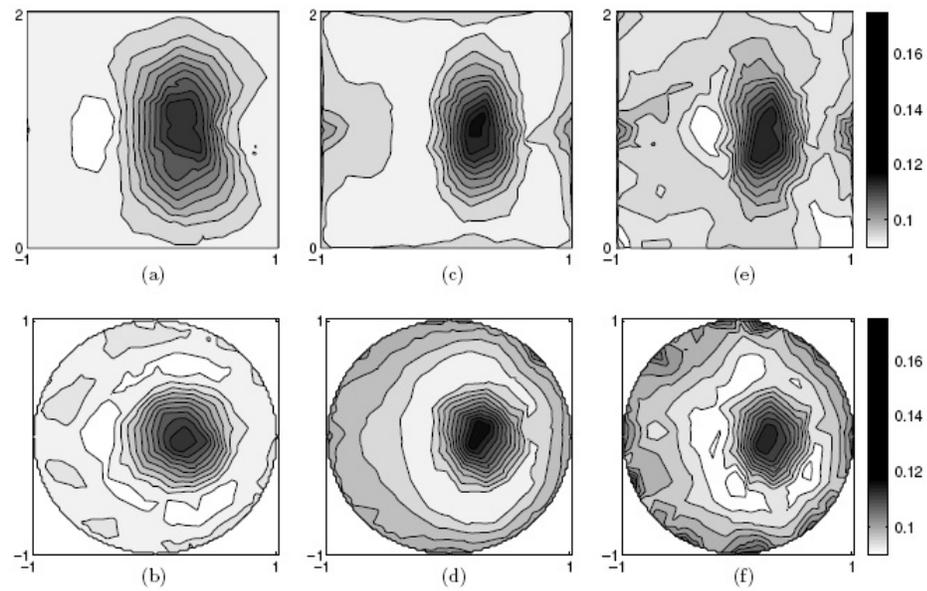


Figure 3-4: Cross sections of the reconstructed absorption coefficient in the plane $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the quasi-Newton lm-BFGS method for the unconstrained optimization and the ALM for problem 1 with different noise levels. The target optical properties are $\sigma_a = 0.2 \text{ cm}^{-1}$ in the inclusion and $\sigma_a = 0.1 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to the reconstruction with unconstrained minimization approach; (c) and (d) correspond to the ALM reconstruction with noise free data; (e) and (f) correspond to the ALM reconstruction with 15 dB added noise.

Impact of noise level

Next we test the impact of noise in the synthetic measurement data on the reconstruction results. An example of a reconstruction performed with data with a signal-to-noise ratio $\chi = 15$ dB is shown in Fig. 3-4 (e) and (f) next to the images obtained with noise-free data (Fig. 3-4 (c) and (d)). As expected the quality of reconstruction decreases with a decreased signal-to-noise ratio. The correlation coefficient drops from 0.76 in the case of noise-free data to 0.63, while the deviation factor increases from 0.69 to 0.85 (see Tab. 3.2). Note that as the noise level increases (and χ decreases), the number of necessary ALM iterations and CPU time increases. The results for $\chi = 20$ dB (the fourth column in Tab. 3.2) are in between the values for 15 dB and noise-free data ($\chi = \infty$). However, even for $\chi = 25$ dB which is a very reasonable noise level for currently available instrumentation, the 9.1 *h* CPU time is still over 11 times faster than the time it takes to complete unconstrained optimization with noise-free data. Only if the signal-to-noise ratio drops below 10 dB, does the augmented Lagrangian method fail to converge to any solution.

Impact of the regularization parameter

In the next series of numerical experiments we test the performance of the ALM for different values of the regularization parameter β . In these experiments 20 dB Gaussian noise is added and the parameter β is equal to 10, 200 and 500×10^{-10} . As we can see, when the regularization parameter is small (Fig. 3-5 (a) and (b) and Tab. 3.2), the impact of noise is more significant, whereas for a large regularization parameter the absolute value of the absorption coefficient in the center of the inhomogeneity is found less accurately (Fig. 3-5 (e) and (f)). This is reflected in the deviation factor $\delta(\sigma_a^e, \sigma_a^r)$ which first drops from 0.97 (in the case of a small value of the regularization parameter $\beta = 10 \times 10^{-1}$) to 0.79 ($\beta = 200 \times 10^{-10}$), but then increases slightly to 0.81 for $\beta = 500 \times 10^{-10}$ (see columns 3-5 in Tab. 3.2). This suggests that there is an optimal value of β between 10 and 500×10^{-10} , for which $\delta(\sigma_a^e, \sigma_a^r)$ becomes smallest. On the other hand, the correlation factor $\rho(\sigma_a^e, \sigma_a^r)$ keeps increasing (improving) as

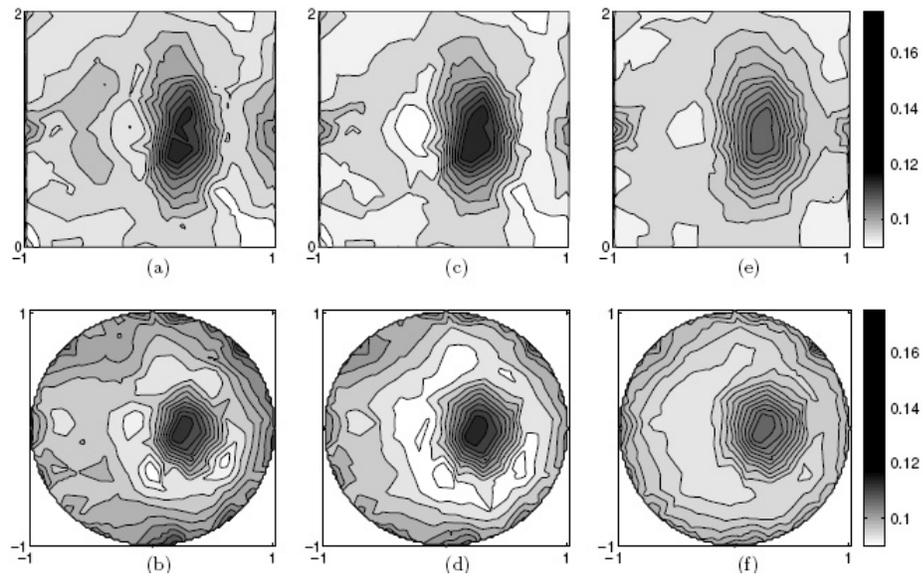


Figure 3-5: Cross sections of the reconstructed absorption coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the ALM for problem 1 with different regularization parameters. The target optical properties are $\sigma_a = 0.2 \text{ cm}^{-1}$ in the inclusion and $\sigma_a = 0.1 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to the ALM reconstruction with $\beta = 10 \times 10^{-10}$; (c) and (d) correspond to the ALM reconstruction with $\beta = 200 \times 10^{-10}$; (e) and (f) correspond to the ALM reconstruction with $\beta = 500 \times 10^{-10}$.

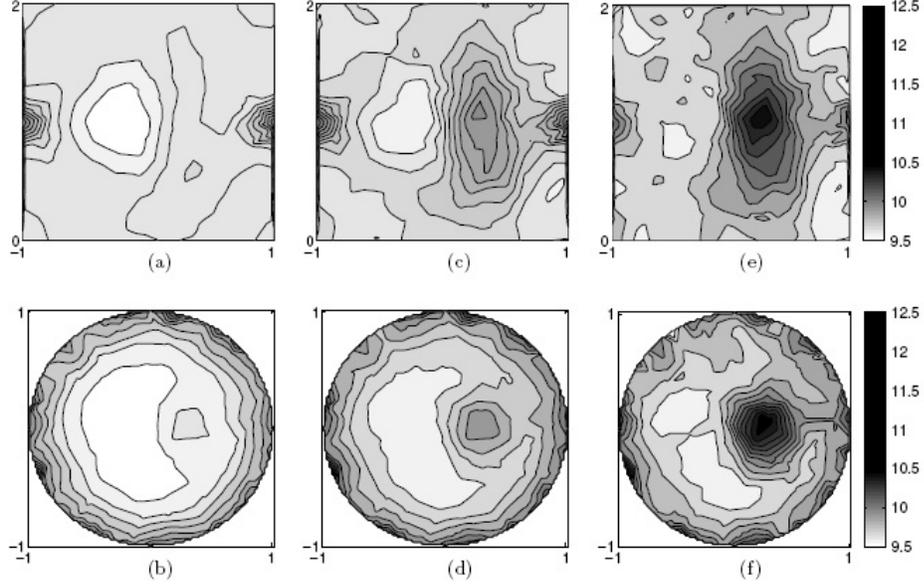


Figure 3-6: Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the augmented Lagrangian method for problem 2. The target optical properties are $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to the reconstruction after 50 iterations of the ALM; (c) and (d) correspond to the reconstruction after 200 iterations of the ALM; (e) and (f) correspond to the reconstruction at convergence (498 iterations).

β increases. It is notable that for a rather large range of values of β “reasonable” reconstruction results can be obtained. In the cases considered in this study, values of β up to 2000×10^{-10} allowed perturbation to be located, even though with incorrect absolute values. It appears that for each particular application (e.g., breast, brain or joint imaging), some preliminary studies should be performed to find optimal regularization parameters.

3.3.3 Reconstruction of scattering coefficients

In the second example, we show the reconstruction of the scattering coefficient with an anisotropic factor $g = 0.5$. We use synthetic measurements with $\chi = 20 \text{ dB}$ and $\beta = 500 \times 10^{-10}$. A large number of discrete ordinates are necessary to capture anisotropy effects as compared to the case of problem 1 where isotropic scattering is assumed. This leads to a total number of 5181696 constraints in the constrained

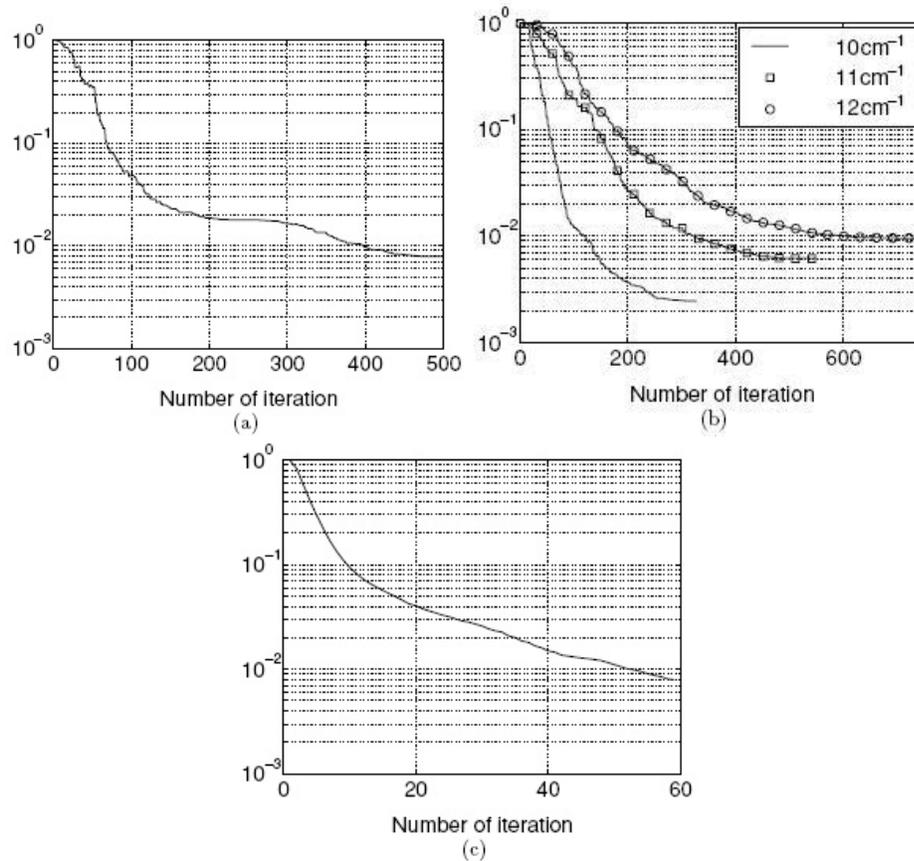


Figure 3-7: Convergence history of $E(\hat{\mathbf{U}}^k)/E(\hat{\mathbf{U}}^0)$ for σ_s reconstruction (in \log_{10} scale). (a) The augmented Lagrangian method, $\chi = 20$ dB and $\beta = 500 \times 10^{-10}$ dB; (b) The augmented Lagrangian method with different initial guesses, $\chi = \infty$ dB and $\beta = 300 \times 10^{-10}$; (c) the lm-BFGS unconstrained optimization method with no noise.

optimization formulation. The results of the reconstruction are presented in Fig. 3-6 and the convergence history is given in Fig. 3-7 (a) and Tab. 3.3. We observe that the reconstruction algorithm converges slower than for σ_a , which is consistent with the results presented in chapter 2. Note that the error function $E(\hat{\mathbf{U}}^k)$ can actually increase on some iterations. This can be best explained by the fact that the ALM tries to minimize the augmented Lagrangian function $\mathcal{L}_K(\Sigma, \hat{\mathbf{U}}; \hat{\boldsymbol{\lambda}})$, not only the error function $E(\hat{\mathbf{U}}^k)$. The acceleration rate we obtained as compared to the unconstrained optimization method is approximately 18.

Iteration number	50	200	498
Signal to noise ratio χ (dB)	20	20	20
Correlation $\rho(\sigma_s^e, \sigma_s^r)$	-0.12	0.42	0.67
Deviation $\delta(\sigma_s^e, \sigma_s^r)$	1.12	0.97	0.81

Table 3.3: Quality of reconstruction of the scattering coefficient as a function of ALM iteration step.

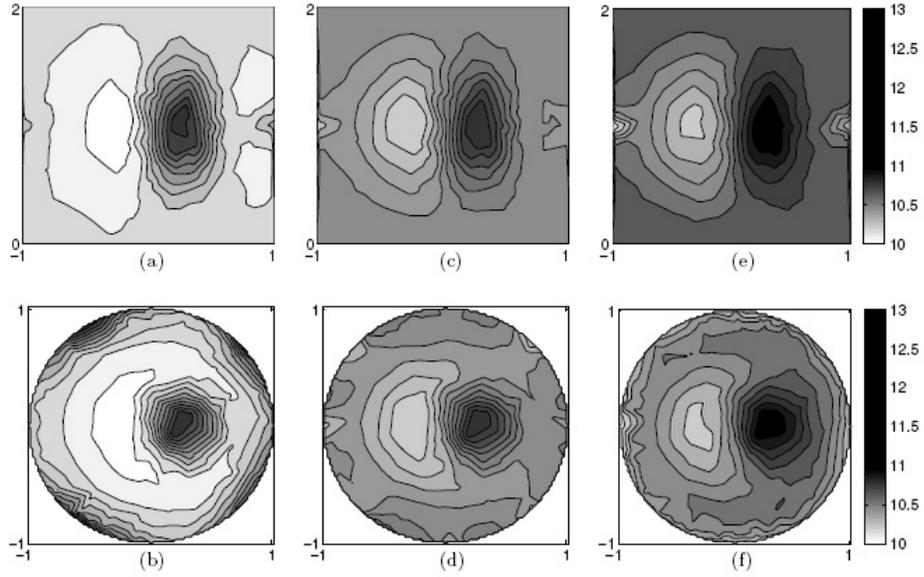


Figure 3-8: Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the augmented Lagrangian method for problem 2 with different initial guesses. The target optical properties are $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to initial guess $\sigma_s^0 = 10 \text{ cm}^{-1}$; (c) and (d) correspond to initial guess $\sigma_s^0 = 11 \text{ cm}^{-1}$; (e) and (f) correspond to initial guess $\sigma_s^0 = 12 \text{ cm}^{-1}$.

Initial guess σ_s^0 (cm^{-1})	10	11	12
Number of ALM iteration to converge	327	549	741
Correlation $\rho(\sigma_s^e, \sigma_s^r)$	0.69	0.66	0.62
Deviation $\delta(\sigma_s^e, \sigma_s^r)$	0.76	1.20	1.34

Table 3.4: Quality of reconstruction of the scattering coefficient as a function of the initial guess.

Impact of the initial guess

We use reconstructions of the scattering coefficient to illustrate convergence of the ALM from different initial guesses σ_s^0 ($\chi = \infty$). In the first example, the initial guess for the scattering coefficient is equal to the background value $\sigma_s^0 = 10 \text{ cm}^{-1}$, in the second it is 10% higher $\sigma_s^0 = 11 \text{ cm}^{-1}$, and in the third it is 20% higher than the background value, $\sigma_s^0 = 12 \text{ cm}^{-1}$. The results of reconstruction are presented in Fig. 3-8 and Tab. 3.4. It can be seen that the correlation factor depends weakly on the initial guess. It only drops by approximately 10% from 0.69 to 0.62, when the initial guess is increased from 10 to 12 cm^{-1} . This just expresses the fact that the location and shape of the inhomogeneity is still rather accurately reconstructed. The absolute values of the optical properties in the image are strongly affected by the initial guess, which is obvious in the images as well as in the values of the deviation factor $\delta(\sigma_s^e, \sigma_s^r)$ which increases from 0.74 to 1.34 when the initial guess is $\sigma_s^0 = 12 \text{ cm}^{-1}$ instead of $\sigma_s^0 = 10 \text{ cm}^{-1}$. Similar results were observed for the case of absorption reconstruction. Overall we found that the ALM will produce images with $\rho(\sigma_s^e, \sigma_s^r) > 0.5$ as long as the initial guess is not more than 40% higher than the actual background value.

Impact of the mesh size

Finally, we compare the reconstructions obtained with different finite-volume meshes, using noise-free synthetic measurement data. Images are displayed in Fig. 3-9, while the corresponding $\rho(\sigma_s^e, \sigma_s^r)$ and $\delta(\sigma_s^e, \sigma_s^r)$ are shown in Tab. 3.5. The images show almost no visible differences. Indeed, refining the mesh from 6747 tetrahedrons to 15615 leads only to slight improvements in the correlation coefficient $\rho(\sigma_s^e, \sigma_s^r)$ (from 0.69 to 0.75) and the deviation factor $\delta(\sigma_s^e, \sigma_s^r)$ (from 0.76 to 0.73). Further mesh refinement does not improve these image quality parameters. This shows that when the mesh is fine enough so that the average distance between the mesh points is of the order of the scattering mean free path ($1/\sigma_s$), the discretization is sufficient to capture most pertinent effects.

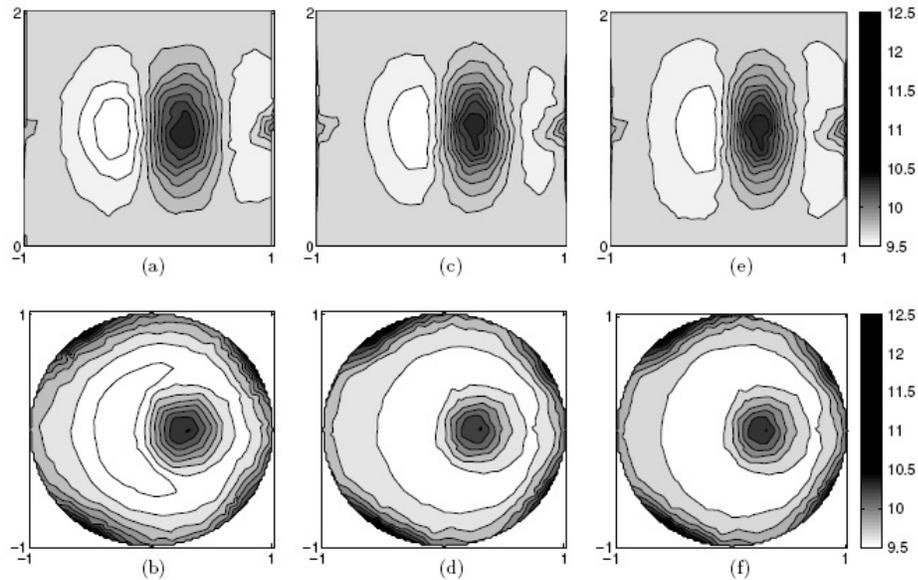


Figure 3-9: Cross sections of the reconstructed scattering coefficient in the planes $y = 0$ ((a), (c) and (e)) and $z = 1$ ((b), (d) and (f)) with the augmented Lagrangian method for problem 2 with different meshes. The target optical properties are $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) and (b) correspond to mesh with 10062 tetrahedrons; (c) and (d) correspond to mesh with 15612 tetrahedrons; (e) and (f) correspond to mesh with 19489 tetrahedrons.

Number of finite volumes	6747	10062	15612	19489
Correlation $\rho(\sigma_s^e, \sigma_s^r)$	0.69	0.72	0.75	0.75
Deviation $\delta(\sigma_s^e, \sigma_s^r)$	0.76	0.75	0.73	0.74

Table 3.5: Quality of reconstruction of the scattering coefficient as a function of the mesh size.

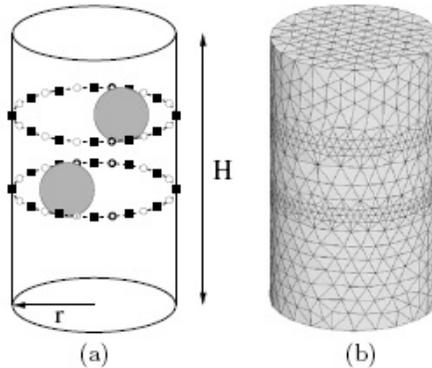


Figure 3-10: Test problem 3 setup. Cylinder height: $H = 5$ cm, radius $r = 1.5$ cm; radius of the embedded small cylinder $r = 0.5$ cm. (a) source-detector layout with 24 sources (\square), 24 detectors (\circ); (b) finite-volume mesh with 13867 tetrahedrons.

Iteration number	50	300	712
Correlation $\rho(\sigma_a^e, \sigma_a^r)$	0.33	0.20	0.29
Deviation $\delta(\sigma_a^e, \sigma_a^r)$	1.14	0.99	0.96
Correlation $\rho(\sigma_s^e, \sigma_s^r)$	0.32	0.07	0.39
Deviation $\delta(\sigma_s^e, \sigma_s^r)$	1.20	1.01	0.95

Table 3.6: Quality of reconstruction of the absorption and scattering coefficients as a function of the ALM iteration step.

3.3.4 Simultaneous reconstruction of two coefficients

As a last example we present a simultaneous reconstruction of both the absorption and scattering coefficients in highly anisotropic scattering media, with $g = 0.9$. In this case, we set the regularization parameter β to 700×10^{-10} . The synthetic measurements for this problem were generated using 80 ordinates and a finite-volume mesh with 23793 tetrahedrons and 20 dB noise was added to the measurements.

Target optical properties represent media with a relatively high absorption and a low scattering, for which the diffusion theory fails to describe the light propagation accurately. The cross sections of the reconstructed absorption and scattering coefficients are presented in Fig. 3-11, and the corresponding values of $\rho(\sigma_a^e, \sigma_a^r)$, $\rho(\sigma_s^e, \sigma_s^r)$ and $\delta(\sigma_a^e, \sigma_a^r)$, $\delta(\sigma_s^e, \sigma_s^r)$ are given in Tab. 3.6. As for problems 1 and 2, the parameters $\rho(\sigma^e, \sigma^r)$ and $\delta(\sigma^e, \sigma^r)$ are computed only in the planes $z = 2.2$ and $z = 3.5$, in which sources and detectors are located.

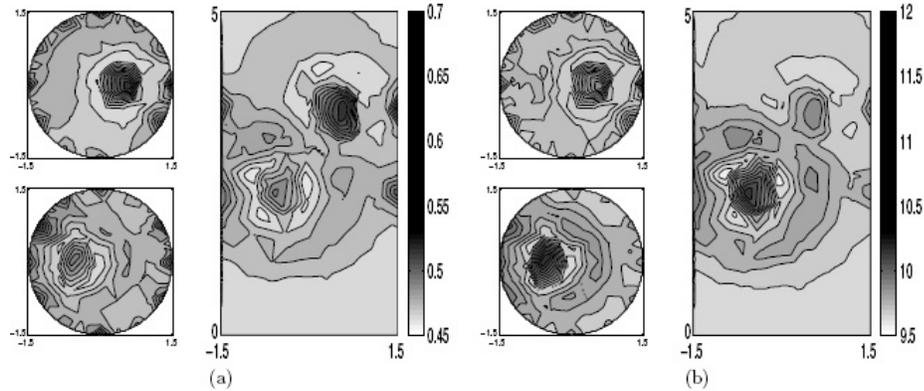


Figure 3-11: Cross sections of the reconstructed absorption and scattering coefficients in the planes $y = 0$, $z = 2.2$ and $z = 3.5$ with the augmented-Lagrangian method for problem 3. The target optical properties are $\sigma_a = 1.0 \text{ cm}^{-1}$, $\sigma_s = 15 \text{ cm}^{-1}$ in the inclusion and $\sigma_a = 0.5 \text{ cm}^{-1}$, $\sigma_s = 10 \text{ cm}^{-1}$ in the background. (a) Reconstruction of σ_a at convergence (712 iterations), left-top: cross section $z = 3.5$, left-bottom: cross section $z = 2.2$, right: cross section $y = 0$. (b) Reconstruction of σ_s at convergence (712 iterations).

3.4 Conclusions and remarks

We have implemented a PDE-constrained optimization method that solves inverse problems of optical tomography using the frequency-domain radiative transport equation as the forward model. By simultaneously updating both radiance and optical properties, the method solves the forward and inverse problems in optical tomography all at once. In this way, the computing time is greatly reduced as compared to traditional unconstrained optimization methods, during which one has to repeatedly solve the forward problem many times. We tested and quantified the performance of the algorithm for various combinations of mesh sizes, noise, regularization parameters, initial guesses, optical properties and measurement geometries. Besides the speed of the code, we compared image qualities by defining a correlation coefficient as well as a deviation factor. In the cases that involve image reconstruction from synthetic measurement data we observe 10– to 30–fold decrease in computing time for the constrained optimization code compared to the unconstrained optimization code. The regularization parameter β has some influence on the computing time, but with reasonable values of β which in our case are of order $10 - 10^3 \times 10^{-10}$, the com-

putational time changes less than 20%. In general, reconstruction of both absorption and scattering together took longer than reconstructions of only the scattering coefficient or only the absorption coefficient. As expected the correlation coefficients and deviation factors worsen as the signal-to-noise ratio decreases. Similarly the deviation factor decreases substantially as the (homogeneous) initial guess is not chosen close to the optical properties of the actual background medium. Interestingly the correlation coefficients is only weakly affected by the initial guess. As long as the optical properties are chosen within 50% of the actual background medium the correlation coefficient changes by only 10-20%. Finally the correlation coefficient and deviation factor do not change once the mesh is fine enough so that the average size of finite volumes becomes less than the average scattering mean free path ($1/\sigma_s$).

Another positive aspect of the augmented Lagrangian method is that it maintains storage requirements that are comparable to requirements encountered in unconstrained optimization methods. The augmented Lagrangian also provides the flexibility of being easily implemented on parallel processors.

Finally, it should be noted that the constrained optimization method introduced in this work does not require that the forward model is the frequency-domain ERT. For example, one can also implement similar codes with the diffusion equation as the forward model of light propagation in tissues or with a system of two ERTs as the forward model as in the case of fluorescence tomography.

Chapter 4

Inverse transport problem in atmospheric remote sensing

In the previous two chapters, we have focused on the development of computational tools to study some of the inverse transport problems in optical tomography applications. In this chapter, we will conduct some theoretical analysis on an inverse transport problem. We will focus on a problem that arises in atmospheric remote sensing. We show that such an inverse problem is an severely ill-posed problem. Special attention has to be paid on inversion algorithms that are designed to do numerical reconstructions just as the case in optical tomography. The presentation of this chapter is based on reference [24].

4.1 Problem statement

The vertical concentrations of atmospheric gases such as carbon monoxide (CO), carbon dioxide (CO₂), and ozone (O₃) play a central role in the energy balance between atmospheric absorption and emission [78, 116, 125], and thus have a strong impact on global climate changes, air quality, and various other meteorological processes. In recent years, spectro-radiometers in Fourier Transform Infrared Spectroscopy (FTIS) have been widely used to monitor the concentration of atmospheric gases. An example of such spectrometer, the Tropospheric Emission Spectrometer (TES) installed on the EOS-Aura spacecraft, will soon measure global three-dimensional distributions of ozone and other gases in the troposphere [28] with unprecedented accuracy. In its

nadir mode, TES will record the spectral radiance from the Earth's atmosphere in the form of line integrals with respect to altitude z . Such measurements can be used to recover the vertical concentration profile of atmospheric gases.

Mathematically, the problem can be formulated as a one-dimensional inverse source problem of a scattering-free transport equation aiming at reconstructing the altitude-dependent gas distribution profiles [38, 53, 94] from wavenumber-dependent boundary radiation measurements. While a lot of work has been done on developing numerical algorithms to address the linear inverse problem [53, 61, 168, 150], comparatively little is known in the literature on more mathematical questions such as uniqueness and stability of the reconstruction. The first part of the chapter addresses this issue.

Under suitable separability assumptions on the absorption coefficients in the transport equation, we show that the gas concentrations can indeed uniquely be determined by radiation measurements—a theoretical underpinning for the reconstruction algorithms we were not able to find in the existing literature—and give an explicit reconstruction procedure. Moreover, we stress that the reconstruction involves the inversion of a Laplace transform, which is known to be a severely ill-posed problem [31, 67]. As a consequence, a somewhat limited amount of information on the profiles can be retrieved from the radiation measurements. Such limitations need to be incorporated in realistic reconstruction methods.

An important objective of the radiation measurements is the detection of relatively thin (compared to the size of the atmosphere) layers such as ozone or dust layers in the Earth's lower atmosphere (the troposphere). Such layers have an important impact on local climate changes and global warming effects to cite a few. Because of the severely ill-posed nature of the inversion problem, such thin layers must be modeled specifically in the inverse problem if they are to be detected. We propose in this paper to *model* such structures as thin inclusions with arbitrary (i.e., not necessarily small) concentration contrast. We perform asymptotic expansions in the thickness of the inclusions to characterize their main impact on the boundary measurements. The technique follows general principles that have been used successfully in many

other fields [9, 19, 44]. The results of the analysis are the following. The location of the inclusions and the product of their thickness with their concentration variations (with respect to the underlying medium assumed to be *known*) can be reconstructed from moderately noisy data. Obtaining more on the inclusion, i.e., both its thickness and its concentration, requires much more accurate data. This provides us with some guidelines in our aim to understand what can versus what cannot be reconstructed from measurements with a given noise level.

Let us note that the *nadir* measurements represent only one modality of TES. Measurements involving directions of incidence other than the vertical one are available and could be incorporated into the model to improve its stability properties. Although this is an important problem, it is not considered further here.

4.2 The mathematical model

We denote by $L(z, \nu)$ the radiation intensity of atmospheric gases at altitude $z \in \mathcal{Z} = (0, Z)$, where Z is the altitude at the “top” of the atmosphere, and wavenumber $\nu \in \mathcal{N} = [\nu_{min}, \nu_{max}]$, where ν_{min} and ν_{max} are the minimum and maximum wavenumbers accessible in real measurements. The radiation source term at the Earth surface is $L(z = 0, \nu)$. The volume source term of radiation is $a(z, \nu)B(z, \nu)$, where $a(z, \nu)$ is the absorption profile of a specific gas in the atmosphere and $B(z, \nu)$ is the Planck function of black-body radiation. The measurements $L(Z, \nu)$ are the radiation intensity on top of the atmosphere $z = Z$. Typically, measurements are available in the wavenumber range of 650 to 2250 cm^{-1} (which corresponds to wavelengths of 15.4 and 4.4 μm , respectively). Thus, ν is in the middle of the thermal infrared region (IR). From the atmospheric radiative transfer theory [78, 116], the transport equation satisfied by $L(z, \nu)$ is:

$$\begin{cases} \frac{\partial L(z, \nu)}{\partial z} + a(z, \nu)L(z, \nu) = a(z, \nu)B(z, \nu), & (z, \nu) \in \mathcal{Z} \times \mathcal{N}, \\ L(0, \nu) = L_0(\nu), & \nu \in \mathcal{N}. \end{cases} \quad (4.1)$$

We assume that $a(z, \nu)$ and $B(z, \nu)$ are positive functions of class $C^0(\mathcal{Z} \times \mathcal{N})$ and $C^1(\mathcal{Z} \times \mathcal{N})$, respectively, and that $L_0(\nu)$ is a positive function of class $C^0(\mathcal{N})$. As usual C^0 is the class of continuous functions and C^1 the class of continuously differentiable functions. The solution $L(z, \nu)$ is then a positive function of class $C^1(\mathcal{Z} \times \mathcal{N})$ [73].

The Planck function $B(z, \nu)$ is given by

$$B(z, \nu) = \frac{2h\nu^3}{c^2(e^{h\nu/kT(z)} - 1)}, \quad (4.2)$$

where h is Planck constant, ν is the wavenumber, k is the Boltzmann constant and c is the speed of light in a vacuum. The temperature profile $T(z)$, assumed here to be of class $C^1(\mathcal{Z})$, is given in Kelvin degrees and is thus always positive. Notice however that $T'(z)$ changes sign on \mathcal{Z} in practice. This will be important in the reconstruction theory. Scattering has been neglected in (4.1), which is an accurate assumption in the “clear sky” environment.

The radiation intensity at the Earth surface is related to the Planck constant of black-body radiation by

$$L_0(\nu) = \varepsilon(\nu)B(0, \nu) \quad (4.3)$$

where $\varepsilon(\nu)$ is the surface emissivity, which we may assume is constant at the Earth’s surface $\varepsilon(\nu) = \varepsilon$ [116].

It is more convenient to work in the sequel with the following quantity

$$H(z, \nu) = L(z, \nu) - B(z, \nu), \quad (4.4)$$

modeling the departure from the black-body radiation equilibrium. One can verify that the equation for $H(z, \nu)$, also of class $C^1(\mathcal{Z} \times \mathcal{N})$, takes the form

$$\begin{cases} \frac{\partial H(z, \nu)}{\partial z} + a(z, \nu)H(z, \nu) = -\frac{\partial B(z, \nu)}{\partial z} \equiv S(z, \nu), & (z, \nu) \in \mathcal{Z} \times \mathcal{N}, \\ H(0, \nu) = \gamma B(0, \nu), & \nu \in \mathcal{N}, \end{cases} \quad (4.5)$$

where $\gamma = \varepsilon - 1$. Upon inverting this first-order ordinary differential equation, we get

$$H(Z, \nu) = H(0, \nu)e^{-\int_0^Z a(\zeta, \nu)d\zeta} + \int_0^Z S(z, \nu)e^{-\int_z^Z a(\zeta, \nu)d\zeta} dz, \quad \nu \in \mathcal{N}. \quad (4.6)$$

Let us define the optical length

$$\alpha(z, \nu) = \int_z^Z a(\zeta, \nu)d\zeta. \quad (4.7)$$

We may then recast the above integral (4.6) as

$$H(Z, \nu) = H(0, \nu)e^{-\alpha(0, \nu)} + \int_0^Z S(z, \nu)e^{-\alpha(z, \nu)} dz, \quad \nu \in \mathcal{N}. \quad (4.8)$$

This is the integral formulation, equivalent to the differential equation (4.1), as it appears in most of the atmospheric inversion literature.

4.3 Uniqueness and ill-posedness of a simplified model

We know that the absorption profile $a(z, \nu)$ depends on both the concentration of atmosphere gases and their absorption properties at specific wavenumbers. To simplify the presentation, we assume in this section that only one gas, such as ozone, contributes to absorption and emission. We then have that

$$a(z, \nu) = c(z)\kappa(z, \nu), \quad (4.9)$$

where $c(z) \in C^0(\mathcal{Z})$ is the unknown (non-negative) concentration profile for the gas and $\kappa(z, \nu) \in C^0(\mathcal{Z} \times \mathcal{N})$ is the (positive) spectral absorption/emission coefficient (also called spectral line shape). Although more complicated models of $\kappa(z, \nu)$ can be considered, we focus here on the so-called Lorentzian line shape of $\kappa(z, \nu)$. It is

valid in the lower atmosphere and takes the form

$$\kappa(z, \nu) \equiv \kappa_L(z, \nu) = Q_{\nu_0} \frac{1}{\pi} \frac{\alpha_L(z)}{(\nu - \nu_0)^2 + \alpha_L^2(z)} \equiv Q_{\nu_0} f(z, \nu - \nu_0), \quad (4.10)$$

where ν_0 is the center of the band of wavenumber we are interested in and $Q_{\nu_0} = \int_{\mathcal{N}} \kappa(z, \nu) d\nu$ is the line strength. The function $\alpha_L \in C^0(\mathcal{Z})$ is called the Lorentz half-width. It is roughly given by

$$\alpha_L(z) \sim T(z)^{-1/2}, \quad z \in \mathcal{Z}, \quad (4.11)$$

where $T(z)$ is the temperature profile of the atmosphere and $f(z, \nu - \nu_0) \in C^0(\mathcal{Z} \times \mathcal{N})$ is the shape factor of a spectral line [78]. The Lorentzian line shape describes how a gas absorbs and emits radiance in a narrow band of wavenumbers centered at ν_0 .

The inverse (retrieval) problem in atmosphere imaging is to assume that the radiation term $B(z, \nu)$ in (4.1) and absorption coefficient $\kappa(z, \nu)$ are known and to reconstruct as much as possible about $c(z)$ from radiation measurements $L(Z, \nu) = H(Z, \nu) + B(Z, \nu)$.

4.3.1 The case of a single gas

The purpose of this section is to show that the reconstruction of $c(z)$ from $L(Z, \nu)$ is uniquely determined (in a slightly simplified setting) and is a *severely ill-posed* problem (see [67]) in the sense that, in the absence of regularization, noise in the data is more amplified during the inversion procedure than what would result from an arbitrary number of differentiations.

We do not have a complete theory for general absorption coefficient $\kappa(z, \nu)$. Rather we make the following *assumption* on the shape factor $f(z, \nu - \nu_0)$:

$$f(z, \nu - \nu_0) = \mu(\nu - \nu_0)g(z), \quad (4.12)$$

where $g(z) \in C^0(\mathcal{Z})$ is uniformly bounded from below by a positive constant and $\mu(\nu - \nu_0) \in C^0(\mathcal{N})$ is a positive function whose range $\mathcal{M} = \mu(\mathcal{N} - \nu_0)$ is an interval

in \mathbb{R}^+ . In the above expression (4.10) this would correspond to replacing κ by its approximation

$$\kappa_L(z, \nu) \approx Q_{\nu_0} \frac{1}{\pi} \frac{\alpha_L(z)}{(\nu - \nu_0)^2 + \bar{\alpha}^2}, \quad (4.13)$$

with $\mu(\nu)$ and $g(z)$ given by

$$\mu(\nu) = \frac{1}{\pi} \frac{\bar{\alpha}}{\nu^2 + \bar{\alpha}^2}, \quad (4.14)$$

and

$$g(z) = \frac{\alpha_L(z)}{\bar{\alpha}}. \quad (4.15)$$

Here $\bar{\alpha}$ is a constant. The range of μ is given by $\mathcal{M} = [\frac{\bar{\alpha}}{\pi(\bar{\nu}^2 + \bar{\alpha}^2)}, \frac{1}{\pi\bar{\alpha}}] \subset \mathbb{R}^+$ with $\bar{\nu} := \max(|\nu_{min} - \nu_0|, |\nu_{max} - \nu_0|)$.

The separability assumption on the shape factor is not totally unreasonable for the Lorentzian line shape (4.10). Following (4.11) and choosing $\bar{\alpha} \sim \bar{T}^{-1/2}$, where \bar{T} the average of $T(z)$ over \mathcal{Z} , we obtain from realistic temperature profiles that $\|\frac{\alpha_L - \bar{\alpha}}{\alpha_L}\|_{L^\infty(\mathcal{Z})} < 0.15$, which implies that at any given wavenumber ν , $\|\frac{\kappa(\cdot, \nu) - \kappa_L(\cdot, \nu)}{\kappa(\cdot, \nu)}\|_{L^\infty(\mathcal{Z})} < 0.15$, where κ and κ_L are given by (4.10) and (4.13), respectively. The maximal error is attained when $\nu = \nu_0$ and decays quite fast away from ν_0 because of the dominance of the term $(\nu - \nu_0)^2$ over $\bar{\alpha}_L^2$. Therefore, the separability assumption serves as a faithful benchmark in understanding the theoretical and numerical aspects of reconstructions based on more detailed and accurate physical models.

We also simplify the behavior of the source terms with respect to wavenumber ν and approximate the Planck function (4.2) as follows

$$B(z, \nu) \approx \frac{2k\nu^2}{c^2} T(z). \quad (4.16)$$

We verify that in the infrared (IR) region of interest, this expansion is quite accurate as $h\nu/kT$ is on the order of at most 10^{-3} in practice. The temperature $T(z)$ is also assumed to be of class $C^1(\mathcal{Z})$.

Accounting for the above simplifications and using the change of variables $\nu \rightarrow \mu$ and $H(z, \nu)c^2/(2k\nu^2) \rightarrow D(z, \mu)$, we obtain, still denoting by $a(z, \mu)$ and $\kappa(z, \mu)$ the absorption coefficients in the new variables, that $D(z, \mu)$ satisfies the following

equation,

$$\begin{cases} \frac{\partial D(z, \mu)}{\partial z} + a(z, \mu)D(z, \mu) = -\frac{\partial T(z)}{\partial z} \equiv S(z), & (z, \mu) \in \mathcal{Z} \times \mathcal{M}, \\ D(0, \mu) = \gamma T(0), & \mu \in \mathcal{M}. \end{cases} \quad (4.17)$$

After defining the *rescaled* optical length by

$$\alpha(z) = \int_z^Z Q_{\nu_0} c(\zeta) g(\zeta) d\zeta, \quad (4.18)$$

equation (4.17) can be inverted as

$$D(Z, \mu) = D(0)e^{-\mu\alpha(0)} + \int_0^Z S(z)e^{-\mu\alpha(z)} dz, \quad \mu \in \mathcal{M}. \quad (4.19)$$

Here $D(Z, \mu)$ is the measurement for $\mu \in \mathcal{M}$. The positive function $g(z)$ and the temperature profile $T(z)$ (hence $S(z)$) are *known* a priori.

The inverse problem for (4.17) is then:

(IP) Determine the positive function $c(z) \in C^0(\mathcal{Z})$ from the measurements $D_m(\mu) = D(Z, \mu)$ for $\mu \in \mathcal{M}$.

As we have already mentioned, several numerical methods have been devised for solving the above inverse problem; see [53, 61, 168] and the monograph [150] and references therein. Many techniques are based on Bayesian inversion techniques [61, 150]. In this paper we concentrate on the mathematical analysis of the continuous (non-discretized) inverse problem (IP). Our main result is the following:

Theorem 4.3.1. *Let us assume that $S(z)$ is a continuous function on \mathcal{Z} , which vanishes at a finite (possibly zero) number of points. Then there is a unique strictly positive function $c(z) \in C^0(\mathcal{Z})$ solving (IP).*

Proof. The gas concentration profile $c(z)$ and the quantity $g(z)$ in (4.15) are both positive functions on $\mathcal{Z} = (0, Z)$, so $\alpha'(z) = -Q_{\nu_0} c(z)g(z) < 0$ on \mathcal{Z} . We can then perform the change of variables $z \mapsto \alpha(z)$ and define the continuously differentiable inverse map $\alpha \mapsto z(\alpha)$. The above inequality implies that $z'(\alpha) < 0$. The transform

(4.19) may thus be recast as

$$D(Z, \mu) = D(0)e^{-\mu\alpha(0)} + \int_0^{\alpha(0)} S(z(\alpha)) \left| \frac{dz}{d\alpha} \right| e^{-\mu\alpha} d\alpha, \quad \mu \in \mathcal{M}. \quad (4.20)$$

Here we have used that $\alpha(Z) = 0$. The data $D(Z, \mu)$ is thus the *Laplace transform* of the distribution

$$h(\alpha) \equiv -z'(\alpha)S(z(\alpha)) + D(0)\delta(\alpha - \alpha(0)) \quad (4.21)$$

The above distribution has support in $[0, \alpha(0)]$. Since it is compactly supported, its Fourier transform $\hat{h}(\zeta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{i\zeta\alpha} h(\alpha) d\alpha$ is an analytic function in ζ [181]. The latter is known for values of ζ such that $\zeta = i\mu$, $\mu \in \mathcal{M}$ since then $D(Z, \mu) = \hat{h}(-i\zeta)$. It is thus sufficient to know $D(Z, \mu)$ on a set with at least one accumulation point to uniquely define $\hat{h}(\zeta)$ for all $\zeta \in \mathbb{C}$ by analytic continuation [55, 156]. This in turn uniquely defines the function $h(\alpha)$. Since \mathcal{M} is an interval in our model, we can thus reconstruct $\alpha(0)$, $D(0)$, and $-z'(\alpha)S(z(\alpha))$ on $(0, \alpha(0))$ from the measurements $D(Z, \mu)$.

We now reconstruct $\alpha(z)$ on $(0, Z)$ from the above measurements. We present two similar methods. Let us first introduce the function $\tilde{T}(\alpha)$ of class $C^1(0, \alpha(0))$ defined by $\tilde{T}(\alpha) = T(z(\alpha))$. We verify that $-z'(\alpha)S(z(\alpha)) = \tilde{T}'(\alpha)$. By integration, and since $\tilde{T}(0) = T(Z)$ is known, we uniquely reconstruct $\tilde{T}(\alpha)$ on $(0, \alpha(0))$. Since $z'(\alpha) < 0$, we deduce that $\tilde{T}'(\alpha) = T'(z(\alpha))z'(\alpha)$ and $T'(z)$ vanish at the same singular points (in their respective variables). If there is no such point, then $T(z)$ is a homeomorphism on $(0, Z)$ (it is bijective, continuous, and maps open sets to open sets since $|T'(z)| > 0$ on the interval; it thus admits a continuous inverse) with inverse $z(T)$, from which we deduce $z(\alpha) = z(\tilde{T}(\alpha))$ on $(0, Z)$; whence its inverse $\alpha(z)$. Otherwise, we call the singular points α_k and z_k , $1 \leq k \leq N$, respectively, with $\alpha(z_k) = \alpha_k$. We also note the endpoints $\alpha_0 = 0$, $\alpha_{N+1} = \alpha(0)$, $z_0 = Z$, and $z_{N+1} = 0$. The points α_k are determined by the data since $\tilde{T}(\alpha)$ is known, and the points z_k are determined by the knowledge of $T(z)$. On each interval (z_{j+1}, z_j) , $T(z)$ is a homeomorphism with inverse function $z(T)$ (for the same reasons as above). We thus obtain $z(\alpha) = z(\tilde{T}(\alpha))$ on (z_{j+1}, z_j) ,

whence $\alpha(z)$ on (α_j, α_{j+1}) . Varying $0 \leq j \leq N$, this allows us to reconstruct $\alpha(z)$ on the whole interval $(0, Z)$.

Another (very similar) way of looking at the reconstruction is to recast (4.21) for $z \in \mathcal{Z}$ as

$$\frac{dz(\alpha)}{d\alpha} = -\frac{h(\alpha)}{S(z(\alpha))}. \quad (4.22)$$

The above (nonlinear) ordinary differential equation for $z(\alpha)$ holds at all but possibly a finite number of points in \mathcal{Z} by assumption on $S(z)$ and can be extended by continuity to the whole interval since $z(\alpha)$ is a C^1 diffeomorphism. We thus uniquely recover the diffeomorphism $z(\alpha)$ from (4.22) since $h(\alpha)$ is continuous on $(0, \alpha(0))$ as can be seen in (4.21) and provided that $S(z)$ is a Lipschitz function (hence the above ordinary differential equation admits a unique solution; this proof requires a little more regularity than the previous one). This also uniquely defines its inverse $\alpha(z)$. Once α is reconstructed we uniquely reconstruct $c(z)$ by differentiating formula (4.18). This complete the proof. \square

Both the analytic continuation [128] and the inversion of the Laplace transform [31] are known to be *severely ill-posed* problems. The reconstruction of the concentration profiles from the boundary measurements is therefore severely ill-posed, even if we had access to data on $\mathcal{M} = \mathbb{R}^+$ (in which case the reconstruction would still rely on inverting a Laplace transform). In practice, this means that only a handful of parameters modeling the concentration profile can realistically be reconstructed from the measured data provided that those data contain only high frequency noise.

The assumption that the temperature gradient $S(z)$ may vanish at a finite number of points allows us to account for non-invertible temperature profiles (i.e., $z(T)$ may be a multi-valued function; the assumption on $S(z)$ implies that it can only take a finite number of values). The temperature profiles are not invertible in practice, see Fig.4-1(a), so we need to account for this situation. The assumption however cannot be substantially relaxed. If $S(z)$ vanishes on an interval, then $\alpha'(z)$ cannot be reconstructed on this interval since (4.21) and (4.22) are no longer equivalent. The measurements at $z = Z$ provide no information on $\alpha(z)$ on the intervals where $S(z)$

vanishes. This implies non-uniqueness of the gas profile reconstruction; see also our numerical simulation at the end of section 4.5.1.

4.3.2 The case of multiple gases

We now extend the results obtained in the preceding section to the multiple-gas case. Let us assume that our composite gas consists of M different species and the absorption spectra of the composite gas contains N ($N \geq M$) well-separated bands centered at ν_i , $i = 1, 2, \dots, N$, respectively. By “well-separated” we mean that for wavenumbers ν in the j th band, we have $|\nu - \nu_j| \ll |\nu_k - \nu_j|$, for all $k \neq j$. We assume moreover that the absorption coefficient for gas i can be written as

$$\kappa_i(z, \nu) = Q_{ij} \mu(\nu - \nu_j) c_i(z) g_i(z), \quad (4.23)$$

where Q_{ij} is the line intensity for gas i in the wavenumber band j . Our main assumption is that the function $\mu(\nu)$ with range \mathcal{M} is universal to all gases and takes the form (4.14), i.e.,

$$\mu(\nu) = \frac{1}{\pi} \frac{\bar{\alpha}}{\nu^2 + \bar{\alpha}^2}, \quad (4.24)$$

where $\bar{\alpha}$ is a constant. With these assumptions we have the following (relatively straightforward) generalization of the single gas case:

Theorem 4.3.2. *Under the assumptions of Theorem 4.3.1, there exists a unique set of positive profiles $c_i(z)$, $i = 1, \dots, M$, such that $D(Z, \mu) \equiv D_m(Z, \mu)$ provided that assumption (4.23) and (4.24) hold and the matrix Q_{ij} has rank M .*

Proof. With the above assumptions, we can write the total absorption map for the composite gas in the vicinity of band j as

$$a(z, \nu) \equiv a_j(z, \nu) = \mu(\nu - \nu_j) \sum_{i=1}^M Q_{ij} c_i(z) g_i(z). \quad (4.25)$$

After defining

$$\alpha_j = \sum_{i=1}^M Q_{ij} \int_z^Z c_i(\zeta) g_i(\zeta) d\zeta, \quad (4.26)$$

we obtain a similar expression as before for the measurements $D_j(Z, \mu)$ in the j th frequency band:

$$D_j(Z, \mu) = D(0)e^{-\mu\alpha_j(0)} + \int_0^Z S(z)e^{-\mu\alpha_j(z)} dz. \quad (4.27)$$

For the same reason as in one gas case, since $\alpha'_j < 0$, we can perform the change of variables $z \rightarrow \alpha_j(z)$. Defining the inverse map $\alpha_j \mapsto z(\alpha_j)$, we obtain

$$D_j(Z, \mu) = D(0)e^{-\mu\alpha_j(0)} + \int_0^{\alpha_j(0)} S(z(\alpha_j)) \left| \frac{dz}{d\alpha_j} \right| e^{-\mu\alpha_j} d\alpha_j. \quad (4.28)$$

We can thus regard $D(Z, \mu)$ as the *Laplace transform* of the distribution

$$h(\alpha_j) \equiv D(0)\delta(\alpha_j - \alpha_j(0)) + S(z(\alpha_j)) \left| \frac{dz}{d\alpha_j} \right|. \quad (4.29)$$

Then, by the same argument as in the single gas case, we can uniquely reconstruct $\alpha_j(z)$. According to (4.26), we can uniquely recover

$$\sum_{i=1}^M Q_{ij} c_i(z) g_i(z) = \alpha'_j(z), \quad j = 1, \dots, N. \quad (4.30)$$

The above inversion can be performed in each of the N absorption bands, after which we arrive at the following system of equations for $c_i(z)g_i(z)$:

$$\begin{pmatrix} Q_{11} & \cdots & Q_{1M} \\ \vdots & \ddots & \vdots \\ Q_{N1} & \cdots & Q_{NM} \end{pmatrix} \begin{pmatrix} c_1 g_1 \\ \vdots \\ c_M g_M \end{pmatrix} (z) = \begin{pmatrix} \alpha'_1 \\ \vdots \\ \alpha'_N \end{pmatrix} (z). \quad (4.31)$$

Since the matrix (Q_{ij}) has rank M , the above system admits at most one solution, is invertible when $M = N$, and provides the unique solution if the source terms $\alpha'_j(z)$ are compatible. This implies that we can uniquely determine the concentration profiles $c_i(z)$ from the measured data and concludes the proof. \square

In many practical situations, the matrix (Q_{ij}) is indeed invertible (with $N = M$).

Such examples can be seen in the reference [125], where it appears that the matrix Q is often diagonally dominant. A diagonally dominant square matrix, i.e., such that $|Q_{ii}| > \sum_{j \neq i}^M |Q_{ij}|$ for all $i = 1, \dots, M$, is known to be invertible [176].

4.4 Small inclusions

We have seen in earlier sections that the reconstruction of concentration profiles from radiation measurements was a severely ill-posed problem. This implies that only very few coefficients parameterizing the concentration can be reconstructed from the measurements provided that noises contained in data have only high frequency components. Therefore localized inclusions such as ozone or dust layers, whose detection is important in many applications, may be poorly reconstructed unless their presence is explicitly parameterized.

We proposed here to model such layers as localized inclusions of small thickness and arbitrary concentration variations compared to the underlying medium that will be assumed to be *known*. The problem of reconstructing localized diffusive or absorbing inhomogeneities has been extensively studied in medical imaging problems [9, 19, 44]. We now consider such a model in profile retrieval and carry out a similar analysis.

4.4.1 The case of a single gas

Let us start with the case of a single inclusion composed of a single gas. We assume that the background profile $c_0(z)$ is known. The true profile is given by

$$c(z) = c_0(z) + \delta c(z). \quad (4.32)$$

The assumption on $\delta c(z)$ is *not* that it is small in $L^\infty(\mathcal{Z})$ but rather that it is small in $L^1(\mathcal{Z})$ and of “small” support. We assume that $\delta c(z)$ takes the (arbitrarily large) value δc on an interval centered at $z = z_0$ and of size δz and takes the value 0

elsewhere:

$$\delta c(z) = \delta c \chi_{I_z}(z), \quad I_z = \left[z_0 - \frac{\delta z}{2}, z_0 + \frac{\delta z}{2} \right].$$

Here, $\chi_{I_z}(z)$ is the indicatrix function of the interval I_z .

Let us denote by $\alpha_0(z)$ the optical length corresponding to the background profile c_0 only. We then observe from equation (4.19) that

$$\begin{aligned} D[c_0 + \delta c](Z, \mu) &= D(0)e^{-\mu\alpha_0(0)} \exp\left(-\mu\delta c \int_{I_z} g(\zeta)d\zeta\right) \\ &+ \int_0^Z S(z)e^{-\mu\alpha_0(z)} \exp\left(-\mu\delta c \int_{I_z \cap (z, Z)} g(\zeta)d\zeta\right) dz. \end{aligned} \quad (4.33)$$

By hypothesis, $D[c_0](Z, \mu)$ is known and we thus have access by approximating $D[c_0 + \delta c](Z, \mu) - D[c_0](Z, \mu)$, to first order in δz , to the following quantity

$$\mu \mapsto \mu\delta c\delta z g(z_0) \left[D(0)e^{-\mu\alpha_0(0)} + \int_0^{z_0} S(z)e^{-\mu\alpha_0(z)} dz \right]. \quad (4.34)$$

Taking the ratio at two different values of μ gives a functional $\mathcal{F}(z_0)$. It is straightforward to check that $z_0 \mapsto \mathcal{F}(z_0)$ is a smooth function. On each interval such that $\mathcal{F}'(z_0) \neq 0$ we can thus uniquely reconstruct z_0 in a stable way. In practical applications, $\mathcal{F}'(z_0)$ may vanish at a finite number of points so that the function $\mathcal{F}(z_0)$ is not monotone. The point z_0 is then not uniquely reconstructed. However it can be uniquely reconstructed when we know *a priori* on which interval z_0 belongs where $\mathcal{F}(z_0)$ does not vanish. Once z_0 is known, we easily obtain $\delta c\delta z$ from the above expression.

Consequently, provided that we have a sufficiently accurate knowledge of the background and that the term $\delta c\delta z$ is sufficiently small, we can reconstruct in a stable way the location of the inclusion z_0 and $\delta c\delta z$, which for want of a better word we will call its *strength*. Moreover this can ideally be performed from only two measurements corresponding to two different values of μ .

We now verify by asymptotic expansions that the first term allowing us to separate δz from δc is of order δz^3 . Indeed, upon carrying out a higher-order Taylor expansion

in (4.33) we deduce that

$$\begin{aligned}
D[c_0 + \delta c](Z, \mu) - D[c_0](Z, \mu) &= \left(-\mu \delta c \delta z g_0 + \frac{1}{2} \mu^2 g_0^2 \delta c^2 \delta z^2 \right. \\
&\quad \left. - \frac{1}{6} (\mu g_0'' \delta c + \mu^3 g_0^3 \delta c^3) \delta z^3 \right) \left[D(0) e^{-\mu \alpha_0(0)} + \int_0^{z_0} S(z) e^{-\mu \alpha_0(z)} dz \right] \\
&\quad + O(\delta z^4),
\end{aligned} \tag{4.35}$$

where $g_0 = g(z_0)$ and $g_0'' = g''(z_0)$.

So when the noise in the data is of order $O(\delta z^4)$ or higher, we can reconstruct z_0 , δz and δc as we have access to $\delta c \delta z^3$ in the term of order $O(\delta z^3)$, at least provided that $g''(z_0)$ does not vanish. However when the noise in the data is of order $O(\delta z^3)$ or larger, all we can possibly reconstruct from the measurements is the location z_0 and the product $\delta c \delta z$. This corresponds to knowing the total amount of ozone variation in the layer but not the respective thickness and concentration variation. If the noise in the data is larger than $\delta c \delta z$, then even this information cannot be retrieved unless a more careful statistical model is considered.

We now consider a case where the location z_0 cannot be recovered uniquely. We deduce from (4.34) that all the information we have access to about z_0 is contained in $g(z_0) (D(0) e^{-\mu \alpha_0(0)} + \int_0^{z_0} S(z) e^{-\mu \alpha_0(z)} dz) \equiv G(z_0)$. Both $g(z)$ and $S(z)$ are related to the temperature profile $T(z)$. Suppose that $T(z)$ is constant on an interval I so that $S(z)$ vanishes on I . Then we verify that $G(z)$ is constant on I , which means that z_0 cannot be reconstructed uniquely when the inclusion is located in a region of constant temperature. Notice that the hypotheses of Theorem 4.3.1 are not satisfied in this case. Consequently, if one tries to recover z_0 by a gradient-based optimization technique such as a Newton or Conjugate Gradient method [180], the gradient of objective functional (for instance $\mathcal{F}(z_0)$) with respect to z_0 will vanish for $z_0 \in I$; see the numerical simulations in section 4.5.

4.4.2 The case of multiple gases

Let us now briefly extend the analysis in the case of M gases assuming the existence of M (to simplify) separated wavenumber bands as described in (4.23). The asymptotic analysis is based on formula (4.27). As in the single-gas case, we assume that the profile for each gas is a superposition of a known background and localized variations of arbitrary contrast. More precisely, we have

$$c_i(z) = c_{0i}(z) + \delta c_i(z), \quad i = 1, \dots, M, \quad (4.36)$$

where c_{0i} is the background concentration profile for species i and where the fluctuations are modeled by

$$\delta c_i(z) = \delta c_i \chi_{I_{z_i}}(z), \quad I_{z_i} = \left[z_i - \frac{\delta z_i}{2}, z_i + \frac{\delta z_i}{2} \right], \quad i = 1, \dots, M.$$

As before, $\chi_{I_{z_i}}(z)$ is the indicatrix function of the interval I_{z_i} . We assume also that all thicknesses δz_i are of the same order $O(\delta z)$. We assume here that each gas may have strong fluctuations in only one layer. The generalization to multiple layers is straightforward and is not considered. This may be accounted for in the present theory by stipulating that several indices $1 \leq j \leq M$ correspond to the same gas.

Upon inserting the above approximation into formula (4.27), we obtain

$$\begin{aligned} D[c_{01} + \delta c_1, \dots, c_{0M} + \delta c_M](Z, \mu) = & \\ D(0)e^{-\mu\alpha_j^0(z)} \exp\left(-\mu \sum_i^M Q_{ij} \delta c_i \int_{I_{z_i}} g(\zeta) d\zeta\right) & \\ + \int_0^Z S(z)e^{-\mu\alpha_j^0(z)} \exp\left(-\mu \sum_i^M Q_{ij} \delta c_i \int_{I_{z_i} \cap (z, Z)} g(\zeta) d\zeta\right) dz, & \quad (4.37) \end{aligned}$$

where α_j^0 denotes the optical length α_j in (4.26) defined with the background profile.

Upon performing a second-order Taylor expansion in the above formula, we get

$$\begin{aligned}
& D_j[c_{01} + \delta c_1, \dots, c_{0M} + \delta c_M](Z, \mu) - D_j[c_{01}, \dots, c_{0M}](Z, \mu) \\
= & -\mu \sum_{i=1}^M \left(Q_{ij} \delta c_i \delta z_i g_i(z_i) \bar{S}(\mu, z_i) \right) \\
& + \mu^2 \sum_{i=1}^M \left(Q_{ij} \delta c_i \delta z_i g_i(z_i) \right)^2 \bar{S}(\mu, z_i) \\
& + \mu^2 \sum_{i=1}^M \sum_{i \neq k=1}^M Q_{ij} \delta c_i \delta z_i g_i(z_i) Q_{kj} \delta c_k \delta z_k g_k(z_k) \bar{S}(\mu, \min(z_i, z_k)) \\
& + O(\delta z^3),
\end{aligned} \tag{4.38}$$

where we have defined the averaged source term

$$\bar{S}(\mu, z_i) \equiv \int_0^{z_i} S(z) e^{-\mu \alpha_j^0(z)} dz + D(0) e^{-\mu \alpha_j^0(0)}. \tag{4.39}$$

Higher order terms can be obtained similarly although their expression becomes much more cumbersome. Notice that we recover (4.34) when $M = 1$. Suppose that the error in the measured data is of order $O(\delta z^2)$. Then we only have access to the information

$$\mu \mapsto \mu \sum_{i=1}^M Q_{ij} \delta c_i \delta z_i g_i(z_i) \bar{S}(\mu, z_i). \tag{4.40}$$

Assuming that the matrix $(Q_{ij})_{i,j}$ is a square invertible matrix, we can reconstruct from measurements in M well-separated bands the quantities defined by

$$p_i \equiv \mu \delta c_i \delta z_i g_i \bar{S}(\mu, z_i). \tag{4.41}$$

This information has the same structure as in the single-gas case. From a minimum of two measurements, we can reconstruct the location z_i . An accuracy of order $O(\delta z)$ in the data then allows us to reconstruct the strength of the i th inclusion $\delta c_i \delta z_i$. The same products appear in the terms proportional to δz^2 . Therefore an accuracy in the data of order $O(\delta z^3)$ is again necessary to estimate δz_i and δc_i separately.

4.5 Numerical reconstructions

We present in this section several numerical simulations that illustrate the theory developed in the preceding section. The atmosphere thickness is normalized to $Z = 1$. We first concentrate on the single gas case and then consider an example with a mixture of two gases. All the data are synthetic and the cases considered academic. However we have chosen temperature and concentration profiles that are qualitatively very similar to those analyzed in [38].

4.5.1 The case of a single gas

We start with the single gas model. We illustrate the predictions of the asymptotic

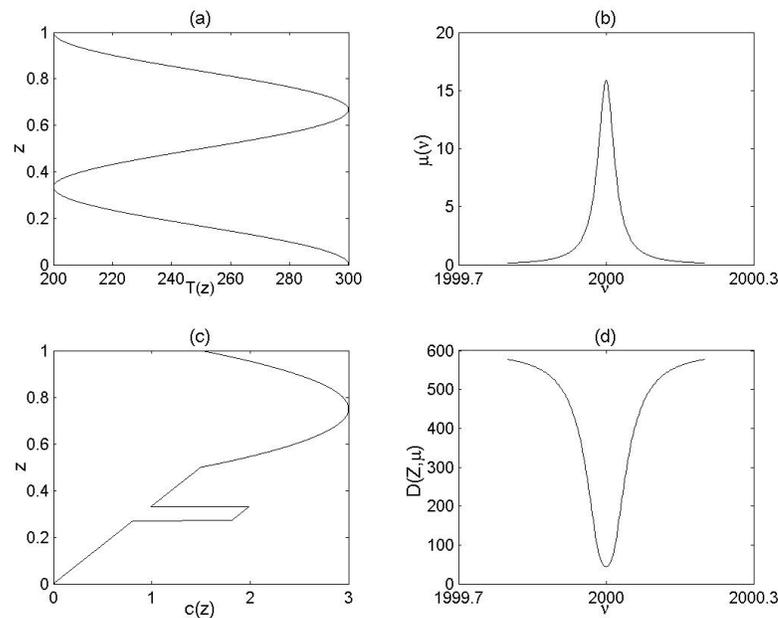


Figure 4-1: Profiles used in the calculation. (a) Temperature profile as a function of z . (b) Rescaled absorption as a function of wavenumber. (c) Ozone concentration as a function of z . (d) Data $D(Z, \mu(\nu))$ as a function of wavenumber ν .

expansions that different properties of the localized inclusions can be retrieved from the measured data depending on the noise level. The concentration profile is given

by

$$c_0(z) = \begin{cases} 3z & z \in (0, 0.5] \\ 3.0 - 20(z - 0.75)^2 & z \in (0.5, 1.0), \end{cases} \quad (4.42)$$

which is a simplified model for the concentration profile of ozone in the atmosphere below 40 kilometers. A thin inclusion is located at $z_0 = 0.3$. The characteristics of that inclusion are that $\delta z = 0.06$ and $\delta c = 1.0$. The temperature profile is modeled by

$$T(z) = 250 + 50 \sin(3\pi z + \frac{\pi}{2}), \quad z \in (0, 1), \quad (4.43)$$

which qualitatively resembles the observed profiles. Fig.4-1 shows the concentration profile, the temperature distribution, the absorption line shape used in the calculations, and the solution of equation (4.17) with respect to wavenumber ν .

The location and characteristics of the inclusion are reconstructed by two methods. In the first method, we minimize the error of the forward model to the true data by using a full search algorithm. This can be done because only three parameters need to be recovered in this case. More precisely we search on a $101 \times 101 \times 401$ uniformly distributed mesh for $(z_0, \delta z, \delta c)$ in the parameter space $[0.25, 0.35] \times [0.01, 0.11] \times [0.80, 1.20]$. We look for the minimum of the least-square error functional

$$E(z_0, \delta z, \delta c) = \int_{\mathcal{M}} \left(D(Z, \mu) - D_m(\mu) \right)^2 d\mu, \quad (4.44)$$

where $D_m(\mu)$ represents the measurement data. In the numerical simulations, we take $Q_0 = \bar{\alpha} = 2 \times 10^{-2}$, $\mathcal{M} = \frac{1}{\pi} [-\frac{1}{101}, 1]$, which means that the wavenumbers either belong to $[1999.8, 2000]$ or to $[2000, 2000.2]$ by symmetry. We use 200 wavenumbers in each band.

We show in Fig.4-2 the distribution of the error in parameter space. In Fig.4-2(a), we present the function at $z_0 = 0.3$ (dashed lines), and the function at $z_0 = 0.3$ and $\delta z \delta c = 0.06$ (thick solid line). We observe that the functional varies quite substantially in the direction of increase (or decrease) of $\delta z \delta c$ but remains almost constant in the orthogonal direction (i.e., along curves where $\delta z \delta c$ is constant). Finding the curve where $\delta c \delta z$ is minimal can thus be achieved even with quite substantial noise in the

data. Finding the global minimum of the functional, which is necessary to separately reconstruct δc and δz , requires much more accurate data. Fig.4-2 (b) show that the functional with respect to z_0 , the location of the inclusion, at δc and δz fixed to their exact value, is quite well-behaved. This indicates that z_0 can also be reconstructed in quite a stable way.

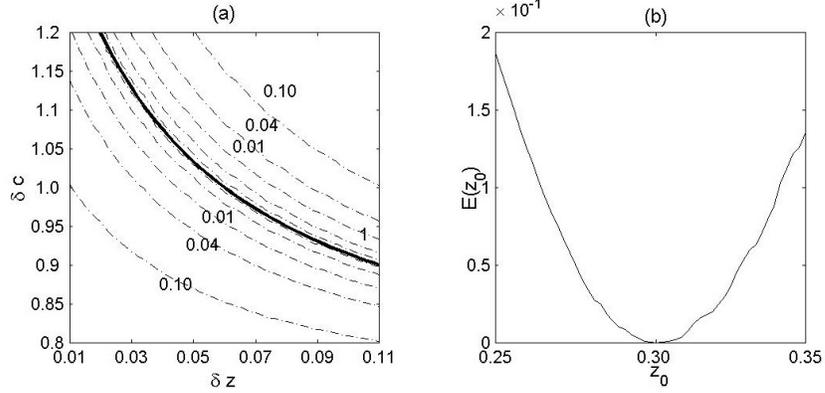


Figure 4-2: Cross section of the error functional in the parameter space. (a) functional at $z_0 = 0.3$; (b) functional at $\delta z = 0.06$ and $\delta c = 1.0$.

Tab.4.1 lists the parameters recovered by the full search algorithm. The accuracy in the recovery of δz and δc decreases as the noise level increases. The location of the inclusion z_0 and the product $\delta z \delta c$ can be obtained satisfactorily even with a relatively high noise level of around 1%. However, at this level of noise, the reconstruction of δz and δc is no longer reliable with relative errors as high as 20%. In all our simulations, a noise level of $x\%$ means that a uniformly distributed random number between $-x\%$ and $x\%$ has been added.

As the number of parameters increases, full search algorithms are not tractable. We have repeated the preceding reconstruction by using the Conjugate Gradient (CG) method [180] to minimize the least square error functional (4.44),

$$\min_{z_0, \delta z, \delta c} E. \quad (4.45)$$

The initial guess for the parameters are $z_0 = 0.27$, $\delta z = 0.07$ and $\delta c = 1.2$. The

Noise	z_0 (error)	δz (error)	δc (error)	$\delta z \delta c$ (error)
0.00%	0.3000 (0.0%)	0.0600 (0.0%)	1.0000 (0.0%)	0.0600 (0.0%)
0.05%	0.3000 (0.0%)	0.0590 (1.7%)	1.0180 (1.8%)	0.0597 (0.5%)
0.10%	0.2970 (1.0%)	0.0570 (6.0%)	1.0430 (4.3%)	0.0592 (1.4%)
1.00%	0.2950 (1.7%)	0.0490 (18.%)	1.1870 (18.%)	0.0581 (3.1%)

Table 4.1: Characteristics of the inclusion reconstructed by a full search algorithm. The true values are $z_0 = 0.3$, $\delta z = 0.06$, $\delta c = 1.0$, hence $\delta z \delta c = 0.06$. The numbers in parentheses denote the relative error in percentage between the reconstructed parameters and their true values.

results are list in Tab.4.2 and are very similar to those obtained with the full search algorithm. The CG algorithm was found to be relatively robust with respect to the choice of the initial guess.

Noise	z_0 (error)	δz (error)	δc (error)	$\delta z \delta c$ (error)
0.00%	0.3000 (0.0%)	0.0600 (0.0%)	1.0000 (0.0%)	0.0601 (0.2%)
0.05%	0.3000 (0.0%)	0.0591 (1.7%)	1.0177 (1.8%)	0.0602 (0.3%)
0.10%	0.2971 (1.0%)	0.0572 (4.7%)	1.0386 (3.9%)	0.0594 (1.0%)
1.00%	0.2952 (1.6%)	0.0492 (18.%)	1.1811 (18.%)	0.0581 (3.1%)

Table 4.2: Same as Tab. 4.1 (with the *same* noisy measurements) except that the Conjugate Gradient algorithm is used in the optimization process.

Let us now consider the special case where uniqueness in the reconstruction of z_0 is not guaranteed. This happens when the temperature gradient vanishes on an interval \mathcal{I} including the inclusion's location. The temperature profile is now chosen to be

$$T(z) = \begin{cases} 250 + 50 \sin(3\pi z + \frac{\pi}{2}) & z \in (0, 0.1) \cup (0.5, 1.0) \\ 240 & z \in [0.1, 0.5], \end{cases} \quad (4.46)$$

and the background concentration profile is given by

$$c_0(z) = \begin{cases} 4z & z \in (0, 0.5] \\ 2.5 - 8(z - 0.75)^2 & z \in (0.5, 1.0). \end{cases} \quad (4.47)$$

A small inclusion is placed at $z_0 = 0.25$. The width of the the inclusion is $\delta z = 0.08$ and the concentration variation $\delta c = 1.2$.

Reconstructions from data at different noise levels by the Conjugate Gradient

method are presented in Tab.4.3. The gradients have been computed by using a finite difference approximation. We found numerically that while we can recover δc and δz almost perfectly with exact simulated data, the exact location z_0 is not retrieved if our initial guess lies within \mathcal{I} . The initial guess for the data in Tab.4.3 was chosen to be $z_0 = 0.28$, $\delta z = 0.07$, and $\delta c = 1.0$.

Noise	z_0 (error)	δz (error)	δc (error)	$\delta z \delta c$ (error)
0.00%	0.280 (12.%)	0.080 (0.0%)	1.202 (0.2%)	0.0962(0.2%)
0.05%	0.280 (12.%)	0.081 (1.3%)	1.180 (1.7%)	0.0956(0.4%)
0.10%	0.280 (12.%)	0.083 (3.8%)	1.144 (4.7%)	0.0950(1.1%)
1.00%	0.280 (12.%)	0.091 (14.%)	1.026 (15.%)	0.0933(2.8%)

Table 4.3: Characteristics of the inclusion reconstructed by the Conjugate Gradient algorithm when the inclusion is placed in a region with vanishing temperature gradient. The real values for those variables are $z_0 = 0.25$, $\delta z = 0.08$, $\delta c = 1.20$ and $\delta z \delta c = 0.096$. The numbers in parentheses denote the relative error in percentage between the reconstructed parameters and their true values.

4.5.2 The case of two gases

Let us now consider the case of two gases. We use (4.42) and (4.47) as the background profiles for the two gases, respectively. The characteristics for the two small inclusions are the following: $(z_1, \delta z_1, \delta c_1) = (0.30, 0.06, 1.00)$ and $(z_2, \delta z_2, \delta c_2) = (0.25, 0.08, 1.20)$. We simulate the data using 800 wavenumbers uniformly distributed in two band centered at $\nu_1 = 1500 \text{ cm}^{-1}$ and $\nu_2 = 2000 \text{ cm}^{-1}$, respectively. The absorption kernel has the form given in (4.13) with parameters given by $\bar{\alpha} = 2 \times 10^{-2}$, and $(Q_{11}, Q_{12}, Q_{21}, Q_{22}) = (2.0, 1.0, 1.0, 2.0) \times 10^{-2}$.

Gas	z_i (error)	δz_i (error)	δc_i (error)	$\delta z_i \delta c_i$ (error)
$i = 1$	0.200 (0.0%)	0.060 (0.0%)	1.002 (0.2%)	0.0601 (0.2%)
$i = 2$	0.250 (0.0%)	0.081 (1.3%)	1.190 (0.8%)	0.0964 (0.4%)

Table 4.4: Characteristics of the inclusions in the two-particle model reconstructed from noise free data. The initial guess is $z_1 = 0.32$, $\delta z_1 = 0.05$, $\delta c_1 = 0.8$ and $z_2 = 0.28$, $\delta z_2 = 0.10$, $\delta c_1 = 1.0$. The numbers in parentheses denote the relative error in percentage between the reconstructed parameters and their true values.

We perform three sets of numerical experiments with noise free data, data with 0.1% noise and data with 1% noise, respectively. The results are list in Tab.4.4,

Tab.4.5 and Tab.4.6, respectively. The initial guess is $z_1 = 0.32$, $\delta z_1 = 0.05$, $\delta c_1 = 0.8$ and $z_2 = 0.28$, $\delta z_2 = 0.10$, $\delta c_1 = 1.0$. We found that the initial guess on the positions may be chosen relatively far away from the true values, while the guess on the other two parameters should be close to the true value in order for the CG algorithm to converge.

Gas	z_i (error)	δz_i (error)	δc_i (error)	$\delta z_i \delta c_i$ (error)
$i = 1$	0.200 (0.0%)	0.059 (1.7%)	1.012 (1.2%)	0.060 (0.5%)
$i = 2$	0.251 (0.4%)	0.082 (2.5%)	1.150 (4.2%)	0.094 (1.8%)

Table 4.5: Same as Tab.4.4 with 0.10% noise.

Notice in Tab.4.4 that the parameters for both inclusions are recovered very accurately in the absence of noise. This is quite similar to the one particle case. The only noticeable difference numerically is that a much wider range of wavenumbers is necessary in the case of multiple particles to ensure convergence. This is in agreement with theory, which indicates that the number of measurements should scale at least linearly with the number of retrieved gas profiles.

At moderate levels of noise, we can still recover the positions of the inclusions and their strength $\delta c \delta z$, but not δc and δz separately. We also observed in our simulations that, as the noise level increases, we even lose the information about z_0 and $\delta z \delta c$. The only quantity which seems numerically to be accurately reconstructed is then p introduced in (4.41).

Gas	z_i (error)	δz_i (error)	δc_i (error)	$\delta z_i \delta c_i$ (error)
$i = 1$	0.204 (2.0%)	0.053 (12.%)	1.081 (8.1%)	0.057 (4.5%)
$i = 2$	0.261 (4.4%)	0.089 (11.%)	1.038 (14.%)	0.092 (3.8%)

Table 4.6: Same as Tab. 4.4 with 1% noise.

4.6 Conclusions and remarks

Under some separation assumptions on the spectral emission coefficient, we have shown that the concentration profiles of single or multiple gases could uniquely be

reconstructed from radiation measurements. Moreover we have shown that the reconstruction invokes the inversion of a Laplace transform (at best) and is therefore a severely ill-posed problem. The assumptions on the emission coefficient necessary to obtain an explicit formula are technical and should not modify the general conclusion that the reconstruction problem is *severely ill-posed* even in more general settings.

To reconstruct localized strong fluctuations such as ozone layers in the troposphere, we have presented an asymptotic model, which assesses the type of information that can be reconstructed based on the quality of the measured data. For instance, we show that with moderate noise levels, we can reconstruct the location of the inclusion and the product of its thickness with its concentration variation (with respect to the background). We have shown that the reconstruction of both the thickness and the concentration variation requires much more accurate data.

We have conducted numerical experiments on academic though qualitatively faithful benchmarks that corroborate the theory. Our main conclusion is that the reconstruction of the thickness and the concentration of ozone layers in the troposphere requires extremely accurate data. In our setting, possible errors in the reconstruction of the background are treated as noise in the measured data. This assumption certainly needs improvement. Yet the method of asymptotic expansions presented in this chapter provides a systematic framework to evaluate the type of information that can be retrieved on localized inclusions from measured data with a given noise level.

Chapter 5

Comparison of transport and diffusion reconstructions in small domains

Transport equations are accurate in modeling the propagation of the particles in scattering media [62, 106, 147]. However, as we have seen in chapter 2 and chapter 3, transport models are quite expensive computationally. Diffusion equations are often preferred in practical applications in optical tomography. However, diffusion approximations are not always valid. The objective of this chapter is to compare reconstructions based on the diffusion equation with those based on the radiative transport equation in media of small size. We show that in this case diffusion-based reconstructions are in general less accurate than transport-based reconstructions. The presentation of this chapter is based on reference [148].

5.1 Problem statement

It is generally believed that the propagation of near infra-red light in tissues is best modeled by the radiative transport equation, which is formulated in phase space, i.e., the space of positions and directions, and thus computationally very expensive. One has to discretize both the spatial and the angular variables to find numerical solutions. To lower computational cost, it is preferable in many applications to replace the transport equation by its diffusion approximation, which models the spatial density

of photons.

The application of the diffusion equation in optical tomography, however, has its limitations. Essentially, the derivation of the diffusion equation from the radiative transport equation is only valid when the underlying tissues are highly scattering and weakly absorbing. The diffusion approximation fails to accurately model light propagation in regions with small or vanishing scattering coefficients, such as the cerebrospinal fluid layers in the human head. In this case, one either has to generalize the classical diffusion equation [23], couple diffusion with transport equations [17, 22, 59], or solve the ERT directly [1, 42, 62, 106, 147]. Another situation where diffusion approximation does not work well, and somewhat related to the previous one, is in modeling light propagation in media of small volumes. Examples are imaging of rheumatoid arthritis in finger joints [92, 108, 131, 145, 184], or imaging of small animals [35, 80, 89]. In these cases, because of the small optical distance between sources and detectors, the diffusion approximation is too crude to approximate the “transport” behavior of photons. How these errors in the diffusion approximation influence optical tomographic reconstructions has not been studied yet.

The objective here is precisely to compare reconstructions based on the diffusion equation with those based on the radiative transport equation in circumstances where diffusion equation does not approximate the transport equation very well. Focusing on the problem of small domains, we show that considerable differences between reconstructions with transport and diffusion models. We show that diffusion-based reconstructions are in general less accurate.

5.1.1 Transport and diffusion approximations

Let us recall that the radiative transport equation describes the photon density in the phase space $X = \Omega \times S^2$, i.e., as a function of both position $\mathbf{x} \in \Omega \subset \mathbb{R}^3$ and propagation direction $\boldsymbol{\theta} \in S^2$. Here Ω is the spatial domain and S^2 the unit sphere in \mathbb{R}^3 . When the intensity of the light source is modulated with frequency ω , i.e., is of the form $f(\mathbf{x}, \boldsymbol{\theta})e^{i\omega t}$, where t is the time variable, the frequency domain radiative

transport equation takes the form [11, 146, 147]:

$$\begin{aligned} \left(\frac{i\omega}{v} + \boldsymbol{\theta} \cdot \nabla + \sigma_t(\mathbf{x})\right)u(\mathbf{x}, \boldsymbol{\theta}) - \sigma_s(\mathbf{x}) \int_{S^2} k(\boldsymbol{\theta} \cdot \boldsymbol{\theta}')u(\mathbf{x}, \boldsymbol{\theta}')d\mu(\boldsymbol{\theta}') &= 0 && \text{in } X \\ u(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x}, \boldsymbol{\theta}) &&& \text{on } \Gamma_-, \end{aligned} \quad (5.1)$$

where $i = \sqrt{-1}$ and v is the speed of light in the medium. The functions $\sigma_t(\mathbf{x})$ and $\sigma_s(\mathbf{x})$ are the total absorption (extinction) coefficient and scattering coefficient, respectively. The solution $u(\mathbf{x}, \boldsymbol{\theta})$ is the radiant power per unit solid angle per unit area perpendicular to the direction of propagation at \mathbf{x} in the direction $\boldsymbol{\theta}$. The boundary sets Γ_{\pm} are defined by

$$\Gamma_{\pm} = \{(\mathbf{x}, \boldsymbol{\theta}) \in \partial\Omega \times S^2 \text{ s.t. } \pm \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) > 0\}, \quad (5.2)$$

with $\boldsymbol{\nu}(\mathbf{x})$ the outward unit normal to Ω at $\mathbf{x} \in \partial\Omega$. For more details on the above radiative transport equation, see chapter 2.

The above radiative transport equation is a microscopic model for light propagation in tissues. Numerical solutions of this model are very expensive because both spatial and angular discretizations have to be performed. It is thus preferable in many applications to replace the transport equation by the less expensive physical-space diffusion equation. The diffusion equation describes light propagation at the macroscopic level, where the unknown quantity is the angularly averaged photon flux.

The approximation of the radiative transport equation by the diffusion equation has been well-documented; see for example [58, 65]. There, it is shown that when absorption is sufficiently low and scattering sufficiently large, the transport process can be modeled macroscopically with the following diffusion equation:

$$\begin{aligned} \frac{i\omega}{v}U(\mathbf{x}) - \nabla \cdot \mathcal{D}\nabla U + \sigma_a(\mathbf{x})U(\mathbf{x}) &= 0, && \text{in } \Omega \\ U + 3\epsilon L_3\boldsymbol{\nu}(\mathbf{x}) \cdot \mathcal{D}\nabla U &= \Lambda(f)(\mathbf{x}), && \text{on } \partial\Omega. \end{aligned} \quad (5.3)$$

Here $U(\mathbf{x})$ is the angularly-averaged photon flux at \mathbf{x} , an approximation of the quantity $\int_{S^2} u(\mathbf{x}, \boldsymbol{\theta})d\boldsymbol{\theta}$ in the transport equation. $\sigma_a(\mathbf{x})$ is the absorption coeffi-

cient which describe rate of absorption at \mathbf{x} . It corresponds to $\sigma_t(\mathbf{x}) - \sigma_s(\mathbf{x})$ in transport equation. The diffusion tensor $\mathcal{D}(\mathbf{x})$ is symmetric and positive definite. It is given by $\mathcal{D}(\mathbf{x}) = 1/(3(\sigma_a(\mathbf{x}) + \sigma'_s(\mathbf{x})))$ in the Henyey Greenstein case, where $\sigma'_s(\mathbf{x}) = (1 - g)\sigma_s(\mathbf{x})$.

The operator Λ is a linear form that maps any incoming angular distribution f to a real number. Explicit expressions can be found in [49, 58] in simple cases. In this paper, we will always use isotropic source terms $f(\mathbf{x}, \boldsymbol{\theta}) = f(\mathbf{x})$ for which we obtain $\Lambda(f) = f$.

A very important aspect in the derivation of diffusion equations is to correctly account for photon leakage at the domain boundary [17, 58, 83]. This requires a detailed boundary layer analysis for the transport equation [58], which shows that leakage should be modeled by the above Robin-type boundary conditions for diffusion, where L_3 is the extrapolation length. Explicit expressions for the extrapolation length are only known in simple cases, for instance when scattering is isotropic ($g = 0$), where $L_3 \approx 0.7104$ [17, 49, 58].

The small parameter ϵ is called the transport mean free path. It is defined as $\epsilon = 3\mathcal{D} = 1/(\sigma_a + \sigma'_s)$. The transport mean free path measures the average distance is takes for photons to be substantially deflected from their original direction by scattering. In the limit where the transport mean free path ϵ go to zero, the error between the diffusion solution $U(\mathbf{x})$ and the transport flux $\int_{S^2} u(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}$ is of order ϵ^2 in regions sufficiently far from the boundary [58].

In the time dependent case, let us note that the diffusion approximation may not be valid for short times [58]. This implies that even in situations of highly scattering and low absorption, the diffusion approximation may not be accurate for high modulation frequencies ω . Numerical evidence for this statement can be found in [72].

5.2 Reconstruction methods

Optical tomography aims at reconstructing $\sigma_a(\mathbf{x})$ and $\sigma_s(\mathbf{x})$ in the transport and diffusion equations from boundary measurements. Our goal here is to quantify the errors in the reconstructions obtained by using the diffusion equation in situations where it is not an accurate approximation to the radiative transport equation. To do this, we assume that the data are generated by the physically accurate transport model. We then consider two reconstructions.

A. Transport reconstruction. In this reconstruction, the radiative transport equation is used as the model for light propagation. The predicted current measurements at the domain boundary are then calculated using

$$J_T(\mathbf{x}) = \int_{S_+^2} \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) u(\mathbf{x}, \boldsymbol{\theta}) d\boldsymbol{\theta}. \quad (5.4)$$

B. Diffusion reconstruction. Here the diffusion equation is used as the light propagation model. The predicted boundary current measurements corresponding (5.4) is computed according to

$$J_D(\mathbf{x}) = -\boldsymbol{\nu}(\mathbf{x}) \cdot \mathcal{D}\nabla U. \quad (5.5)$$

In this paper, we focus on the reconstruction of the absorption coefficient σ_a only and assume σ_s to be known. The reconstruction of σ_a is quite useful in many practical applications, such as e.g. the monitoring of the oxygenation of tissues [177].

5.2.1 Reconstruction algorithms

As usual, we solve the reconstruction problem by minimizing the mismatch between model predictions and measured data for several source-detector pairs:

$$\min_{\sigma_a(\mathbf{x}) \in [\sigma_a^{\min}, \sigma_a^{\max}]} \mathcal{F}_\beta := \mathcal{F}(\sigma_a) + \frac{\beta}{2} \int_{\Omega} \nabla \sigma_a \cdot \nabla \sigma_a d\mathbf{x}, \quad (5.6)$$

where the last term is a Tikhonov regularization functional with regularization parameter β , and where the mismatch functional is defined as

$$\mathcal{F}(\sigma_a(\mathbf{x})) \equiv \frac{1}{2} \sum_{q=1}^{N_q} \|J^q(\mathbf{x}) - z_q(\mathbf{x})\|_{L^2(\partial\Omega)}^2. \quad (5.7)$$

Here σ_a^{min} and σ_a^{max} are physical lower and upper bounds imposed on σ_a . N_q is the number of sources used and $z_q(\mathbf{x})$ denotes the current measurements corresponding to source q . As stated before, we assume that the current data $z_q(\mathbf{x})$ are generated by the transport equation.

We use the limited memory BFGS algorithm that we have developed in [147] to solve the minimization problem (5.6). To use it for the diffusion case, we use the diffusion equation as the forward model and correspondingly modify the gradient calculations for the objective function. We adopt a very similar adjoint state method for the gradient calculation. We refer to [147] for details of the BFGS algorithm for the transport reconstructions.

5.2.2 Discretization of forward models

To calculate model predictions for the minimization algorithm, we numerically solve the radiative transport equation (5.1) by discretizing it using the discrete ordinate method for the angular variable and a finite volume method for the spatial variable [146, 147]. We refer to our earlier work [146, 147] for some numerical tests on the finite volume discretization of the transport equation and related reconstruction results.

The diffusion equation (5.3) is discretized by using a similar finite volume method. Finite volume methods [68] ensure the conservation of mass (or momentum, energy) in a discrete sense, which is important in transport and diffusion calculations. They also have the advantage of easily handling complicated geometries by arbitrary triangulations, which we need in tomographic applications.

We denote by \mathcal{M} a mesh of \mathbb{R}^n consisting of polyhedral bounded convex subsets

of \mathbb{R}^n . \mathcal{M} covers our computational domain Ω . Let $\mathcal{C} \in \mathcal{M}$ be a *control cell*, that is an element of the mesh \mathcal{M} , $\partial\mathcal{C}$ its boundary, and $V_{\mathcal{C}}$ its volume. We assume that the unknown quantity, for example $U(\mathbf{x})$, is constant in \mathcal{C} and denote the value of $U(\mathbf{x})$ on \mathcal{C} by $U^{\mathcal{C}}$.

Integrating the diffusion equation (5.3) over cell \mathcal{C} and using the divergence theorem, we obtain the following equations

$$-\int_{\partial\mathcal{C}} \mathbf{n}_{\mathcal{C}} \cdot \mathcal{D}\nabla U d\gamma(\mathbf{x}) + (\sigma_a^{\mathcal{C}} - \frac{i\omega}{v})V_{\mathcal{C}}U^{\mathcal{C}} = 0, \quad (5.8)$$

where $\mathbf{n}_{\mathcal{C}}(\mathbf{x})$ denotes the outward normal to $\partial\mathcal{C}$ at point $\mathbf{x} \in \partial\mathcal{C}$, $d\gamma(\mathbf{x})$ denotes the surface measure on $\partial\mathcal{C}$, and $\sigma_a^{\mathcal{C}}$ is the value of σ_a on cell \mathcal{C} .

Now we have to approximate the flux through the boundary of \mathcal{C} , i.e., the first integral term in equation (5.8). Let $\{\mathcal{C}_i\}_{i=1}^I$ be the set of neighboring cells of \mathcal{C} . We denote by $S_{\mathcal{C},i}$ the common edge of cell \mathcal{C} and \mathcal{C}_i , i.e., $S_{\mathcal{C},i} = \partial\mathcal{C} \cap \partial\mathcal{C}_i$. We then have

$$-\int_{\partial\mathcal{C}} \mathbf{n}_{\mathcal{C}}(\mathbf{x}) \cdot \mathcal{D}\nabla U d\gamma(\mathbf{x}) = -\sum_i \int_{S_{\mathcal{C},i}} \mathbf{n}_{\mathcal{C}}(\mathbf{x}) \cdot \mathcal{D}\nabla U d\gamma(\mathbf{x}). \quad (5.9)$$

The flux $\int_{S_{\mathcal{C},i}} \mathbf{n}_{\mathcal{C}}(\mathbf{x}) \cdot \mathcal{D}\nabla U d\gamma(\mathbf{x})$ can be approximated by various numerical schemes. In this work, we take a first-order scheme:

$$F^{\mathcal{C},i} := -\int_{S_{\mathcal{C},i}} \mathbf{n}_{\mathcal{C}}(\mathbf{x}) \cdot \mathcal{D}\nabla U d\gamma(\mathbf{x}) = \frac{\mathcal{D}^{\mathbf{nn}} + \mathcal{D}_i^{\mathbf{nn}}}{2} |S_{\mathcal{C},i}| (U^{\mathcal{C}} - U^{\mathcal{C}_i}) / \Delta, \quad (5.10)$$

where $|S_{\mathcal{C},i}|$ is the measure of $S_{\mathcal{C},i}$, Δ is the distance between the center of \mathcal{C} and \mathcal{C}_i . The notation $\mathcal{D}^{\mathbf{nn}}$ denote the \mathbf{nn} component of the diffusion tensor in the $(\boldsymbol{\tau}, \mathbf{n})$ coordinate system on surface $S_{\mathcal{C},i}$. In isotropic scattering case, it is just \mathcal{D} .

When $S_{\mathcal{C},i} \subset \partial\Omega$, we have

$$F^{\mathcal{C},i} := -\int_{S_{\mathcal{C},i}} \mathbf{n}_{\mathcal{C}}(\mathbf{x}) \cdot \mathcal{D}\nabla U d\gamma(\mathbf{x}) = \frac{|S_{\mathcal{C},i}|}{nL_n} (U^{\mathcal{C}} - f). \quad (5.11)$$

We then obtain a full discretization of the diffusion equation

$$\sum_i F^{c,i} + (\sigma_a^c - \frac{i\omega}{v})V_c U^c = 0. \quad (5.12)$$

Let N denote the total number of control cells. After collecting the discretized diffusion equation (5.12) on all control cells, we arrive at the following system of complex-valued algebraic equations

$$\mathbf{A}\mathbf{U} = \mathbf{G}, \quad (5.13)$$

where $\mathbf{A} \in \mathbb{C}^{N \times N}$. The boundary source $\Lambda(f)(\mathbf{x})$, which comes into the discretized system via the flux approximation (5.10) is denoted by \mathbf{G} .

5.3 Numerical results

We provide in this section several numerical experiments where we compare the reconstructions with diffusion and transport equations as the models for photon propagation.

5.3.1 Setup for the reconstructions

For our numerical experiments we consider the three-dimensional cylindrical domain:

$$\Omega := \{(\mathbf{x}, z) : |\mathbf{x}| < 1; 0 < z < 2\}$$

with boundary $\partial\Omega := \{(\mathbf{x}, z) : |\mathbf{x}| = 1; 0 < z < 2\} \cup \{(\mathbf{x}, z) : |\mathbf{x}| < 1; z = 0\} \cup \{(\mathbf{x}, z) : |\mathbf{x}| < 1; z = 2\}$. Here for simplicity, we have used the notation $\mathbf{x} \equiv (x, y)$. We embed a small cylindrical inclusion

$$\Omega_c = \{(\mathbf{x}, z) : |\mathbf{x} - (0.5, 0)| < 0.2, 0.2 < z < 1.8\},$$

into the domain. We show in Fig. 5-1 the XZ cross section of the domain at $y = 0$ and the XZ cross section at $z = 1$. Four point sources are placed on the surface

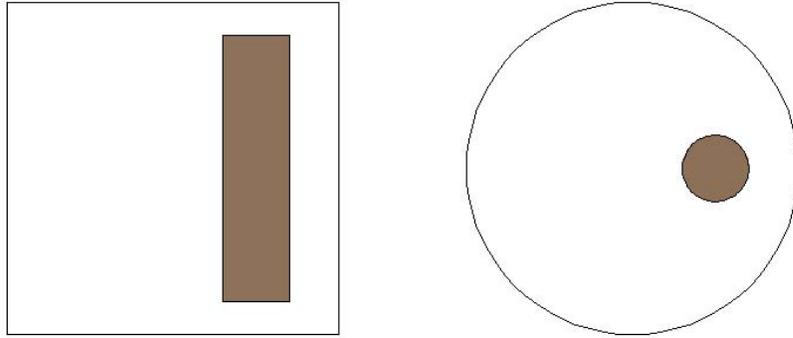


Figure 5-1: XZ ($y = 0$) and XY ($z = 1$) cross-sections of the computational domain.

of the cylinder at $(-1, 0, 1)$, $(-1, 0, 1)$, $(-1, 0, 1)$ and $(-1, 0, 1)$, respectively. All sources are isotropic such that we can use the same description in diffusion as in transport reconstructions. We place seven layers of detectors with z -coordinate given by $z_i = i * 0.25$, $1 \leq i \leq 7$. On each layer, 32 detectors are uniformly distributed on the domain boundary. We partition the domain into 19452 tetrahedral elements. A level symmetric discrete ordinate set [113] has been used to discretize the angular variable in the transport equation. A total number of 120 directions (corresponding to the discretization S_{10}) is used.

All synthetic data are generated with a discretization about twice as fine (in spatial variables) as the discretization used in the inversions to limit the so-called inverse crimes [54].

In the following sections, we consider reconstructions based on *noisy* data. Noise is added to the synthetic data in the following manner. Let $z_k \in \mathbb{C}$ be the k^{th} exact data. We decompose it as $z_k = r_k e^{i\theta_k}$, with r_k a non-negative real number. Then r_k and θ_k are corrupted by noise as $r_k^\delta = (1 + \delta * \text{rand}_{kr})r_k$ and $\theta_k^\delta = (1 + \delta * \text{rand}_{k\theta})\theta_k$. Here, rand_{kr} and $\text{rand}_{k\theta}$ for all measurements $1 \leq k \leq K$ are independent identically distributed random variable on $(-1, 1)$ with uniform distribution (thus with variance $2/3$). The noisy k^{th} data is then defined by $z_k^\delta = r_k^\delta e^{i\theta_k^\delta}$. Note that the same value of δ is chosen here to model noise on the phase and on the amplitude. This allows for a simpler presentation of the effects of noise on the reconstructions. Note also that noise is chosen multiplicative both on the phase and the intensity. Whereas multiplicative noise on the intensity is rather classical, our choice of multiplicative

noise on the phase may be justified as follows. What is measured in practice is the phase shift with respect to the modulation of the source term. In the absence of scattering, photons thus accumulate a phase equal to $\omega l/c$, where l is the traveled distance from the source. In practical optical tomography with modulated sources, ω is at most 1GHz so that $\omega/c \lesssim 3$. Since distances are on the order of a few centimeters, phase shifts $\omega l/c \lesssim 0.1$ in the absence of scattering are a fraction of π . The multiplicative noise on the phase thus implies that errors on the phase shift are larger on measurements away from the source, where phase shift is significant, than in the vicinity of the source, where it is close to 0.

The quality of the reconstructions is measured as follows. Denote by M^e and M^r the exact and reconstructed absorption coefficients, respectively. We then define the relative l^2 error between M^e and M^r by:

$$\mathcal{E}_{l^2} = \frac{\|M^r - M^e\|_{l^2}}{\|M^e\|_{l^2}}. \quad (5.14)$$

5.3.2 Diffusive media of small size

We compare reconstructions with diffusion and transport models in media of small size. Because the media are relatively small, the optical separation between the sources and the detectors is also relatively small. Photons undergo only a small number of scattering events between a source and a detector. It is well known that the diffusion approximation to the transport equation becomes less accurate in describing particles propagation when such small tissue volumes are considered. What we want to study in the section is how these inaccuracies affect the tomographic reconstructions in such media.

We consider the following setup. The background optical properties in the big cylinder is given by $\sigma_a = 0.1 \text{ cm}^{-1}$, while $\sigma_a = 0.2 \text{ cm}^{-1}$ for the inclusion, which is twice as high the absorption coefficient of the background. We set the scattering coefficient $\sigma_s = 100 \text{ cm}^{-1}$ for the whole domain. The anisotropic factor for the scattering kernel is $g = 0.9$ so that $\sigma'_s = 10 \text{ cm}^{-1}$. The modulation frequency is set to $\omega = 0.0$ (steady state).

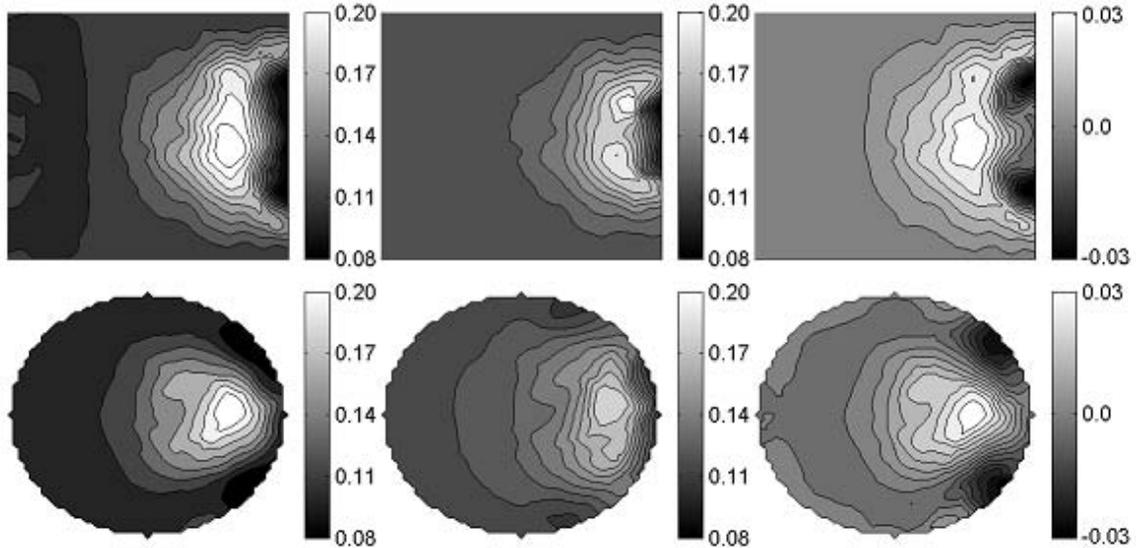


Figure 5-2: Top row: XZ ($y = 0$) cross-sections of the reconstructed absorption coefficients with transport equation (left), diffusion equation (middle), and their difference (right). Bottom row: same as top row but for XY ($z = 1$) cross sections. The reconstruction are done with noise-free data.

We show in Fig. 5-2 cross-sections of the reconstructed absorption coefficient based on transport and diffusion equations with noise-free transport data. To stress the difference in the reconstruction, we also plot the difference of the reconstruction in Fig. 5-2.

We first observe that structures in the z -direction are not well reconstructed in either case. This is because light sources are all located on the $z = 1$ plane. Few photons propagate sufficiently far along the z -direction. We have verified that adding sources on other planes along the z -axis makes the reconstructions better, as expected.

Our second observation is that the location of the inhomogeneity is found by both transport and diffusion reconstructions. However, the transport-based reconstruction provides a more accurate value of the actual optical properties of the inclusion. The diffusion reconstruction over-estimates the volumes of the absorption coefficient in some places and under estimates the coefficients in the inclusion. This can be best seen on the right figures in Fig. 5-2 where we plot the difference between transport- and diffusion-based reconstructions.

Our numerical examples show that the difference between the transport and diffu-

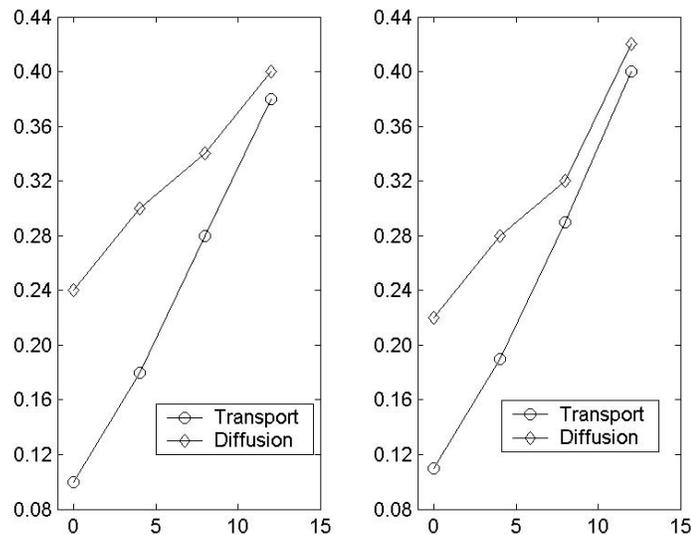


Figure 5-3: Quality in transport and diffusion reconstructions using data with different noise levels (in percentage). Left: reconstructions with scattering coefficient $\sigma_s = 100\text{cm}^{-1}$; Right: reconstructions with scattering coefficient $\sigma_s = 150\text{cm}^{-1}$.

sion reconstructions becomes less prominent as noise contained in the measured data increases. This can be seen on the left in Fig. 5-3, where we plot the quantitative error in the reconstructions as a function of noise level. As noise level reaches a certain value, above 12% in this case, the difference between transport and diffusion reconstructions becomes almost indistinguishable.

We have performed a second group of simulations where we increase the background scattering coefficient to $\sigma_s = 150\text{ cm}^{-1}$ so that $\sigma'_s = 15\text{ cm}^{-1}$. The reconstruction quality from noisy data is shown in Fig. 5-3. We see that although the medium is now 50% more diffusive, reconstructions based on the diffusion approximation do not significantly improve, because the domain still remains relatively thin optically. For typical values of the absorption and scattering parameters in tissues, the diffusion approximation is not very accurate in small domains such as fingers or small animals.

The last point we stress here is that transport-based reconstructions are computationally much more costly than diffusion-based reconstructions. Typically, we observe that diffusion reconstructions are about 60 times faster than transport reconstructions. Although the computational speed really depends on how one discretize

the problem, we have observed in most cases an acceleration factor of at least 40 in diffusion reconstructions.

5.3.3 Effects of modulation frequency

As we have remarked before, in the frequency domain, the diffusion approximation works only for not-to-high modulation frequency [72]. Essentially, one has to scale the modulation frequency ω to ω/ϵ as one scale the absorption coefficient in order to derive the correct diffusion approximation in the limit of small mean free paths. This is similar to the scaling of the time variable in time-dependent case considered e.g. in [58]. In practice, however, relatively high modulation frequencies need to be used to obtain a significant phase shift that can be measured.

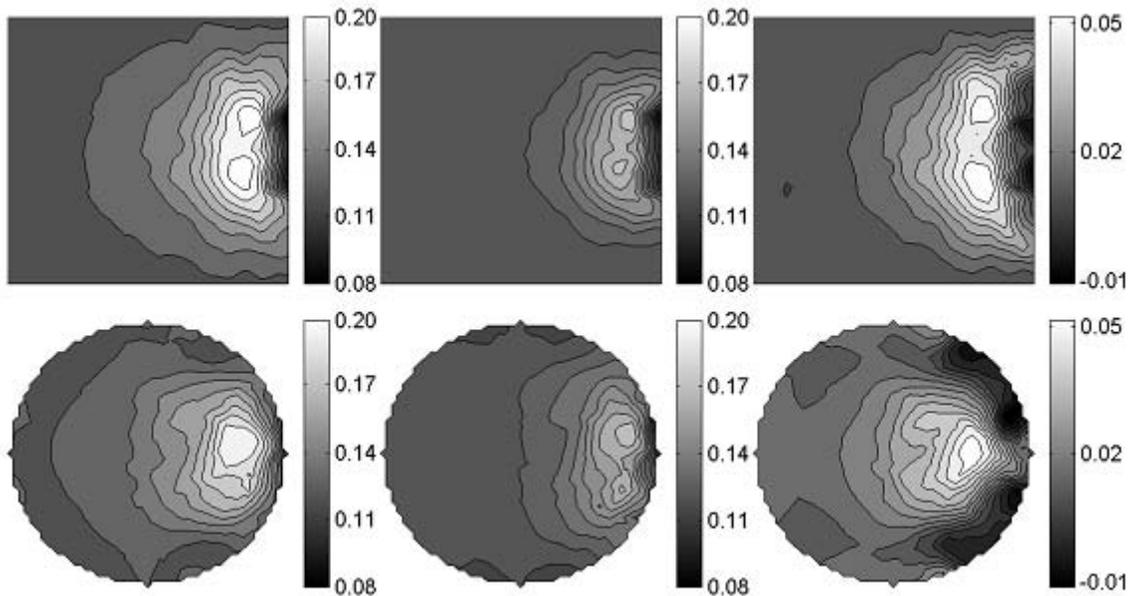


Figure 5-4: Top row: XZ ($y = 0$) cross-sections of reconstructed absorption coefficients with transport equation (left), diffusion equation (middle), and their difference (right). Bottom row: same as top row but for XY ($z = 1$) cross sections. The reconstruction are done with noise-free data. Modulation frequency for the sources is $\omega = 0.8\text{GHz}$.

We consider here the same reconstructions as those of the last section though with measured data obtained at different modulation frequencies. We show in Fig. 5-4 reconstructions with modulation frequency of $\omega = 0.8\text{GHz}$. Again we observe that

the shape on the z -direction is not well reconstructed in both cases. The location of the inhomogeneity is found by both the transport and the diffusion reconstructions.

As in the previous section, the transport-based reconstruction provides more accurate values of the actual optical properties of the inclusion. The differences of the two have been plotted on the right figures in Fig. 5-4.

As expected, the difference between diffusion- and transport-based results increases with the modulation frequency, as can be seen by comparing results in Fig. 5-4 and Fig. 5-2. This can also be seen from Fig. 5-5 where we plot the quality of reconstructions against the modulation frequency. Four modulation frequencies has been considered. They are 0.2 GHz, 0.4GHz, 0.6 GHz and 0.8 GHz. The quality of transport reconstructions slightly increases as the modulation frequency increases, but the quality of the diffusion reconstructions decreases as the modulation frequency increases.

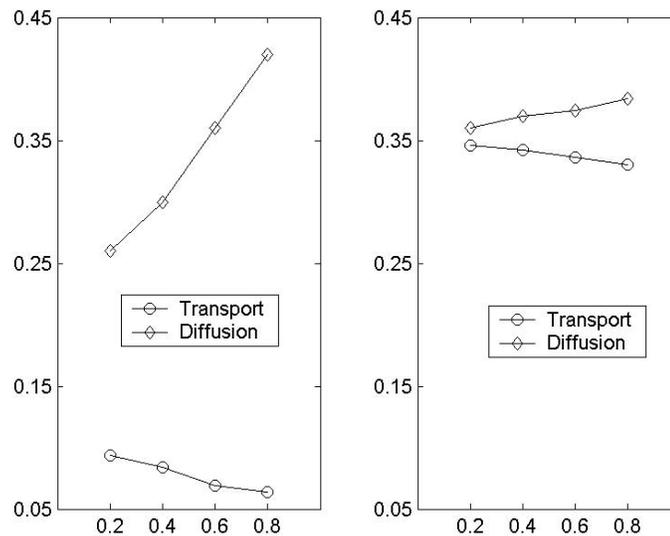


Figure 5-5: Quality of reconstructions as functions of modulation frequencies (in unit of GHz). Left: reconstructions with noise-free data; Right: reconstructions with 12% noise in the data.

As noise in the data increases, the difference between diffusion- and transport-based reconstructions becomes smaller. We show in the right plot of Fig. 5-5 the quality of reconstructions with 12% multiplicative noise. Although there is still a

difference between the two reconstructions (especially in the high frequency cases), the difference is much smaller than in the case of noise-free data.

Computationally, increasing the modulation frequency results in an increase of the computational time used to solve the inverse problem. This is due to the fact that modulation frequency appears on the off-diagonal elements of the matrices derived from the discretization of the equations. Increasing the frequency increases the condition number of the matrices. However, even in the high frequency situation, we still observe that transport-based reconstructions are about 50 times slower than diffusion-based reconstructions.

5.3.4 The impact of the extrapolation length

The choice of the extrapolation length in the diffusion equation has a significant influence on the solution of the diffusion equation, especially near the boundary; see for example [17, 58] and references there in. We study in this section the effect of the extrapolation length on the quality of the reconstructions.

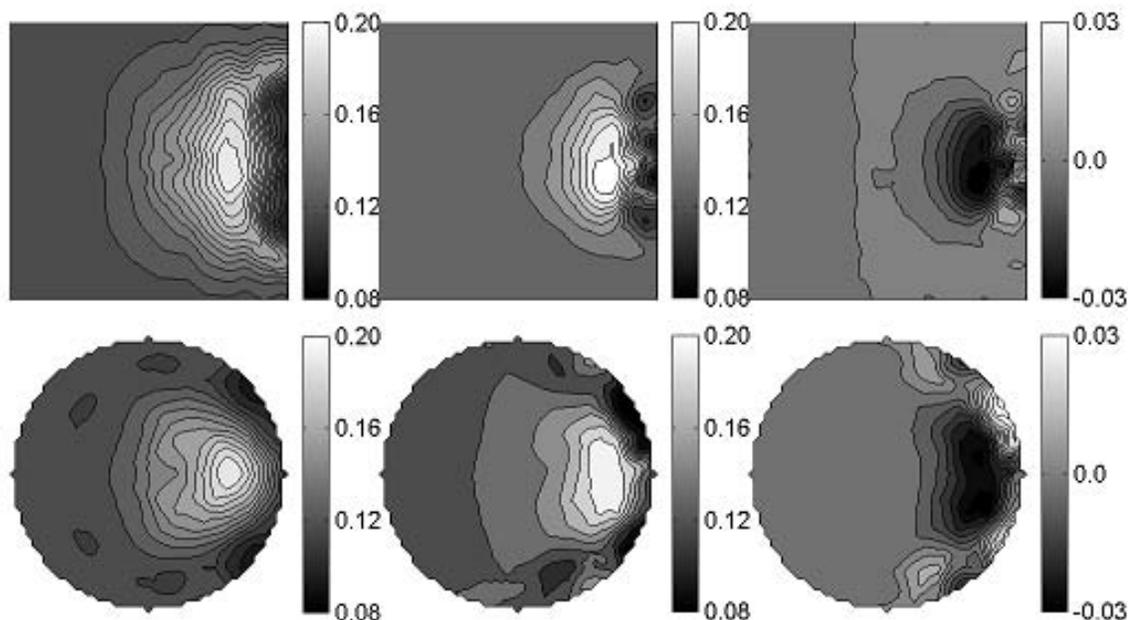


Figure 5-6: Top row: XZ ($y = 0$) cross-sections of reconstructed absorption coefficients with transport equation (left), diffusion equation (middle), and their difference (right). Bottom row: same as top row but for XY ($z = 1$) cross sections. The reconstruction are done with noise-free data.

All simulations in this section are done with isotropic scattering. In other words, we have set the anisotropic factor to $g = 0$. The scattering coefficient is set to be $\sigma_s = 10 \text{ cm}^{-1}$. We show in Fig. 5-6 the reconstructions using the transport equation and diffusion equation with extrapolation length $L_3 = 0$. The data here are noise-free.

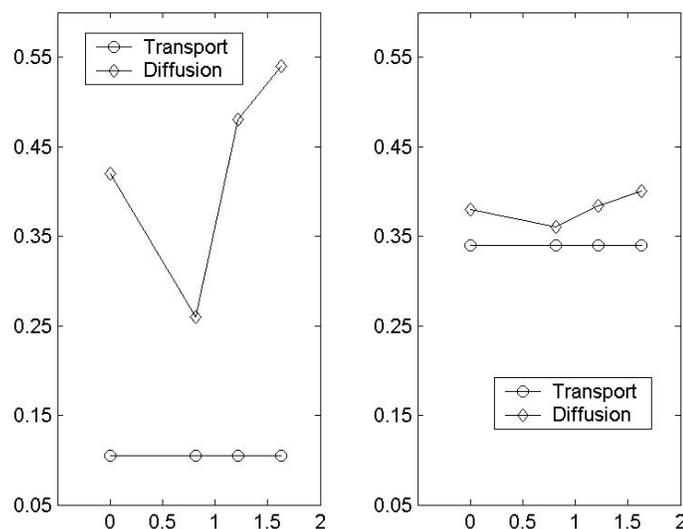


Figure 5-7: Quality of reconstructions as functions of extrapolation length. Left: reconstructions with noise-free data; Right: reconstructions with 12% noise in the data. Transport reconstructions are shown here just as a reference.

We see from Fig. 5-6 that there is a significant amount of overshooting in the diffusion reconstructions. The quality of the diffusion-based reconstruction also decreases when very large extrapolation lengths are used. In Fig. 5-7 we compare the quality of reconstructions by diffusion equations with various extrapolation lengths. Although the least difference between diffusion and transport reconstructions may not happen exactly at the place of right extrapolation length, it does happen when a value close to the right value is chosen. We thus conclude that that extrapolation length does have a significant impact on the quality of reconstructions.

As usual, noise in the data plays a significant role. The difference between transport-based and diffusion-based is already very small when the noise level reaches 12%. One would expect that as noise increases, the difference would become indistinguishable again. Computationally, we observe that transport-based reconstructions

are still about 50 times slower than diffusion-based reconstructions again.

5.3.5 Diffusive media with void regions

The last case we want to discuss is the situation when non-scattering void regions are present in the domain. It has been shown in various situations that when void region presents in scattering media, diffusion equations fail to approximate transport accurately [17, 23, 90]. Special attention has to be paid when using diffusion equations in this situation.

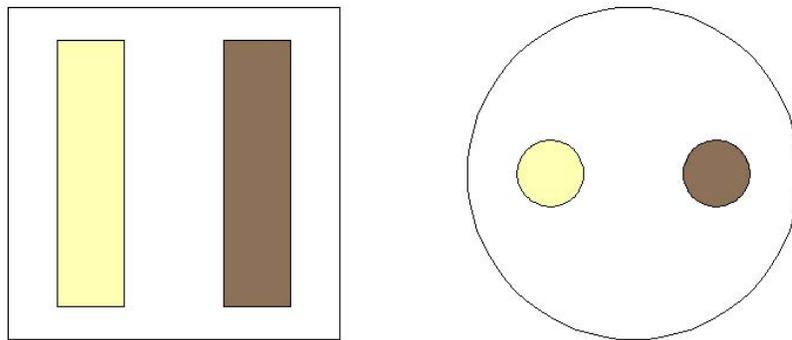


Figure 5-8: XZ ($y = 0$) and XY ($z = 1$) cross-sections of the computational domain with a void inclusion.

We again consider here the reconstruction of the absorption coefficient in the cylinder and assume that $\sigma_s(\mathbf{x})$ is known. We embed a void cylindrical inclusion centered at $(-0.5, 0)$ in the media. It is of the same size as the absorbing inclusion; see Fig. 5-8 for the geometrical setting. Void means that the scattering and absorption coefficients vanish in that region. As in the last section, we set the scattering coefficient $\sigma_s = 10 \text{ cm}^{-1}$, and anisotropy factor $g = 0$ in the rest of the domain. In the diffusion equation, we replace the diffusion coefficient in the void region by its surrounding diffusion coefficient. In the absence of a better guess, this is better than evaluating the diffusion coefficient $\mathcal{D} = 1/(3\sigma_t)$ as being infinite, though better choices yet may be available; see e.g. [18, 23].

We show in Fig. 5-9 results from reconstructions with noise-free data. The transport reconstruction looks quite similar to the one in Fig. 5-2. The diffusion reconstruction however, looks very different. The diffusion model generates spurious absorption

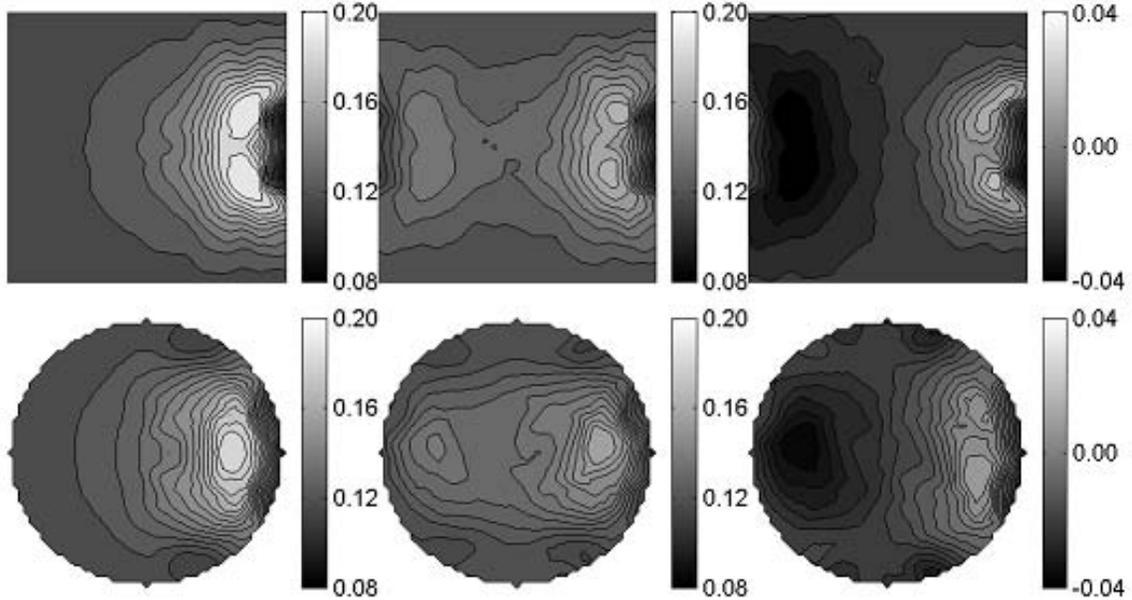


Figure 5-9: Top row: XZ ($y = 0$) cross-sections of reconstructed absorption coefficients with transport equation (left), diffusion equation (middle), and their difference (right). Bottom row: same as top row but for XY ($z = 1$) cross sections. The reconstruction are done with noise-free data. A void region is embedded in the domain.

at the location of the void to compensate for the wrong transport of photons in that area. Whereas voids have little effect on the absorption reconstruction with the transport model (provided that we know where the void is), they further degrade diffusion reconstructions unless the void region is modeled appropriately; see [12, 18, 23, 59, 71].

The transport and diffusion reconstructions in the presence of a void have been performed with different noise levels in the measured data. The quality of the reconstructions is plotted against the noise level in Fig. 5-10. As noise increases, the difference between transport and diffusion equations decreases. One can expect that when the noise in the data reaches a certain level, here about 12% of multiplicative noise, the difference between the transport and diffusion reconstructions may become indistinguishable. Similar results for reconstructions with a background scattering coefficient $\sigma_s = 15 \text{ cm}^{-1}$ are shown in the right panel of Fig. 5-10.

From the viewpoint of computational cost, the transport-based reconstructions become slightly faster here because of the presence of the void region. But it is far from

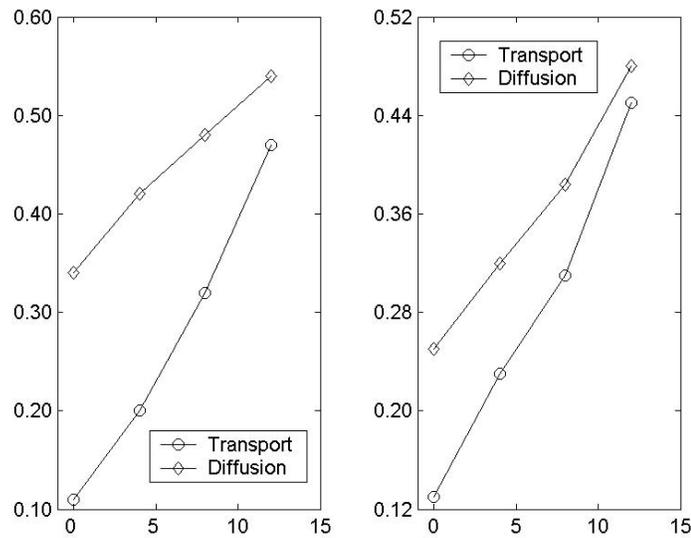


Figure 5-10: Quality of transport and diffusion reconstructions using data with different noise levels in the presence of a void. Left: reconstructions with scattering coefficient $\sigma_s = 10 \text{ cm}^{-1}$; Right: reconstructions with scattering coefficient $\sigma_s = 15 \text{ cm}^{-1}$. The anisotropy factor $g = 0$ in both cases.

being comparable to diffusion reconstructions. We still observe that transport-based reconstructions are about 40 times slower than diffusion-based reconstructions.

5.4 Conclusions and remarks

We have conducted a comparative study of optical tomographic reconstructions based on transport and diffusion models in media of small (optical) volume. We have shown that diffusion-based reconstructions were significantly less accurate in such geometries. Although both the diffusion and transport reconstructions are usually able to locate an inhomogeneity buried in the media, transport-based reconstructions provide more accurate values of the optical properties, in this paper the absorption, of inclusions.

We have quantified the role on the reconstruction of such factors as the source modulation frequency, the extrapolation length used in the diffusion model, and the presence of void regions. Whereas increasing the source modulation frequency im-

proves transport solutions, it usually degrades the diffusion solutions. We have seen that poorly modeled void inclusions in the diffusion model generated spurious absorbing inclusions at the void location. In each situation, we have quantified the errors made by the transport and diffusion reconstructions.

We have characterized the effects of noise in the measured data. When multiplicative noise reaches a certain level, about 12% in our simulations, the transport and diffusion reconstructions becomes almost indistinguishable. However, given that state-of-the-art optical imaging system show noise levels of typically less than 5% or even 1%, the benefits of transport-based reconstructions can be realized with most of the currently available systems [89, 92, 143, 161, 177]. Computationally however, the diffusion-based reconstructions are always extremely favorable. We have consistently observed that transport-based reconstructions were about 50 times more expensive than diffusion-based reconstructions. Because of their much more accurate properties in the presence of moderate noise levels, we recommend the use of transport-based reconstructions in small optical domains in spite of their computational cost. Applications for such reconstructions are imaging of joint diseases in human fingers [92, 108, 131, 145, 184] and monitoring of brain activity and tumor growth in small animals [35, 80, 89].

Chapter 6

Generalized diffusion approximation and its validations

There are two important cases in typical optical tomography applications where classical diffusion approximations do not work well. The first case is when the tissue of interest is relatively small, as we have pointed out in the previous chapter. The other case is when the highly scattering tissue is large but there are extended non-scattering regions embedded in the tissue. Diffusion approximation does not hold in those non-scattering regions but holds in the rest of the tissue. In this case, if one still want to use diffusion equations, one has to modify the diffusion equation to take into account the effect of these non-scattering regions. This is a modeling problem and is the subject of study of this chapter. The presentation of this chapter is based on reference [23].

6.1 Problem statement

Among other applications, optical tomography is being considered as an interesting technique to image tumors in human head and monitor cerebral oxygenation in neonates. Since the most of the tissues in human head are highly scattering and low absorbing, one would expect that diffusion equations can be used as the light propagation model these applications. This almost true except for the fact that there exists in the head a thin layer filled with cerebrospinal fluid. This layer is almost collision-less and absorption-less. Diffusion models perform very poorly in such lay-

ers [13, 59, 90, 149]. These models have to be modified if one wants to image the oxygenation in the head of neonates for instance.

One could certainly solve phase-space transport equations instead of the inaccurate diffusion equations [85, 90]. There exists a large literature on numerical techniques that allow us to use coarser schemes (modeling transport or diffusion equations) in the regions where multiple scattering makes the simulation relatively straightforward and finer schemes in the vicinity of the clear layer where transport effects must be calculated accurately [22, 26, 77, 112, 175].

Because clear layers are thin in practice, an alternative solution exists to solving transport equations. Arridge and his collaborators have developed several hybrid models that would solve a diffusion equation where the tissues are highly scattering and model the transport behavior in the clear layer [13, 59, 149]. Similar models were developed using an approach based on the asymptotic expansion of transport equations by one of the authors [17, 18]. The models that come out of this research have the following common features. 1) They are diffusion equations with matching conditions at the boundary of the layer that account for the guiding effect of the clear layer. 2) These matching conditions take the form of non-local interface conditions for the photon density and current. 3) These models are quite accurate in practice, both to solve forward and inverse problems. 4) Their computational cost is much lower than that of full transport. 5) However it is often significantly higher than the cost of classical diffusion and the models are quite complicated to implement in practice.

In this chapter we build on the asymptotic expansion techniques in [17, 18] to propose a new model that accurately models the clear layer effects for a computational cost and an implementation by finite element method that is essentially the same as that of classical diffusion. The model is obtained by localizing the interface conditions at the clear layer. This local interface condition models a tangential diffusion process that accounts for the propagation of photons along the clear layer. In variational form, this diffusion process is about as easy to solve as the classical volume diffusion process.

Let us start with the following steady-state linear radiative transfer equation

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla u(\mathbf{x}, \boldsymbol{\theta}) + \sigma_a(\mathbf{x})u(\mathbf{x}, \boldsymbol{\theta}) + Q(u)(\mathbf{x}, \boldsymbol{\theta}) &= S(\mathbf{x}), & \text{in } \Omega \times S^{n-1} \\ u(\mathbf{x}, \boldsymbol{\theta}) &= g(\mathbf{x}, \boldsymbol{\theta}) & \text{on } \Gamma_- = \{(\mathbf{x}, \boldsymbol{\theta}) \in \partial\Omega \times S^{n-1} \text{ s.t. } \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) < 0\}. \end{aligned} \quad (6.1)$$

Here, $u(\mathbf{x}, \boldsymbol{\theta})$ is the photon flux intensity at point $\mathbf{x} \in \Omega$, where Ω is a subset in \mathbb{R}^n , with direction of propagation $\boldsymbol{\theta} \in S^{n-1}$, where S^{n-1} is the unit sphere in \mathbb{R}^n . The space dimension is $n = 3$ in practice. In this paper we consider $n = 2$ because it is computationally simpler and $n = 3$. The source of photons at the boundary of the domain is given by $g(\mathbf{x}, \boldsymbol{\theta})$ and the outward unit normal to the domain is $\boldsymbol{\nu}(\mathbf{x})$ at $\mathbf{x} \in \partial\Omega$. The volume source of photons is given by $S(\mathbf{x})$. The absorption coefficient is denoted by $\sigma_a(\mathbf{x})$, and the scattering operator Q is defined as

$$Q(u)(\mathbf{x}, \boldsymbol{\theta}) = \sigma_s(\mathbf{x}) \left(u(\mathbf{x}, \boldsymbol{\theta}) - \int_{S^{n-1}} u(\mathbf{x}, \boldsymbol{\theta}') d\mu(\boldsymbol{\theta}') \right). \quad (6.2)$$

Here, $\sigma_s(\mathbf{x})$ is the scattering coefficient and $d\mu$ is the surface measure on S^{n-1} normalized so that $\int_{S^{n-1}} d\mu(\boldsymbol{\theta}) = 1$.

In two space dimensions, we parameterize $\boldsymbol{\theta} = (\cos \theta, \sin \theta)$ and have

$$\int_{S^1} u(\mathbf{x}, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) = \frac{1}{2\pi} \int_0^{2\pi} u(\mathbf{x}, \theta) d\theta,$$

identifying $u(\mathbf{x}, \boldsymbol{\theta})$ with $u(\mathbf{x}, \theta)$. In three space dimensions, we parameterize $\boldsymbol{\theta} = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta)$ and have

$$\int_{S^2} u(\mathbf{x}, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) = \frac{1}{4\pi} \int_0^{2\pi} \int_0^\pi u(\mathbf{x}, \theta, \phi) \sin \theta d\theta d\phi,$$

identifying $u(\mathbf{x}, \boldsymbol{\theta})$ with $u(\mathbf{x}, \theta, \phi)$.

The optical tomography problem consists of reconstructing $\sigma_a(\mathbf{x})$ and $\sigma_s(\mathbf{x})$ from boundary measurements $u(\mathbf{x}, \boldsymbol{\theta})$ for $\mathbf{x} \in \partial\Omega$ and $\boldsymbol{\theta} \in S^{n-1}$. This is quite a difficult problem both in theory and in practice [11, 52, 62, 106, 130, 173]. This problem is also very expensive computationally because the radiative transfer equations are posed in the phase space, with a minimum of three spatial variables and two angular

variables in practical calculations. They are therefore often replaced by their diffusion approximation, which does not involve any angular variable.

Diffusion approximations are valid in the regime of high scattering $\sigma_s \gg 1$ and small absorption $\sigma_a \ll 1$. We can then approximate the solution $u(\mathbf{x}, \boldsymbol{\theta})$ by

$$u(\mathbf{x}, \boldsymbol{\theta}) = U(\mathbf{x}) - \frac{1}{\sigma_s(\mathbf{x})} \boldsymbol{\theta} \cdot \nabla U(\mathbf{x}) + \text{smaller order terms}, \quad (6.3)$$

where $U(\mathbf{x})$ is the solution to the following diffusion equation

$$\begin{aligned} -\nabla \cdot \mathcal{D}(\mathbf{x}) \nabla U(\mathbf{x}) + \sigma_a(\mathbf{x}) U(\mathbf{x}) &= S(\mathbf{x}) && \text{in } \Omega \\ U(\mathbf{x}) + \varepsilon n L_n \mathcal{D}(\mathbf{x}) \boldsymbol{\nu}(\mathbf{x}) \cdot \nabla U(\mathbf{x}) &= g(\mathbf{x}) && \text{on } \partial\Omega, \end{aligned} \quad (6.4)$$

where we assume that $g(\mathbf{x}, \boldsymbol{\theta}) = g(\mathbf{x})$ does not depend on $\boldsymbol{\theta}$ to simplify, and where the diffusion coefficient is defined by

$$\mathcal{D}(\mathbf{x}) = \frac{1}{n(\sigma_a(\mathbf{x}) + \sigma_s(\mathbf{x}))}, \quad n = 2, 3. \quad (6.5)$$

The extrapolation length L_n accounts for the leakage of photons at the domain boundary. Approximate values are $L_2 = 0.8164$ and $L_3 = 0.7104$ for isotropic scattering [17, 21, 58, 118]. Diffusion equations are very well studied both mathematically and physically and can be justified by various means [11, 58, 111].

The diffusive regime is valid in most human tissues, where absorption is relatively small and scattering quite large, with typical values of the order of $\sigma_a = 0.1\text{cm}^{-1}$ and $\sigma_s = 20\text{cm}^{-1}$. This correspond to an absorption mean free path of 10cm and a scattering mean free path of 0.05cm. Notice that $\mathcal{D}(\mathbf{x})$ and $\sigma_a(\mathbf{x})$ in (6.4) are then of comparable order.

The presence of cerebrospinal fluid in the human head prevents the use of the classical diffusion equation (6.4). The reason is that this fluid is optically clear: photons propagate along straight lines almost scattering-free in such fluids. This creates a guiding effect that the diffusion equation (6.4) cannot capture. Several works exist to understand and fix this problem [17, 18, 59, 71, 149]. The main idea

consists of using the diffusion equation where it is valid and coupling it with local transport in the non-scattering regions. An asymptotic analysis [18] justifies such an approach for thin clear inclusions. Several such hybrid models have been analyzed numerically [17]. This analysis shows the adequacy and robustness of the models. The main difficulty is that their numerical implementation is still difficult and their cost significantly higher than that of classical diffusion (6.4), although much lower than that of the full transport equation (6.1). It is the objective of this paper to further simplify the hybrid model and obtain a scheme that is both accurate and computationally efficient.

6.2 Generalized diffusion model

Following an earlier asymptotic derivation [18], we propose here what we believe is the simplest model that captures both the diffusive behavior outside of the clear layer and the guiding effect within the clear layer. It is based on solving a diffusion equation with local jump conditions at the clear layer.

6.2.1 Notation and Geometry.

The geometry of the clear layer Ω^C is as follows. We define Σ as a closed smooth surface embedded in Ω and

$$\Omega^C = \{\mathbf{y} \in \Omega \text{ s.t. } \mathbf{y} = \mathbf{x} + t\boldsymbol{\nu}(\mathbf{x}), \quad \text{where } \mathbf{x} \in \Sigma \text{ and } |t| < L\}. \quad (6.6)$$

Here L is a fixed sufficiently small number and $\boldsymbol{\nu}(\mathbf{x})$ is the outward normal to (the volume inside) Σ at $\mathbf{x} \in \Sigma$. We denote by Σ^E and Σ^I the outer and inner surfaces of Ω^C and assume that these surfaces are smooth; see Fig. 6-1. We define $\boldsymbol{\nu}_C(\mathbf{x})$ as the outward unit normal to Ω^C at a point $\mathbf{x} \in \partial\Omega^C = \Sigma^E \cup \Sigma^I$. For $\mathbf{x} \in \Sigma$, we define

$$\mathbf{x}^E = \mathbf{x} + L\boldsymbol{\nu}(\mathbf{x}) \in \Sigma^E \quad \text{and} \quad \mathbf{x}^I = \mathbf{x} - L\boldsymbol{\nu}(\mathbf{x}) \in \Sigma^I. \quad (6.7)$$

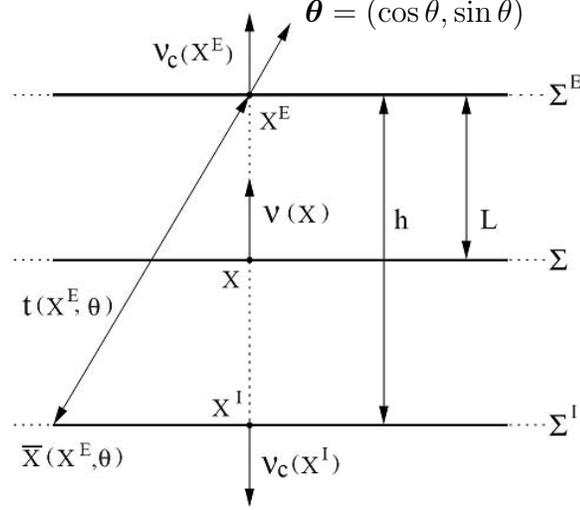


Figure 6-1: Local geometry of the clear layer.

It is useful to see \mathbf{x}^E and \mathbf{x}^I as functions of $\mathbf{x} \in \Sigma$. Notice that the outward normal to Ω^C at $\mathbf{x}^E \in \Sigma^E$ is $\boldsymbol{\nu}_C(\mathbf{x}^E) = \boldsymbol{\nu}(\mathbf{x})$ and the outward normal at $\mathbf{x}^I \in \Sigma^I$ is $\boldsymbol{\nu}_C(\mathbf{x}^I) = -\boldsymbol{\nu}(\mathbf{x})$.

The solution operator to the radiative transfer equation in Ω^C is denoted by \mathcal{R}^C . Let us define

$$\Gamma_{\pm}^C = \{(\mathbf{x}, \boldsymbol{\theta}) \in \partial\Omega^C \times S^{n-1} \text{ s.t. } \pm \boldsymbol{\theta} \cdot \boldsymbol{\nu}_C(\mathbf{x}) > 0\},$$

and consider the problem inside the layer

$$\begin{aligned} \boldsymbol{\theta} \cdot \nabla v(\mathbf{x}, \boldsymbol{\theta}) + \sigma_a(\mathbf{x})v(\mathbf{x}, \boldsymbol{\theta}) + Q(v)(\mathbf{x}, \boldsymbol{\theta}) &= 0, & \text{in } \Omega^C \times S^{n-1} \\ v(\mathbf{x}, \boldsymbol{\theta}) &= g(\mathbf{x}, \boldsymbol{\theta}) & \text{on } \Gamma_-^C. \end{aligned}$$

We then define \mathcal{R}^C as the operator that maps $g(\mathbf{x}, \boldsymbol{\theta})$ on Γ_-^C to $v(\mathbf{x}, \boldsymbol{\theta})|_{\Gamma_+^C} = \mathcal{R}^C g(\mathbf{x}, \boldsymbol{\theta})$, the restriction to the transport solution $v(\mathbf{x}, \boldsymbol{\theta})$ to the outgoing surface (in the phase space) Γ_+^C . Such an operator is well defined in suitably chosen weighted L^p spaces [58].

We now define the operator \mathcal{R}_1^C by

$$\mathcal{R}_1^C = \mathcal{R}^C - \mathcal{I}, \tag{6.8}$$

where the near-identity operator \mathcal{I} is defined from Γ_-^C to Γ_+^C by

$$\mathcal{I}u(\mathbf{x}, \boldsymbol{\theta}) = \begin{cases} u(\mathbf{x} + 2L\boldsymbol{\nu}(\mathbf{x}), \boldsymbol{\theta}), & \text{when } \mathbf{x} \in \Sigma^I \\ u(\mathbf{x} - 2L\boldsymbol{\nu}(\mathbf{x}), \boldsymbol{\theta}), & \text{when } \mathbf{x} \in \Sigma^E. \end{cases}$$

This near-identity operator \mathcal{I} is merely a translation from the inner boundary to the outer boundary and vice versa. This is an approximation to what happens to most photons that cross the clear layer: since the clear layer is optically thin (because σ_s is small in Ω^C) and most photons reach the clear layer at \mathbf{x}^I with an incidence angle far from orthogonal to $\boldsymbol{\nu}(\mathbf{x})$, they do not propagate for a long time in the clear layer and exit it at a point near \mathbf{x}^E . For those relatively rare photons that reach the clear layer with a direction almost orthogonal to $\boldsymbol{\nu}(\mathbf{x})$, the exit point will no longer be \mathbf{x}^E . This behavior is captured by \mathcal{R}_1^C and modifies the local current equilibrium.

6.2.2 Generalized diffusion equation with non-local interface conditions.

It was shown [18, p.1687] that a good approximation of $u(\mathbf{x}, \boldsymbol{\theta})$, the solution to (6.1), was given by $U(\mathbf{x})$ solution of

$$\begin{aligned} -\nabla \cdot \mathcal{D}(\mathbf{x})\nabla U(\mathbf{x}) + \sigma_a(\mathbf{x})U(\mathbf{x}) &= S(\mathbf{x}) && \text{in } \Omega \setminus \Omega^C \\ U(\mathbf{x}) + \varepsilon n L_n \mathcal{D}(\mathbf{x})\boldsymbol{\nu}(\mathbf{x}) \cdot \nabla U(\mathbf{x}) &= g(\mathbf{x}) && \text{on } \partial\Omega \\ U(\mathbf{x}^E) &= U(\mathbf{x}^I) && \text{on } \Sigma \\ \boldsymbol{\nu}(\mathbf{x}) \cdot \mathcal{D}(\mathbf{x}^E)\nabla U(\mathbf{x}^E) - \boldsymbol{\nu}(\mathbf{x}) \cdot \mathcal{D}(\mathbf{x}^I)\nabla U(\mathbf{x}^I) &= KU(\mathbf{x}) && \text{on } \Sigma, \end{aligned} \tag{6.9}$$

where the integral operator K is given by

$$\begin{aligned} KU(\mathbf{x}) &= \int_{\Gamma_+(\mathbf{x}^E)} \boldsymbol{\theta} \cdot \boldsymbol{\nu}_C(\mathbf{x}^E) (\mathcal{R}_1^C U)(\mathbf{x}^E, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) \\ &+ \int_{\Gamma_+(\mathbf{x}^I)} \boldsymbol{\theta} \cdot \boldsymbol{\nu}_C(\mathbf{x}^I) (\mathcal{R}_1^C U)(\mathbf{x}^I, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}). \end{aligned} \tag{6.10}$$

We have defined $\Gamma_+(\mathbf{x}) = \{\boldsymbol{\theta} \in S^{n-1} \text{ s.t. } \boldsymbol{\theta} \cdot \boldsymbol{\nu}_C(\mathbf{x}) > 0\}$ and have implicitly used that \mathbf{x}^E and \mathbf{x}^I defined in (6.7) are functions of $\mathbf{x} \in \Sigma$. Notice that this diffusion

problem is posed on $\Omega \setminus \Omega^C$. What happens inside the layer Ω^C is modeled by the operator \mathcal{R}_1^C in the definition of K . The two jump conditions in (6.9) indicate the boundary conditions satisfied by U at the boundary $\partial\Omega^C$. It was shown [18] that the above problem (6.9) was well-posed provided that the thickness of the layer L was sufficiently small. Numerical simulations based on (6.9) and on similar generalized diffusion models [17, 18] have shown the accuracy of the approximation.

The physical interpretation of the jump conditions is the following. The jump of the total flux vanishes, $U(\mathbf{x}^E) = U(\mathbf{x}^I)$, because the clear layer is not sufficiently thick to modify this equilibrium. However, it is sufficiently large to modify the current balance. The difference of currents crossing the interfaces of the clear layer is balanced by the current of photons inside the clear layer. The latter is modeled by $KU(\mathbf{x})$. As a minor remark, let us mention that the asymptotic expansion [18] involves an additional Jacobian term corresponding to the map \mathcal{I} . Since \mathcal{I} is near-identity, we have replaced the Jacobian by 1. Accounting for this Jacobian does not change the limiting equations that will be obtained below.

6.2.3 Localization of the interface conditions.

We now aim at further simplifying (6.9) by replacing the nonlocal operator K in (6.10) by its local approximation. In doing so, we will model the clear layer Ω^C by a local jump condition for the diffusion solution at Σ . We assume that the clear layer is totally non-scattering, i.e. that $\sigma_s(\mathbf{x}) = 0$ for $\mathbf{x} \in \Omega^C$. This assumption is fairly accurate in practice. All the results we present below are not significantly modified when the layer is weakly scattering; see our remarks at the end of the section.

Let us consider the two-dimensional case $n = 2$. Let $(\mathbf{x}, \boldsymbol{\theta}) \in \Gamma_+^C$. We define $t(\mathbf{x}, \boldsymbol{\theta})$ as the time it takes to travel from Γ_-^C to \mathbf{x} in direction $-\boldsymbol{\theta}$ (with unit speed). We also define $\bar{\mathbf{x}} = \bar{\mathbf{x}}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{x} - t(\mathbf{x}, \boldsymbol{\theta})\boldsymbol{\theta}$, the starting point on Γ_-^C . Since the clear layer is non-scattering, we obtain by solving the free transport equation along its characteristics that

$$\mathcal{R}^C U(\mathbf{x}, \boldsymbol{\theta}) = e^{-\sigma_a t(\mathbf{x}, \boldsymbol{\theta})} U(\bar{\mathbf{x}}),$$

assuming that absorption is constant in Ω^C . Let us consider a point $\mathbf{x}^I \in \Sigma^I$ such that $\bar{\mathbf{x}}(\mathbf{x}^I, \theta) \in \Sigma^E$ for all θ such that $\boldsymbol{\nu}_C(\mathbf{x}^I) \cdot \boldsymbol{\theta} > 0$. This means that the photons reaching \mathbf{x}^I all come from the other interface Σ^E . We then have that the contribution to the current of photons crossing Σ^I is given by

$$\begin{aligned} J^I &= \int_{\Gamma_+(\mathbf{x}^I)} \boldsymbol{\theta} \cdot \boldsymbol{\nu}_C(\mathbf{x}^I) (\mathcal{R}_1^C U)(\mathbf{x}^I, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}) \\ &= \frac{1}{2\pi} \int_0^\pi \sin \theta \left(e^{-\sigma_a t(\mathbf{x}^I, \theta)} U(\bar{\mathbf{x}}^I) - U(\mathbf{x}^E) \right) d\theta, \end{aligned}$$

where θ is chosen so that $0 \leq \theta \leq \pi$ spans $\Gamma_+(\mathbf{x}^I)$. Notice that both $\bar{\mathbf{x}}^I$ and \mathbf{x}^E belong to Σ^E . Locally around \mathbf{x}^E we can parameterize Σ^E by the arc-length distance $s(\bar{\mathbf{x}}^I) \equiv s(\theta, \mathbf{x}^E)$ to \mathbf{x}^E . When the curvature of Σ is positive, all points $\bar{\mathbf{x}}^I$ are close to \mathbf{x}^E since the clear layer is thin. We can thus use the Taylor expansion

$$\begin{aligned} U(\bar{\mathbf{x}}^I) &= U(\mathbf{x}^E) + s(\theta; \mathbf{x}^E) \frac{\partial U}{\partial s}(\mathbf{x}^E) + \frac{1}{2} s^2(\theta; \mathbf{x}^E) \frac{\partial^2 U}{\partial s^2}(\mathbf{x}^E) + \text{smaller terms} \\ &= U(\mathbf{x}^E) + \frac{\partial}{\partial s} \left(\frac{s^2(\theta; \mathbf{x}^E)}{2} \frac{\partial U}{\partial s} \right) (\mathbf{x}^E) + \text{smaller terms.} \end{aligned} \tag{6.11}$$

Similarly, we have

$$e^{-\sigma_a t(\mathbf{x}^I, \theta)} = 1 - \sigma_a t(\mathbf{x}^I, \theta) + \text{smaller terms.} \tag{6.12}$$

We finally obtain the following approximation

$$\begin{aligned} J^I &= -\sigma_a^I(\mathbf{x}^I) U(\mathbf{x}^E) + b^I(\mathbf{x}^E) \frac{\partial U}{\partial s}(\mathbf{x}^E) + d^I(\mathbf{x}^E) \frac{\partial^2 U}{\partial s^2}(\mathbf{x}^E) + \dots \\ &= -\sigma_a^I(\mathbf{x}^I) U(\mathbf{x}^E) + \frac{\partial}{\partial s} \left(d^I(\mathbf{x}^E) \frac{\partial U}{\partial s} \right) (\mathbf{x}^E) + \dots, \end{aligned} \tag{6.13}$$

where

$$\begin{aligned} \sigma_a^I(\mathbf{x}^I) &= \sigma_a \frac{1}{2\pi} \int_0^\pi t(\mathbf{x}^I, \theta) \sin \theta d\theta \\ b^I(\mathbf{x}^E) &= \frac{1}{2\pi} \int_0^\pi s(\theta; \mathbf{x}^E) \sin \theta d\theta \\ d^I(\mathbf{x}^E) &= \frac{1}{2\pi} \int_0^\pi \frac{1}{2} s^2(\theta; \mathbf{x}^E) \sin \theta d\theta. \end{aligned} \tag{6.14}$$

Notice that b^I vanishes when the surface Σ is symmetrical about \mathbf{x} since then

$s(\pi - \theta; \mathbf{x}^E) = -s(\theta; \mathbf{x}^E)$. This justifies that the asymptotic expansion is pushed to second order in (6.11). The local approximation to the contribution $KU(\mathbf{x}^E) - J^I$ can be obtained in a similar manner. Its calculation is slightly more complicated since it involves two contributions coming from photons that entered Ω^C through Σ^I and Σ^E .

Adding the contributions from the two layer boundaries and sending the thickness of the clear layer to zero (thus identifying \mathbf{x}^E and \mathbf{x}^I with $\mathbf{x} \in \Sigma$), we obtain that

$$KU(\mathbf{x}) = \frac{\partial}{\partial s} \left(d^C(\mathbf{x}) \frac{\partial U}{\partial s} \right) (\mathbf{x}) - \sigma_a^C(\mathbf{x}) U(\mathbf{x}) + \text{smaller terms.} \quad (6.15)$$

The diffusion coefficient d^C is positive and the absorption coefficient σ_a^C is non-negative. This implies that the asymptotic limit of the operator K is negative in the sense that neglecting smaller order terms and integrating by parts, $\int_{\Sigma} (KU)(\mathbf{x}) U(\mathbf{x}) dS(\mathbf{x}) \leq 0$ for smooth functions $U(\mathbf{x})$, where $dS(\mathbf{x})$ is the surface measure on Σ .

The above procedure can be generalized to the three dimensional case without any theoretical difficulty, although the local parameterization of the surfaces Σ^E and Σ^I and the calculation of the travel times $t(\mathbf{x}, \theta, \phi)$ and currents in (6.10) become more complicated.

6.2.4 Tangential diffusion coefficient for circular layers.

In the rest of this paper, we assume that the surface Σ is a circle of radius R in two space dimensions $n = 2$ and a sphere of radius R in three space dimensions $n = 3$; see Fig. 6-2. We also assume to simplify that the clear layer is non-absorbing, i.e. $\sigma_a(\mathbf{x}) = 0$ in Ω^C . We then obtain that

$$KU(\mathbf{x}^E) = d^C \Delta_{\perp} U(\mathbf{x}^E) + \text{small terms,} \quad (6.16)$$

where Δ_{\perp} is the Laplace-Beltrami operator for the sphere when $n = 3$ (i.e., the Laplace operator in the tangent plane to the sphere) and $\Delta_{\perp} = \frac{\partial^2}{\partial s^2}$ for the circle

when $n = 2$. The diffusion coefficient is given in two space dimensions by

$$\begin{aligned}
d^C &= d_{\text{ex-ex}}^C + d_{\text{ex-in}}^C + d_{\text{in-ex}}^C, \\
d_{\text{ex-ex}}^C &= \frac{1}{2\pi} \int_0^{\theta_0} \sin \theta (R+L)^2 (2\theta)^2 d\theta \\
d_{\text{ex-in}}^C &= \frac{1}{2\pi} \int_{\theta_0}^{\pi/2} \sin \theta (R-L)^2 \left(\theta - \arccos \left(\frac{R+L}{R-L} \cos \theta \right) \right)^2 d\theta \\
d_{\text{in-ex}}^C &= \frac{1}{2\pi} \int_0^{\pi/2} \sin \theta (R+L)^2 \left(-\theta + \arccos \left(\frac{R-L}{R+L} \cos \theta \right) \right)^2 d\theta.
\end{aligned} \tag{6.17}$$

Here, θ_0 is the limiting angle below which incident particles enter the clear layer through the upper surface and exit it through the same upper surface. It is defined by

$$\theta_0 = \arccos \left(\frac{R-L}{R+L} \right). \tag{6.18}$$

The three components of d^C are the contributions of photons that exit (enter) the clear layer through the upper (upper) surface ($d_{\text{ex-ex}}^C$), the upper (lower) surface ($d_{\text{ex-in}}^C$), and the lower (upper) surface ($d_{\text{in-ex}}^C$), respectively. Thus $d_{\text{in-ex}}^C$ is given by $d^I(\mathbf{x}^E)$ in (6.14) and the two other contributions correspond to the photons crossing the clear layer through Σ^E .

A similar expression can be calculated for the tangential diffusion coefficient in three space dimensions. We have not reproduced this lengthy expression here.

6.2.5 Generalized diffusion model with local interface conditions.

With these approximations, the generalized diffusion model takes then the following form in the limit of vanishing thickness of the clear layer:

$$\begin{aligned}
-\nabla \cdot \mathcal{D}(\mathbf{x}) \nabla U(\mathbf{x}) + \sigma_a(\mathbf{x}) U(\mathbf{x}) &= S(\mathbf{x}) && \text{in } \Omega \setminus \Sigma \\
U(\mathbf{x}) + \varepsilon n L_n \mathcal{D}(\mathbf{x}) \boldsymbol{\nu}(\mathbf{x}) \cdot \nabla U(\mathbf{x}) &= g(\mathbf{x}) && \text{on } \partial\Omega \\
U(\mathbf{x}^+) &= U(\mathbf{x}^-) && \text{on } \Sigma \\
\boldsymbol{\nu}(\mathbf{x}) \cdot \mathcal{D}(\mathbf{x}^+) \nabla U(\mathbf{x}^+) - \boldsymbol{\nu}(\mathbf{x}) \cdot \mathcal{D}(\mathbf{x}^-) \nabla U(\mathbf{x}^-) &= d^C \Delta_{\perp} U(\mathbf{x}) && \text{on } \Sigma.
\end{aligned} \tag{6.19}$$

For $\mathbf{x} \in \Sigma$, we have defined $\mathbf{x}^\pm = \mathbf{x} \pm 0\nu(\mathbf{x})$. This equation is much simpler to solve than (6.9) because the jump conditions are now *local* on Σ . Notice that we have replaced \mathbf{x}^E and \mathbf{x}^I by \mathbf{x} since the layer is sufficiently thin. Also the diffusion equation is now posed on $\Omega \setminus \Sigma$ instead of $\Omega \setminus \Omega^C$. The flux of photons $U(\mathbf{x})$ is continuous across the interface Σ . The current $\nu \cdot \nabla U$ is however discontinuous and its jump is given by $d^C \Delta_\perp U(\mathbf{x})$, which is also continuous since only derivatives along the interface Σ are considered.

The numerical implementation of (6.19) is also relatively straightforward. Indeed let us consider the variational formulation of (6.19). Upon multiplying (6.19) by a test function $w(\mathbf{x})$ and integrating by parts using the Gauss formula, we obtain that

$$\begin{aligned} & \int_{\Omega} \left(\mathcal{D}(\mathbf{x}) \nabla U(\mathbf{x}) \cdot \nabla w(\mathbf{x}) + \sigma_a(\mathbf{x}) U(\mathbf{x}) w(\mathbf{x}) \right) d\mathbf{x} \\ & + \int_{\Sigma} d^C \nabla_\perp U(\mathbf{x}) \cdot \nabla_\perp w(\mathbf{x}) dS(\mathbf{x}) + \int_{\partial\Omega} \frac{1}{\varepsilon n L_n} U(\mathbf{x}) w(\mathbf{x}) dS(\mathbf{x}) \\ & = \int_{\Omega} S(\mathbf{x}) w(\mathbf{x}) d\mathbf{x} + \int_{\partial\Omega} \frac{1}{\varepsilon n L_n} g(\mathbf{x}) w(\mathbf{x}) dS(\mathbf{x}). \end{aligned} \quad (6.20)$$

Here ∇_\perp is the gradient operator along the surface Σ and dS is the surface measure on Σ and $\partial\Omega$. Since the diffusion coefficients $\mathcal{D}(\mathbf{x})$ and $d^C(\mathbf{x})$ are positive, we obtain that the above equation is well-posed. Moreover its discretization by finite element method (Galerkin projection) is straightforward thanks to the variational formulation (6.20) [39]. A similar variational formulation was also used to solve (6.9) [17]. Notice that (6.19) is however considerably simpler to solve as the calculation and integration of the response operator \mathcal{R}_1^C in (6.10) is replaced by a single tangential diffusion coefficient d^C .

6.2.6 Remarks on the mathematical model.

The derivation of (6.19) can be justified rigorously by using the asymptotic expansions and techniques developed in an earlier work [18]. We present the main results below and refer to that work for additional details.

Denoting by ε the mean free path, i.e., the main distance of propagation of the photons between successive collisions, the scaling of the clear layer such that the

operator KU is of order $O(1)$ is given by $L^2|\ln L| \approx \varepsilon$. Discarding logarithmic terms, this means that the clear layer must be approximately of size $\sqrt{\varepsilon} \gg \varepsilon$. When the clear layer is much smaller than $\sqrt{\varepsilon}$, the guiding effects can be neglected as a first approximation and classical diffusion equations (6.4) are asymptotically valid. When the clear layer is much larger than $\sqrt{\varepsilon}$, it is too large for the diffusion equilibrium $U(\mathbf{x}^E) = U(\mathbf{x}^I)$ to hold. In effect, a nonlocal equilibrium arises, which imposes that the flux of photons is asymptotically constant inside the layer. This case was analyzed [17, 18] both theoretically and numerically.

When the clear layer has the correct scaling, $L^2|\ln L| \approx \varepsilon$, and the curvature of the surface Σ is uniformly positive (is a uniformly positive definite matrix in three space dimensions), we can show [18] that the error between $u(\mathbf{x}, \boldsymbol{\theta})$ and $U(\mathbf{x})$ is of order $\sqrt{\varepsilon}$. The error is no longer of order ε as in the case of classical diffusion [58]. For typical mean free paths of order 10^{-3} - 10^{-2} , the error will therefore possibly be of the order of a few percents. To further quantify this error term, we propose several numerical simulations in the following section.

When the clear layer is no longer scattering-free, the distance traveled by the photons when they cross the clear layer decreases as fewer photons travel collision-less parallel to the layer boundary. This implies that the tangential diffusion coefficient also decreases. However the final form of the generalized diffusion equation is not modified by weakly scattering layers. In the limit of strongly scattering layers, the tangential diffusion coefficient vanishes. This simply corresponds to the validity of classical diffusion, where the interface conditions are continuity of the flux intensity and current.

More general geometries such as oscillatory clear layers can also be considered [149]. Oscillations will also reduce the value of the tangential diffusion coefficient as photons are forced to exit the clear layer more rapidly by the geometry. Although further theoretical and numerical studies are necessary to adapt the proposed method to more complex geometries, we believe that the tangential diffusion process is a rather stable limiting process to model the guiding effect in clear and not-so-clear layers. All we have to do is to find an average surface Σ and then the tangential diffusion coefficient

that generalizes (6.14).

Let us finally mention that we restrict ourselves here to the steady-state transport equation with isotropic scattering. The generalization of the results presented below to anisotropic scattering is straightforward as long as the diffusion approximation can be justified. Time dependent and frequency harmonic equations also can be treated similarly so long as the variations in time of the source terms are slow compared with the characteristic mean free time, i.e., the mean time between successive collisions of the photons with the underlying medium. For time dependent equations, the term $c^{-1}\frac{\partial u}{\partial t}$ need be added in front of (6.1), (6.4), and the main result (6.19). In the time harmonic case, $c^{-1}i\omega u$ is added instead. Here c is the light speed and ω the modulation frequency of the source term.

6.3 Validation of the model with forward simulations

In this section we solve (6.19) numerically and compare its solution to the transport solution $u(\mathbf{x}, \boldsymbol{\theta})$ obtained by a Monte Carlo algorithm. Numerical simulations are performed both for the two and three dimensional problems with circular and spherical clear layers, respectively.

We assume that $S(\mathbf{x}) = 0$ and that $g(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_0)$, where $\mathbf{x}_0 \in \partial\Omega$ is a point on the boundary of the domain where a constant source emits light isotropically. The transport and diffusion solutions are compared by looking at the exiting currents at the boundary of the domain $\partial\Omega \setminus \{\mathbf{x}_0\}$. The transport and diffusion currents are given by

$$J_T(\mathbf{x}) = \int_{S^{n-1}} \boldsymbol{\theta} \cdot \boldsymbol{\nu}(\mathbf{x}) u(\mathbf{x}, \boldsymbol{\theta}) d\mu(\boldsymbol{\theta}), \quad J_D(\mathbf{x}) = -\mathcal{D}(\mathbf{x}) \boldsymbol{\nu}(\mathbf{x}) \cdot \nabla U(\mathbf{x}), \quad (6.21)$$

respectively. These currents correspond to the information that is available in physical experiments.

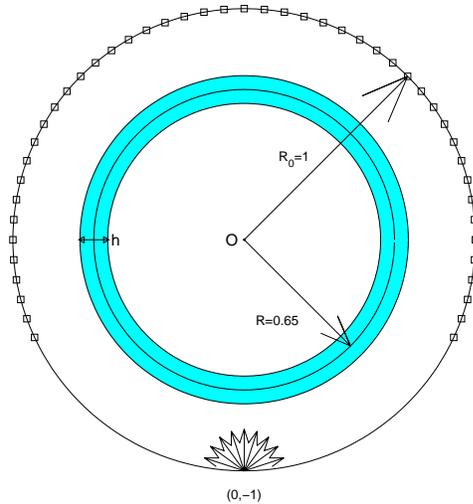


Figure 6-2: Geometry of the two-dimensional setting and cross-section of the geometry of the three-dimensional setting with azimuthal symmetry.

6.3.1 Two dimensional numerical simulations.

In two space dimensions, the domain Ω is the unit disc $\Omega = \{\mathbf{x} \in \mathbb{R}^2, |\mathbf{x}| < 1\}$. The surface Σ modeling the clear layer is a circle of radius R . Photons enter the domain at the point $\mathbf{x}_0 = (0, -1)$. In the numerical experiments we have chosen $R = 0.65$. The thickness of the clear layer is given by $h = 2L$; see Fig. 6-2. We consider several values of h . The scattering cross section $\sigma_s(\mathbf{x})$ is chosen constant and equal to 10^2 in $\Omega \setminus \Omega^C$ and vanishes in Ω^C . This implies that the total size of the domain is of the order of 100 mean free paths. In all our simulations the mean free path is 0.01. We assume that there is no absorption to simplify. In the absence of clear layer, the problem would be very much in the regime of validity of diffusion.

The transport equation (6.1) is solved by Monte Carlo method [167]. Particles start at \mathbf{x}_0 with uniformly chosen initial direction and propagate inside the domain until they exit it. The outgoing current J_T is calculated accordingly. A number of particles of $8 \cdot 10^7$ has been used to obtain a sufficiently small statistical variance.

The generalized diffusion model (6.19) is solved by Fourier decomposition after passage to polar coordinates. The jump of derivatives at the clear layer can easily be accounted for in this setting. We thus obtain a quasi-analytic expression for J_D . This is the reason why we have chosen circular clear layers.

The current at the domain boundary is discretized into 2×36 cells of size 5 degrees each (or $\pi/36$). Cell 1 corresponds to the vicinity of the source and cell 36 corresponds to the vicinity of the upper point $(0, 1)$. The symmetry about $x = 0$ is used in the calculations. The number of particles is such that at least $15 \cdot 10^3$ particles exit through each cell. This ensures a statistical relative error of less than 10^{-2} by the law of large numbers, which is below or comparable to the error expected from the diffusion model for a mean free path of 10^{-2} . The lowest density is obviously obtained in the upper cell 36 and in the absence of clear layers. The corresponding diffusive flux is obtained by averaging the flux given by the Fourier expansion in each cell.

We now compare the transport and diffusion exiting currents of photons for several sizes of the clear layer. The thickness of the clear layer varies between 1 and 7 mean free paths. According to theory clear layers of the order of the mean free path are too small to significantly modify the solution obtained in the absence of clear layer. Clear layers on the order of the square root of the mean free path (10 here) however have a significant effect on the solution. Because the mean free path here is still relatively large (in the sense that its square root is not very small) we start seeing effects for clear layers of roughly 2 – 3 mean free paths.

h	0.01	0.02	0.03	0.04	0.05	0.06	0.07
$d_{\text{ex-ex}}^C$	0.0064	0.026	0.058	0.10	0.16	0.23	0.32
$d_{\text{in-ex}}^C$	0.0029	0.0093	0.018	0.028	0.039	0.051	0.062
$d_{\text{in-in}}^C$	0.0031	0.011	0.021	0.036	0.053	0.073	0.096
d_{theory}^C	0.0124	0.0455	0.0971	0.166	0.253	0.355	0.475
$d_{\text{best fit}}^C$	0.0129	0.0465	0.0983	0.167	0.253	0.356	0.474
$E_{\text{GDM}} (\%)$	1.17	1.56	1.43	1.09	0.81	0.56	0.60
$E_{\text{BF}} (\%)$	0.73	0.65	0.57	0.49	0.46	0.47	0.46
$E_{\text{DI}} (\%)$	3.3	10.2	17.7	24.5	30.2	35.3	39.8
$E_{\text{DI2}} (\%)$	5.7	11.8	18.2	17.8	18.1	17.9	17.8

Table 6.1: Tangential diffusion coefficients and relative root mean square error (L^2 norm) between the Monte Carlo simulations and the various diffusion models for several thicknesses of the clear layer. The errors E_{GDM} , E_{BF} , E_{DI} , and E_{DI2} represent the relative root mean square error (in percentage) between the Monte Carlo simulations and the generalized diffusion model obtained using d_{theory}^C , the generalized diffusion model obtained by best fit, the classical diffusion equation, and the generalized diffusion model with tangential diffusion coefficient 1.5 times larger than d_{theory}^C , respectively.

Tab. 6.1 present the results of the numerical experiments. For thicknesses h between 1 and 7 mean free paths, the tangential diffusion coefficient defined in (6.17) is given in row 4 and its three components in rows 1 to 3. The coefficient obtained by best fitting (in the least-square sense) the generalized diffusion model to the Monte Carlo data on the outgoing density between angles 60 and 180 (cells 12 to 36) is given in row 5. We observe that the theoretical coefficient is quite close to the best fit. This is confirmed by looking at the errors made by the different models. The relative L^2 norm between the Monte Carlo simulations and the various models between cells 12 and 36 is given in rows 6 to 9 for the generalized diffusion model, the best fit from data, the classical diffusion model with no clear layer (i.e., the diffusion coefficient is taken constant and equal to $1/200$ on the whole domain), and a generalized diffusion model where the tangential diffusion coefficient has been chosen very large, respectively. By very large, we mean a tangential diffusion coefficient 1.5 times larger than its theoretical value. This solution corresponds to overestimating the guiding effect of the clear layer as we would obtain by using a diffusion approximation with a large diffusion coefficient given by (6.5) inside the clear layer. It is known that the correct solution is then not obtained [17, 18, 71, 149]. The different models are also compared in Figure 6-3 for four different thicknesses h . The viewgraphs confirm the error estimates of Tab. 6.1.

6.3.2 Interpretation of results.

Let us first state that the generalized diffusion model successfully deals with the guiding effects caused by the presence of a clear layer. The relative root mean square error between transport and this diffusion model does not exceed two percents. The accuracy degrades in the vicinity of the source term (not shown) but this is classical of diffusion approximations and is independent of the clear layer. The diffusion model obtained from (6.17) is almost as accurate as the best fit model. Classical diffusion, where the clear layer is replaced by a diffusive medium, is accurate when the clear layer is thin. However the error becomes unacceptable in practice (about 10%) even for thicknesses of the order of two mean free paths. The guiding effect is neglected

and the upward propagation of photons is clearly underestimated. The opposite effect arises when the tangential diffusion coefficient is chosen too large. As we have already mentioned, this is similar to using the diffusion model (6.5) also inside the clear layer, which gives higher a diffusion coefficient than is physically correct and overestimates the guiding effect.

Let us conclude with a short comment on the theoretical diffusion coefficient. It is not difficult to show that asymptotically as $h \rightarrow 0$, $d_{\text{ex-ex}}^C$ is a term of order h^2 whereas the other contributions $d_{\text{in-ex}}^C$ and $d_{\text{in-in}}^C$ are terms of order $h^2 |\ln h| \gg h^2$. We have observed this behavior numerically for values of h of order $10^{-4} - 10^{-3}$. However in the cases shown here, where the mean free path is of order 10^{-2} , the term $d_{\text{ex-ex}}^C$, although asymptotically smaller than the other contributions, actually dominates in the calculation of the theoretical diffusion coefficient.

6.3.3 Three dimensional numerical simulations.

Let us now consider the three-dimensional case. The domain is now a sphere of radius 1, the clear layer a corona of thickness h centered at $R = 0.65$, and the source is at position $(0, 0, -1)$. The transport equations are still solved by Monte Carlo method and the diffusion equation by projection onto spherical harmonics.

We did not estimate the theoretical tangential diffusion coefficients that generalizes (6.17) to the three-dimensional case. This coefficient could certainly be calculated analytically or computed numerically by assessing how far photons can go on average by crossing the clear layer. Rather we would like to stress another advantage we see in (6.19) as a model in optical tomography, where the photon measurements at the boundary are used to image the diffusion and absorption properties of the domain on the other side of the clear layer (i.e., close to the origin $(0, 0, 0)$). We claim that the clear layer can be modeled by a possibly spatially dependent tangential diffusion coefficient provided that we have an a priori knowledge of its location. In other words we claim that the inverse problem based on simulating the full clear layer (in transport then) and the inverse problem based on replacing the clear layer by a tangential diffusion process (with a priori unknown strength) will give similar reconstructions.

This of course comes at the expense of also reconstructing the value of the tangential diffusion coefficient from the measured data. This claim corresponds to showing that the best diffusion fit yields a good approximation to the transport solution. We have seen that this is the case in two dimensions.

We now present results that confirm that this is also the case in three dimensions. The number of particles used in the calculations is $2 \cdot 10^7$. Such calculations are already quite long because particles stay longer inside the domain in three dimensions than in two dimensions. Moreover the number of particles exiting the domain in the upper part of the sphere is also smaller than in two dimensions. This renders our numerical simulations less accurate than in two dimensions. This has the advantage at least of mimicking more closely noisy measurements and thus we consider them as an interesting benchmark. The numerical results are presented in Fig. 6-4. They certainly show that validity of (6.19) as an accurate model to simulate the guiding effect. The root mean square error between the transport solution and the best fit generalized diffusion model is of the order of 2–3%. This error is moreover mostly due to random fluctuations. Other diffusion models that do not correctly account for this effect introduce too large errors to be really considered as practical for the purpose of inversion of physical properties from boundary measurements [17, 18, 71, 149]. For instance classical diffusion in our simulations is as far as 50% off the transport solution.

6.4 Conclusions and remarks

We propose a generalized diffusion model that accounts for the multiple scattering of photons in highly scattering media (classical diffusion regime) and well as for the near-collision-less propagation of the same photons in clear layers (purely transport regime resulting in a guiding effect).

This model can be mathematically derived from the phase space radiative transport equation as a small mean-free-path limit. It captures the guiding effect of photons in the clear layer quite well. Moreover it has almost the same cost as classical

diffusion, which completely fails to model the clear layer effects, and a lower cost than previously derived generalized diffusion equations, which are already much less expensive than full transport solutions. The reason for this lower cost is that the nonlocal interface conditions of the latter diffusion models are replaced by their best local approximation. This best local approximation takes the form of a tangential diffusion process.

The strength of this diffusion process can be calculated analytically or numerically provided that one has access to the geometry of the clear layer. When this geometry is unknown or only partially known, we have shown numerically that the diffusion process that best fits the impact of the clear layer gives boundary measurements that are visually indistinguishable from the measurements obtained by solving the full transport equations. We believe that the generalized diffusion model can thus safely be used in optical tomography as an accurate approximation of the forward model.

The analysis in this chapter tells us that if we know the location and geometry of the clear layer, we can use the generalized diffusion model to replace the radiative transport model of light propagation. In practice, if one does not know the information about the clear layers, can one still use the generalized diffusion equation? We will give partial answer to this question in next chapter.

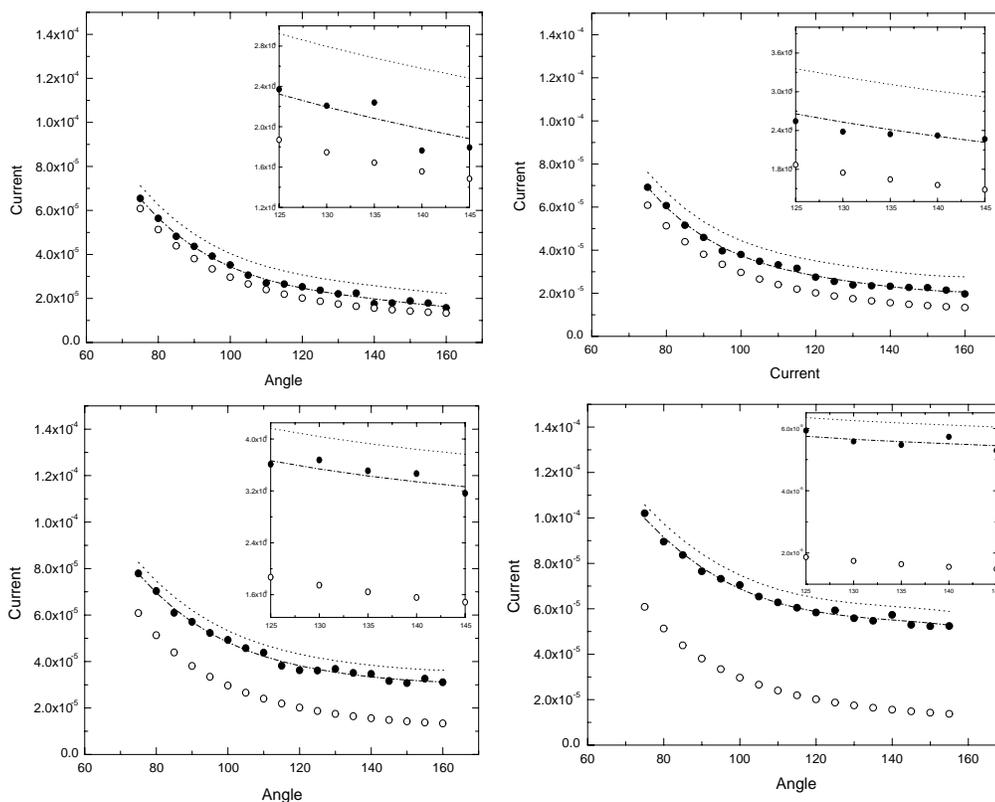


Figure 6-3: Plots of the current between cells 14 (70 degrees) and 36 (180 degrees) at the boundary of the unit disc (two-dimensional simulation) for the Monte Carlo solution and the different diffusion models. The thickness of the clear layer in mean free path is 2, 3, 5, and 7 for the top-left, top-right, bottom-left, bottom-right figures, respectively. In each sub-figure, the Monte Carlo simulation is represented by solid circles, the classical diffusion by empty circles, the generalized diffusion model with theoretical tangential diffusion coefficient by the solid line, the generalized diffusion model with best fit by the dash-dotted lines, and the generalized model with large tangential diffusion coefficient by the dotted line. The inset represents a magnification of the above results between angles 125 and 145.

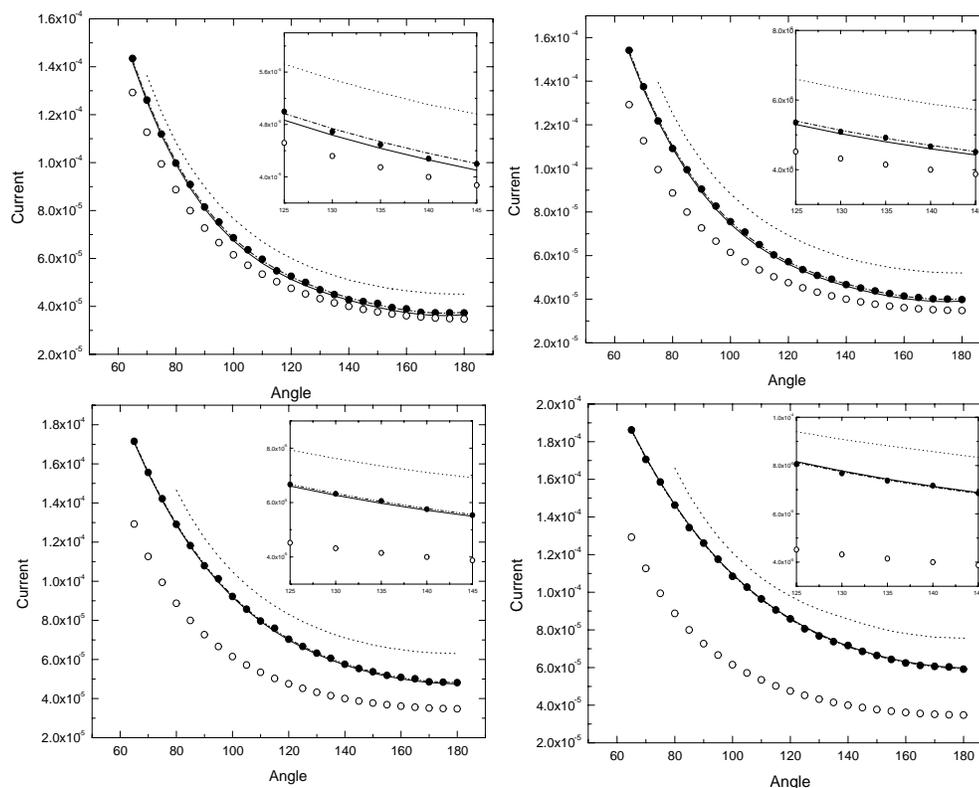


Figure 6-4: Plots of the current between cells 15 (75 degrees) and 32 (160 degrees) at the boundary of the unit sphere (three-dimensional simulation with azimuthal symmetry) for the Monte Carlo solution and the different diffusion models. The thickness of the clear layer in mean free path is 1, 2, 4, and 6 for the top-left, top-right, bottom-left, bottom-right figures, respectively. In each sub-figure, the Monte Carlo simulation is represented by solid circles, the classical diffusion by empty circles, the generalized diffusion model with theoretical tangential diffusion coefficient by the solid line, the generalized diffusion model with best fit by the dash-dotted lines, and the generalized model with large tangential diffusion coefficient by the dotted line. The inset represents a magnification of the above results between angles 125 and 145.

Chapter 7

Surface identifications by shape sensitivity analysis and the level set method

We continue our analysis on the generalized diffusion model in this chapter. As we have seen before, a tangential diffusion supported on a co-dimension one surface correctly account for the effect of extended non-scattering regions in highly scattering media. A natural question then to ask is: if we don't know *a priori* where those non-scattering region locates, can we detect them from boundary measurements? This problem reduces to the reconstruction of the singular surface in the generalized diffusion model from boundary measurements, which will be the problem to be analyzed in this chapter. The presentation of this chapter is based on reference [25].

7.1 The singular surface problem

The identification of unknown surfaces or interfaces in physical problems governed by partial differential equations has been an active field of research recently [20, 69, 98]. Apart from the fields of shape optimization and optimal design [7, 137], such problems emerge in applications such as optical tomography [20, 64], inverse scattering [117, 159] and, more generally, parameter identification in partial differential equations [47]. Most works in the current literature deal with the reconstruction of interfaces that separate regions with different contrasts from boundary or far-field measurements,

typically interfaces across which one of the constitutive parameters in the partial differential equation jumps.

In the inverse interface problem we discuss here, the role of the interface is not to separate regions with different physical coefficients but rather to be the support of a tangential diffusion process as we have seen in the previous chapter. Such a process may also model thin areas characterized by very high values of the diffusion coefficient, as in the modeling of cracks of thickness $\epsilon \ll 1$ and conductivity of order ϵ^{-1} in impedance tomography, as shown in [95].

In the absence of general analytic formulae, the inverse interface problem is usually solved by minimizing an objective function that measures the mismatch between the model predictions and the measurements. A central element in the minimization procedure is the calculation of the gradient of the objective function with respect to the variations in the shape of the interface. This is the shape sensitivity analysis [84, 166]. Another important element in the minimization procedure is a numerical tool that is used to advect the interface once a suitable descent direction has been obtained by shape sensitivity analysis. As in the pioneering work by Santosa [159] and subsequent works mentioned in the review paper [40], the level set method [40, 135] may be used to that purpose. This chapter generalizes the combination of a shape sensitivity analysis and level set method to the reconstruction of surfaces supporting singular diffusion processes from boundary measurements.

7.1.1 Forward model

Let $\Omega \subset \mathbb{R}^n$ ($n = 2, 3$) be a domain with Lipschitz boundary $\Gamma (\equiv \partial\Omega)$ and $\Sigma \subset \Omega$ a closed, non self-intersecting, interface of class C^2 embedded in Ω and separating it into interior (Ω_I) and exterior (Ω_E) parts, so that we may write $\Omega = \Omega_I \cup \Omega_E \cup \Sigma$. We also require that Σ stay away from $\partial\Omega$, i.e., $d(\Sigma, \Gamma) > C$ for some positive constant C . The geometry of interest is depicted in Fig. 7-1 in the two-dimensional setting. We consider the following elliptic partial differential equation in Ω with interface

condition on Σ :

$$\begin{aligned}
-\nabla \cdot \mathcal{D}(\mathbf{x})\nabla u(\mathbf{x}) + a(\mathbf{x})u(\mathbf{x}) &= 0 && \text{in } \Omega \setminus \Sigma, \\
\mathcal{D}(\mathbf{x})\boldsymbol{\nu}(\mathbf{x}) \cdot \nabla u(\mathbf{x}) &= g(\mathbf{x}) && \text{on } \Gamma, \\
[u] &= 0 && \text{on } \Sigma, \\
[\mathbf{n} \cdot \mathcal{D}\nabla u] &= -\nabla_{\perp} \cdot d(\mathbf{x})\nabla_{\perp} u(\mathbf{x}) && \text{on } \Sigma.
\end{aligned} \tag{7.1}$$

The scalar (to simplify) diffusion coefficients $\mathcal{D}(\mathbf{x})$ and $d(\mathbf{x})$ are uniformly positive; the absorption coefficient $a(\mathbf{x})$ is assumed to be smooth and bounded from above and below by positive constants, i.e., $0 < c_1 < a(\mathbf{x}) < c_2 < \infty$; $\mathbf{n}(\mathbf{x})$ is the outward unit normal vector to Ω_I at $\mathbf{x} \in \Sigma$ and $\boldsymbol{\nu}(\mathbf{x})$ is the outward unit outer normal vector to Ω at $\mathbf{x} \in \Gamma$. The tangential differential operator ∇_{\perp} is the restriction of ∇ to Σ , so that for a sufficiently smooth function $\phi(\mathbf{x})$ defined on Ω , we have $\nabla_{\perp}\phi(\mathbf{x}) = \nabla\phi(\mathbf{x}) - (\mathbf{n}(\mathbf{x}) \cdot \nabla\phi(\mathbf{x}))\mathbf{n}(\mathbf{x})$ for $\mathbf{x} \in \Sigma$. The symbol $\nabla_{\perp} \cdot \nabla_{\perp}$ denotes the Laplace-Beltrami operator on Σ . The jump conditions across the interface Σ are defined by

$$[u] = u(\mathbf{x}^+) - u(\mathbf{x}^-), \quad [\mathbf{n} \cdot \mathcal{D}\nabla u] = \mathbf{n} \cdot \mathcal{D}\nabla u(\mathbf{x}^+) - \mathbf{n} \cdot \mathcal{D}\nabla u(\mathbf{x}^-),$$

with

$$u(\mathbf{x}^{\pm}) = \lim_{t \rightarrow 0^+} u(\mathbf{x} \pm t\mathbf{n}(\mathbf{x})), \quad \nabla u(\mathbf{x}^{\pm}) = \lim_{t \rightarrow 0^+} \nabla u(\mathbf{x} \pm t\mathbf{n}(\mathbf{x})).$$

Equation (7.1) models a background diffusion-absorption process in the domain Ω with a tangential diffusion process supported on the surface Σ [23, 95].

The problem described in (7.1) is well-posed in the following Hilbert space:

$$H_{\Sigma}^1(\Omega) := \left\{ u(\mathbf{x}) : u \in H^1(\Omega), \text{ such that } \int_{\Sigma} |\nabla_{\perp} u|^2 d\sigma < \infty \right\}, \tag{7.2}$$

where $H^1(\Omega)$ is the usual Sobolev space of L^2 functions in the domain Ω whose first-order partial derivatives also in $L^2(\Omega)$ [5, 58]. In other words, $H_{\Sigma}^1(\Omega)$ consists of functions in $H^1(\Omega)$ with tangential gradient on Σ in $L^2(\Sigma)$. One can verify that

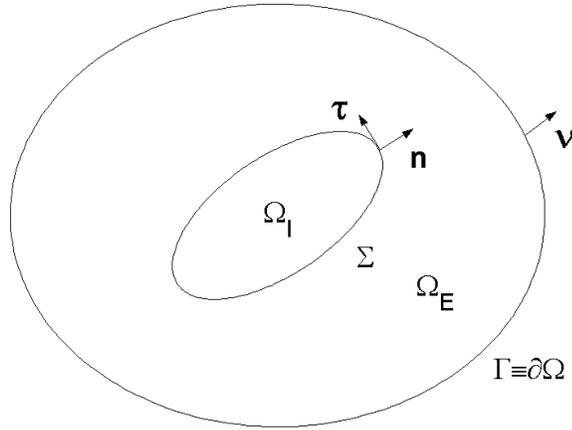


Figure 7-1: Geometric setting of the problem in the two-dimensional setting with $\Omega = \Omega_I \cup \Omega_E \cup \Sigma$.

$H_{\Sigma}^1(\Omega)$ is a Hilbert space equipped with the scalar product:

$$(u, v)_{H_{\Sigma}^1} = \int_{\Omega} (uv + \nabla u \cdot \nabla v) d\mathbf{x} + \int_{\Sigma} \nabla_{\perp} u \cdot \nabla_{\perp} v d\sigma(\mathbf{x}),$$

where $d\sigma(\mathbf{x})$ denote the Lebesgue measure on Σ , and a natural norm

$$\|u\|_{H_{\Sigma}^1} = \sqrt{(u, u)_{H_{\Sigma}^1}}.$$

Upon multiplying (7.1) by a test function $\phi(\mathbf{x}) \in H_{\Sigma}^1(\Omega)$ and integrating by parts, we obtain that

$$S(u, \phi) = f_g(\phi), \quad (7.3)$$

where the bilinear form $S(\cdot, \cdot)$ is defined by

$$\begin{aligned} S(u, \phi) := & \int_{\Omega} \mathcal{D}(\mathbf{x}) \nabla u(\mathbf{x}) \cdot \nabla \phi(\mathbf{x}) d\mathbf{x} + \int_{\Omega} a(\mathbf{x}) u(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x} \\ & + \int_{\Sigma} d(\mathbf{x}) \nabla_{\perp} u(\mathbf{x}) \cdot \nabla_{\perp} \phi(\mathbf{x}) d\sigma(\mathbf{x}), \end{aligned} \quad (7.4)$$

and the linear form $f_g(\phi)$ by

$$f_g(\phi) := \int_{\Gamma} g(\mathbf{x}) \phi(\mathbf{x}) d\sigma(\mathbf{x}).$$

Note that S is symmetric, i.e., $S(u, \phi) = S(\phi, u)$. Because the diffusion coefficients $\mathcal{D}(\mathbf{x})$ and $d(\mathbf{x})$ and the absorption coefficient $a(\mathbf{x})$ are positive and bounded, one can verify that the bilinear form S is coercive. It then follows from Lax-Milgram theory [57, 76] that if $g \in H^{-1/2}(\Gamma)$, then (7.1) admits a unique solution $u \in H^1_\Sigma$ with trace on Γ , $u|_\Gamma \in H^{1/2}(\Gamma)$; see also [20].

7.1.2 Inverse surface problem

A practically useful inverse problem related to equation (7.1) consists of reconstructing the interface Σ from knowledge of u at the boundary Γ . The Neumann to Dirichlet (NtD) operator, which maps the incoming flux g to u on the boundary [96] is defined as:

$$\Lambda_\Sigma : \begin{array}{ll} H^{-1/2}(\Gamma) & \longmapsto H^{1/2}(\Gamma) \\ g(\Gamma) & \longmapsto u|_\Gamma. \end{array}$$

This operator obviously depends on the geometry of Σ . The inverse interface problem of (7.1) may then be formulated as:

(IP) Determine the interface Σ from knowledge of the Neumann to Dirichlet operator Λ_Σ .

If all the other coefficients in (7.1) are known, it is shown in [20] that knowledge of the local Neumann to Dirichlet map uniquely determines the interface Σ . Let us denote by $\Gamma_g \subset \Gamma$ the part of the boundary where non-zero boundary current are applied and measurements are taken. In other words, we replace the boundary condition of (7.1) by

$$\mathcal{D}(\mathbf{x})\nu(\mathbf{x}) \cdot \nabla u(\mathbf{x}) = \begin{cases} g(\mathbf{x}), & \text{on } \Gamma_g \\ 0, & \text{on } \Gamma \setminus \Gamma_g. \end{cases}$$

Denoting by $\Lambda_{\Sigma}^{\Gamma_g}$ the local Neumann to Dirichlet operator for the new problem, which implies that u is measured only on Γ_g . Then we have the following uniqueness result:

Proposition 7.1.1 ([20]). *Let $\Lambda_{\Sigma_1}^{\Gamma_g}$ and $\Lambda_{\Sigma_2}^{\Gamma_g}$ be the local NtD maps associated with interfaces Σ_1 and Σ_2 , respectively. Suppose that the functions $\mathcal{D}(\mathbf{x})$, $d(\mathbf{x})$ and $a(\mathbf{x})$*

are known and satisfy the above mentioned regularity assumptions. Then $\Lambda_{\Sigma_1}^{\Gamma_g} = \Lambda_{\Sigma_2}^{\Gamma_g}$ implies that $\Sigma_1 = \Sigma_2$.

The objective of this paper is to design a numerical method to reconstruct the singular interface Σ from knowledge of Λ_Σ or $\Lambda_\Sigma^{\Gamma_g}$. Our method is based on classical numerical optimization techniques. We convert the reconstruction problem to a regularized nonlinear least square problem:

$$\mathcal{F}_\alpha(\Sigma) := \frac{1}{2} \|u - u_m^\delta\|_{L^2(\Gamma)}^2 + \alpha \int_\Sigma d\sigma(\mathbf{x}) \rightarrow \min_{\Sigma \in \Pi}. \quad (7.5)$$

Here u_m^δ denotes a noisy measurement of u on the domain boundary Γ with noise level δ , while Π denotes the space of admissible surfaces Σ . The first term in the objective functional $\mathcal{F}_\alpha(\Sigma)$ evaluates the discrepancy between the measured and predicted data, while the second term is a regularization term with parameter α . The choice of set Π is critical to the existence of minimizers to the functional $\mathcal{F}_\alpha(\Sigma)$. If we assume that Π consists of interfaces such that $\int_\Sigma d\sigma(\mathbf{x})$ is the $n - 1$ dimensional Hausdorff measure of Σ , which turns out to be the perimeter of the inner domain Ω_I in two dimensions, we can then view the reconstruction of Σ as the identification of the domain Ω_I penalized by its perimeter. By techniques such as those of [8, 56], the existence of minimizer to functional \mathcal{F}_α should follow from the lower semicontinuity of $\mathcal{F}_\alpha(\Sigma)$ with respect to Ω_I (thus Σ) in either the space of sets with finite parameter or the space of simply-connected, Hausdorff measurable compact sets. For our analysis below, we need interfaces that are at least of class C^2 such that the mean curvature of the interfaces can be defined in the classical way. It is however not clear to us so far that a minimizer of \mathcal{F}_α exists in such a class of interfaces. Our analysis in the following sections are thus based on the assumption that a regular minimizer does exist.

In many applications, such as the reconstruction of clear layers in optical tomography, we may have *a priori* information about the location of the singular interface, whence constraints on the size of Π , which may simplify the inverse problem. We do not consider this situation here.

7.1.3 Comparison with the reconstruction of inclusions

It is instructive to compare the reconstruction of singular surfaces as they are described in the preceding section with the more classical problem of the reconstruction of interfaces separating regions characterized by different diffusion coefficients; see [98, 117, 159]. In the latter works, the inclusion is characterized by a constant diffusion coefficient that differs from the constant background diffusion coefficient. The inclusion is then reconstructed by minimizing the functional (7.5). The construction of velocity fields allowing us to minimize (7.5) is not modified when the inclusion's and background diffusion coefficients are allowed to be (not necessarily constant) smooth functions, so long as the difference between these functions does not vanish. More precisely, we consider the following model for the inclusions:

$$\begin{aligned}
 -\nabla \cdot \mathcal{D}(\mathbf{x})\nabla u(\mathbf{x}) + a(\mathbf{x})u(\mathbf{x}) &= 0 && \text{in } \Omega \\
 \mathcal{D}(\mathbf{x})\boldsymbol{\nu}(\mathbf{x}) \cdot \nabla u(\mathbf{x}) &= g(\mathbf{x}) && \text{on } \Gamma \\
 [u] &= 0 && \text{on } \Sigma \\
 [\mathbf{n} \cdot \mathcal{D}\nabla u] &= 0 && \text{on } \Sigma,
 \end{aligned} \tag{7.6}$$

where the diffusion coefficient $\mathcal{D}(\mathbf{x})$ jumps across the interface Σ

$$\mathcal{D}(\mathbf{x}) = \begin{cases} \mathcal{D}_0(\mathbf{x}) + \delta\mathcal{D}(\mathbf{x}) &\equiv \mathcal{D}_I(\mathbf{x}), & \mathbf{x} \in \Omega_I \\ \mathcal{D}_0(\mathbf{x}) &\equiv \mathcal{D}_E(\mathbf{x}), & \mathbf{x} \in \Omega_E, \end{cases} \tag{7.7}$$

with $\mathcal{D}(\mathbf{x})$ uniformly bounded from above and below by positive constants and $\delta\mathcal{D}(\mathbf{x})$ strictly positive or strictly negative. The case where \mathcal{D}_0 and $\delta\mathcal{D}$ are constant has been studied in [98, 117, 159]. The behavior of the solution $u(\mathbf{x})$ to (7.1) with $d(\mathbf{x}) > 0$ is very similar to the behavior of solution of model (7.6) with $\delta\mathcal{D}(\mathbf{x}) > 0$. In section 7.5, we will give a more quantitative numerical comparison between the two models.

7.2 Shape sensitivity analysis

In order to solve the surface reconstruction problem by minimization of the functional $\mathcal{F}_\alpha(\Sigma)$ in (7.5), it is essential to compute the variation of $\mathcal{F}_\alpha(\Sigma)$ with respect to a small perturbation in Σ . This involves computing the sensitivity of the diffusion solution with respect to deformations in the shape. This is the shape sensitivity analysis described in the shape optimization literature [166].

The main novelty of the paper is to carry out the shape sensitivity analysis in the presence of a singular interface. Unlike the model (7.6) treated in [98, 117, 159], the current jumps across the interface Σ in (7.1). This significantly modifies the shape sensitivity analysis and the important relationship between shape and material derivatives; see below. Let us also mention that many geometries have been addressed in the shape optimization literature [60, 70, 165, 166]. Because of the specificity of problem (7.1), none of them may be applied directly, although similarities in the methodology and mathematical machinery are easily drawn.

The framework for the shape sensitivity analysis is the following. We perturb the interface Σ according to the map $\mathbf{F}_t : \mathbb{R}^n \rightarrow \mathbb{R}^n$ (the parameter $t \in \mathbb{R}^+$ is a small positive real number) defined by:

$$\mathbf{F}_t(\mathbf{x}) = \mathbf{x} + t\mathbf{V}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^n. \quad (7.8)$$

Here $\mathbf{V}(\mathbf{x}) : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a vector field of class C^1 with compact support in the domain Ω so that each point on the boundary of Ω remains invariant under the perturbation \mathbf{F}_t . We denote this as $\mathbf{V} \in C_0^1(\Omega; \mathbb{R}^n)$. Under this perturbation, points $\mathbf{x} \in \Omega$ are mapped to $\mathbf{x} + t\mathbf{V}(\mathbf{x})$. However, the whole domain Ω remains invariant in the sense that $\Omega = \mathbf{F}_t(\Omega)$.

We denote by Σ_t the image of Σ under the perturbation, and denote by $u_t(\mathbf{x})$ the solution of problem (7.1) with Σ replaced by the perturbed interface Σ_t . The variation of u with respect to variations in the interface Σ is called the *shape derivative* of u with respect to Σ . More precisely:

Definition 7.2.1 (Shape derivative). *Let $u \in H_{\Sigma}^1$ and $u_t \in H_{\Sigma_t}^1$ be solutions of problem (7.1) with interface Σ and Σ_t , respectively. Assume that $\mathbf{V} \in C_0^1(\Omega; \mathbb{R}^n)$ be a vector field given in (7.8). If the limit*

$$u'(\Sigma; \mathbf{V}) := \lim_{t \rightarrow 0} \frac{u_t - u}{t} \quad (7.9)$$

exists in the strong (weak) topology of some Banach space of functions $B(\Omega)$, then we call $u'(\Sigma; \mathbf{V})$ the strong (weak) shape derivative of u in direction \mathbf{V} .

We refer to Rem. 7.2.9 below for a remark on the choice of a Banach space and a topology. The calculation of $u'(\Sigma; \mathbf{V})$ is greatly simplified by the introduction of a *material derivative* [166]:

Definition 7.2.2 (Material derivative). *Let $u \in H_{\Sigma}^1$, $u_t \in H_{\Sigma_t}^1$ and \mathbf{V} be given as in definition 7.2.1, and define $u^t = u_t \circ \mathbf{F}_t$. If the limit*

$$\dot{u}(\Sigma; \mathbf{V}) := \lim_{t \rightarrow 0} \frac{u^t - u}{t} \quad (7.10)$$

exists in the strong (weak) topology of some Banach space of functions $B(\Omega)$, we call $\dot{u}(\Sigma; \mathbf{V})$ the strong (weak) material derivative of u in direction \mathbf{V} .

We also refer to Rem. 7.2.9 for the choice of a Banach space and a topology. The material derivative thus quantifies the variations of u with respect to changes in the geometry for a moving (Lagrangian) coordinate system. The shape and material derivatives introduced in Defs. 7.2.1 and 7.2.2, respectively, are not independent from each other. More precisely, we have [166]:

$$u'(\Sigma; \mathbf{V}) = \dot{u}(\Sigma; \mathbf{V}) - \mathbf{V} \cdot \nabla u, \quad (7.11)$$

provided that both $\dot{u}(\Sigma; \mathbf{V})$ and $\mathbf{V} \cdot \nabla u$ make sense. This relation tells us that in order to compute the shape derivative of u , we can compute the material derivative first and then use (7.11) to obtain the shape derivative.

7.2.1 The material derivatives

Before we compute the material derivatives of u for model (7.1), we need to introduce some notation. We will denote by $(\cdot, \cdot)_{(X)}$ the inner product of space $L^2(X)$:

$$(x, y)_{(X)} := \int_X x \cdot y d\mu,$$

with $d\mu$ the Lebesgue measure on a domain X . For any vector quantity \mathbf{Y} on the interface, we use $Y_n \mathbf{n} \equiv (\mathbf{n} \cdot \mathbf{Y})\mathbf{n}$ and $\mathbf{Y}_\perp \equiv \mathbf{Y} - (\mathbf{n} \cdot \mathbf{Y})\mathbf{n}$ to denote the normal and tangential components of \mathbf{Y} , respectively.

We now examine the variations of the solution to the diffusion equation (7.1) when the interface Σ_t varies. We first observe that u_t satisfies the following relation:

$$(\mathcal{D}\nabla u_t, \nabla \phi_t)_{(\Omega)} + (au_t, \phi_t)_{(\Omega)} + (d\nabla_\perp u_t, \nabla_\perp \phi_t)_{(\Sigma_t)} = f_g(\phi_t), \quad (7.12)$$

for all $\phi_t \in H_{\Sigma_t}^1(\Omega)$. We introduce

$$J_t = \det(\mathbf{DF}_t) \quad \text{and} \quad \mathbf{A}_t = \mathbf{DF}_t^{-1} \mathbf{DF}_t^{-*}, \quad (7.13)$$

with the superscript $*$ denoting the transpose operation and superscript $^{-*}$ denoting the transpose of the inverse. The Jacobi matrix of the transformation \mathbf{F}_t is denoted by \mathbf{DF}_t . The strong continuity of the (matrix) functions J_t , \mathbf{A}_t , and \mathbf{F}_t and the following identities can be verified [166]

$$(\nabla u_t) \circ \mathbf{F}_t = (\mathbf{DF}_t^{-*}) \nabla u^t, \quad J_t|_{t=0} = 1, \quad \mathbf{A}_t|_{t=0} = \mathbf{I} \quad (7.14)$$

$$\frac{d}{dt} \mathbf{F}_t|_{t=0} = \mathbf{V}, \quad \frac{d}{dt} (\mathbf{DF}_t)|_{t=0} = \mathbf{DV}, \quad \frac{d}{dt} (\mathbf{DF}_t^{-1})|_{t=0} = -\mathbf{DV} \quad (7.15)$$

$$J'_0 \equiv \frac{dJ_t}{dt}|_{t=0} = \nabla \cdot \mathbf{V}, \quad \mathbf{A}'_0 \equiv \frac{d\mathbf{A}_t}{dt}|_{t=0} = -(\mathbf{DV} + (\mathbf{DV})^*). \quad (7.16)$$

Here \mathbf{I} is the identity matrix.

We now replace $\nabla_\perp u_t$ on the interface Σ by $\nabla u_t^+ - (\mathbf{n}_t \cdot \nabla u_t^+) \mathbf{n}_t$. We could also replace it by $\nabla u_t^- - (\mathbf{n}_t \cdot \nabla u_t^-) \mathbf{n}_t$ and will show that the final result does not depend

on the chosen expression, as it should; see for example (7.34) below. We thus recast (7.12) as

$$\begin{aligned} & (\mathcal{D}\nabla u_t, \nabla \phi_t)_{(\Omega)} + (a u_t, \phi_t)_{(\Omega)} + (d\nabla u_t^+, \nabla \phi_t^+)_{(\Sigma_t)} \\ & \quad - (d\mathbf{n}_t \cdot \nabla u_t^+, \mathbf{n}_t \cdot \nabla \phi_t^+)_{(\Sigma_t)} = f_g(\phi_t). \end{aligned} \quad (7.17)$$

Performing the change of variables $\mathbf{x} \mapsto \mathbf{F}_t(\mathbf{x})$ in the above equality yields

$$S_t(u^t, \phi^t) = f_g(\phi^t), \quad (7.18)$$

where $\phi^t = \phi_t \circ \mathbf{F}_t$ and $S_t(u^t, \phi^t)$ are given by

$$\begin{aligned} S_t(u^t, \phi^t) & \equiv (\mathcal{D}_{\mathbf{F}_t} J_t \mathbf{A}_t \nabla u^t, \nabla \phi^t)_{(\Omega)} + (a_{\mathbf{F}_t} J_t u^t, \phi^t)_{(\Omega)} \\ & \quad + (d_{\mathbf{F}_t} \omega_t \mathbf{A}_t \nabla u^{+t}, \nabla \phi^{+t})_{(\Sigma)} - (d_{\mathbf{F}_t} \pi_t \mathbf{A}_t \mathbf{n} \cdot \nabla u^{+t}, \mathbf{A}_t \mathbf{n} \cdot \nabla \phi^{+t})_{(\Sigma)}, \end{aligned} \quad (7.19)$$

with $\mathcal{D}_{\mathbf{F}_t} \equiv \mathcal{D} \circ \mathbf{F}_t$, $a_{\mathbf{F}_t} \equiv a \circ \mathbf{F}_t$ and $d_{\mathbf{F}_t} \equiv d \circ \mathbf{F}_t$. The functions ω_t and π_t are defined as

$$\omega_t = J_t \|\mathbf{D}\mathbf{F}_t^{-*} \cdot \mathbf{n}\|_{\mathbb{R}^n}, \quad \pi_t = \frac{J_t}{\|\mathbf{D}\mathbf{F}_t^{-*} \cdot \mathbf{n}\|_{\mathbb{R}^n}} \quad (7.20)$$

with $\|\cdot\|_{\mathbb{R}^n}$ denoting the Euclidean norm in \mathbb{R}^n , and verify

$$\omega_0 = 1, \quad \pi_0 = 1 \quad (7.21)$$

$$\omega'_0 \equiv \frac{d\omega_t}{dt} \Big|_{t=0} = \nabla \cdot \mathbf{V} - \mathbf{n}^* \mathbf{D}\mathbf{V}\mathbf{n} \equiv \operatorname{div}_{\Sigma} \mathbf{V} \quad (7.22)$$

$$\pi'_0 \equiv \frac{d\pi_t}{dt} \Big|_{t=0} = \nabla \cdot \mathbf{V} + \mathbf{n}^* \mathbf{D}\mathbf{V}\mathbf{n}. \quad (7.23)$$

Choosing the test function ϕ^t in (7.3), we then deduce from (7.18) and (7.3) that

$$S(u, \phi^t) = S_t(u^t, \phi^t). \quad (7.24)$$

On the other hand, we have for all $\phi, \psi \in H_{\Sigma}^1(\Omega)$, the following result

$$\begin{aligned}
S_t(\psi, \phi) - S(\psi, \phi) &= ((\mathcal{D}_{\mathbf{F}_t} - \mathcal{D})J_t\mathbf{A}_t\nabla\psi, \nabla\phi)_{(\Omega)} + (\mathcal{D}(J_t\mathbf{A}_t - \mathbf{I})\nabla\psi, \nabla\phi)_{(\Omega)} + \\
&\quad ((a_{\mathbf{F}_t} - a)J_t\psi, \phi)_{(\Omega)} + (a(J_t - 1)\psi, \phi)_{(\Omega)} + ((d_{\mathbf{F}_t} - d)\omega_t\mathbf{A}_t\nabla\psi^+, \nabla\phi^+)_{(\Sigma)} \\
&\quad + (d(\omega_t\mathbf{A}_t - \mathbf{I})\nabla\psi^+, \nabla\phi^+)_{(\Sigma)} - ((d_{\mathbf{F}_t} - d)\pi_t\mathbf{A}_t\mathbf{n} \cdot \nabla\psi^+, \mathbf{A}_t\mathbf{n} \cdot \nabla\phi^+)_{(\Sigma)} \\
&\quad - (d(\pi_t\mathbf{A}_t - \mathbf{I})\mathbf{n} \cdot \nabla\psi^+, \mathbf{A}_t\mathbf{n} \cdot \nabla\phi)_{(\Sigma)} - (d\mathbf{n} \cdot \nabla\psi^+, (\mathbf{A}_t - \mathbf{I})\mathbf{n} \cdot \nabla\phi^+)_{(\Sigma)}, \quad (7.25)
\end{aligned}$$

which implies that

$$\begin{aligned}
|S_t(\psi, \phi) - S(\psi, \phi)| &\leq \frac{C_1(t)}{2}(\|\nabla\psi\|_{L^2(\Omega)}^2 + \|\nabla\phi\|_{L^2(\Omega)}^2) + \frac{C_2(t)}{2}(\|\psi\|_{L^2(\Omega)}^2 + \|\phi\|_{L^2(\Omega)}^2) \\
&\quad + \frac{C_3(t)}{2}(\|\nabla\psi^+\|_{L^2(\Sigma)}^2 + \|\nabla\phi^+\|_{L^2(\Sigma)}^2) + \frac{C_4(t)}{2}(\|\mathbf{n} \cdot \nabla\psi^+\|_{L^2(\Sigma)}^2 + \|\mathbf{n} \cdot \nabla\phi^+\|_{L^2(\Sigma)}^2), \quad (7.26)
\end{aligned}$$

with $C_1(t)$, $C_2(t)$, $C_3(t)$ and $C_4(t)$ given by

$$\begin{aligned}
C_1(t) &= \|(\mathcal{D}_{\mathbf{F}_t} - \mathcal{D})J_t\mathbf{A}_t\|_{L^\infty(\Omega)} + \|\mathcal{D}(J_t\mathbf{A}_t - \mathbf{I})\|_{L^\infty(\Omega)}, \\
C_2(t) &= \|(a_{\mathbf{F}_t} - a)J_t\|_{L^\infty(\Omega)} + \|a(J_t - 1)\|_{L^\infty(\Omega)}, \\
C_3(t) &= \|(d_{\mathbf{F}_t} - d)\omega_t\mathbf{A}_t\|_{L^\infty(\Omega)} + \|d(\omega_t\mathbf{A}_t - \mathbf{I})\|_{L^\infty(\Omega)}, \\
C_4(t) &= \|(d_{\mathbf{F}_t} - d)\pi_t\mathbf{A}_t\|_{L^\infty(\Omega)}\|\mathbf{A}_t\|_{L^\infty(\Omega)} \\
&\quad + \|d(\pi_t\mathbf{A}_t - \mathbf{I})\|_{L^\infty(\Omega)}\|\mathbf{A}_t\|_{L^\infty(\Omega)} + \|\mathbf{A}_t - \mathbf{I}\|_{L^\infty(\Omega)}. \quad (7.27)
\end{aligned}$$

Here the norms $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^\infty}$ are the usual ones defined on vector (matrix) functions. Because of the strong continuity of \mathbf{A}_t , J_t , ω_t and π_t (as functions of t), we deduce the following result on S_t :

Lemma 7.2.3. *The bilinear form S_t is continuous with respect to the perturbation parameter t in (7.8) at $t = 0$, which means*

$$\lim_{t \rightarrow 0^+} S_t(\cdot, \cdot) = S(\cdot, \cdot). \quad (7.28)$$

Let us recast the identity (7.24) as the following relation

$$T_1 + T_2 + T_3 - T_4 - T_5 = 0 \quad (7.29)$$

where the terms T_k are given by

$$\begin{aligned} T_1 &= \left(\frac{\mathcal{D}_{\mathbf{F}_t} - \mathcal{D}}{t} J_t \mathbf{A}_t \nabla u^t + \mathcal{D} J_t \mathbf{A}_t \frac{\nabla u^t - \nabla u}{t} + \mathcal{D} \frac{J_t \mathbf{A}_t - \mathbf{I}}{t} \nabla u, \nabla \phi^t \right)_{(\Omega)} \\ T_2 &= \left(\frac{a_{\mathbf{F}_t} - a}{t} J_t u^t + a J_t \frac{u^t - u}{t} + a \frac{J_t - 1}{t} u, \phi^t \right)_{(\Omega)} \\ T_3 &= \left(\frac{d_{\mathbf{F}_t} - d}{t} \omega_t \mathbf{A}_t \nabla u^{+t} + d \frac{\omega_t \mathbf{A}_t - \mathbf{I}}{t} \nabla u^{+t} + d \frac{\nabla u^{+t} - \nabla u^+}{t}, \nabla \phi^{+t} \right)_{(\Sigma)} \\ T_4 &= \left(\frac{d_{\mathbf{F}_t} - d}{t} \pi_t \mathbf{A}_t \mathbf{n} \cdot \nabla u^{+t} + d \frac{\pi_t \mathbf{A}_t - \mathbf{I}}{t} \mathbf{n} \cdot \nabla u^{+t}, \mathbf{A}_t \mathbf{n} \cdot \nabla \phi^{+t} \right)_{(\Sigma)} \\ T_5 &= \left(d \frac{\mathbf{n} \cdot \nabla u^{+t} - \mathbf{n} \cdot \nabla u^+}{t}, \mathbf{A}_t \mathbf{n} \cdot \nabla \phi^{+t} \right)_{(\Sigma)} + \left(d \mathbf{n} \cdot \nabla u^+, \frac{\mathbf{A}_t - \mathbf{I}}{t} \mathbf{n} \cdot \nabla \phi^{+t} \right)_{(\Sigma)}. \end{aligned}$$

Thanks to the continuity of S_t at $t = 0$, we can take a limit $t \rightarrow 0$ in (7.29) and obtain the following equation for the material derivative of u :

$$\begin{aligned} & (\mathbf{V} \cdot \nabla \mathcal{D} \nabla u, \nabla \phi)_{(\Omega)} + (\mathcal{D} \nabla \dot{u}, \nabla \phi)_{(\Omega)} + (\mathcal{D} (J'_0 \mathbf{I} + \mathbf{A}'_0) \nabla u, \nabla \phi)_{(\Omega)} + (\mathbf{V} \cdot \nabla a u, \phi)_{(\Omega)} + (a \dot{u}, \phi)_{(\Omega)} \\ & + (a J'_0 u, \phi)_{(\Omega)} + (\mathbf{V}_\perp \cdot \nabla_\perp d \nabla u^+, \nabla \phi^+)_{(\Sigma)} + (d (\omega'_0 \mathbf{I} + \mathbf{A}'_0) \nabla u^+, \nabla \phi^+)_{(\Sigma)} + (d \nabla \dot{u}^+, \nabla \phi^+)_{(\Sigma)} \\ & - (\mathbf{V}_\perp \cdot \nabla_\perp d \mathbf{n} \cdot \nabla u^+, \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} - (d (\pi'_0 \mathbf{I} + \mathbf{A}'_0) \mathbf{n} \cdot \nabla u^+, \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} \\ & - (d \mathbf{n} \cdot \nabla \dot{u}^+, \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} - (d \mathbf{n} \cdot \nabla u^+, \mathbf{A}'_0 \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} = 0. \quad (7.30) \end{aligned}$$

Using the expressions for \mathbf{A}'_0 , J'_0 , ω'_0 and π'_0 , we can show that the following simplifications are possible:

$$\begin{aligned} & (d (\omega'_0 \mathbf{I} + \mathbf{A}'_0) \nabla u^+, \nabla \phi^+)_{(\Sigma)} - (d (\pi'_0 \mathbf{I} + \mathbf{A}'_0) \mathbf{n} \cdot \nabla u^+, \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} \\ & - (d \mathbf{n} \cdot \nabla u^+, \mathbf{A}'_0 \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} = (d \operatorname{div}_\Sigma \mathbf{V} \nabla_\perp u, \nabla_\perp \phi)_{(\Sigma)} + (d \mathbf{A}'_0 \nabla_\perp u, \nabla_\perp \phi)_{(\Sigma)}, \quad (7.31) \end{aligned}$$

where we have replaced $\nabla u^+ - (\mathbf{n} \cdot \nabla u^+) \mathbf{n}$ by $\nabla_\perp u$. The quantity $\operatorname{div}_\Sigma \mathbf{V}$ is defined

in (7.22). It can also be shown that

$$\begin{aligned} (\mathbf{V}_\perp \cdot \nabla_\perp d\nabla u^+, \nabla \phi^+)_{(\Sigma)} - (\mathbf{V}_\perp \cdot \nabla_\perp d\mathbf{n} \cdot \nabla u^+, \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} \\ = (\mathbf{V}_\perp \cdot \nabla_\perp d\nabla_\perp u, \nabla_\perp \phi)_{(\Sigma)}, \end{aligned} \quad (7.32)$$

and

$$(d\nabla \dot{u}^+, \nabla \phi^+)_{(\Sigma)} - (d\mathbf{n} \cdot \nabla \dot{u}^+, \mathbf{n} \cdot \nabla \phi^+)_{(\Sigma)} = (d\nabla_\perp \dot{u}, \nabla_\perp \phi)_{(\Sigma)}. \quad (7.33)$$

We can thus simplify (7.30) as

$$\begin{aligned} S(\dot{u}, \phi) = & -(\mathbf{V} \cdot \nabla \mathcal{D} \nabla u, \nabla \phi)_{(\Omega)} - (\mathcal{D}(J'_0 \mathbf{I} + \mathbf{A}'_0) \nabla u, \nabla \phi)_{(\Omega)} - (\nabla \cdot (a \mathbf{V}) u, \phi)_{(\Omega)} \\ & - (\mathbf{V}_\perp \cdot \nabla_\perp d\nabla_\perp u, \nabla_\perp \phi)_{(\Sigma)} - (d \operatorname{div}_\Sigma \mathbf{V} \nabla_\perp u, \nabla_\perp \phi)_{(\Sigma)} - (d \mathbf{A}'_0 \nabla_\perp u, \nabla_\perp \phi)_{(\Sigma)}. \end{aligned} \quad (7.34)$$

We summarize the above results in the following theorem:

Theorem 7.2.4. *Let $\mathcal{D}(\mathbf{x})$, $a(\mathbf{x})$ and $d(\mathbf{x})$ be functions of class C^1 . Then the material derivative $\dot{u} \in H^1_\Sigma(\Omega)$ of the solution $u \in H^1_\Sigma(\Omega)$ to (7.1) in direction \mathbf{V} is the unique solution to (7.34). Moreover, we verify that*

$$[\dot{u}] = 0, \quad \text{on } \Sigma. \quad (7.35)$$

The condition (7.35) comes from the third identity in (7.1).

7.2.2 The shape derivative

The shape derivative of u can be computed by using (7.11). However, before we proceed to computing it, we stress that u' can no longer be an element of H^1_Σ . The jump of the normal derivative of u across the interface Σ causes a discontinuity of the tangential derivative of u' across the interface according to formula (7.11), i.e., $\nabla_\perp u'(\mathbf{x}^+) \neq \nabla_\perp u'(\mathbf{x}^-)$. Let us introduce the following Hilbert space

$$Z^1_\Sigma(\Omega) := \left\{ v(\mathbf{x}) : v \in H^1(\Omega_I) \otimes H^1(\Omega_E), \text{ s.t. } \int_\Sigma |\nabla_\perp v^+|^2 d\sigma + \int_\Sigma [v]^2 d\sigma < \infty \right\}. \quad (7.36)$$

We also define $\kappa(\mathbf{x})$ as the mean curvature of Σ (seen as a $n - 1$ manifold embedded in \mathbb{R}^n) at $\mathbf{x} \in \Sigma$. We now state the main result of this paper, which allows us to characterize the shape derivative of u :

Theorem 7.2.5. *Assume that $\mathcal{D}(\mathbf{x})$, $a(\mathbf{x})$ and $d(\mathbf{x})$ are functions of class C^1 . Then the shape derivative $u' \in Z_{\Sigma}^1(\Omega)$ of the solution $u \in H_{\Sigma}^1(\Omega)$ to (7.1) in direction \mathbf{V} , is the unique solution of*

$$\begin{aligned} & (\mathcal{D}\nabla u', \nabla\phi)_{(\Omega)} + (au', \phi)_{(\Omega)} + (d\nabla_{\perp}u'^+, \nabla_{\perp}\phi)_{(\Sigma)} \\ & \quad = -(d\operatorname{div}_{\Sigma}\mathbf{V}_{\perp}\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} - (d\kappa V_n\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} \\ & \quad - (\mathbf{V}_{\perp} \cdot \nabla_{\perp}d\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} + (\mathbf{V} \cdot \nabla u^+, \nabla_{\perp} \cdot d\nabla_{\perp}\phi)_{(\Sigma)} \\ & \quad \quad - (d\mathbf{A}'_0\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} + (\mathbf{V}_{\perp} \cdot \nabla_{\perp}\phi, \nabla \cdot d\nabla_{\perp}u)_{(\Sigma)}, \end{aligned} \quad (7.37)$$

for all $\phi \in H_{\Sigma}^1(\Omega)$. Moreover, the jump of u' across Σ is given by

$$[u'] = -[\mathbf{V} \cdot \nabla u]. \quad (7.38)$$

We remark that thanks to the above jump conditions, (7.37) still holds if the following substitutions are performed:

$$\begin{aligned} (d\nabla_{\perp}u'^+, \nabla_{\perp}\phi)_{(\Sigma)} & \rightarrow (d\nabla_{\perp}u'^-, \nabla_{\perp}\phi)_{(\Sigma)} \\ (\mathbf{V} \cdot \nabla u^+, \nabla_{\perp} \cdot d\nabla_{\perp}\phi)_{(\Sigma)} & \rightarrow (\mathbf{V} \cdot \nabla u^-, \nabla_{\perp} \cdot d\nabla_{\perp}\phi)_{(\Sigma)}. \end{aligned} \quad (7.39)$$

We also remark that the source term (right-hand side) in (7.37) only involves terms defined on Σ . This is natural, for all other constitutive parameters of (7.1) are kept independent of t , and should be contrasted with the results obtained in (7.34) for the material derivative in Lagrangian coordinates.

Proof of Theorem 7.2.5. First, replacing \dot{u} in (7.35) by $u' + \mathbf{V} \cdot \nabla u$ yields the jump

condition of u' across the interface, (7.38). Similar replacements in (7.34) lead to

$$\begin{aligned}
& (\mathcal{D}\nabla u', \nabla\phi)_{(\Omega)} + (au', \phi)_{(\Omega)} + (d\nabla_{\perp}u'^+, \nabla_{\perp}\phi)_{(\Sigma)} = -(\mathcal{D}\nabla(\mathbf{V} \cdot \nabla u), \nabla\phi)_{(\Omega)} \\
& \quad - ((\mathbf{V} \cdot \nabla\mathcal{D})\nabla u, \nabla\phi)_{(\Omega)} - (\mathcal{D}(J'_0\mathbf{I} + \mathbf{A}'_0)\nabla u, \nabla\phi)_{(\Omega)} - (a\mathbf{V} \cdot \nabla u, \phi)_{(\Omega)} \\
& \quad - (\nabla \cdot (a\mathbf{V})u, \phi)_{(\Omega)} - ((\mathbf{V}_{\perp} \cdot \nabla_{\perp}d)\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} - (d\operatorname{div}_{\Sigma}\mathbf{V}\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} \\
& \quad \quad - (d\nabla_{\perp}(\mathbf{V} \cdot \nabla u^+), \nabla_{\perp}\phi)_{(\Sigma)} - (d\mathbf{A}'_0\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)}. \quad (7.40)
\end{aligned}$$

We then verify by integrations by parts that

$$\begin{aligned}
& (\mathcal{D}(J'_0\mathbf{I} + \mathbf{A}'_0)\nabla u, \nabla\phi)_{(\Omega)} = (\mathcal{D}\nabla \cdot \mathbf{V}\nabla u, \nabla\phi)_{(\Omega)} - (\mathcal{D}\nabla(\mathbf{V} \cdot \nabla u), \nabla\phi)_{(\Omega)} \\
& \quad + (\mathcal{D}(\mathbf{V} \cdot \nabla)\nabla u, \nabla\phi)_{(\Omega)} - (\mathcal{D}(\nabla u \cdot \nabla)\mathbf{V}, \nabla\phi)_{(\Omega)}. \quad (7.41)
\end{aligned}$$

This implies the following:

$$\begin{aligned}
& (\mathcal{D}\nabla(\mathbf{V} \cdot \nabla u), \nabla\phi)_{(\Omega)} + (\mathbf{V} \cdot \nabla\mathcal{D}\nabla u, \nabla\phi)_{(\Omega)} + (\mathcal{D}(J'_0\mathbf{I} + \mathbf{A}'_0)\nabla u, \nabla\phi)_{(\Omega)} \\
& \quad = (\nabla \cdot (\mathcal{D}\nabla u), \mathbf{V} \cdot \nabla\phi)_{(\Omega)} - (\mathbf{V}_{\perp} \cdot \nabla_{\perp}\phi, \nabla \cdot d\nabla_{\perp}u)_{(\Sigma)}. \quad (7.42)
\end{aligned}$$

The terms on the boundary $\Gamma = \partial\Omega$ vanish because \mathbf{V} has compact support in Ω .

Thanks to the above identity, (7.40) may be recast as

$$\begin{aligned}
& (\mathcal{D}\nabla u', \nabla\phi)_{(\Omega)} + (au', \phi)_{(\Omega)} + (d\nabla_{\perp}u'^+, \nabla_{\perp}\phi)_{(\Sigma)} \\
& \quad = -(\nabla \cdot (\mathcal{D}\nabla u), \mathbf{V} \cdot \nabla\phi)_{(\Omega)} - (\mathbf{V} \cdot \nabla au, \phi)_{(\Omega)} - (a\mathbf{V} \cdot \nabla u, \phi)_{(\Omega)} - (aJ'_0u, \phi)_{(\Omega)} \\
& \quad - ((\mathbf{V}_{\perp} \cdot \nabla_{\perp}d)\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} - (d\operatorname{div}_{\Sigma}\mathbf{V}\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} + (\mathbf{V} \cdot \nabla u^+, \nabla_{\perp} \cdot d\nabla_{\perp}\phi)_{(\Sigma)} \\
& \quad \quad - (d\mathbf{A}'_0\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)} + (\mathbf{V}_{\perp} \cdot \nabla_{\perp}\phi, \nabla \cdot d\nabla_{\perp}u)_{(\Sigma)}. \quad (7.43)
\end{aligned}$$

Further integrations by parts in (7.1) allow us to show that

$$-(\nabla \cdot (\mathcal{D}\nabla u), \mathbf{V} \cdot \nabla\phi)_{(\Omega)} - (\mathbf{V} \cdot \nabla au, \phi)_{(\Omega)} - (a\mathbf{V} \cdot \nabla u, \phi)_{(\Omega)} - (aJ'_0u, \phi)_{(\Omega)} = 0.$$

These lengthy calculations and combined with the following result [166, proposition

2.57]

$$\operatorname{div}_{\Sigma} \mathbf{V} = \operatorname{div}_{\Sigma} \mathbf{V}_{\perp} + \kappa V_n$$

finally yield (7.37). \square

Remark 7.2.6. In some applications (including the analysis of clear layers in optical tomography [18, 23]), it may be necessary to generalize the above calculations to the situation where the tangential diffusion coefficient d depends on the geometry of the interface; for instance via its curvature. In that case, we have to impose that $d(\mathbf{x})$, assumed to be known, is shape differentiable with respect to Σ_t . Theorem 7.2.5 then still holds provided that we add the term $-(d' \nabla_{\perp} u, \nabla_{\perp} \phi)_{\Sigma}$ to the right hand side in (7.37). Although this may not be as relevant practically, similar generalizations are possible to the case where $\mathcal{D}(\mathbf{x})$ and $a(\mathbf{x})$ also depend on the geometry of the interface.

The calculation of the material and shape derivatives of the solution u to (7.1) can also be done with model (7.6). We provide the following result without detailing its derivation. Similar results when \mathcal{D} and $\delta\mathcal{D}$ are constant can be found in [98, 117, 166].

Theorem 7.2.7. *Assume that $\mathcal{D}(\mathbf{x})$ and $a(\mathbf{x})$ are functions of class C^1 . Then the material derivative $\dot{u} \in H^1$ of the solution u to equation (7.6) is the unique solution to*

$$\begin{aligned} & (\mathcal{D} \nabla \dot{u}, \nabla \phi)_{(\Omega)} + (a \dot{u}, \phi)_{(\Omega)} \\ & = -(\mathbf{V} \cdot \nabla \mathcal{D} \nabla u, \nabla \phi)_{(\Omega)} - (\mathcal{D} \mathbf{A}'_0 \nabla u, \nabla \phi)_{(\Omega)} - (\nabla \cdot (a \mathbf{V}) u, \phi)_{(\Omega)} \end{aligned} \quad (7.44)$$

for all $\phi \in H^1_{\Sigma}(\Omega)$. The shape derivative of $u' \in H^1$ of $u \in H^1$ then satisfies

$$(\mathcal{D} \nabla u', \nabla \phi)_{(\Omega)} + (a u', \phi)_{(\Omega)} = -(\delta \mathcal{D} V_n \nabla_{\perp} u, \nabla_{\perp} \phi)_{(\Sigma)} \quad (7.45)$$

for all $\phi \in H^1_{\Sigma}(\Omega)$.

The proof of this theorem is very similar to that of Thm. 7.2.5 except that we

have to replace identity (7.42) by

$$\begin{aligned} & (\mathcal{D}\nabla(\mathbf{V} \cdot \nabla u), \nabla\phi)_{(\Omega)} + ((J'_0\mathbf{I} + \mathbf{A}'_0)\mathcal{D}\nabla u, \nabla\phi)_{(\Omega)} + (\mathbf{V} \cdot \nabla\mathcal{D}\nabla u, \nabla\phi)_{(\Omega)} \\ & = (\nabla \cdot (\mathcal{D}\nabla u), \mathbf{V} \cdot \nabla\phi)_{(\Omega)} + (\delta\mathcal{D}V_n\nabla_{\perp}u, \nabla_{\perp}\phi)_{(\Sigma)}. \end{aligned} \quad (7.46)$$

Remark 7.2.8. The method based on the map in (7.8) that we have adopted in the paper is not the only choice for shape sensitivity analysis. An *a priori* more general method called the speed (or velocity) method consists of defining the transform \mathbf{F}_t by $\mathbf{F}_t = \mathbf{X}(\mathbf{x}, t)$ with $\mathbf{X}(\mathbf{x}, t)$ the solution of the following equation:

$$\begin{aligned} \dot{\mathbf{X}}(t, \mathbf{x}) &= \mathbf{V}(t, \mathbf{X}(t, \mathbf{x})), \\ \mathbf{X}(0, \mathbf{x}) &= \mathbf{x}. \end{aligned} \quad (7.47)$$

It has been shown that the velocity method and the transform method used in this paper are actually equivalent in the sense that under sufficient regularity conditions, it is possible to associate a unique velocity field to a given transform \mathbf{F}_t and vice versa [166, sec. 2.9].

Remark 7.2.9. The calculations obtained in the preceding two sections show that the Banach space $B(\Omega)$ may be chosen as the Hilbert $H^1_{\Sigma}(\Omega)$ in the definition 7.2.2 of the material derivative for model (7.1) and as $H^1(\Omega)$ for model (7.6); this is because $[\dot{u}] = 0$ across Σ . In both cases, thanks to estimates of the form $C_k(t) \leq Ct$ for a constant C in (7.27), we can show that convergence occurs for the strong topology.

The definition of the space $B(\Omega)$ in definition 7.2.1 is the same for model (7.6). It is however more complicated for model (7.1). Because u' jumps across Σ , it is not an element of $H^1(\Omega)$, let alone $H^1_{\Sigma}(\Omega)$. We can however choose $B(\Omega) = L^2(\Omega)$ and observe that convergence in (7.9) is strong in that space. The singular interface model (7.1) introduces singularities that are not present in the inclusion model (7.6).

7.3 Choosing the direction of descent

The analysis presented in the last section enables us to compute the sensitivity of the error functional (7.48) to geometric changes in the interface. Since the vector field $\mathbf{V}(\mathbf{x})$ in (7.8) has compact support, the boundary Γ stays unaffected by perturbations in the interface. We can thus obtain the Eulerian derivative of the error functional as

$$d\mathcal{F}_\alpha(\Sigma) := \lim_{t \rightarrow 0} \frac{\mathcal{F}_\alpha(\Sigma_t) - \mathcal{F}_\alpha(\Sigma)}{t} = (u - u_m^\delta, u')_{(\Gamma)} + \alpha(\kappa(\mathbf{x}), V_n)_{(\Sigma)}. \quad (7.48)$$

The second term comes from [98, 166]:

$$\left(\frac{d}{dt} \int_{\Sigma_t} d\sigma_t(\mathbf{x}) \right) \Big|_{\Sigma} = \int_{\Sigma} \kappa(\mathbf{x}) V_n d\sigma(\mathbf{x}). \quad (7.49)$$

We recall that $\kappa(\mathbf{x})$ is the mean curvature of the interface Σ at $\mathbf{x} \in \Sigma$.

Since we want the error functional (7.5) to decrease as the interface moves, we need to find a vector field \mathbf{V} such that $d\mathcal{F}_\alpha(\Sigma) \leq 0$. Let us denote by w the solution to the following adjoint equation

$$\begin{aligned} -\nabla \cdot \mathcal{D}(\mathbf{x}) \nabla w(\mathbf{x}) + a(\mathbf{x})w(\mathbf{x}) &= 0 && \text{in } \Omega \setminus \Sigma \\ \mathcal{D}(\mathbf{x}) \boldsymbol{\nu}(\mathbf{x}) \cdot \nabla w(\mathbf{x}) &= u - u_m^\delta && \text{on } \Gamma \\ [w] &= 0 && \text{on } \Sigma \\ [\mathbf{n} \cdot \mathcal{D} \nabla w] &= -\nabla_\perp \cdot d(\mathbf{x}) \nabla_\perp w(\mathbf{x}) && \text{on } \Sigma. \end{aligned} \quad (7.50)$$

Upon multiplying (7.50) by u' , performing an integration by parts and taking into account the fact that u' jumps across the interface, we obtain that

$$\begin{aligned} (\mathcal{D} \nabla w, \nabla u')_{(\Omega)} + (aw, u')_{(\Omega)} + (d \nabla_\perp w, \nabla_\perp u'^+)_{(\Sigma)} \\ = (u - u_m^\delta, u')_{(\Gamma)} - ([u'], \mathcal{D} \mathbf{n} \cdot \nabla w^-)_{(\Sigma)}. \end{aligned} \quad (7.51)$$

We also observe that the solution of (7.50) belongs to H_Σ^1 . Replacing the test function

ϕ in (7.37) by w , we obtain

$$\begin{aligned}
& (\mathcal{D}\nabla u', \nabla w)_{(\Omega)} + (au', w)_{(\Omega)} + (d\nabla_{\perp} u'^+, \nabla_{\perp} w)_{(\Sigma)} \\
&= -(d\operatorname{div}_{\Sigma} \mathbf{V}_{\perp} \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} - (d\kappa V_n \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} + (\mathbf{V} \cdot \nabla u^+, \nabla_{\perp} \cdot d\nabla_{\perp} w)_{(\Sigma)} \\
&\quad - (\mathbf{V}_{\perp} \cdot \nabla_{\perp} d\nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} - (d\mathbf{A}'_0 \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} + (\mathbf{V}_{\perp} \cdot \nabla_{\perp} w, \nabla \cdot d\nabla_{\perp} u)_{(\Sigma)}.
\end{aligned} \tag{7.52}$$

The above equations (7.51) and (7.52) imply that

$$\begin{aligned}
(u - u_m^{\delta}, u')_{(\Gamma)} &= ([u'], \mathcal{D}\mathbf{n} \cdot \nabla w^-)_{(\Sigma)} - (d\operatorname{div}_{\Sigma} \mathbf{V}_{\perp} \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} \\
&\quad - (d\kappa V_n \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} - (\mathbf{V}_{\perp} \cdot \nabla_{\perp} d\nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} + (\mathbf{V} \cdot \nabla u^+, \nabla_{\perp} \cdot d\nabla_{\perp} w)_{(\Sigma)} \\
&\quad - (d\mathbf{A}'_0 \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} + (\mathbf{V}_{\perp} \cdot \nabla_{\perp} w, \nabla \cdot d\nabla_{\perp} u)_{(\Sigma)}.
\end{aligned} \tag{7.53}$$

Since the tangential component of \mathbf{V} does not affect the evolution of the interface [160, 166], we can assume that the vector field $\mathbf{V}|_{\Sigma}$ is normal to Σ , i.e., $\mathbf{V}_{\perp}|_{\Sigma} = \mathbf{0}$. Then a combination of (7.48) and (7.53) yields

$$\begin{aligned}
d\mathcal{F}_{\alpha}(\Sigma) &= (V_n \nabla_{\perp} \cdot d\nabla_{\perp} u, \mathbf{n} \cdot \nabla w^-)_{(\Sigma)} + (V_n \mathbf{n} \cdot \nabla u^+, \nabla_{\perp} \cdot d\nabla_{\perp} w)_{(\Sigma)} \\
&\quad - (V_n d\kappa \nabla_{\perp} u, \nabla_{\perp} w)_{(\Sigma)} + (\alpha\kappa, V_n)_{(\Sigma)}.
\end{aligned} \tag{7.54}$$

Using the interface conditions in (7.1) and (7.50) we can further simplify the above equality as

$$d\mathcal{F}_{\alpha}(\Sigma) = (V_n, -d\kappa \nabla_{\perp} u \cdot \nabla_{\perp} w - \mathbf{n} \cdot \nabla u^+ \mathbf{n} \cdot \mathcal{D}\nabla w^+ + \mathbf{n} \cdot \nabla u^- \mathbf{n} \cdot \mathcal{D}\nabla w^+ + \alpha\kappa)_{(\Sigma)}. \tag{7.55}$$

It remains to choose \mathbf{V} such that $d\mathcal{F}_{\alpha}(\Sigma) \leq 0$. For the singular surface model (7.1) and the model of inclusion (7.6), we show the following result.

Proposition 7.3.1. *For the model in (7.1), the functional $\mathcal{F}_{\alpha}(\Sigma)$ given in (7.5) will*

not increase if the interface moves according to a vector field characterized by

$$V_n = d\kappa \nabla_{\perp} u \cdot \nabla_{\perp} w + \mathbf{n} \cdot \nabla u^+ \mathbf{n} \cdot \mathcal{D} \nabla w^+ - \mathbf{n} \cdot \nabla u^- \mathbf{n} \cdot \mathcal{D} \nabla w^- - \alpha \kappa, \quad (7.56)$$

where u and w solve (7.1) and (7.50), respectively. For the model given by (7.6), the functional $\mathcal{F}_{\alpha}(\Sigma)$ (7.5) is non-increasing if the interface Σ moves in the direction

$$V_n = -\left(\delta \mathcal{D} \nabla_{\perp} u \cdot \nabla_{\perp} w + \alpha \kappa\right), \quad (7.57)$$

where u solves (7.6) and w solves the adjoint problem:

$$\begin{aligned} -\nabla \cdot \mathcal{D}(\mathbf{x}) \nabla w(\mathbf{x}) + a(\mathbf{x})w(\mathbf{x}) &= 0 && \text{in } \Omega \\ \mathcal{D}(\mathbf{x}) \boldsymbol{\nu}(\mathbf{x}) \cdot \nabla w(\mathbf{x}) &= u - u_m^{\delta} && \text{on } \Gamma \\ [w] &= 0 && \text{on } \Sigma \\ [\mathbf{n} \cdot \mathcal{D} \nabla w] &= 0 && \text{on } \Sigma, \end{aligned} \quad (7.58)$$

with the diffusion coefficient $\mathcal{D}(\mathbf{x})$ given by (7.7).

Note that (7.57) is the well-known result for the inverse obstacle problem obtained by shape sensitivity analysis [29, 98, 117, 159]. Allowing the diffusion coefficient \mathcal{D} to be spatially dependent in model (7.6) does not modify the choice of a velocity field. In the inverse problem for singular surfaces, both the geometry of the surface (via its mean curvature κ) and the tangential diffusion process it carries, enter non-trivially in the choice of the vector field given in (7.56).

7.4 Level set implementation

Once the direction of descent has been chosen, we need an efficient way to move the interface along that direction. We use here the level set method [135, 138] to do so. The level set method represents interfaces as the zero level sets of level set functions and then moves of the interfaces implicitly by solving a Hamilton-Jacobi equation for the level set functions. The application of the level set method to shape optimization

problem has been pioneered in [159] and further studied in [7, 29, 47, 98]. We refer to the recent monographs [135, 136, 163] and their references therein for a detailed account of the method and its many applications.

7.4.1 Representing and moving interfaces

Let Σ_t be an evolution interface in $\Omega \subset \mathbb{R}^n$ viewed as the zero level set of a function $\psi(\mathbf{x}, t)$:

$$\Sigma_t := \{\mathbf{x} : \mathbf{x} \in \Omega, \text{ such that } \psi(\mathbf{x}, t) = 0\}.$$

To track the position of the interface Σ_t , we evaluate the derivative of $\psi(\mathbf{x}(t), t) = 0$ with respect to t to obtain

$$\frac{\partial \psi}{\partial t} + \dot{\mathbf{x}}(t) \cdot \nabla \psi = \frac{\partial \psi}{\partial t} + \mathbf{V} \cdot \nabla \psi = 0, \quad (7.59)$$

where \mathbf{V} is the velocity field at the interface. Since the tangential velocity does not affect the evolution of the interface [160], we can choose $\mathbf{V}_\perp = \mathbf{0}$. Using the fact that the normal vector of the interface can be written as $\mathbf{n}(\mathbf{x}) = \nabla \psi / |\nabla \psi|$, we arrive at

$$\frac{\partial \psi}{\partial t} + V_n |\nabla \psi| = 0. \quad (7.60)$$

This is a nonlinear transport equation of the Hamilton-Jacobi form. Let us now suppose that we know an approximate position for the interface and the normal velocity V_n at a given “time step”. Then by solving this Hamilton-Jacobi equation, we can compute the position of the interface at the following “time step”.

7.4.2 Implementation of the level set method

The level set method is implemented numerically as follows. We focus on the two-dimensional setting to simplify the calculations.

Algorithm:

- L1. We choose an initial level set function $\psi^0(\mathbf{x})$, such that the interface can be

represented as $\Sigma_0 = \{\mathbf{x} : \mathbf{x} \in \Omega, \psi^0(\mathbf{x}) = 0\}$, and set $k = 0$;

- L2. We solve the state equation (7.1) (resp. (7.6)) with the interface $\Sigma_k = \{\mathbf{x} : \mathbf{x} \in \Omega, \psi^k(\mathbf{x}) = 0\}$;
- L3. We compare the solution with given measurements. If a stopping criteria is satisfied, we stop the calculation. Otherwise:
- L4. We solve the adjoint equation (7.50) (resp. 7.58)) to compute the normal velocity V_n on Σ_k by (7.56) (resp. (7.57)). We extend the velocity field to a computational tube around Σ by using (7.64) below;
- L5. We move the interface Σ_k to a new interface Σ_{k+1} by updating the Hamilton-Jacobi equation (7.60) by one time step Δt ;
- L6. We re-initialize the level set function according to equation (7.67) if necessary;
- L7. We set $k := k + 1$ and go back to step L2.

Here are additional details about the implementation. The Hamilton-Jacobi equation (7.60) has been discretized by using the following first-order scheme [135]

$$\frac{\psi_{i,j}^{n+1} - \psi_{i,j}^n}{\Delta t} + \max(V_{i,j}^n, 0)H_+ + \min(V_{i,j}^n, 0)H_- = 0, \quad (7.61)$$

where the superscript n and subscript i, j denote time and space grid point, respectively. The numerical Hamiltonians H_+ and H_- are given by

$$\begin{aligned} H_+ &= \sqrt{\max(a, 0)^2 + \min(b, 0)^2 + \max(c, 0)^2 + \min(d, 0)^2} \\ H_- &= \sqrt{\min(a, 0)^2 + \max(b, 0)^2 + \min(c, 0)^2 + \max(d, 0)^2} \end{aligned} \quad (7.62)$$

with

$$\begin{aligned} a &\equiv D_x^- \psi_{i,j}^n := \frac{\psi_{i,j}^n - \psi_{i-1,j}^n}{\Delta x}, & b &\equiv D_x^+ \psi_{i,j}^n := \frac{\psi_{i+1,j}^n - \psi_{i,j}^n}{\Delta x}, \\ c &\equiv D_y^- \psi_{i,j}^n := \frac{\psi_{i,j}^n - \psi_{i,j-1}^n}{\Delta y}, & d &\equiv D_y^+ \psi_{i,j}^n := \frac{\psi_{i,j+1}^n - \psi_{i,j}^n}{\Delta y}. \end{aligned} \quad (7.63)$$

The time step Δt is chosen so small as to satisfy the CFL stability condition. The surface Σ_k is updated to Σ_{k+1} after each iteration of the Hamilton-Jacobi equation and the vector field is updated according to (7.56).

The vector field V_n in (7.56) is only defined at the interface Σ . We need to extend it in the neighborhood of Σ to solve the Hamilton Jacobi equations. This is done by using the following two way extrapolation equation [50]

$$V_t + S(\psi) \frac{\nabla \psi}{|\nabla \psi|} \cdot \nabla V = 0, \quad (7.64)$$

where the sign function is defined as

$$S(\psi) = \begin{cases} -1 & \text{if } \psi < 0 \\ 0 & \text{if } \psi = 0 \\ +1 & \text{if } \psi > 0. \end{cases} \quad (7.65)$$

A detailed discussion can be found in reference [135, P.76]. The equation (7.64) is solved as follows [140]:

$$\begin{aligned} \frac{V_{ij}^{n+1} - V_{ij}^n}{\Delta t} + \max(S_{ij} n_{ij}^x, 0) D_x^- V_{ij}^n + \min(S_{ij} n_{ij}^x, 0) D_x^+ V_{ij}^n \\ + \max(S_{ij} n_{ij}^y, 0) D_y^- V_{ij}^n + \min(S_{ij} n_{ij}^y, 0) D_y^+ V_{ij}^n = 0, \end{aligned} \quad (7.66)$$

over a time interval of roughly 5 – 10 times Δt , where $D_x^\pm V_{ij}^n$ and $D_y^\pm V_{ij}^n$ are finite differences defined as in (7.63). The sign function $S(\psi)$ is approximated by $\frac{\psi}{\sqrt{\psi^2 + \delta^2}}$ with δ a small regularization parameter. The directions $\hat{n} = (n^x, n^y) = (\frac{\psi_x}{\sqrt{\psi_x^2 + \psi_y^2}}, \frac{\psi_y}{\sqrt{\psi_x^2 + \psi_y^2}})$ are computed by a central difference scheme.

Finally, we comment on the re-initialization process (step *L6* in the above algorithm). The level set function may become very flat or very steep near the interface Σ . To avoid this, we replace the level set function $\psi(\mathbf{x}, t)$ by $d(\mathbf{x}, t)$ which is the value of the signed distance from \mathbf{x} to Σ . The quantity $d(\mathbf{x}, t)$ satisfies the Eikonal equation $|\nabla d| = 1$, and is the steady state solution of the following re-initialization equation

$$\begin{aligned} \frac{\partial \psi}{\partial t} + S(\psi_0)(|\nabla \psi| - 1) &= 0 & \text{in } \Omega \times (0, +\infty) \\ \psi(\mathbf{x}, t) &= \psi_0 & \text{in } \Omega \times \{0\}. \end{aligned} \quad (7.67)$$

A stationary solution of (7.67) is obtained by choosing t large enough; see [135, 171]. Here we approximate the function $S(\psi_0)$ by $\frac{\psi_0}{\sqrt{\psi_0^2 + |\nabla\psi_0|^2 \Delta x^2}}$ as suggested in [140]. The numerical scheme for equation (7.67) is given by [140, p.427]:

$$\frac{\psi_{ij}^{n+1} - \psi_{ij}^n}{\Delta t} + \max(S_{ij}, 0)(H_+ - 1) + \min(S_{ij}, 0)(H_- - 1), \quad (7.68)$$

where H_{\pm} are defined as in (7.62). In the examples shown in the next section, we reinitialize the level set function every ten time steps.

7.5 Numerical simulations

In this section, we numerically invert the singular surface problem (7.1) and the inclusion's support problem (7.6) by using shape derivative analysis and the level set method. We consider the two-dimensional setting and the domain $\Omega = (-1, 1) \times (-1, 1)$. This domain is discretized by a uniform 401×401 grid on which all the Hamilton-Jacobi equations are solved by using the finite difference schemes described above and the elliptic equations (7.1) and (7.50) during the iterative process are solved by the finite element method on rectangular elements [39] and a nonlinear conjugate gradient solver. All the numerical minimizations of the error functional (7.48) presented in this paper are performed with the optimal choice of the regularization parameter α obtained by the Morozov discrepancy principle [67].

The synthetic data are calculated by solving (7.1) and (7.50) by a finite element method on an unstructured triangulation with approximately the same number of nodes as the uniform grid mentioned above. The only common nodes of the two set of meshes are the boundary nodes where the measurements are taken. We have checked that the systematic error between the solutions on the uniform mesh and the fined unstructured mesh is far below 0.05%. The synthetic measurements have been obtained by a different numerical procedure than what is being used in the reconstruction algorithm to limit the occurrence of “inverse crimes”, where the minimization of the un-penalized functional (7.48) with $\alpha = 0$ may return the correct answer for the

wrong reasons; see [54] for an account of this problem.

In all simulations, we have chosen the diffusion coefficients to be $(\mathcal{D}, d)=(1.0, 0.3)$ in model (7.1) and $(\mathcal{D}_0, \delta\mathcal{D})=(1.0, 3.0)$ in model (7.6). The absorption coefficient $a = 0$ in both models. The values taken by these parameters have a significant impact on the reconstruction. This will be discussed briefly at the end of this section.

7.5.1 Reconstructions of ellipses

We start with the simple example where Ω_I is an ellipse. Note that in real applications such as optical imaging of human brain, we may be allowed to approximate clear layers by such simple convex interfaces. The ellipse we want to reconstruct is given in polar coordinate by

$$\Sigma = \left\{ (r, \theta) : \left(\frac{r \cos \theta}{a} \right)^2 + \left(\frac{r \sin \theta}{b} \right)^2 = 1 \right\}, \quad (7.69)$$

with a and b the semi-major and semi-minor axis length, respectively. We test our algorithm with different values for (a, b) .

To characterize the error in the reconstruction, we introduce the following Fourier decomposition of $r(\theta)$:

$$r(\theta) = \sum_{k=-N}^N c_k e^{-ik\theta}, \quad (7.70)$$

where we have chosen $N = 20$ in the following calculations. The complexity of the curve will be measured by the magnitude of the Fourier coefficients c_k and their decay rate as k increases. Let \tilde{c}_k be the Fourier coefficients of a reconstructed interface. We then define the ϵ_0 and ϵ_{-1} errors between the original and the reconstructed interfaces as

$$\epsilon_0 = \left(\sum_{k=-N}^N |c_k - \tilde{c}_k|^2 \right)^{1/2}, \quad \text{and} \quad \epsilon_{-1} = \left(\sum_{k=-N}^N (1 + k^2)^{-1} |c_k - \tilde{c}_k|^2 \right)^{1/2}, \quad (7.71)$$

respectively.

The reconstruction results from different additive noise levels in the case $(a, b) = (0.8, 0.4)$ are given in Fig. 7-2. The left column of Fig. 7-2 shows the reconstructions

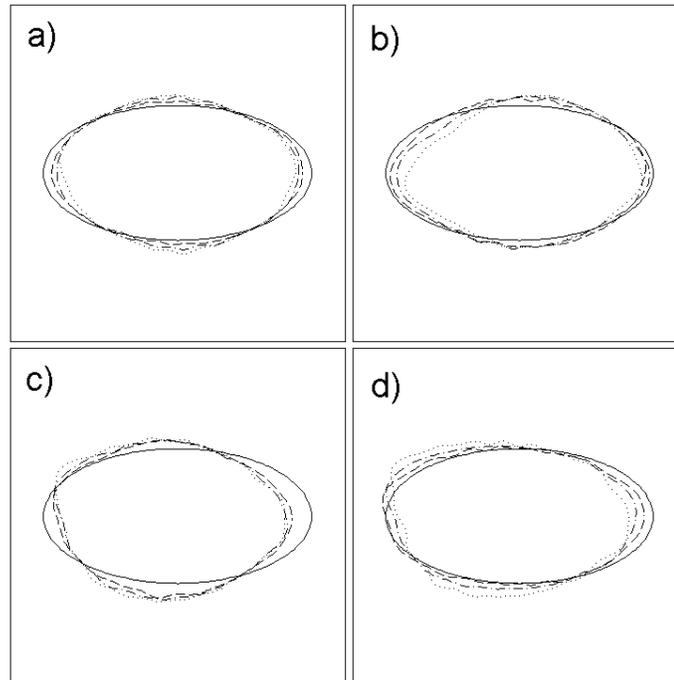


Figure 7-2: Reconstruction of the elliptic interface (7.69) with synthetic data at different noise levels for full (top row) and local (bottom row; see text for description) Neumann to Dirichlet measurements. We have $(a, b) = (0.8, 0.4)$. The reconstructions in (a) and (c) are done with the model in (7.1), while those in (b) and (d) are done with the model in (7.6). The lines in the pictures denote real interfaces (solid), reconstructions from data with 0.5% additive noise (dashed), reconstructions from data with 1% additive noise (dash-dotted) and reconstructions from data with 2% additive noise (dotted), respectively. The initial guess is given by the circle $\Sigma_0 = \{(r, \theta) : (r \cos \theta)^2 + (r \sin \theta)^2 = 0.8^2\}$ in all the simulations.

for the model (7.1) from full and local Neumann-to-Dirichlet measurements. In the latter case, measurements are only taken on the left side ($x = -1$) of the boundary. We have used the MATLAB contour function to plot the zero level set (characterizing the interface Σ) of the level set function. All the simulations have been implemented in Fortran 77. The same reconstructions have been performed for the model (7.6) and the results are shown in the right column of Fig. 7-2.

We list in Tab. 7.1 and Tab. 7.2 the errors in the reconstructions of ellipses of different aspect ratios using model (7.1) with full and partial Neumann to Dirichlet measurements, respectively. Note that the closer the aspect ratio $\frac{b}{a}$ is to 1.0, the less Fourier modes are needed to accurately represent $r(\theta)$. From these tables we see

Cases		0.5%	1.0%	2.0%
$(a, b) = (0.8, 0.4)$	(x_0, y_0)	(0.001, 0.000)	(-0.001, -0.002)	(-0.001, -0.001)
	ϵ_0	0.037	0.057	0.078
	ϵ_{-1}	0.008	0.011	0.015
$(a, b) = (0.8, 0.6)$	(x_0, y_0)	(0.002, 0.000)	(-0.002, 0.002)	(-0.003, 0.001)
	ϵ_0	0.011	0.020	0.031
	ϵ_{-1}	0.003	0.004	0.006
$(a, b) = (0.8, 0.8)$	(x_0, y_0)	(-0.000, -0.002)	(0.017, 0.006)	(0.000, 0.009)
	ϵ_0	0.005	0.015	0.017
	ϵ_{-1}	0.001	0.004	0.004

Table 7.1: Errors in the reconstructions of ellipses (7.69) with different values of (a, b) using model (7.1) with full measurements. The center of original interfaces $(x_0, y_0) = (0, 0)$.

Cases		0.5%	1.0%	2.0%
$(a, b) = (0.8, 0.4)$	(x_0, y_0)	(-0.039, -0.013)	(-0.047, -0.017)	(-0.057, -0.016)
	ϵ_0	0.076	0.098	0.104
	ϵ_{-1}	0.018	0.026	0.040
$(a, b) = (0.8, 0.6)$	(x_0, y_0)	(-0.015, 0.008)	(-0.013, 0.008)	(0.013, 0.009)
	ϵ_0	0.035	0.052	0.076
	ϵ_{-1}	0.014	0.018	0.020
$(a, b) = (0.8, 0.8)$	(x_0, y_0)	(-0.030, 0.010)	(-0.031, 0.006)	(-0.045, 0.004)
	ϵ_0	0.019	0.029	0.048
	ϵ_{-1}	0.007	0.011	0.019

Table 7.2: Same as Tab. 7.1 except that the reconstructions are obtained from partial measurements.

that as the aspect ratio increases, the reconstructions get more and more sensitive to the presence of noise in the data. In the reconstructions from full data, the center of the curves is relatively stably reconstructed even in the presence of significant noise. In the case of local measurements on part of the boundary, the reconstructed center of the ellipse is biased towards the part of the boundary where the boundary measurements are taken.

7.5.2 Reconstruction of more complicated surfaces

The reconstructions in the above section are all done with ellipses, which are convex interfaces. The Fourier coefficients of those interfaces decay relatively fast as k

increases. The reconstruction of such curves is thus not very difficult because the superposition of very few low-order Fourier modes can approximate the original interface quite accurately and those low-order Fourier modes can be stably reconstructed from data with even moderately high noise level. We now reconstruct more complicated interfaces the representation of which require higher-order Fourier modes. Since high order modes are more sensitive to the presence of noise in the data, we expect such interfaces to be harder to reconstruct. For simplicity, we reconstruct here star-shaped interfaces given by

$$\Sigma = \{(r, \theta) : r^2 + 0.3r \sin(N\theta) = 0.6^2\}. \quad (7.72)$$

Several choices for N are considered in the reconstructions below.

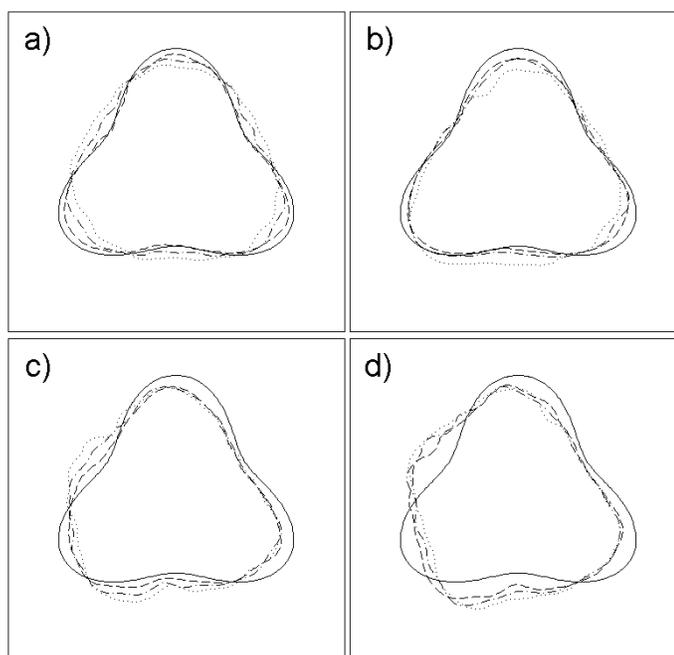


Figure 7-3: Reconstruction of the star-shaped interface (7.72) from synthetic data with different noise levels in the case of full (top row) and local (bottom row) Neumann to Dirichlet measurements. The interface parameter is $N = 3$. The reconstructions in (a) and (c) are for model (7.1), while those in (b) and (d) are for model (7.6). The lines in the pictures denote real interfaces (solid), reconstructions with 0.1% noise (dashed), reconstructions with 0.3% noise (dash-dotted) and reconstructions with 0.5% noise (dotted), respectively. The initial guess is the circle $\Sigma_0 = \{(r, \theta) : (r \cos \theta)^2 + (r \sin \theta)^2 = 0.8^2\}$.

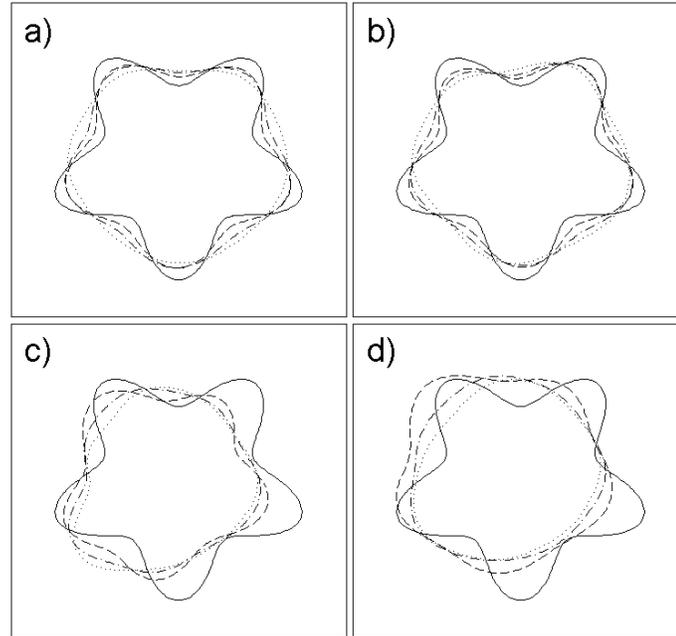


Figure 7-4: Same as in Fig. 7-3 except that $N = 5$.

We show in Figs. 7-3 and 7-4 reconstructions with $N = 3$ and $N = 5$, respectively, using synthetic data at different noise levels for full (top row) and local (bottom row) Neumann to Dirichlet measurements. Again, we use only the left side ($x = -1$) of the boundary for the local measurements.

The latter reconstructions are more sensitive to noise in the data than those in the preceding section although the centers of the interfaces are always relatively well reconstructed when full measurements are available.

Cases		0.5%	1.0%	2.0%
Model (7.1), $N = 3$	Full	(0.000, 0.0110)	(-0.005, 0.017)	(-0.003, 0.022)
	Local	(-0.029, -0.015)	(-0.046, -0.023)	(-0.053, -0.031)
Model (7.6), $N = 3$	Full	(-0.001, 0.003)	(-0.003, -0.017)	(-0.013, -0.046)
	Local	(-0.054, -0.028)	(-0.056, -0.028)	(-0.063, -0.031)
Model (7.1), $N = 5$	Full	(0.000, -0.007)	(0.001, -0.018)	(0.002, -0.018)
	Local	(-0.061, 0.013)	(-0.067, 0.019)	(-0.084, 0.018)
Model (7.6), $N = 5$	Full	(0.010, -0.006)	(0.012, -0.012)	(0.017, -0.015)
	Local	(-0.065, +0.022)	(-0.099, 0.022)	(-0.081, 0.021)

Table 7.3: Reconstructed centers for the cases presented in Figs. 7-3 ($N = 3$) and 7-4 ($N = 5$).

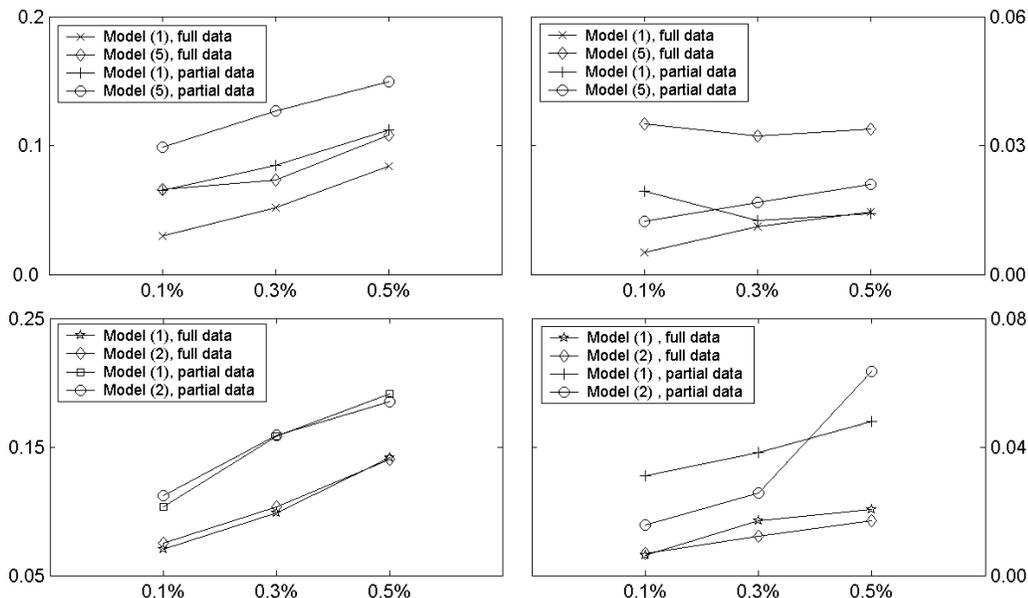


Figure 7-5: Errors in the reconstructions of (7.72) for different noise levels and different parameters N . Upper left: ϵ_0 and $N = 3$; Upper right: ϵ_{-1} and $N = 3$; Bottom left: ϵ_0 and $N = 5$; Bottom right: ϵ_{-1} and $N = 5$.

We observe in our numerical experiments that the ratios of the parameters, d/\mathcal{D} and $\mathcal{D}_0/\delta\mathcal{D}$, have important effects on the reconstruction results. The bigger the ratio, the more stable the reconstruction. This is simply because the effect of the interface on the boundary measurements increases. Note however that when the ratio $d/\delta\mathcal{D}$ is large, the conjugate gradient method used to calculate the solution of (7.1) converges very slowly. This is because the conditioning number of the finite element matrix in model (7.1) significantly increases when the ratio increases. For this reason, we have chosen the values $(\mathcal{D}, d)=(1.0, 0.3)$ to save computational time. Larger values of d would require to find an efficient preconditioner if solutions are to be obtained in a reasonable computational time. Indeed our simulations, based on the Morozov discrepancy principle [67] to find the optimal regularization parameter α , are very demanding computationally.

7.6 Conclusions and remarks

We have considered the reconstruction of singular surfaces in diffusion models arising in optical and electrical impedance tomography. We have performed a shape sensitivity analysis to describe the effects of variations in the surface on the boundary measurements. We have obtained that such effects primarily depended on the mean curvature of surface and the value of the tangential diffusion process supported on the surface. This is in contrast to the classical case of discontinuous diffusion coefficients across an interface.

We have introduced a level set method to evolve the surface so as to minimize an error functional. We have shown numerically that the reconstruction of the low-order Fourier modes of the interface can be achieved quite accurately from moderately noisy data. Higher frequency modes require less noisy data. The reconstructions can be done from either full or local Neumann to Dirichlet measurements although full measurements obviously provide more accurate reconstructions.

The major drawback of the current method is that it requires the diffusion coefficient $d(\mathbf{x})$ to be known. Generalizations, for instance along the lines of the works [47], to reconstructions of both the interface and the tangential diffusion coefficient need to be addressed. Note that in such a context, the coefficient $d(\mathbf{x})$ will depend on the geometric properties of the interface Σ (see Rem. 7.2.6 and [18, 23, 95]).

Chapter 8

Summary

We have studied in this thesis some direct and inverse problems related to the transport and diffusion equations. We focused on the applications of the inverse problems in optical tomography and atmospheric remote sensing.

In chapter 2 we formulated the inverse problem in optical tomography as a regularized least square problem based on the frequency-domain radiative transport equation. The transport equation is used as the forward model of light propagation in biological tissues. In the inversion procedure, the forward model is discretized by using a spatial finite volume method and an angular discrete ordinates method. We solve the regularized least square problem by using a limited-memory Quasi-Newton method with BFGS type updating rule for the Hessian matrix, and have incorporated positivity constraints and L^∞ bounds on the optical parameters. Numerical reconstructions based on synthetic data provide results that are in agreement with the expected reconstructions. Notably, the crosstalk between the two optical parameters is significantly reduced in frequency-domain reconstructions.

To speed up the reconstructions, we have implemented in chapter 3 an augmented Lagrangian method to solve inverse problems of optical tomography. The method formulates inverse transport problems as PDE-constrained optimization problems. By simultaneously updating both radiance and optical properties, the method solves the forward and inverse problems in optical tomography *all at once*. In this way, the computing time is greatly reduced as compared to traditional unconstrained optimization methods, during which one has to repeatedly solve the forward problem many times.

We tested and quantified the performance of the algorithm for various combinations of mesh sizes, noise, regularization parameters, initial guesses, optical properties and measurement geometries. In the cases that involve image reconstruction from synthetic measurement data we observe 10– to 30–fold decrease in computing time for the constrained optimization code compared to the unconstrained optimization code.

In chapter 4 we consider an inverse transport problem in atmospheric remote sensing where we try to reconstruct the concentration profiles of atmospheric gases from wavenumber-dependent boundary radiation measurement taken by space-borne infrared spectrometer. We showed in simplified situations that although the problem does admit a unique solution, it is severely ill-posed. Because of the severe ill-posedness, instead of attempting to reconstruct the whole concentration profile, one should really focus on feature reconstruction. We proposed an explicit procedure based on asymptotic analysis to reconstruct localized structures in the profile.

When scattering is extremely high and absorption is extremely low, transport-based reconstructions are computationally very expensive. Diffusion models are preferred in this case to serve as the forward model of light propagations. In many practical problems where diffusion equations are used, the conditions for diffusion approximation to be accurate are however not very well satisfied. In chapter 5 we have performed detail comparison between diffusion and transport-based reconstructions in those situations to characterize the errors we made in reconstructions when diffusion approximation are used. We found that factors such as size of the media to be reconstructed, void regions, modulation frequency and extrapolation length play huge roles in the quality of reconstructed image if classical diffusion equations are used as the forward model of light propagation. If the transport equations are believed to be the right models for light propagation, then the error one made by using the diffusion equation can be quite high. We thus propose to use transport equations in those situations.

In chapter 6, we propose a generalized diffusion model that accounts for the multiple scattering of photons in highly scattering media (classical diffusion regime) and well as for the near-collision-less propagation of the same photons in clear layers

(purely transport regime resulting in a guiding effect). This model can be mathematically derived from the phase space radiative transport equation as a small mean-free-path limit. It captures the guiding effect of photons in the clear layer quite well. Moreover it has almost the same cost as classical diffusion, which completely fails to model the clear layer effects, and a lower cost than previously derived generalized diffusion equations, which are already much less expensive than full transport solutions. The reason for this lower cost is that the nonlocal interface conditions of the latter diffusion models are replaced by their best local approximation. This best local approximation takes the form of a tangential diffusion process. The strength of this diffusion process can be calculated analytically or numerically provided that one has access to the geometry of the clear layer. When this geometry is unknown or only partially known, we have shown numerically that the diffusion process that best fits the impact of the clear layer gives boundary measurements that are visually indistinguishable from the measurements obtained by solving the full transport equations. We believe that the generalized diffusion model can thus safely be used in optical tomography as an accurate approximation of the forward model.

In chapter 7, we considered the reconstruction of singular surfaces in the generalized diffusion equations. We have performed a shape sensitivity analysis to describe the effects of variations in the surface on the boundary measurements. We have obtained that such effects primarily depended on the mean curvature of surface and the value of the tangential diffusion process supported on the surface. This is in contrast to the classical case of discontinuous diffusion coefficients across an interface.

We have introduced a level set method to evolve the surface so as to minimize an error functional. We have shown numerically that the reconstruction of the low-order Fourier modes of the interface can be achieved quite accurately from moderately noisy data. Higher frequency modes require less noisy data. The reconstructions can be done from either full or local Neumann to Dirichlet measurements although full measurements obviously provide more accurate reconstructions.

Bibliography

- [1] G. S. ABDOULAEV AND A. H. HIELSCHER, *Three-dimensional optical tomography with the equation of radiative transfer*, J. Electron. Imaging, 12 (2003), pp. 594–601.
- [2] G. S. ABDOULAEV, K. REN, AND A. H. HIELSCHER, *Optical tomography as a PDE-constrained optimization problem*, Inverse Problems, 21 (2005), pp. 1507–1530.
- [3] P. D. ACTON AND H. KUNG, *Small animal imaging with high resolution single photon emission tomography*, Nuclear Medicine and Biology, 30 (2003), pp. 889–895.
- [4] M. L. ADAMS AND E. W. LARSEN, *Fast iterative methods for discrete-ordinates particle transport calculations*, Prog. Nucl. Energy, 40 (2002), pp. 3–150.
- [5] R. A. ADAMS, *Sobolev Spaces*, Academic Press, New York, 1975.
- [6] V. AGOSHKOV, *Boundary Value Problems for the Transport Equations*, Birkhauser, Boston, 1998.
- [7] G. ALLAIRE, F. JOUVE, AND A. M. TOADER, *Structural optimization using sensitivity analysis and a level set method*, J. Comput. Phys., 194 (2004), pp. 363–393.
- [8] L. AMBROSIO AND G. BUTTAZZO, *An optimal design problem with perimeter penalization*, Calc. Var., 1 (1993), pp. 55–69.
- [9] H. AMMARI AND H. KANG, *Reconstruction of Small Inhomogeneities from Boundary Measurements*, Lecture Notes in Mathematics, Springer, Berlin, 2004.
- [10] R. ARONSON, R. L. BARBOUR, J. LUBOWSKY, AND H. GRABER, *Application of transport theory to infra-red medical imaging*, in Modern Mathematical methods in Transport Theory, W. Greenberg and J. Polewczak, eds., Birkhäuser, 1991.
- [11] S. R. ARRIDGE, *Optical tomography in medical imaging*, Inverse Problems, 15 (1999), pp. R41–R93.

- [12] S. R. ARRIDGE, H. DEGHANI, M. SCHWEIGER, AND E. OKADA, *The finite element model for the propagation of light in scattering media: A direct method for domains with non-scattering regions*, *Med. Phys.*, 27 (2000), pp. 252–264.
- [13] S. R. ARRIDGE AND J. C. HEBDEN, *Optical imaging in medicine: II. modelling and reconstruction*, *Phys. Med. Biol.*, 42 (1997), pp. 841–853.
- [14] S. R. ARRIDGE AND W. R. B. LIONHEART, *Nonuniqueness in diffusion-based optical tomography*, *Opt. Lett.*, 23 (1998), pp. 882–884.
- [15] G. BAL, *Inverse problems for homogeneous transport equations: I. the one-dimensional case*, *Inverse Problems*, 16 (2000), pp. 997–1011.
- [16] —, *Inverse problems for homogeneous transport equations: II. the multidimensional case*, *Inverse Problems*, 16 (2000), pp. 1013–1028.
- [17] —, *Particle transport through scattering regions with clear layers and inclusions*, *J. Comp. Phys.*, 180 (2002), pp. 659–685.
- [18] —, *Transport through diffusive and non-diffusive regions, embedded objects, and clear layers*, *SIAM J. Appl. Math.*, 62 (2002), pp. 1677–1697.
- [19] —, *Optical tomography for small volume absorbing inclusions*, *Inverse Problems*, 19 (2003), pp. 371–386.
- [20] —, *Reconstructions in impedance and optical tomography with singular interfaces*, *Inverse Problems*, 21 (2005), pp. 113–131.
- [21] G. BAL, V. FREILIKHER, G. PAPANICOLAOU, AND L. RYZHIK, *Wave transport along surfaces with random impedance*, *Phys. Rev. B*, 62 (2000), pp. 6228–6240.
- [22] G. BAL AND Y. MADAY, *Coupling of transport and diffusion models in linear transport theory*, *Math. Model. Numer. Anal.*, 36 (2002), pp. 69–86.
- [23] G. BAL AND K. REN, *Generalized diffusion model in optical tomography with clear layers*, *J. Opt. Soc. Am. A*, 20 (2003), pp. 2355–2364.
- [24] —, *Atmospheric concentration profile reconstructions from radiation measurements*, *Inverse Problems*, 21 (2005), pp. 153–168.
- [25] —, *Reconstruction of singular surfaces by shape sensitivity analysis and level set method*, *Math. Models Methods Appl. Sci.*, 8 (2006), pp. 1347–1373.
- [26] G. BAL AND X. WARIN, *Discrete ordinate methods in xy -geometry with spatially varying angular discretization*, *Nuclear Science and Engineering*, 127 (1997), pp. 169–181.

- [27] R. BARRETT, M. BERRY, T. F. CHAN, J. DEMMEL, J. DONATO, J. DONGARRA, V. EIJKHOUT, R. POZO, C. ROMINE, AND H. VAN DER VORST, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, Philadelphia, 2nd ed., 1994.
- [28] R. BEER, T. GLAVICH, AND D. RIDER, *Tropospheric emission spectrometer for the earth observing system's Aura satellite*, *Applied Optics*, 40 (2001), pp. 2356–2367.
- [29] H. BEN AMEUR, M. BURGER, AND B. HACKL, *Level set methods for geometric inverse problem in linear elasticity*, *Inverse Problems*, 20 (2004), pp. 673–696.
- [30] D. A. BENARON, S. R. HINTZ, A. VILLRINGER, D. BOAS, A. KLEINSCHMIDT, J. FRAHM, C. HIRTH, H. OBRIG, J. C. VAN HOUTEN, E. L. KERMIT, W. F. CHEONG, AND D. K. STEVENSON, *Noninvasive functional imaging of human brain using light*, *J. of Cerebral Blood Flow and Metabolism*, 20 (2000), pp. 469–477.
- [31] M. BERTERO AND E. R. PIKE, *Signal processing for linear instrumental systems with noise: A general theory with illustrations from optical imaging and light scattering problems*, in *Handbook of Statistics V*, N. K. Bose and C. R. Rao, eds., Elsevier, Amsterdam, 1993, pp. 1–46.
- [32] D. P. BERTSEKAS, *Constrained Optimization and Lagrange Multiplier Methods*, Academic Press, New York, 1982.
- [33] L. BIEGLER, O. GHATTAS, M. HEINKENSCHLOSS, AND B. V. BLOEMENWAANDERS, eds., *Large-Scale PDE-Constrained Optimization*, *Lecture Notes in Computational Science and Engineering*, Springer-Verlag, Berlin, 2003.
- [34] A. BLUESTONE, G. ABDOULAEV, C. SCHMITZ, R. BARBOUR, AND A. H. HIELSCHER, *Three-dimensional optical-tomography of hemodynamics in the human head*, *Optics Express*, 9 (2001), pp. 272–286.
- [35] A. Y. BLUESTONE, M. STEWART, B. LEI, I. S. KASS, J. LASKER, G. ABDOULAEV, AND A. H. HIELSCHER, *Three-dimensional optical tomographic brain imaging in small animals, part I: Hypercapnia*, *Journal of Biomedical Optics*, 9 (2004), pp. 1046–1062.
- [36] D. A. BOAS, D. H. BROOKS, E. L. MILLER, C. A. DIMARZIO, M. KILMER, R. J. GAUDETTE, AND Q. ZHANG, *Imaging the body with diffuse optical tomography*, *IEEE Signal Processing Magazine*, 18 (2001), pp. 57–75.
- [37] D. A. BOAS, K. CHEN, D. GREBERT, AND M. A. FRANCESCHINI, *Improving the diffuse optical imaging spatial resolution of the cerebral hemodynamic response to brain activation in humans*, *Optics Lett.*, 29 (2004), pp. 1506–1508.

- [38] K. BOWMAN, J. WORDEN, T. STECK, AND H. M. WORDEN, *Capturing time and vertical variability of tropospheric ozone: A study using TES nadir retrievals*, J. Geophys. Res., 107 (2002), pp. 4723–4734.
- [39] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, New York, 2002.
- [40] M. BURGER AND S. J. OSHER, *A survey on level set methods for inverse problems and optimal design*, European J. Appl. Math., 16 (2005), pp. 263–301.
- [41] R. H. BYRD, P. LU, J. NOCEDAL, AND C. ZHU, *A limited memory algorithm for bound constrained optimization*, SIAM J. Sci. Comput., 16 (1995), pp. 1190–1208.
- [42] W. CAI, M. XU, AND R. R. ALFANO, *Three-dimensional radiative transfer tomography for turbid media*, IEEE Journal of Selected Topics in Quantum Electronics, 9 (2003), pp. 189–198.
- [43] K. M. CASE AND P. F. ZWEIFEL, *Linear Transport Theory*, Addison-Wesley Publishing Company, Reading, 1967.
- [44] D. J. CEDIO-FENGYA, S. MOSKOW, AND M. S. VOGELIOUS, *Identification of conductivity imperfections of small diameter by boundary measurements. continuous dependence and computational reconstruction*, Inverse Probl., 14 (1998), pp. 553–595.
- [45] M. T. CHAHINE, *Inverse problems in radiative transfer: Determination of atmospheric parameters*, J. Atmos. Sci., 27 (1970), pp. 960–967.
- [46] ———, *Remote sensing of cloud parameters*, J. Atmos. Sci., 39 (1982), pp. 159–170.
- [47] T. F. CHAN AND X.-C. TAI, *Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients*, J. Comput. Phys., 193 (2003), pp. 40–66.
- [48] B. CHANCE, R. R. ALFANO, B. J. TROMBERG, AND A. KATZIR, eds., *Optical Tomography and Spectroscopy of Tissue*, vol. V, Bellingham, WA, 2003, SPIE.
- [49] S. CHANDRASEKHAR, *Radiative Transfer*, Dover, New York, 1960.
- [50] S. CHEN, B. MERRIMAN, S. OSHER, AND P. SMEREKA, *A simple level set method for solving Stefan problems*, J. Comput. Phys., 135 (1997), pp. 8–29.
- [51] M. CHENEY, D. ISAACSON, AND J. NEWELL, *Electrical impedance tomography*, SIAM Rev., 41 (1999), pp. 85–101.
- [52] M. CHOULLI AND P. STEFANOV, *Reconstruction of the coefficients of the stationary transport equation from boundary measurements*, Inverse Probl., 12 (1996), pp. L19–L23.

- [53] C. CLERBAUX, J. HADJI-LAZARO, S. PAYAN, C. CAMY-PEYRET, J. WANG, D. P. EDWARDS, AND M. LUO, *Retrieval of CO from nadir remote-sensing measurements in the infrared by use of four different inversion algorithms*, Appl. Opt., 41 (2002), pp. 7068–7078.
- [54] D. COLTON AND R. KRESS, *Inverse Acoustic and Electromagnetic Scattering Theory*, Springer-Verlag, New York, 1998.
- [55] J. B. CONWAY, *Functions in One Complex Variable I*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1978.
- [56] G. DAL MASO AND R. TOADER, *A model for the quasi-static growth of brittle fractures: existence and approximation results*, Arch. Ration. Mech. Anal., 162 (2002), pp. 101–135.
- [57] R. DAUTRAY AND J.-L. LIONS, *Mathematical Analysis and Numerical Methods for Science and Technology, Vol II*, Springer-Verlag, Berlin, 1993.
- [58] —, *Mathematical Analysis and Numerical Methods for Science and Technology, Vol VI*, Springer-Verlag, Berlin, 1993.
- [59] H. DEGHANI, S. R. ARRIDGE, M. SCHWEIGER, AND D. T. DELPY, *Optical tomography in the presence of void regions*, J. Opt. Soc. Am. A, 17 (2000), pp. 1659–1670.
- [60] F. DESAINT AND J.-P. ZOLÉSIO, *Shape derivatives for the Laplace-Beltrami equation*, in Partial Differential Equation Methods in Control and Shape Analysis, G. D. Prato and J.-P. Zolésio, eds., Marcel Dekker, Inc., New York, 1997.
- [61] A. DOICU, F. SCHREIER, AND M. HESS, *Iterative regularization methods for atmospheric remote sensing*, J. Quant. Spectrosc. Radiat. Transfer, 83 (2004), pp. 47–61.
- [62] O. DORN, *A transport-backtransport method for optical tomography*, Inverse Problems, 14 (1998), pp. 1107–1130.
- [63] —, *Scattering and absorption transport sensitivity functions for optical tomography*, Optics Express, 7 (2000), pp. 492–506.
- [64] —, *Shape reconstruction in scattering media with void using a transport model and level sets*, Canad. Appl. Math. Quart., 10 (2001), pp. 239–275.
- [65] R. ELALOUI, R. CARMINATI, AND J. GREFFET, *Time-dependent transport through scattering media: from radiative transfer to diffusion*, J. Opt. A: Pure Appl. Opt., 4 (2002), pp. S103–S108.
- [66] R. A. ELLIOTT, T. DURACZ, N. J. MCCORMICK, AND D. R. EMMONS, *Experimental test of a time-dependent inverse radiative transfer algorithm for estimating scattering parameters*, J. Opt. Soc. Am. A, 5 (1988), pp. 366–373.

- [67] H. W. ENGL, M. HANKE, AND A. NEUBAUER, *Regularization of Inverse Problems*, Kluwer Academic Publishers, Dordrecht, 1996.
- [68] R. EYMARD, T. GALLOUET, AND R. HERBIN, *Finite volume methods*, in Handbook of Numerical Analysis VII, P. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 2000.
- [69] W. FANG AND K. ITO, *Identification of contact regions in semiconductor transistors by level-set methods*, J. Comput. Appl. Math., 159 (2003), pp. 399–410.
- [70] J. FERCHICHI AND J.-P. ZOLÉSIO, *Shape sensitivity for the Laplace-Beltrami operator with singularities*, J. Differ. Equations, 196 (2004), pp. 340–384.
- [71] M. FIRBANK, S. R. ARRIDGE, M. SCHWEIGER, AND D. T. DELPY, *An investigation of light transport through scattering bodies with non-scattering regions*, Phys. Med. Biol., 41 (1996), pp. 767–783.
- [72] M. FRANCOEUR, R. VAILLON, AND D. R. ROUSSE, *Theoretical analysis of frequency and time-domain methods for optical characterization of absorbing and scattering media*, J. Quant. Spectrosc. Radiat. Transfer, 93 (2005), pp. 139–150.
- [73] K. O. FRIEDRICHS, *Advanced Ordinary Differential Equations*, Taylor & Francis, New York, 1985.
- [74] L. FUKSHANSKY, N. FUKSHANSKY-KAZARINOVA, AND A. M. V. REMISOWSKY, *Estimation of optical properties in a living tissue by solving the inverse problem of the multiflux radiative transfer*, Appl. Opt., 30 (1991), pp. 3145–3153.
- [75] O. GHATTAS AND J. H. BARK, *Optimal control of two- and three-dimensional incompressible Navier-Stokes flows*, J. Comput. Phys., 136 (1997), pp. 231–244.
- [76] D. GILBARG AND N. S. TRUDINGER, *Elliptic Partial Differential Equations of Second Order*, Springer-Verlag, Berlin, 2000.
- [77] F. GOLSE, S. JIN, AND C. D. LEVERMORE, *The convergence of numerical transfer schemes in diffusive regimes I: The discrete-ordinate method*, SIAM J. Numerical Analysis, 36 (1999), pp. 1333–1369.
- [78] R. M. GOODY AND Y. L. YUNG, *Atmospheric Radiation: Theoretical Basis*, Oxford University Press, New York, second ed., 1989.
- [79] H. L. GRABER, J. CHANG, R. ARONSON, AND R. L. BARBOUR, *A perturbation model for imaging in dense scattering media: Derivation and evaluation of imaging operators*, in Proc. SPIE, vol. IS11, Int. Soc. Opt. Eng., 1993, pp. 121–143.

- [80] E. E. GRAVES, R. WEISSLEDER, AND V. NTZIACHRISTOS, *Fluorescence molecular imaging of small animal tumor models*, Current Molecular Medicine, 4 (2004), pp. 419–430.
- [81] E. HABER, U. ASCHER, AND D. OLDENBURG, *On optimization techniques for solving nonlinear inverse problems*, Inverse Problems, 16 (2000), pp. 1263–1280.
- [82] P. C. HANSEN AND D. P. O’LEARY, *The use of l-curve in the regularization of discrete ill-posed problems*, SIAM Journal on Scientific Computing, 14 (1993), pp. 1487–1503.
- [83] R. C. HASKELL, L. O. SVAASAND, T.-T. TSAY, T.-C. FENG, M. S. MCADAMS, AND B. J. TROMBERG, *Boundary conditions for the diffusion equation in radiative transfer*, J. Opt. Soc. Am. A, 11 (1994), pp. 2727–2741.
- [84] J. HASLINGER AND P. NEITTANMAAKI, *Finite Element Approximation for Optimal Shape, Material and Topology Design*, John Wiley and Sons, Chichester, New York, second ed., 1996.
- [85] C. K. HAYAKAWA, J. SPANIER, F. BEVILACQUA, A. K. DUNN, J. S. YOU, B. J. TROMBERG, AND V. VENUGOPALAN, *Perturbation monte carlo methods to solve inverse photon migration problems in heterogeneous tissues*, Opt. Lett., 26 (2001), pp. 1335–1337.
- [86] J. C. HEBDEN, S. R. ARRIDGE, AND D. T. DELPY, *Optical imaging in medicine: I. experimental techniques*, Phys. Med. Biol., 42 (1997), pp. 825–840.
- [87] L. G. HENYEVY AND J. L. GREENSTEIN, *Diffuse radiation in the galaxy*, Astrophys. J., 90 (1941), pp. 70–83.
- [88] A. HIELSCHER, A. BLUESTONE, G. ABDOULAEV, A. KLOSE, J. LASKER, M. STEWART, U. NETZ, AND J. BEUTHAN, *Near-infrared diffuse optical tomography*, Disease Markers, 18 (2002), pp. 313–337.
- [89] A. H. HIELSCHER, *Optical tomographic imaging of small animals*, Current Opinion in Biotechnology, 16 (2005), pp. 79–88.
- [90] A. H. HIELSCHER, R. E. ALCOUFFE, AND R. L. BARBOUR, *Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissue*, Phys. Med. Biol., 43 (1998), pp. 1285–1302.
- [91] A. H. HIELSCHER AND S. BARTEL, *Use of penalty terms in gradient-based iterative reconstruction schemes for optical tomography*, J. Biomed. Optics, 183 (2001), pp. 183–192.
- [92] A. H. HIELSCHER, A. D. KLOSE, A. SCHEEL, B. MOA-ANDERSON, M. BACKHAUS, U. NETZ, AND J. BEUTHAN, *Sagittal laser optical tomography for imaging of rheumatoid finger joints*, Physics in Medicine and Biology, 49 (2004), pp. 1147–1163.

- [93] E. M. C. HILLMAN, J. C. HEBDEN, M. SCHWEIGER, H. DEGHANI, F. E. W. SCHMIDT, D. T. DELPY, AND S. R. ARRIDGE, *Time resolved optical tomography of the human forearm*, Phys. Med. Biol., 46 (2001), pp. 1117–1130.
- [94] X. HUANG, Y. L. YUNG, AND J. S. MARGOLIS, *Use of high-resolution measurements for the retrieval of temperature and gas-concentration profiles from outgoing infrared spectra in the presence of cirrus clouds*, Appl. Opt., 42 (2003), pp. 2155–2165.
- [95] P. H. HUNG AND E. SÁNCHEZ-PALENCIA, *Phénomènes de transmission à travers des couches minces de conductivité élevée*, J. Math. Anal. Appl., 47 (1974), pp. 284–309.
- [96] V. ISAKOV, *Inverse Problems for Partial Differential Equations*, Springer-Verlag, New York, 1998.
- [97] K. ITO AND K. KUNISCH, *The augmented lagrangian method for parameter estimation in elliptic systems*, SIAM J. Control Optim., 28 (1990), pp. 137–157.
- [98] K. ITO, K. KUNISCH, AND Z. LI, *Level set function approach to an inverse interface problem*, Inverse Problems, 17 (2001), pp. 1225–1242.
- [99] H. JIANG, K. D. PAULSEN, U. L. ÖSTERBERG, B. W. POGUE, AND M. S. PATTERSON, *Simultaneous reconstruction of optical absorption and scattering maps turbid media from near-infrared frequency-domain data*, Opt. Lett., 20 (1995), pp. 2128–2130.
- [100] C. T. KELLEY, *Iterative Methods for Optimization*, Frontiers in Applied Mathematics, Society of Industrial and Applied Mathematics, Philadelphia, 1999.
- [101] A. KIENLE, F. K. FORSTER, AND R. HIBST, *Influence of the phase function on determination of the optical properties of biological tissue by spatially resolved reflectance*, Optics Lett., 26 (2001), pp. 1571–1573.
- [102] A. D. KIM AND J. B. KELLER, *Light propagation in biological tissue*, J. Opt. Soc. Am. A, 20 (2003), pp. 92–98.
- [103] A. KIRSCH, *An Introduction to the Mathematical Theory of Inverse Problems*, Springer-Verlag, New York, 1996.
- [104] A. D. KLOSE, , U. NETZ, J. BEUTHAN, AND A. H. HIELSCHER, *Optical tomography using the time-independent equation of radiative transfer. part 1: Forward model*, J. Quant. Spectrosc. Radiat. Transfer, 72 (2002), pp. 691–713.
- [105] A. D. KLOSE AND A. H. HIELSCHER, *Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer*, Med. Phys., 26 (1999), pp. 1698–1707.

- [106] —, *Optical tomography using the time-independent equation of radiative transfer. part 2: Inverse model*, J. Quant. Spectrosc. Radiat. Transfer, 72 (2002), pp. 715–202.
- [107] —, *Quasi-Newton methods in optical tomographic image reconstruction*, Inverse Problems, 19 (2003), pp. 387–409.
- [108] A. D. KLOSE, A. H. HIELSCHER, K. M. HANSON, AND J. BEUTHAN, *Three-dimensional optical tomography of a finger joint model for diagnostic of rheumatoid arthritis*, in Proc. SPIE, vol. 3566, Int. Soc. Opt. Eng., 1998, pp. 151–160.
- [109] E. W. LARSEN, *Solution of the inverse problem in multigroup transport theory*, J. Math. Phys., 22 (1981), pp. 158–160.
- [110] —, *Solution of three dimensional inverse transport problems*, Transport Theory and Statistical Physics, 17 (1988), pp. 147–167.
- [111] E. W. LARSEN AND J. B. KELLER, *Asymptotic solution of neutron transport problems for small mean free paths*, J. Math. Phys., 15 (1974), pp. 75–81.
- [112] E. W. LARSEN, J. E. MOREL, AND W. F. M. JR., *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes*, J. Comp. Phys., 69 (1987), pp. 283–324.
- [113] E. E. LEWIS AND W. F. MILLER, *Computational Methods of Neutron Transport*, American Nuclear Society, La Grange Park, IL, 1993.
- [114] A. LI, E. MILLER, M. KILMER, T. BRUKILACCHIO, T. CHAVES, J. STOTT, Q. ZHANG, T. WU, M. CHORLTON, R. MOORE, D. KOPANS, AND D. BOAS, *Tomographic optical breast imaging guided by 3-D mammography*, Applied Optics, 42 (2003), pp. 5181–5190.
- [115] E. H. LIEB AND M. LOSS, *Analysis*, American Mathematical Society, Providence, RI, Second ed., 2001.
- [116] K. LIOU, *An Introduction to Atmospheric Radiation*, Academic Press, Amsterdam, second ed., 2002.
- [117] A. LITMAN, D. LESSELIER, AND F. SANTOSA, *Reconstruction of a two dimensional binary obstacle by controlled evolution of a level-set*, Inverse Problems, 14 (1998), pp. 685–704.
- [118] J. M. LUCK AND T. M. NIEUWENHUIZEN, *Light scattering from mesoscopic objects in diffusive media*, Eur. Phys. J. B, 7 (1999), pp. 483–500.
- [119] V. MARKEL AND J. SCHOTLAND, *Effects of sampling and limited data in optical tomography*, App. Phys. Lett., 81 (2002), pp. 1180–1182.

- [120] V. A. MARKEL, V. MITAL, AND J. C. SCHOTLAND, *Inverse problem in optical diffusion tomography. III. Inverse formulas and singular-value decomposition*, J. Opt. Soc. Am A, 20 (2003), pp. 890–902.
- [121] V. A. MARKEL, J. A. O’SULLIVAN, AND J. C. SCHOTLAND, *Inverse problem in optical diffusion tomography. IV. Nonlinear inversion formulas*, J. Opt. Soc. Am A, 20 (2003), pp. 903–912.
- [122] V. A. MARKEL AND J. C. SCHOTLAND, *Inverse problem in optical diffusion tomography. I. Fourier-Laplace inverse formula*, J. Opt. Soc. Am. A, 18 (2001), pp. 1336–1347.
- [123] —, *Inverse problem in optical diffusion tomography. II. Role of boundary conditions*, J. Opt. Soc. Am A, 19 (2002), pp. 558–566.
- [124] T. O. MCBRIDE, B. W. POGUE, U. L. ÖSTERBERG, AND K. D. PAULSEN, *Separation of absorption and scattering heterogeneities in NIR tomographic imaging of tissue*, in OSA Technical Digest, Washington DC, 2000, Optical Society of America. Biomedical Topical Meetings.
- [125] E. J. MCCARTNEY, *Absorption and Emission by Atmospheric Gases: The Physical Processes*, John Wiley & Sons, New York, 1983.
- [126] N. J. MCCORMICK, *Recent developments in inverse scattering transport methods*, Transport Theory and Statistical Physics, 13 (1984), pp. 15–28.
- [127] —, *Inverse radiative transfer problems: A review*, Nuclear Science and Engineering, 112 (1992), pp. 185–198.
- [128] K. MILLER, *Least-squares method for ill-posed problems with a prescribed bound*, SIAM J. Math. Anal., 1 (1970), pp. 52–74.
- [129] M. MOKHTAR-KHARROUBI, ed., *Mathematical Topics in Neutron Transport Theory: New Aspects*, Advances in Mathematics for Applied Sciences, World Scientific, Singapore, 1997.
- [130] G. J. MÜLLER, ed., *Medical Optical Tomography: functional imaging and optical technologies*, SPIE Series Vol. IS11, Optical Society of American, 1993.
- [131] U. NETZ, J. BEUTHAN, H. CAPIUS, H. KOCH, A. KLOSE, AND A. HIELSCHER, *Imaging of rheumatoid arthritis in finger joints by sagittal optical tomography*, Medical Laser Application, 16 (2001), pp. 306–310.
- [132] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, Springer-Verlag, New York, 1999.
- [133] M. A. O’LEARY, D. A. BOAS, B. CHANCE, AND A. G. YODH, *Experimental images of heterogeneous turbid media by frequency domain diffusion photon tomography*, Opt. Lett., 20 (1995), pp. 426–428.

- [134] C. OROZCO AND O. GHATTAS, *Massively parallel aerodynamic shape optimization*, Computing Systems in Engineering, 1-4 (1992), pp. 311–320.
- [135] S. OSHER AND R. FEDKIW, *Level Set Methods and Dynamic Implicit Surfaces*, Springer-Verlag, New York, 2002.
- [136] S. OSHER AND N. PARAGIOS, eds., *Geometric Level Set Methods in Imaging, Vision, and Graphics*, Springer-Verlag, New York, 2003.
- [137] S. J. OSHER AND F. SANTOSA, *Level set methods for optimization problems involving geometry and constraints I. frequencies of a two-density inhomogeneous drum*, J. Comput. Phys., 171 (2001), pp. 272–288.
- [138] S. J. OSHER AND J. A. SETHIAN, *Front propagation with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulations*, J. Comput. Phys., 79 (1988), pp. 12–49.
- [139] B. W. PATTON AND J. P. HOLLOWAY, *Application of preconditioned GMRES to the numerical solution of the neutron transport equation*, Annals of Nuclear Energy, 29 (2002), pp. 109–136.
- [140] D. PENG, B. MERRIMAN, S. OSHER, H. ZHAO, AND M. KANG, *A PDE-based fast local level set method*, J. Comput. Phys., 155 (1999), pp. 410–438.
- [141] B. POGUE, T. MCBRIDE, J. PREWITT, U.L.ÖSTEBERG, AND K. PAULSEN, *Spatially variant regularization improves diffuse optical tomography*, Applied Optics, 38 (1999), pp. 2950–2960.
- [142] B. W. POGUE, M. S. PATTERSON, H. JIANG, AND K. D. PAULSEN, *Initial assessment of a simple system for frequency domain diffuse optical tomography*, Phys. Med. Biol., 40 (1995), pp. 1709–1729.
- [143] B. W. POGUE, S. P. POPLACK, T. O. MCBRIDE, W. A. WELLS, K. S. OSTERMAN, U. L. OSTERBERG, AND K. D. PAULSEN, *Quantitative hemoglobin tomography with diffuse near-infrared spectroscopy: pilot results in the breast*, Radiology, 218 (2001), pp. 261–266.
- [144] B. W. POGUE, M. TESTORF, T. O. MCBRIDE, U. L. ÖSTERBERG, AND K. D. PAULSEN, *Instrumentation and design of frequency-domain diffuse optical imager for breast cancer detection*, Opt. Express, 1 (1997), pp. 391–403.
- [145] V. PRAPAVAT, W. RUNGE, J. MANS, A. KRAUSE, J. BEUTHAN, AND G. MÜLLER, *Development of a finger joint phantom for the optical simulation of early stages of rheumatoid arthritis*, Biomedizinische Technik, 42 (1997), pp. 319–326.
- [146] K. REN, G. S. ABDOULAEV, G. BAL, AND A. H. HIELSCHER, *Algorithm for solving the equation of radiative transfer in the frequency domain*, Optics Lett., 29 (2004), pp. 578–580.

- [147] K. REN, G. BAL, AND A. H. HIELSCHER, *Frequency domain optical tomography based on the equation of radiative transfer*, SIAM J. Sci. Comput., 28 (2006). In Press.
- [148] —, *Transport- and diffusion-based optical tomography in small domains: A comparative study*, Applied Optics, (2006). Submitted.
- [149] J. RIPOLL, M. NIETO-VESPERINAS, S. R. ARRIDGE, AND H. DEGHANI, *Boundary conditions for light propagation in diffusive media with nonscattering regions*, J. Opt. Soc. Am. A, 17 (2000), pp. 1671–1682.
- [150] C. RODGERS, *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific, Singapore, 2000.
- [151] R. ROY AND E. M. SEVICK-MURACA, *Truncated Newton's optimization scheme for absorption and fluorescence optical tomography (part I and part II)*, Opt. Express, 4 (1999), pp. 353–382.
- [152] —, *Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I theory and formulation*, Opt. Express, 4 (1999), pp. 353–371.
- [153] —, *Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part II reconstruction from synthetic measurements*, Opt. Express, 4 (1999), pp. 372–382.
- [154] —, *Active constrained truncated newton method for simple-bound optical tomography*, J. Opt. Soc. Am. A, 17 (2000), pp. 1627–1641.
- [155] —, *Three-dimensional unconstrained and constrained image-reconstruction techniques applied to fluorescence, frequency-domain photon migration*, Applied Optics, 40 (2001), pp. 2206–2215.
- [156] W. RUDIN, *Real and Complex Analysis*, McGraw-Hill, New York, Second ed., 1974.
- [157] Y. SAAD, *Iterative Methods for Sparse Linear Systems*, SIAM, Philadelphia, 2nd ed., 2003.
- [158] Y. SAAD AND M. H. SCHULTZ, *GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems*, SIAM J. Sci. Stat. Comput., 7 (1986), pp. 856–869.
- [159] F. SANTOSA, *A level-set approach for inverse problems involving obstacles*, ESAIM: Control, Optimization and Calculus of Variations, 1 (1996), pp. 17–33.
- [160] G. SAPIRO, *Geometric Partial Differential Equations and Image Analysis*, Cambridge University Press, New York, NY, 2001.

- [161] C. H. SCHMITZ, M. LÖCKER, J. M. LASKER, A. H. HIELSCHER, AND R. L. BARBOUR, *Instrumentation for fast functional optical tomography*, Rev. of Scientific Instrumentation, 73 (2002), pp. 429–439.
- [162] M. SCHWEIGER, S. ARRIDGE, , AND D. DELPY, *Application of the finite-element method for the forward and inverse models in optical tomography*, J. Math. Imaging Vision, 3 (1993), pp. 263–283.
- [163] J. A. SETHIAN, *Level Set Methods and Fast Marching Methods*, Cambridge University Press, Cambridge, MA, 1999.
- [164] A. SHENOY, M. HEINKENSCHLOSS, AND E. M. CLIFF, *Airfoil design by an all-at-once method*, Int. J. Comput. Fluid Dynamics, 11 (1998), pp. 3–25.
- [165] J. SIMON, *Differentiation with respect to the domain in boundary value problems*, Numer. Funct. Anal. Optim., 2 (1980), pp. 649–687.
- [166] J. SOKOLOVSKI AND J. P. ZOLÉSIO, *Introduction to Shape Optimization*, Springer, Berlin, Heidelberg, 1992.
- [167] J. SPANIER AND E. M. GELBARD, *Monte Carlo Principles and Neutron Transport Problems*, Addison-Wesley, Reading, Mass., 1969.
- [168] T. STECK, *Methods for determining regularization for atmospheric retrieval problems*, Appl. Opt., 41 (2002), pp. 1788–1797.
- [169] P. STEFANOV, *Inverse problems in transport theory*. Inside Out: Inverse Problems and Applications, Vol (47), MSRI publications, 2003.
- [170] P. STEFANOV AND G. UHLMANN, *Optical tomography in two dimensions*, Methods Appl. Anal., 10 (2003), p. 19.
- [171] M. SUSSMAN, P. SMEREKA, AND S. OSHER, *A levelset approach for computing solutions to incompressible two-phase flow*, J. Comp. Phys., vol. 114 (1994), pp. 146–159.
- [172] J. SYLVESTER AND G. UHLMANN, *Global uniqueness theorem for an inverse boundary value problem*, Ann. Math., 125 (1987), pp. 153–169.
- [173] A. TAMASAN, *An inverse boundary value problem in two-dimensional transport*, Inverse Problems, 18 (2002), pp. 209–219.
- [174] P. TARONI, G. DANESINI, A. TORRICELLI, A. PIFFERI, L. SPINELLI, AND R. CUBEDDU, *Clinical trial of time-resolved scanning optical mammography at 4 wavelengths between 683 and 975 nm*, J. Biomedical Optics, 9 (2004), pp. 464–473.
- [175] M. D. TIDRIRI, *Asymptotic analysis of a coupled system of kinetic equations*, C. R. Acad. Sci. Paris Série I Math., 328 (1999), pp. 637–642.

- [176] L. N. TREFETHEN AND D. BAU, *Numerical Linear Algebra*, SIAM, Philadelphia, 1997.
- [177] B. J. TROMBERG, N. SHAH, R. LANNING, A. CERUSSI, J. ESPINOZA, T. PHAM, L. SVAASAND, AND J. BUTLER, *Non-invasive in vivo characterization of breast tumors using photon migration spectroscopy*, *Neoplasia*, 2 (2000), pp. 26–40.
- [178] G. UHLMANN, *Developments in inverse problems since Calderón's fundamental paper*, in *Harmonic Analysis and Partial Differential Equations*, M. Christ, C. E. Kenig, and C. Sadosky, eds., *Chicago Lectures in Mathematics*, The Chicago University Press, Chicago, 1999.
- [179] C. R. VOGEL, *Non-convergence of the L-curve regularization parameter selection method*, *Inverse Problems*, 12 (1996), pp. 535–547.
- [180] ———, *Computational Methods for Inverse Problems*, *Frontiers in Applied Mathematics*, SIAM, Philadelphia, 2002.
- [181] A. VRETBLAD, *Fourier Analysis and Its Applications*, *Graduate Texts in Mathematics*, Springer-Verlag, New York, 2003.
- [182] A. P. WANG AND S. UENO, *An inverse problem in a three-dimensional radiative transfer*, *Astrophys. Space Sci.*, 155 (1989), pp. 105–111.
- [183] A. J. WELCH AND M. J. C. VAN-GEMERT, *Optical-thermal Response of Laser Irradiated Tissue*, Plenum Press, New York, 1995.
- [184] Y. XU, N. IFTIMIA, H. JIANG, L. KEY, AND M. BOLSTER, *Three-dimensional diffuse optical tomography of bones and joints*, *J. Biomedical Optics*, 7 (2002), pp. 88–92.
- [185] A. G. YODH AND B. CHANCE, *Spectroscopy and imaging with diffusing light*, *Physics Today*, 48 (1995), pp. 34–40.
- [186] C. ZHU, R. H. BYRD, P. LU, AND J. NOCEDAL, *L-BFGS-B-FORTRAN subroutines for large-scale bound constrained optimization*, *ACM Trans. Math. Software*, 23 (1997), pp. 550–560.