Department of Mathematics and Systems Analysis

Efficient reconstruction algorithms for three-dimensional tomographic imaging

Helle Majander

Aalto University

DOCTORAL DISSERTATIONS

Aalto University publication series
DOCTORAL DISSERTATIONS 181/2016

Efficient reconstruction algorithms for three-dimensional tomographic imaging

Helle Majander

A doctoral dissertation completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall E of the school on 11 November 2016 at 12.

Aalto University
School of Science
Department of Mathematics and Systems Analysis

Supervising professor
Prof. Nuutti Hyvönen

Thesis advisor
Prof. Nuutti Hyvönen

Preliminary examiners
Prof. Samuli Siltanen, University of Helsinki, Finland
Dr. Marta M. Betcke, University College London, United Kingdom

Opponent
Prof. Kim Knudsen, Technical University of Denmark, Denmark

Aalto University publication series
DOCTORAL DISSERTATIONS 181/2016

ISBN 978-952-60-7007-0 (printed)
ISBN 978-952-60-7006-3 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-60-7006-3

Unigrafia Oy
Helsinki 2016

Finland

441 697
Printed matter

Aalto University

Abstract

Aalto University, P.O. Box 11000, FI-00076 Aalto www.aalto.fi

Author
Helle Majander

Name of the doctoral dissertation
Efficient reconstruction algorithms for three-dimensional tomographic imaging

Publisher School of Science

Unit Department of Mathematics and Systems Analysis

Series Aalto University publication series DOCTORAL DISSERTATIONS 181/2016

Field of research Mathematics

Manuscript submitted 14 June 2016 **Date of the defence** 11 November 2016

Permission to publish granted (date) 1 September 2016 **Language** English

☐ **Monograph** ☒ **Article dissertation** ☐ **Essay dissertation**

Abstract

This thesis considers nonlinear parameter estimation problems arising from tomographic imaging modalities governed by elliptic partial differential equations. These are ill-posed inverse problems and hence their solution requires regularization or, in the Bayesian framework, incorporation of prior information about the to-be-reconstructed spatially varying parameter. In particular, if the parameter is known to have distinct inclusions in a constant background, we quantify such information by assuming that after discretization the parameter follows an edge-enhancing prior distribution. Moreover, we study how to recover from different kinds of errors in the data, since even small ones can be enough to ruin the reconstruction for an ill-posed tomographic imaging problem.

We consider the solution of the investigated inverse problem to be the maximum a posteriori estimate for the parameter of interest, which can be found by solving a minimization problem. We propose to search for the minimizer by an iterative algorithm based on combining linearizations of the forward model, lagged diffusivity steps and a priorconditioned Krylov subspace method (LSQR). By presenting examples from electrical impedance tomography (EIT), diffuse optical tomography (DOT) and quantitative photoacoustic tomography (QPAT), we demonstrate that such a method can be implemented efficiently enough to be feasible for solving large-scale three-dimensional problems. In addition, we use a conformal invariance result for the complete electrode model (CEM) of EIT to compensate for geometric modeling errors.

Keywords parameter estimation problem, tomographic imaging, edge-enhancing regularization, modeling errors, priorconditioning, LSQR, electrical impedance tomography, complete electrode model, diffuse optical tomography, quantitative photoacoustic tomography

ISBN (printed) 978-952-60-7007-0 **ISBN (pdf)** 978-952-60-7006-3

ISSN-L 1799-4934 **ISSN (printed)** 1799-4934 **ISSN (pdf)** 1799-4942

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2016

Pages 156 **urn** http://urn.fi/URN:ISBN:978-952-60-7006-3

Aalto-yliopisto

Tiivistelmä

Aalto-yliopisto, PL 11000, 00076 Aalto www.aalto.fi

Tekijä
Helle Majander

Väitöskirjan nimi
Tehokkaita rekonstruktioalgoritmeja kolmiulotteiseen tomografiseen kuvantamiseen

Julkaisija Perustieteiden korkeakoulu

Yksikkö Matematiikan ja systeemianalyysin laitos

Sarja Aalto University publication series DOCTORAL DISSERTATIONS 181/2016

Tutkimusala Matematiikka

Käsikirjoituksen pvm 14.06.2016 **Väitöspäivä** 11.11.2016

Julkaisuluvan myöntämispäivä 01.09.2016 **Kieli** Englanti

☐ **Monografia** ☒ **Artikkeliväitöskirja** ☐ **Esseeväitöskirja**

Tiivistelmä

Tässä väitöskirjassa käsitellään tomografisiin kuvantamismenetelmiin liittyviä epälineaarisia parametriestimointiongelmia, joiden taustalla on elliptinen osittaisdifferentiaaliyhtälö. Kyseessä on huonosti asetettu käänteisongelma, jonka ratkaisu vaatii regularisointia tai bayesiläisessä viitekehyksessä prioritietoa rekonstruoitavasta paikkariippuvasta parametrista. Jos erityisesti tiedetään, että parametri koostuu inkluusioista vakioarvoisessa taustassa, tämä tieto kvantifioidaan olettamalla, että diskretoinnin jälkeen parametri noudattaa reunoja korostavaa priorijakaumaa. Väitöskirjassa käsitellään myös erityyppisten mittaus- ja mallinnusvirheiden huomiointia, sillä huonosti asetettujen tomografisten kuvantamisongelmien tapauksessa pienetkin epätarkkuudet voivat pilata rekonstruktion.

Tarkastellun inversio-ongelman ratkaisuna pidetään tässä työssä posteriorijakauman maksimikohtaa, joka voidaan löytää ratkaisemalla tietty minimointiongelma. Minimoijaa etsitään iteratiivisella algoritmilla, joka perustuu suoran mallin linearisoinnin, viivästetyn diffusiivisuusaskeleen ja pohjustetun Krylov-aliavaruusmenetelmän (LSQR) yhdistämiseen. Käsittelemällä impedanssitomografiaan (EIT), diffusiiviseen optiseen tomografiaan (DOT) ja kvantitatiiviseen fotoakustiseen tomografiaan (QPAT) liittyviä esimerkkejä osoitetaan, että esitelty algoritmi voidaan toteuttaa riittävän tehokkaasti suuren mittakaavan kolmiulotteisten ongelmien ratkaisemiseksi. Lisäksi erästä EIT:n täydellisen elektrodimallin konformi-invarianssitulosta käytetään geometristen mallinnusvirheiden kompensointiin.

Avainsanat parametriestimointi, tomografinen kuvantaminen, reunoja korostava regularisointi, mallivirheet, pohjustaminen, LSQR, impedanssitomografia, täydellinen elektrodimalli, diffusiivinen optinen tomografia, kvantitatiivinen fotoakustinen tomografia

ISBN (painettu) 978-952-60-7007-0 **ISBN (pdf)** 978-952-60-7006-3

ISSN-L 1799-4934 **ISSN (painettu)** 1799-4934 **ISSN (pdf)** 1799-4942

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2016

Sivumäärä 156 **urn** http://urn.fi/URN:ISBN:978-952-60-7006-3

Preface

This work has been carried out during the years 2012–2016 at the Department of Mathematics and Systems Analysis of Aalto University, although during the last two years I have been hosted at Centre de Mathématiques Appliquées of École Polytehnique in France. I acknowledge the Academy of Finland, the Foundation for Aalto University Science and Technology as well as the Emil Aaltonen Foundation for funding during these years.

I want to express my gratitude to the supervisor and advisor of this thesis, Prof. Nuutti Hyvönen, for his patient guidance and support during my doctoral studies. I also thank the preliminary examiners, Prof. Samuli Siltanen and Dr. Marta M. Betcke, for taking the time to review the manuscript, and Prof. Kim Knudsen for agreeing to act as my opponent. I owe special thanks also to Prof. Antti Hannukainen, Dr. Tanja Tarvainen, Dr. Lauri Harhanen and Dr. Stratos Staboulis with whom I have had the pleasure of working while writing this thesis.

For encouragement especially in the beginning of my doctoral studies, I want to thank Dr. Harri Hakula and Dr. Antti Rasila, the inverse problems group and the 'coffee room gang' at Aalto. I am also grateful for Prof. Houssem Haddar, his DéFI-team and other friends at l'X for welcoming me to France and making my stay here unforgettable. I thank my close friend Hanna, my mother Karin and my sister Aini for their patience and continuous support even from far away. Finally, my warmest thanks to Atte for being there for me.

Palaiseau, September 12, 2016,

Helle Majander

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I L. Harhanen, N. Hyvönen, H. Majander and S. Staboulis. Edge-enhancing reconstruction algorithm for three-dimensional electrical impedance tomography. *SIAM Journal on Scientific Computing*, 37(1), B60–B78, February 2015.

II A. Hannukainen, L. Harhanen, N. Hyvönen and H. Majander. Edge-promoting reconstruction of absorption and diffusivity in optical tomography. *Inverse Problems*, 32(1), 015008, 19 pages, January 2016.

III A. Hannukainen, N. Hyvönen, H. Majander and T. Tarvainen. Efficient inclusion of total variation type priors in quantitative photoacoustic tomography. *SIAM Journal on Imaging Sciences*, 9(3), 1132–1153, August 2016.

IV N. Hyvönen, H. Majander and S. Staboulis. Compensation for geometric modeling errors by electrode movement in electrical impedance tomography. *arXiv:1605.07823*, 22 pages, September 2016.

Author's Contribution

Publication I: "Edge-enhancing reconstruction algorithm for three-dimensional electrical impedance tomography"

The implementation of the algorithm (excluding the forward solver) is mainly due to the author. She also participated in the writing process. In particular, she designed and reported the numerical experiments.

Publication II: "Edge-promoting reconstruction of absorption and diffusivity in optical tomography"

The implementation of the algorithm (excluding the forward solver) and a large part of the writing process are due to the author. In particular, she designed and reported the numerical experiments.

Publication III: "Efficient inclusion of total variation type priors in quantitative photoacoustic tomography"

The author participated in the implementation of the matrix-free approach; the other parts of the algorithm (excluding the forward solver) are due to her. The author also substantially contributed to the writing process. In particular, she designed and reported the numerical experiments.

Publication IV: "Compensation for geometric modeling errors by electrode movement in electrical impedance tomography"

The author is responsible for modifying the forward solver of II and III for EIT. She also substantially contributed to the writing process. In particular, she designed and reported the numerical experiments.

1. Introduction

After Wilhelm Röntgen's discovery of X-rays in 1895, the world of medicine experienced a major change. Before that, the only way to *see* inside a patient was by a surgery or some other invasive method, but with the new technology, a doctor could make an *exterior* measurement to get an image of what the patient looked like from *inside*. Conventional X-ray imaging developed into computerized tomography (CT) and was soon accompanied by other non-invasive imaging modalities, such as ultrasonography, positron emission tomography (PET) and magnetic resonance imaging (MRI). Newer methodologies include optical tomography, electrical impedance and capacitance tomography, and elastography as well as various hybrid methods. For a thorough history of the development of diagnostic tomographic imaging, see [37]. In addition to diagnostics, different imaging methods have several other applications, for example in non-destructive testing of materials, and imaging of deep geologic structures, the floor of an ocean or the space.

The shared aim of all tomographic imaging modalities is to recover internal information based on interaction between the input signal and the imaged object. The type of the recovered information depends on what kind of signal is employed, in what way it is applied, and how the measurements are recorded. For example in X-ray imaging, the target is exposed to electromagnetic radiation in X-ray frequencies, and the distinction of tissues is based on their capability of absorbing different amounts of photons from the radiation. Furthermore, although conventional X-ray imaging and CT use the same type of input signals, due to the different application and recording methods, the two modalities yield different information about the target.

Formulating a tomographic imaging task mathematically is a two-fold problem. First, assuming that the internal parameter of interest (the to-

be-recovered information) is known, the *forward problem* is to make use of the physical laws related to the employed imaging method to describe how the recorded measurement depends on the parameter. The resulting mathematical model is then used in the second part, which is the *inverse problem* of estimating the value of the internal parameter based on some actual recorded measurements. While specifying the forward model is by no means a trivial task, it is beyond the scope of this thesis. We will consider some well-established forward models and focus on the parameter estimation problem.

Being an inverse problem, estimation of the internal parameter is not a straightforward problem to solve. The difficulty arises from the fact that such problems are usually *ill-posed*. According to the definition given by Jacques Hadamard in 1920's, a *well-posed* problem has a unique solution that depends continuously on the data [43]. Conversely, an ill-posed problem may not have an exact solution at all or there may be several (possibly an infinite number) of them. Most importantly, even if a solution exists, its dependence on the data is certainly not continuous in any reasonable metric. Hence, to solve such a problem, it is necessary to search for an approximate solution and to use additional *prior* information that helps to define which one of all the possible approximate solutions is likely to be the desired one. Furthermore, all the errors have to be carefully accounted for, since otherwise even small ones may ruin the solution. Broadly speaking, the possible inaccuracies can be divided into two categories: *model errors* result from imperfections in the forward model compared to the physical reality whereas *measurement errors* correspond to data uncertainties, which can depend for example on the used instrumentation.

In addition to the difficulties arising from the ill-posedness, reconstructing the parameter of interest in a tomographic imaging problem often introduces numerical challenges. Although the final images produced by many practical imaging modalities are two-dimensional, the actual problems are usually inherently three-dimensional. Even when all the related mathematical concepts can be formulated similarly in both spatial dimensions, the computational cost often increases substantially with the dimension. For this reason, we are particularly interested in developing algorithms that are also capable of handling three-dimensional problems efficiently.

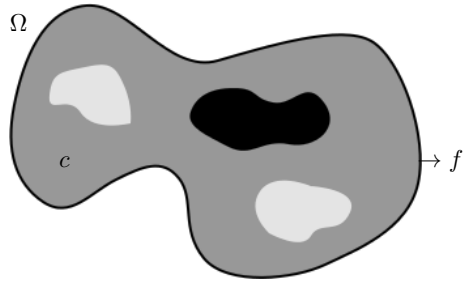

Figure 1.1. The set-up of a general parameter estimation problem (inverse boundary value problem): determine the value of a parameter c living in a domain Ω, based on the (boundary) data f.

Parameter estimation problems

In order to handle the considered problems in a more concrete manner, we introduce in this section a notation for a general parameter estimation problem, illustrated in Figure 1.1. We denote the domain (the examined body) by $\Omega \subset \mathbb{R}^d$. Although in this work we are mainly interested in the three-dimensional case, $d = 3$, the presented methods also apply to $d = 2$. The parameter of interest is denoted by $c \in C \subset D$, where $D = D(\Omega)$ is the chosen *parameter space*, which is a function space defined over the domain, and C is the set of *admissible parameters* (containing e.g. positivity constraints). Our data (the measurement) is denoted by $f \in F$, where F is the *data space* usually defined on (a part of) the boundary $\partial\Omega$. Notice that we only consider time-independent settings, that is, neither the parameter c nor the measurement f depends on time.

In this configuration, the forward problem is to determine the *forward operator* $\mathcal{F} : C \to F$, the structure of which depends on the application. We restrict our attention to such $\mathcal{F}$ that can be described by an elliptic partial differential equation. In particular, the considered forward operator may be nonlinear. The parameter estimation problem is now described as follows: given the data $f \in F$, find such $c \in C$ that $\mathcal{F}(c) \approx f$. We need to look for an approximate solution, since the problem is ill-posed and in practical applications the available data is always erroneous. More precisely, we consider the usual approach of finding the solution $c \in C$ that minimizes the least-squares functional

$$\Phi(c) = \frac{1}{2}\|\mathcal{F}(c) - f\|_F^2 \tag{1.1}$$

accompanied by suitable *regularization* that reflects our prior information on the parameter. Note that the norm $\|\cdot\|_F$ may depend on the employed

noise model. In this thesis we consider numerical aspects that let us efficiently solve minimization problems of this type with a high-dimensional discretization of the parameter space, a large number of data and in the presence of modeling errors.

The rest of this overview is organized as follows. In Section 2 we go through some numerical techniques related to solving the parameter estimation problem and conclude by introducing a reconstruction algorithm that can be efficiently implemented for relatively large problems. Section 3 is dedicated to describing the three imaging modalities that are considered in the four publications comprising this thesis. The results of the publications are then summarized in Section 4.

2. Numerical methods

In this section we consider the numerical solution of the parameter estimation problem described in Section 1. We begin by discretizing the problem in Section 2.1 and proceed in Section 2.2 to describe the Bayesian framework, which is used to derive the to-be-minimized functional corresponding to (1.1). In Section 2.3 we discuss the handling of modeling errors, and finally in Section 2.4 we briefly go through the minimization algorithms relevant to this thesis.

2.1 Discretization

To keep the explanations simple, we only consider here the case in which all parameters and measurements are real-valued. Notice, however, that the described approaches generalize to the complex-valued setting as well. In particular, this observation is relevant for the set-up of Publication II. Our presentation follows the books [16, 26]. For a more general description of the theoretical aspects considered in this section, we refer to [115].

In a tomographic imaging problem the main parameter of interest is a function depending on the position $x \in \Omega$ and as such, it naturally lives in an *infinite-dimensional* function space that depends on the considered imaging modality. Since our algorithms are designed for handling *finite-dimensional* parameters, we first need to discretize the parameter and the forward model. In other words, we approximate $c \in C \subset D$ by some $\hat{c} \in \widehat{C} \subset \widehat{D}$, where $\widehat{D}$ is a finite-dimensional subspace of D. The approximate parameter can then be represented in terms of a basis $\{\phi_j\}_{j=1}^n$ of $\widehat{D}$ as

$$\hat{c}(x) = \sum_{j=1}^{n} c_j \phi_j(x), \qquad x \in \Omega, \tag{2.1}$$

which is uniquely determined by the vector of coefficients $\boldsymbol{c} = [c_j]_{j=1}^n \in \mathbb{R}^n$; we write $\widehat{D} \cong \mathbb{R}^n$. After discretization, our aim is to recover the vector $\boldsymbol{c}$.

Forward model

To be able to numerically solve the forward problem corresponding to an approximate parameter, we also need to discretize the forward model $\mathcal{F} : C \to F$. To explain how this is done, we split the evaluation of $g = \mathcal{F}(c)$ into two parts using $\mathcal{F} : C \to S \to F$ as follows: for a given $c \in C$, first solve the *state* $u \in S$ from the *state equation* and then use the *measurement operator* $\mathcal{M} : S \to F$ to retrieve the data corresponding to u. Conceptually, the variable $u \in S$ describes completely the state of the system corresponding to a fixed parameter c and the considered input. The measurement operator $\mathcal{M}$ then indicates which (small) part of the information in u can actually be measured. For simplicity, in this section we present the theory assuming that the measurement $g = \mathcal{M}(u)$ is only related to one state $u \in S$. However, in practice the data in (1.1) is often recorded corresponding to several different input signals applied to the target (see Section 3), and hence it is related to several states $u^{(1)}, \ldots, u^{(N)}$. Notice that the generalization to such a case is straightforward by using $g = [\mathcal{M}(u^{(1)}), \ldots, \mathcal{M}(u^{(N)})]$.

The state equation of a (diffuse) tomographic imaging problem is typically governed by an elliptic (or parabolic) partial differential equation (PDE) with appropriate boundary conditions. We assume that there exists a corresponding *variational formulation*:

$$B(u, v) = L(v) \qquad \text{for all } v \in S, \tag{2.2}$$

where $B : S \times S \to \mathbb{R}$ and $L : S \to \mathbb{R}$. Here $B = B(c)$ depends implicitly on the parameter c. Details on how to deduce the formulation (2.2) for a given boundary value problem can be found for example in [16]. The following theorem lists the sufficient conditions under which the variational formulation has a unique solution.

Theorem 1 (Lax–Milgram)**.** *Assume that S is a real Hilbert space. Let $L : S \to \mathbb{R}$ be a continuous linear functional and $B : S \times S \to \mathbb{R}$ a bilinear form for which there exist constants $\beta < \infty$ and $\alpha > 0$ such that*

$$|B(v, w)| \le \beta \|v\|_S \|w\|_S \qquad \textit{for all } v, w \in S \qquad \textit{(boundedness)},$$
$$B(v, v) \ge \alpha \|v\|_S^2 \qquad \textit{for all } v \in S \qquad \textit{(coercivity)}.$$

Then there exists a unique $u \in S$ that solves (2.2)*.*

The *state space* S is typically infinite-dimensional; $u \in S$ can be for example a function of $x \in \Omega$. Hence, we need to make a finite-dimensional

approximation $\hat{u} \in \widehat{S} \subset S$. If $\{\phi_j\}_{j=1}^m$ forms a basis for $\widehat{S} \cong \mathbb{R}^m$, then

$$\hat{u} = \sum_{j=1}^{m} u_j \phi_j \tag{2.3}$$

for some $\boldsymbol{u} = [u_j]_{j=1}^m \in \mathbb{R}^m$. Notice that we often use the same discretization for c and u, i.e. the bases for $\widehat{S}$ and $\widehat{D}$ are the same, and $m = n$, but in general they need not to be. When $\widehat{S}$ is chosen properly, there exists a unique solution also for the approximate problem corresponding to (2.2) and the associated error is characterized by the following theorem.

Theorem 2 (Céa's lemma)**.** *Assume S, L, and B are as in Theorem 1. If $\widehat{S}$ is a finite-dimensional subspace of S, there exists a unique $\hat{u} \in \widehat{S}$ solving*

$$B(\hat{u}, \hat{v}) = L(\hat{v}) \qquad \textit{for all } \hat{v} \in \widehat{S}. \tag{2.4}$$

Furthermore, if $u \in S$ is the unique solution of (2.2)*, then*

$$\|u - \hat{u}\|_S \leq \frac{\beta}{\alpha} \min_{\hat{v} \in \widehat{S}} \|u - \hat{v}\|_S.$$

See for example [16] for proofs of these theorems. Now, using (2.1), (2.3) and letting the *test function* $\hat{v}$ go through all the basis functions $\{\phi_j\}_{j=1}^m$, the approximate state equation (2.4) can be written in a discrete form

$$\boldsymbol{B}\boldsymbol{u} = \boldsymbol{L}, \tag{2.5}$$

where $\boldsymbol{B} = [B(\phi_i, \phi_j)]_{i,j=1}^m \in \mathbb{R}^{m \times m}$ and $\boldsymbol{L} = [L(\phi_j)]_{j=1}^m \in \mathbb{R}^m$. Solving $\boldsymbol{u} = [u_j]_{j=1}^m \in \mathbb{R}^m$ from (2.5) is then conceptually straightforward.

After the approximate state is solved from (2.5) and (2.3), we use the measurement operator to get the (approximate) data, $g = \mathcal{M}(\hat{u})$. The measurement operator may also (implicitly) depend on other infinite-dimensional objects, which are again replaced by suitable finite-dimensional approximations when numerically evaluating $\mathcal{M}(\hat{u})$. In addition, even though physically realistic measurements $g \in F$ are always finite-dimensional, this is not necessarily the case in mathematical models. If F is infinite-dimensional, again, we replace it by a finite-dimensional approximation $g \approx \hat{g} \in \widehat{F} \subset F$. For notational convenience, we often identify the finite-dimensional approximations with the corresponding coefficient vectors and use $\mathcal{F}$ also to denote the discretized forward model i.e. we write $\boldsymbol{g} = \mathcal{F}(\boldsymbol{c})$.

Derivative of the forward model

To minimize a functional of form (1.1) with a differentiation-based algorithm, we also need to approximate the derivative of the map $c \mapsto \mathcal{F}(c)$

evaluated at a given point. There are two possible approaches: either to discretize the model first, and then calculate the *exact derivative of the discretized model*, or to first calculate analytically the *exact derivative of the continuous model*, and then discretize the resulting equations. For example [23] presents some methods for computing the exact derivative of the discrete model $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^k$ at a point $\boldsymbol{c} \in \mathbb{R}^n$, i.e. the Jacobian matrix

$$J_c^{\mathcal{F}}(\boldsymbol{c}) = \left[\frac{\partial \mathcal{F}_i(\boldsymbol{c})}{\partial c_j}\right]_{i,j=1}^{k,n} \in \mathbb{R}^{k \times n}.$$

In this work we take the other approach. We assume that the continuous forward model $\mathcal{F} : C \to F$ is Fréchet differentiable with respect to the parameter $c \in C$ i.e. that it has a *Fréchet derivative* for all $c \in C$, defined as follows.

Definition 1. *Let D and F be Banach spaces and $C \subset D$ an open subset of D. The Fréchet derivative of $\mathcal{F} : C \to F$ at $c \in C$ is the bounded linear map $\mathcal{F}'(c) : D \to F$ that satisfies*

$$\lim_{0 \neq h \to 0} \frac{\|\mathcal{F}(c+h) - \mathcal{F}(c) - (\mathcal{F}'(c))(h)\|_F}{\|h\|_D} = 0.$$

To discretize the Fréchet derivative, notice first that since $(\mathcal{F}'(c))(h) \in F$ it is either already finite-dimensional, $F \cong \mathbb{R}^k$, or we can discretize it in the same way as the measurement $g \in F$. We then write the derivative by its components as $[(\mathcal{F}_i'(c))(h)]_{i=1}^k \in \mathbb{R}^k$. Using the approximation (2.1) for the parameter $c \in C$ and a similar one for the direction $h \in D$, we have

$$(\mathcal{F}'(c))(h) \approx \left[(\mathcal{F}_i'(\hat{c}))(\phi_j)\right]_{i,j=1}^{k,n} \boldsymbol{h} \approx J_c^{\mathcal{F}}(\boldsymbol{c})\boldsymbol{h}.$$

It remains to numerically approximate $(\mathcal{F}_i'(\hat{c}))(\phi_j)$ for all $i = 1, \ldots, k$ and $j = 1, \ldots, n$. In practice, since the forward operator is given as a composition of the measurement operator and the solution to the state equation, the discretization of its derivative reflects back to the discretizations of the state equation and the measurement operator. The details are mostly model-specific and are hence postponed to Section 3.

Finite element method

Finite element method (FEM) is possibly the most important technique for constructing finite-dimensional subspaces for function spaces related to a spatial domain Ω. The idea is to divide the domain into a finite number of *elements* to form *local* finite-dimensional subspaces, which are then combined to construct the *global* one. We demonstrate the idea with a simple example using a triangular mesh and piecewise linear basis functions.

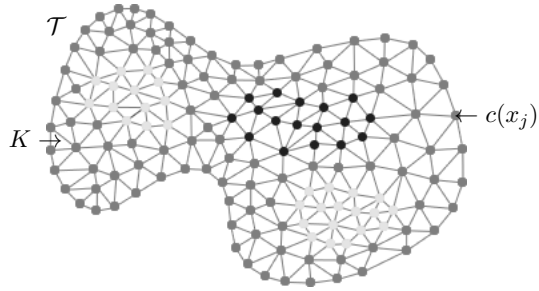

Figure 2.1. An approximation of the parameter $c(x) \approx \sum c(x_j)\phi_j(x_j)$ on a triangular FE mesh $\mathcal{T}$ of the domain Ω.

For a more general approach, detailed description and error estimates we refer to [16, 26].

Let $\mathcal{T}$ be a division of the domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ into triangles (or tetrahedrons), $\overline{\Omega} = \cup_{K \in \mathcal{T}} K$, so that the interiors of distinct elements are disjoint and the vertices and edges (and faces) of $\mathcal{T}$ coincide, i.e. no vertex of any element lies in the interior of an edge. An example of such a division is given in Figure 2.1. To this division $\mathcal{T}$ we associate a space P of continuous functions that are linear on each triangle K. The space P is finite-dimensional and, as an example, $P \subset H^1(\Omega)$. On each element K, a linear function can be uniquely determined by its values at the vertices of the triangle. Hence, collecting the values at all vertices x_i, $i = 1, \ldots, n$ in $\mathcal{T}$, we can uniquely determine any function of P. Furthermore, the piecewise linear functions with

$$\phi_j(x_i) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}, \qquad i, j = 1, \ldots, n \tag{2.6}$$

form a basis for P. The power of FEM lies in the fact that when using these basis functions, the system corresponding to (2.5) becomes sparse. Furthermore, the discrete variables are easy to form. Consider for example the approximation of $c \in C$ in (2.1). With the basis functions (2.6), the coefficients satisfy $c_j = c(x_j)$ and hence the discrete variable $\boldsymbol{c}$ is formed by simple point evaluations of c.

2.2 Bayesian inversion

There are different ways to form the regularized cost functional corresponding to (1.1). Here we present the Bayesian approach along the lines of [58, 101]; for the so-called regularization theory, see for example [36].

In what follows, we consider a discretized parameter $\boldsymbol{c} \in \mathbb{R}^n$ and a discrete (or discretized) measurement $\mathcal{F}(\boldsymbol{c}) \in \mathbb{R}^k$. For infinite-dimensional results see [70, 76, 100] and references therein.

We assume that the measurements are corrupted by additive Gaussian noise, i.e. our data $\boldsymbol{f} \in \mathbb{R}^k$ is of form

$$\boldsymbol{f} = \mathcal{F}(\boldsymbol{c}) + \boldsymbol{\eta}, \tag{2.7}$$

where $\boldsymbol{\eta} \in \mathbb{R}^k$ is (a realization of) a zero-mean Gaussian random variable with a covariance matrix $\Gamma_\eta \in \mathbb{R}^{k\times k}$. Consult for example [58] for other noise models. Assuming that the noise $\boldsymbol{\eta}$ and the parameter $\boldsymbol{c}$ are independent, the *likelihood*, i.e. the probability density of the measurement given the parameter, is also Gaussian with the same covariance Γ_η and with a mean shifted by $\mathcal{F}(\boldsymbol{c})$. To put it explicitly, the probability density function has the form

$$p(\boldsymbol{f} \,|\, \boldsymbol{c}) \;\propto\; \exp\!\left(-\frac{1}{2}\big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)^{\mathrm{T}} \Gamma_\eta^{-1} \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)\right),$$

where we have ignored the coefficient $((2\pi)^k \det(\Gamma_\eta))^{-1/2}$, since it does not depend on $\boldsymbol{c}$.

The information that we have on the parameter $\boldsymbol{c} \in \mathbb{R}^n$ before any measurements — the *a priori* information — is in our case quantified by giving the parameter a *prior* probability density of the form

$$p(\boldsymbol{c}) \;\propto\; \exp\big(-a\,R(\boldsymbol{c})\big), \tag{2.8}$$

where $a > 0$ is a free 'regularization' parameter and R is a function describing the expected behavior of $\boldsymbol{c}$. From the Bayes' formula we now get the posterior density

$$\begin{aligned} p(\boldsymbol{c} \,|\, \boldsymbol{f}) \;&\propto\; p(\boldsymbol{f} \,|\, \boldsymbol{c})\, p(\boldsymbol{c}) \\ &\propto\; \exp\Big(-\frac{1}{2}\big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)^{\mathrm{T}} \Gamma_\eta^{-1} \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big) - a\,R(\boldsymbol{c})\Big), \end{aligned} \tag{2.9}$$

where parameter-independent constants have again been ignored.

In an idealized statistical sense, the posterior probability density function (2.9) is the solution to the inverse problem we are considering. However, in practice we are usually interested in more concrete properties of the parameter $\boldsymbol{c}$, which can be described by estimates calculated from the posterior density. In particular, the so-called *point estimates* give directly reconstructions of the parameter (see e.g. [58] for other estimate types). The most common point estimates are the *conditional mean* (CM) estimate

$$\boldsymbol{c}_{\mathrm{CM}} \;=\; \int_{\mathbb{R}^n} \boldsymbol{c}\, p(\boldsymbol{c} \,|\, \boldsymbol{f})\, d\boldsymbol{c}$$

and the *maximum a posteriori* (MAP) estimate

$$\boldsymbol{c}_{\text{MAP}} \;=\; \arg\max_{\boldsymbol{c}\in\mathbb{R}^n} p(\boldsymbol{c}\,|\,\boldsymbol{f}).$$

The CM estimate can be approximated for example by using Markov chain Monte Carlo (MCMC) techniques (see e.g. [105]). In this work we restrict our attention to the MAP estimate, which can be found by solving a minimization problem (see Section 2.4).

Gaussian priors

The choice of a function $R(\boldsymbol{c})$ for the prior (2.8) can be done in many ways and it is not always clear, which is the best one. We first consider the function

$$R(\boldsymbol{c}) = \frac{1}{2}(\boldsymbol{c}-\boldsymbol{c}_0)^{\mathrm{T}}\Gamma_c^{-1}(\boldsymbol{c}-\boldsymbol{c}_0), \tag{2.10}$$

which gives the parameter of interest a Gaussian prior density with mean $\boldsymbol{c}_0 \in \mathbb{R}^n$ and covariance $\Gamma_c \in \mathbb{R}^{n\times n}$. As an example, set $\boldsymbol{c}_0 = c_0\mathbf{1}$, with $\mathbf{1} = [1,\ldots,1]^{\mathrm{T}} \in \mathbb{R}^n$, and $\Gamma_c = \gamma^2\mathrm{I}$, where $\mathrm{I} \in \mathbb{R}^{n\times n}$ is the identity matrix. Then the coefficients c_j in (2.1) vary from the mean $c_0 \in \mathbb{R}$ with the standard deviation $\gamma \in \mathbb{R}_+$ independently of one another. However, as these coefficients correspond to the node values of the parameter $c(x)$ in the FE discretization, the assumption of independence (and identical distributions) is often unreasonable; usually we expect there to be some correlation between the parameter values evaluated at nodes lying close to one another. A more informative Gaussian prior can be formed, e.g., by introducing the distance-dependent covariance matrix of 'squared exponential type'

$$(\Gamma_c)_{i,j} \;=\; \gamma^2 \exp\left(-\frac{|x_i - x_j|^2}{2\lambda^2}\right), \quad \gamma,\lambda > 0,$$

where $x_i, x_j \in \mathbb{R}^d$ are the FE nodes corresponding to the coefficients c_i, c_j, respectively, γ is the standard deviation for the node values, and λ is the correlation length [72]. A downside of using this type of covariance matrix is that, in principle, its construction requires the calculation of distances between all pairs of nodes in the (unstructured) FE mesh, which can be computationally expensive.

Gaussian densities are the most commonly used family of priors, since they can often give good approximations even to non-Gaussian phenomena and they are relatively easy to handle. In particular, for a *linear* forward operator $\mathcal{F}(\boldsymbol{c}) = \mathcal{F}\boldsymbol{c}$, when both the likelihood and the prior are

Gaussian, the posterior is also Gaussian with explicit expressions for the mean $\overline{c}$ and the covariance matrix [58]. This simplifies the statistical analysis of the posterior, as for example $\boldsymbol{c}_{\mathrm{MAP}} = \boldsymbol{c}_{\mathrm{CM}} = \overline{c}$.

The regularization theory provides also other approaches to choosing a penalty term of form (2.10). By choosing the 'weighting matrix' Γ_c^{-1} appropriately, this kind of regularization can correspond to different norms in the continuous setting. For example with $\boldsymbol{c}_0 = \mathbf{0} \in \mathbb{R}^n$ and $\Gamma_c^{-1} = S \in \mathbb{R}^{n \times n}$ defined by

$$S_{i,j} = \int_\Omega \nabla\phi_i(x) \cdot \nabla\phi_j(x)\, dx, \qquad i,j = 1,\ldots,n,$$

we get the squared L^2-norm for the gradient of the finite-dimensional parameter

$$R(\boldsymbol{c}) \;=\; \frac{1}{2}\boldsymbol{c}^{\mathrm{T}} S \boldsymbol{c} = \frac{1}{2}\int_\Omega |\nabla \hat{c}(x)|^2\, dx. \tag{2.11}$$

Using the regularization term (2.11) corresponds to assuming an (improper) *smoothness prior* [58] for the parameter $\boldsymbol{c}$. In fact, as long as the weight matrix S is positive (semi-)definite, the Tikhonov regularization [106] with this type of penalty term can always be identified with a MAP estimate corresponding to a (possibly improper) Gaussian prior with the inverse covariance matrix S.

Edge-enhancing priors

Even though the use of Gaussian (smoothness) priors may be a computationally inviting choice, in many applications our prior knowledge does not support the assumption of a smoothly behaving parameter. Instead, we want to choose a prior that allows the parameter to exhibit jumps while also restricting it from behaving too radically. To achieve this, consider a prior density of form (2.8) with the function R being a discretized version of the regularizer

$$\mathcal{R}(c) \;:=\; \int_\Omega r\big(|\nabla c(x)|\big)\, dx, \tag{2.12}$$

where $r : \mathbb{R}_+ \to \mathbb{R}_+$ is a continuously differentiable, monotonically increasing function [3]. As an example, with $r(t) = \frac{1}{2}t^2$ we reproduce (2.11). Choosing the function r carefully allows us to construct priors related to nonquadratic *edge-enhancing* regularization functionals that are designed to promote jumps in the parameter c. Note that such a choice corresponds to assuming an *impulse prior* for $|\nabla c(x)|$, see e.g. [58] for more information.

Edge-enhancing regularizers are frequently used in image reconstruction problems. One well known example is the *total variation* (TV) norm, which is in essence the L^1-norm of the gradient, i.e., the $W^{1,1}$ seminorm of the parameter [91]. We get the corresponding prior with the choice

$$r(t) = t \qquad \text{or} \qquad r(t) = \sqrt{t^2 + T^2} \approx |t|,$$

where the latter is a smoothened approximation that uses a small parameter $T > 0$ to ensure the differentiability of the regularizer (2.12). Similar choices can be used to implement the TV^q prior (see e.g. [52]). Another nonquadratic penalty term considered in this thesis was presented by Perona and Malik in [87] with two versions that correspond to the choices

$$r(t) = \frac{1}{2} T^2 \log\left(1 + (t/T)^2\right) \quad \text{and} \quad r(t) = \frac{1}{2} T^2 \left(1 - \exp\left(-(t/T)^2\right)\right).$$

The first one prefers large jumps over small ones and hence gives better contrast, whereas the second favors large constant regions over small ones. In both versions the parameter $T > 0$ controls the threshold below which the discontinuities are considered as noise.

2.3 Model errors

In the previous section we dealt with (additive Gaussian) measurement errors when constructing the to-be-minimized functional. However, so far we have ignored the errors that are caused by imperfections in the forward model, i.e. the model errors. These can result, for example, from unknown auxiliary model parameters such as the shape of the imaged object or other geometric dependencies. In addition, inaccuracies caused by implementational choices such as domain truncation or discretization are considered as modeling errors.

Arguably the most straightforward way to deal with auxiliary model parameters is to include their estimation in the reconstruction algorithm (see e.g. [14, 31, 57, 111]). When the dimension of these additional parameters is reasonably low and the computation of the corresponding derivatives cheap, this approach can be efficient. In some special cases, the errors caused by inaccurate (geometric) model parameters can also be compensated by changing other parameters, whose reconstruction is easier to implement. This procedure has been considered for mismodeled domain shapes in [63, 64] and Publication IV. The implementational errors can in some cases be estimated analytically or numerically (see

e.g. [16, 26, 109]) and possibly be accounted for, e.g. by choosing a suitable stopping condition for an iterative method. Moreover, when using an accurate enough numerical scheme, these errors can be neglected. There are few general methods for dealing with model errors in (diffuse) tomographic problems. In the following sections we briefly describe two well-established approaches.

Approximation error approach

To simplify the presentation, we assume here that the range of the forward operator is finite-dimensional, $\mathcal{F}(c) \in F \cong \mathbb{R}^k$, and we only consider the case of discretization errors. The idea of the approximation error approach [58, 59] is to replace the discretized model (2.7) by the accurate one

$$\boldsymbol{f} \;=\; \mathcal{F}(c) + \boldsymbol{\eta} \;=\; \mathcal{F}(\boldsymbol{c}) + \epsilon(c) + \boldsymbol{\eta}, \tag{2.13}$$

where $\boldsymbol{c} \in \mathbb{R}^n$, $\epsilon(c) = \mathcal{F}(c) - \mathcal{F}(\boldsymbol{c}) \in \mathbb{R}^k$ is the modeling error and $\boldsymbol{\eta} \in \mathbb{R}^k$ is the Gaussian noise vector with zero mean and a covariance $\Gamma_\eta \in \mathbb{R}^{k\times k}$. In case of a linear forward model $\mathcal{F}(c) = \mathcal{F}c$ and Gaussian distributions for the parameter $\boldsymbol{c}$ and the error $\epsilon(c)$, the posterior distribution corresponding to the model (2.13) can be solved exactly [59]. Otherwise, it is usually assumed that the continuous model can be approximated reasonably well (within the measurement accuracy) by a dense discretization, $\mathcal{F}(c) \approx \widetilde{\mathcal{F}}(\tilde{\boldsymbol{c}})$, where $\tilde{\boldsymbol{c}} \in \mathbb{R}^m$, $m > n$, and the distribution for the modeling error $\epsilon(c) \approx \epsilon(\tilde{\boldsymbol{c}}) = \widetilde{\mathcal{F}}(\tilde{\boldsymbol{c}}) - \mathcal{F}(\boldsymbol{c})$ is approximated from a simulated sample.

In order to approximate the distribution of $\epsilon(\tilde{\boldsymbol{c}})$, first draw (e.g. using MCMC) a sample of vectors $\{\tilde{\boldsymbol{c}}^{(j)}\}_{j=1}^N$ distributed according to the prior probability density $p(\tilde{\boldsymbol{c}})$. Then the modeling error distribution can be approximated from the sample

$$\left\{\epsilon\big(\tilde{\boldsymbol{c}}^{(j)}\big) = \widetilde{\mathcal{F}}\big(\tilde{\boldsymbol{c}}^{(j)}\big) - \mathcal{F}\big(P\tilde{\boldsymbol{c}}^{(j)}\big)\right\}_{j=1}^N,$$

where $P : \tilde{\boldsymbol{c}} \mapsto \boldsymbol{c}$ is a linear model reduction operator (e.g. linear interpolation). In practice, it is usually assumed that the Gaussian distribution with the sample-based mean $\overline{\epsilon}$ and covariance Γ_ϵ,

$$\overline{\epsilon} = \frac{1}{N}\sum_{j=1}^N \epsilon\big(\tilde{\boldsymbol{c}}^{(j)}\big), \qquad \Gamma_\epsilon = \frac{1}{N-1}\sum_{j=1}^N \Big(\epsilon\big(\tilde{\boldsymbol{c}}^{(j)}\big) - \overline{\epsilon}\Big)\Big(\epsilon\big(\tilde{\boldsymbol{c}}^{(j)}\big) - \overline{\epsilon}\Big)^{\mathrm{T}},$$

is a reasonable approximation for the distribution of $\epsilon(\tilde{\boldsymbol{c}})$. Then the density of the total error $\epsilon(\tilde{\boldsymbol{c}}) + \boldsymbol{\eta}$ is also Gaussian with the mean $\overline{\epsilon}$ and the

covariance $\Gamma_\epsilon + \Gamma_\eta$. If the model is furthermore simplified by ignoring the dependence between the parameter $\boldsymbol{c}$ and the total error, the reasoning in Section 2.2 can be repeated to determine the posterior density

$$p(\boldsymbol{c} \,|\, \boldsymbol{f}) \,\propto\, \exp\Big(-\frac{1}{2}\big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c}) - \bar{\epsilon}\big)^{\mathrm{T}} \big(\Gamma_\epsilon + \Gamma_\eta\big)^{-1} \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c}) - \bar{\epsilon}\big) - a\, R(\boldsymbol{c})\Big).$$

Notice that the generation of the (large) error sample may be computationally expensive, but this step can be done 'offline' before the actual measurements are performed. The use of this method in the presence of e.g. geometric errors requires a fixed (geometric) parametrization and a good understanding of the possible measurement settings, which may be challenging when real-life set-ups are considered. See [5, 82, 83, 84] for examples of using the approximation error approach to compensate for discretization and geometric errors.

Difference imaging

Instead of reconstructing the value of the parameter $\boldsymbol{c} \in \mathbb{R}^n$ itself, in difference imaging the aim is to estimate the difference $\delta\boldsymbol{c} = \boldsymbol{c}^{(2)} - \boldsymbol{c}^{(1)}$ between parameter values corresponding to e.g. two time instances (see for example [8]). In particular, this approach is only applicable when *two* sets of measurements, $\boldsymbol{f}^{(1)}, \boldsymbol{f}^{(2)}$, are available. Assuming the additive noise model (2.7) and a linear forward operator $\mathcal{F}(\boldsymbol{c}) = \mathcal{F}\boldsymbol{c}$, we have

$$\delta\boldsymbol{f} := \boldsymbol{f}^{(2)} - \boldsymbol{f}^{(1)} = \mathcal{F}\delta\boldsymbol{c} + \delta\boldsymbol{\eta},$$

where $\delta\boldsymbol{\eta} = \boldsymbol{\eta}^{(2)} - \boldsymbol{\eta}^{(1)}$. With $\boldsymbol{\eta}^{(1)}$ and $\boldsymbol{\eta}^{(2)}$ being independent zero mean Gaussian noise vectors, $\delta\boldsymbol{\eta}$ is also Gaussian with zero mean and the covariance $\Gamma_{\eta^{(1)}} + \Gamma_{\eta^{(2)}}$. Then, following similar reasoning as before, we get the posterior density

$$p(\delta\boldsymbol{c} \,|\, \delta\boldsymbol{f}) \,\propto\, \exp\Big(-\frac{1}{2}\big(\delta\boldsymbol{f} - \mathcal{F}\delta\boldsymbol{c}\big)^{\mathrm{T}} \big(\Gamma_{\eta^{(1)}} + \Gamma_{\eta^{(2)}}\big)^{-1} \big(\delta\boldsymbol{f} - \mathcal{F}\delta\boldsymbol{c}\big) - a\, R(\delta\boldsymbol{c})\Big),$$

where the prior (2.8) is assumed for the difference $\delta\boldsymbol{c}$. Since we are now using the difference $\delta\boldsymbol{f}$ of two measurements, it can be argued that (most of) the modeling error is canceled out (see e.g. [9]).

For a nonlinear forward model, the most popular modification of difference imaging is to make a linear approximation

$$\mathcal{F}(\boldsymbol{c}) \approx \mathcal{F}(\boldsymbol{c}^{(0)}) + J^{(0)}(\boldsymbol{c} - \boldsymbol{c}^{(0)}),$$

where $\boldsymbol{c}^{(0)}$ is the reference point for the linearization and $J^{(0)} = J^{\mathcal{F}}_{\boldsymbol{c}}(\boldsymbol{c}^{(0)})$ is the Jacobian matrix of the map $\boldsymbol{c} \mapsto \mathcal{F}(\boldsymbol{c})$ evaluated at $\boldsymbol{c}^{(0)}$. Using this

approximation, the difference imaging approach can be applied as in the linear case.

To avoid the linearization, an alternative method was proposed in [73]. The idea is to write $\boldsymbol{c}^{(2)} = \boldsymbol{c}^{(1)} + \delta\boldsymbol{c}$ and reconstruct $\tilde{\boldsymbol{c}} := \left[(\boldsymbol{c}^{(1)})^{\mathrm{T}}, (\delta\boldsymbol{c})^{\mathrm{T}}\right]^{\mathrm{T}}$ using the model

$$\tilde{\boldsymbol{f}} := \begin{bmatrix} \boldsymbol{f}^{(1)} \\ \boldsymbol{f}^{(2)} \end{bmatrix} = \begin{bmatrix} \mathcal{F}(\boldsymbol{c}^{(1)}) \\ \mathcal{F}(\boldsymbol{c}^{(1)} + \delta\boldsymbol{c}) \end{bmatrix} + \begin{bmatrix} \boldsymbol{\eta}^{(1)} \\ \boldsymbol{\eta}^{(2)} \end{bmatrix} =: \widetilde{\mathcal{F}}(\tilde{\boldsymbol{c}}) + \tilde{\boldsymbol{\eta}}.$$

Here the noise vector $\tilde{\boldsymbol{\eta}}$ is again Gaussian with zero mean and a block diagonal covariance matrix $\Gamma_{\tilde{\eta}} = \mathrm{diag}(\Gamma_{\eta^{(1)}}, \Gamma_{\eta^{(2)}})$ and the posterior density has the form

$$p(\tilde{\boldsymbol{c}} \,|\, \tilde{\boldsymbol{f}}) \;\propto\; \exp\Big(-\frac{1}{2}\big(\tilde{\boldsymbol{f}} - \widetilde{\mathcal{F}}(\tilde{\boldsymbol{c}})\big)^{\mathrm{T}} \Gamma_{\tilde{\eta}}^{-1} \big(\tilde{\boldsymbol{f}} - \widetilde{\mathcal{F}}(\tilde{\boldsymbol{c}})\big) - a\, R(\tilde{\boldsymbol{c}}) \Big),$$

where the prior (2.8) is this time assumed for the composite variable $\tilde{\boldsymbol{c}}$. The downside of this approach is the increased dimension of the corresponding minimization problem. Furthermore, the effect on the modeling errors is not as obvious as in the linear(ized) case. However, it has been experimentally shown that, at least for electrical impedance tomography, the method is robust to several types of model errors [74].

2.4 Minimization schemes

The MAP estimate corresponding to the posterior density (2.9) is equivalent to the minimizer of the Tikhonov-type functional

$$\Phi(\boldsymbol{c}) \;:=\; \frac{1}{2}\big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)^{\mathrm{T}} \Gamma_{\eta}^{-1} \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big) + a\, R(\boldsymbol{c}). \tag{2.14}$$

Obviously, we arrive at a similar functional when considering any of the modified posterior densities presented in Section 2.3. In practice, the problem of finding the *global* minimum of (2.14) is usually simplified to the search for a *local* minimum, which can be identified with the help of the following theorem [53].

Theorem 3. *Assume that $\Phi : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. If $\boldsymbol{c}^* \in \mathbb{R}^n$ is a local minimum of Φ, then $\nabla\Phi(\boldsymbol{c}^*) = 0$ (necessary condition). Assume then that Φ is twice continuously differentiable. If a point $\boldsymbol{c}^* \in \mathbb{R}^n$ satisfies the necessary condition and the Hessian $\nabla^2\Phi(\boldsymbol{c}^*)$ is positive definite, then $\boldsymbol{c}^*$ is a local minimum of Φ (sufficient condition).*

Notice that in certain cases a local minimum is also the global one. In particular, this is the case for (strictly) convex functions. When considering a linear forward operator $\mathcal{F}(\boldsymbol{c}) = \mathcal{F}\boldsymbol{c}$ with a Gaussian prior i.e. the

regularization term (2.10), the convexity of (2.14) is ensured by the following theorems (see e.g. [53] for proofs).

Theorem 4. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix. Then the quadratic form $Q(x) := \frac{1}{2}x^{\mathrm{T}}Ax$, is a strictly convex function if and only if A is positive definite (and a convex function if and only if A is positive semi-definite).*

Theorem 5. *Let $F : \mathbb{R}^k \to \mathbb{R}$ be a (strictly) convex function and $A : \mathbb{R}^n \to \mathbb{R}^k$ an affine map. Then the function $F \circ A : \mathbb{R}^n \to \mathbb{R}$ is (strictly) convex.*

Theorem 6. *Let the functions $F_1, \ldots, F_N : \mathbb{R}^n \to \mathbb{R}$ be (strictly) convex and $t_1, \ldots, t_N \in \mathbb{R}_+$. Then the function $F := \sum_{j=1}^{N} t_j F_j$ is (strictly) convex.*

However, in general even the task of finding a local minimum is not a tractable problem when the dimension of the parameter $\boldsymbol{c} \in \mathbb{R}^n$ is high. Hence, the minimization algorithms are actually designed to find a *critical point* of Φ i.e. a point $\boldsymbol{c}^* \in \mathbb{R}^n$ that is only required to satisfy the necessary condition of Theorem 3. The search for such a point is performed iteratively: starting from an initial guess $\boldsymbol{c}^{(0)}$, the algorithm gives a rule to construct a sequence $\{\boldsymbol{c}^{(l)}\}$ where $\boldsymbol{c}^{(l)} \approx \boldsymbol{c}^*$ when $l \to \infty$. In practice it is certainly not possible to continue the iteration all the way to infinity, which means that the algorithm has to also include a *stopping criterion* (or several). This is described by a function $\mathcal{C} : \boldsymbol{c}^{(l)} \mapsto \{\mathrm{TRUE}, \mathrm{FALSE}\}$, whose evaluation determines if the current iterate can be perceived as an accurate enough approximation of $\boldsymbol{c}^*$.

In the rest of this section, we briefly introduce the minimization schemes that are relevant to this thesis. More thorough considerations and an abundance of different algorithms can be found for example in the text books [11, 53, 80].

Gauss–Newton iteration

When the regularization term in (2.14) is quadratic, i.e. of form (2.10), the minimization of Φ corresponds to a nonlinear least squares problem. Solutions to problems of this type can be efficiently approximated by the Gauss–Newton algorithm. In the following, we derive the algorithm by linearizing $\mathcal{F}$ in (2.14). Note that it can also be derived from the Newton's method by ignoring the second order terms in the Hessian, see e.g. [32].

To use the Gauss–Newton approach, it is necessary to assume that at each step of the iteration we are able to evaluate (approximations of) the forward solution $\mathcal{F}(\boldsymbol{c}^{(l)})$ and the Jacobian $J_{\mathcal{C}}^{\mathcal{F}}(\boldsymbol{c}^{(l)})$ of the map $\boldsymbol{c} \mapsto \mathcal{F}(\boldsymbol{c})$ at

a given point $\boldsymbol{c}^{(l)}$. In practice, these can be obtained as FEM approximations as described in Section 2.1. Denoting the Jacobian $J^{(l)} = J_c^{\mathcal{F}}(\boldsymbol{c}^{(l)})$ for convenience, we can make a linear approximation of the forward solution around $\boldsymbol{c}^{(l)}$,

$$\mathcal{F}(\boldsymbol{c}) \approx \mathcal{F}(\boldsymbol{c}^{(l)}) + J^{(l)}(\boldsymbol{c} - \boldsymbol{c}^{(l)}),$$

and use this to write the to-be-minimized functional (2.14) in the approximate form

$$\Phi^{(l)}(\boldsymbol{c}) := \frac{1}{2}\big(\boldsymbol{y}^{(l)} - J^{(l)}\boldsymbol{c}\big)^{\mathrm{T}}\Gamma_\eta^{-1}\big(\boldsymbol{y}^{(l)} - J^{(l)}\boldsymbol{c}\big) + a\,R(\boldsymbol{c}),$$

where $\boldsymbol{y}^{(l)} = \boldsymbol{f} - \mathcal{F}(\boldsymbol{c}^{(l)}) + J^{(l)}\boldsymbol{c}^{(l)}$. To find a critical point for this approximate functional, we look for a solution that satisfies the necessary condition $\nabla\Phi^{(l)}(\boldsymbol{c}) = 0$, which is equivalent to the equation

$$(J^{(l)})^{\mathrm{T}}\Gamma_\eta^{-1}J^{(l)}\boldsymbol{c} + a\,(\nabla R)(\boldsymbol{c}) = (J^{(l)})^{\mathrm{T}}\Gamma_\eta^{-1}\boldsymbol{y}^{(l)}. \tag{2.15}$$

With the assumption of a quadratic regularization term of form (2.10), we have $(\nabla R)(\boldsymbol{c}) = \Gamma_c^{-1}(\boldsymbol{c} - \boldsymbol{c}_0)$ and (2.15) is equivalent to the normal equation

$$A^{\mathrm{T}}A\,\boldsymbol{d}^{(l)} = A^{\mathrm{T}}\begin{bmatrix} L_\eta\big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c}^{(l)})\big) \\ \sqrt{a}L_c(\boldsymbol{c}_0 - \boldsymbol{c}^{(l)}) \end{bmatrix}, \qquad A = \begin{bmatrix} L_\eta J^{(l)} \\ \sqrt{a}L_c \end{bmatrix} \tag{2.16}$$

with $L_\eta^{\mathrm{T}}L_\eta = \Gamma_\eta^{-1}$ and $L_c^{\mathrm{T}}L_c = \Gamma_c^{-1}$ being Cholesky factorizations of the inverse covariance matrices, and $\boldsymbol{d}^{(l)} = \boldsymbol{c} - \boldsymbol{c}^{(l)}$. The Gauss–Newton iteration is then as follows.

Algorithm 1 (Gauss–Newton)**.** Assume *that the to-be-minimized functional is of form* (2.14) *with the regularizer* (2.10) *and that* a, $\boldsymbol{f}$, $\boldsymbol{c}_0$, Γ_η^{-1} *and* Γ_c^{-1} *are fixed.* Initialize *by calculating Cholesky factorizations* $L_\eta^{\mathrm{T}}L_\eta = \Gamma_\eta^{-1}$ *and* $L_c^{\mathrm{T}}L_c = \Gamma_c^{-1}$ *and setting* $\boldsymbol{c}^{(0)} = \boldsymbol{c}_0$.

Iterate *for* $l = 0, 1, 2, \ldots,$ *while* $\mathcal{C}(\boldsymbol{c}^{(l)}) \neq$ TRUE*:*
1. evaluate $\mathcal{F}(\boldsymbol{c}^{(l)})$ *and* $J^{(l)} = J_c^{\mathcal{F}}(\boldsymbol{c}^{(l)})$*,*
2. solve the descent direction $\boldsymbol{d}^{(l)}$ *from* (2.16)*,*
3. determine the step size $\alpha^{(l)} > 0$ *e.g. by a line search algorithm,*
4. set $\boldsymbol{c}^{(l+1)} = \boldsymbol{c}^{(l)} + \alpha^{(l)}\boldsymbol{d}^{(l)}$.

Notice that the size of the normal equation (2.16) — or of the corresponding least squares problem — solved in step 2 of the algorithm is proportional to the number of nodes in the FE mesh. Hence, even though the solution of the system is conceptually straightforward (see e.g. [41]),

determining the exact solution may lead to a lengthy computation (especially in three spatial dimensions). For the line search mentioned in step 3, we refer to [11, 53] and references therein, but note that the execution of such algorithms may be challenging when the evaluation of a forward solution is time-consuming.

Lagged diffusivity iteration

In [112] a method was proposed for minimizing the functional (2.14) with a nonquadratic regularizer of form (2.12) in case of a linear forward operator $\mathcal{F}(\boldsymbol{c}) = \mathcal{F}\boldsymbol{c}$. The algorithm is based on the fact that the discretized Fréchet derivative of the regularization term (2.12) gives the gradient of $R : \mathbb{R}^n \to \mathbb{R}_+$ the representation $(\nabla R)(\boldsymbol{c}) = G_c \boldsymbol{c}$, where G_c is the parameter-dependent positive semi-definite matrix with the elements

$$(G_c)_{i,j} := \int_\Omega \frac{r'(|\nabla \hat{c}(x)|)}{|\nabla \hat{c}(x)|} \, \nabla\phi_i(x) \cdot \nabla\phi_j(x) \, dx, \quad i,j = 1, \ldots, N. \tag{2.17}$$

The necessary condition $\nabla\Phi(\boldsymbol{c}) = 0$ is then equivalent to

$$A^{\mathrm{T}} A \, \boldsymbol{c} = A^{\mathrm{T}} \begin{bmatrix} L_\eta \boldsymbol{f} \\ 0 \end{bmatrix}, \qquad A = \begin{bmatrix} L_\eta \mathcal{F} \\ \sqrt{a} L_c \end{bmatrix} \tag{2.18}$$

where $L_\eta^{\mathrm{T}} L_\eta = \Gamma_\eta^{-1}$ and $L_c^{\mathrm{T}} L_c = G_c$ are again Cholesky factorizations. The lagged diffusivity iteration for (2.14) is given as follows [112].

Algorithm 2 (Lagged diffusivity)**.** Assume *that the to-be-minimized functional is of form* (2.14) *with a linear forward operator* $\mathcal{F}(\boldsymbol{c}) = \mathcal{F}\boldsymbol{c}$ *and* R *being a discretized version of the regularizer* (2.12)*. Let* a*,* Γ_η^{-1} *and the function* r *related to* R *together with its free parameter(s) be fixed.* Initialize *by calculating a Cholesky factorization* $L_\eta^{\mathrm{T}} L_\eta = \Gamma_\eta^{-1}$ *and by setting* $\boldsymbol{c}^{(0)} = \boldsymbol{c}_0$ *with a suitably chosen initial guess* $\boldsymbol{c}_0$*.*

Iterate *for* $l = 0, 1, 2, \ldots,$ *while* $\mathcal{C}(\boldsymbol{c}^{(l)}) \neq$ TRUE*:*
1. evaluate G_c *according to* (2.17) *with* $\hat{c} = \hat{c}^{(l)} = \sum_{j=1}^n \boldsymbol{c}_j^{(l)} \phi_j$*,*
2. calculate a Cholesky factorization $L_c^{\mathrm{T}} L_c = G_c$*,*
3. set $\boldsymbol{c}^{(l+1)}$ *to be the solution of the normal equation* (2.18)*.*

The name of the method is due to the parameter-dependent 'diffusivity coefficient' in (2.17),

$$D_c(x) = \frac{r'(|\nabla \hat{c}(x)|)}{|\nabla \hat{c}(x)|},$$

lagging one step behind at each iteration. In principle, the normal equation (2.18) can be solved in step 3 in the same way as (2.16). However, with this interpretation, in addition to the computation of the solution

itself, we need to factorize G_c at each round of the iteration, which increases the computational cost of the algorithm. The equation (2.18) can be reformulated in several ways in order to make its solution more efficient (see e.g. [11]). Our choice is to resort to a *preconditioned Krylov subspace method*, which we present in the following sections.

Conjugate gradient method

Our aim is to combine the Gauss–Newton method used in connection with nonlinear forward operators and the lagged diffusivity method designed for nonquadratic regularizers to solve a general minimization problem of form (2.14). However, in order for such a method to be tractable for large (three-dimensional) problems, we need to be able to solve the normal equations (2.16) and (2.18) efficiently.

To this end, consider for an arbitrary matrix A the system

$$A^{\mathrm{T}}A\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}, \tag{2.19}$$

or equivalently the linear least squares minimization problem

$$\arg\min_{\boldsymbol{x}} \|\boldsymbol{b} - A\boldsymbol{x}\|^2. \tag{2.20}$$

We can efficiently approximate the corresponding (minimum norm) solution with the conjugate gradient method originally presented in [50]. This is an iterative method designed specifically for solving large and sparse systems, i.e. such as the ones arising when solving elliptic PDEs. The basic version of the method, formulated as a minimization algorithm, is as follows.

Algorithm 3 (Conjugate gradient method)**.** Assume *that the to-be-minimized functional is of form* $\Phi = \|\boldsymbol{b} - A\boldsymbol{x}\|^2$ *and that the matrix* A *and the vector* $\boldsymbol{b}$ *are fixed.* Initialize *with a suitable initial guess* $\boldsymbol{x}^{(0)} = \boldsymbol{x}_0$*, set the residual* $\boldsymbol{r}^{(0)} = \boldsymbol{b} - A\boldsymbol{x}^{(0)}$ *and the direction* $\boldsymbol{d}^{(0)} = A^{\mathrm{T}}\boldsymbol{r}^{(0)}$.

Iterate *for* $l = 0, 1, 2, \ldots,$ *while* $\mathcal{C}(\boldsymbol{x}^{(l)}) \neq \texttt{TRUE}$*:*
1. set the step size $\alpha^{(l)} = \|A^{\mathrm{T}}\boldsymbol{r}^{(l)}\|^2 / \|A\boldsymbol{d}^{(l)}\|^2$*,*
2. let $\boldsymbol{x}^{(l+1)} = \boldsymbol{x}^{(l)} + \alpha^{(l)}\boldsymbol{d}^{(l)}$ *and* $\boldsymbol{r}^{(l+1)} = \boldsymbol{r}^{(l)} - \alpha^{(l)}A\boldsymbol{d}^{(l)}$*,*
3. set the direction parameter $\beta^{(l)} = \|A^{\mathrm{T}}\boldsymbol{r}^{(l+1)}\|^2 / \|A^{\mathrm{T}}\boldsymbol{r}^{(l)}\|^2$*,*
4. let $\boldsymbol{d}^{(l+1)} = A^{\mathrm{T}}\boldsymbol{r}^{(l+1)} + \beta^{(l)}\boldsymbol{d}^{(l)}$.

For the justification of the algorithm as well as modifications of the basic version, see e.g. [11, 41, 45]. Note that the search directions and step sizes

are chosen so that at the lth iteration we solve (2.20) over

$$\boldsymbol{x} \in \boldsymbol{x}_0 + \mathcal{K}_l(A^{\mathrm{T}}A; A^{\mathrm{T}}\boldsymbol{r}^{(0)}),$$

where $\mathcal{K}_l$ is the lth Krylov subspace related to (2.19) and the initial guess.

Definition 2. *The lth Krylov subspace corresponding to a matrix M and a vector $\boldsymbol{v}$, is defined as*

$$\mathcal{K}_l(M; \boldsymbol{v}) = \operatorname{span}\{\boldsymbol{v}, M\boldsymbol{v}, \ldots, M^{l-1}\boldsymbol{v}\}.$$

A defect of the method is that, if the matrix A in (2.20) is ill-conditioned, the basic version of the algorithm may become numerically unstable. For the analytically equivalent LSQR method with better numerical properties, we refer to [85, 86].

Priorconditioning

Notice that we can reformulate the normal equation (2.16) as

$$(A^{\mathrm{T}}A + aG)\boldsymbol{x} = A^{\mathrm{T}}\boldsymbol{b}, \tag{2.21}$$

with definitions $A = L_\eta J^{(l)}$, $G = \Gamma_c^{-1}$, $\boldsymbol{b} = L_\eta \left(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c}^{(l)}) + J^{(l)}(\boldsymbol{c}^{(l)} - \boldsymbol{c}_0)\right)$ and $\boldsymbol{x} = \boldsymbol{c} - \boldsymbol{c}_0$. Likewise, we can express (2.18) in form (2.21) by defining $A = L_\eta \mathcal{F}$, $G = G_c$, $\boldsymbol{b} = L_\eta \boldsymbol{f}$ and $\boldsymbol{x} = \boldsymbol{c}$. Solving (2.21) is equivalent to the minimization problem

$$\arg\min_{\boldsymbol{x}} \left(\|\boldsymbol{b} - A\boldsymbol{x}\|^2 + a\, \boldsymbol{x}^{\mathrm{T}} G \boldsymbol{x} \right), \tag{2.22}$$

where the regularization term corresponds to the prior of our choice. Instead of using a Cholesky factorization of G to rewrite the equation in form (2.19), we can take the approach presented in [18, 19, 20, 21, 22] and precondition the system with the matrix G — also known as *priorconditioning*. The *formal* idea is to use a factorization $L_c^{\mathrm{T}} L_c = G$ to transform (2.21) into

$$\left((AL_c^{-1})^{\mathrm{T}}(AL_c^{-1}) + a\mathrm{I}\right)\tilde{\boldsymbol{x}} \;=\; (AL_c^{-1})^{\mathrm{T}}\boldsymbol{b}, \qquad \tilde{\boldsymbol{x}} = L_c\boldsymbol{x}, \tag{2.23}$$

where I is the identity matrix. Note that, in regularization terms, priorconditioning corresponds to the transformation of the quadratic Tikhonov functional in (2.22) into the standard form (see e.g. [35, 47, 51]).

When the conjugate gradient method is applied to the transformed equation (2.23) and the initial guess is set to $\boldsymbol{x}_0 = 0$, the solution space for $\boldsymbol{x}$ spanned by the method after l iterations is

$$\boldsymbol{x}^{(l)} \in \mathcal{K}_l \left(G^{-1}A^{\mathrm{T}}A; G^{-1}A^{\mathrm{T}}\boldsymbol{b}\right).$$

In particular, for all l, the solution is of the form $\boldsymbol{x}^{(l)} = G^{-1}\boldsymbol{z}$ for some $\boldsymbol{z} \in \mathbb{R}^n$, and hence the prior information is built into the approximate solution starting from the very first iterates (see also [3, 21, 61]). As a result, the conjugate gradient algorithm applied to the priorconditioned problem (2.23) usually converges significantly faster than when applied to (2.19) — at least as long as the prior model is accurate. In addition, since the prior information is directly incorporated into the Krylov subspace spanning the solution, it is possible to get rid of the free regularization parameter by setting $a = 0$ in (2.23) and using an early stopping rule of the conjugate gradient algorithm for regularization [36, 45]. Finally, it is essential to notice that to solve the priorconditioned problem, the potentially costly factorization $L_c^{\mathrm{T}} L_c = G$ need not actually be carried out: a preconditioned LSQR algorithm that works directly with the inverse G^{-1} was presented in [3]. This approach can be accompanied e.g. by a multigrid method to approximate operations with the inverse when $G = G_c$ (see [17]).

To use the described method, it is necessary to have an invertible priorconditioner G, which causes slight problems when edge-enhancing priors are considered. This can be observed by interpreting $G = G_c$ defined by (2.17) as the system matrix of the FE discretized variational formulation of the elliptic partial differential operator

$$-\nabla \cdot D_c \nabla \tag{2.24}$$

accompanied by the homogeneous Neumann boundary condition. Since neither the operator nor the boundary condition 'sees' an additive constant in a target function, it is clear that G_c is not invertible. Fortunately, there are some straightforward ways to circumvent this problem. For example, G_c can be made invertible by replacing the Neumann boundary condition for (2.24) with a Dirichlet condition on some part of the boundary $O \subset \partial\Omega$ i.e. setting $c(x) = c_0$ with some constant $c_0 \in \mathbb{R}$ when $x \in O$. Invertibility can also be ensured by adding a small homogeneous absorption term to the operator i.e. using $-\nabla \cdot D_c \nabla + A$ with a small $A \in \mathbb{R}_+$ instead of (2.24). Another way to circumvent the problem is to employ suitable matrix decompositions [47] in the spirit of [56].

Combined algorithm

In [3] the lagged diffusivity iteration was combined with a preconditioned LSQR algorithm to introduce an efficient method for solving linear inverse problems with edge-enhancing priors. Here we generalize this algorithm

to nonlinear inverse problems by incorporating the linearization step from the Gauss–Newton iteration into the method. This approach was first presented in Publication I.

Consider the minimization of a functional of form (2.14) with a discretized version of (2.12) as a regularizer. After linearizing the measurement operator as in the Gauss–Newton iteration, we need to solve the equation (2.15). Now the gradient of R can be represented as in the lagged diffusivity iteration, which then gives the normal equation

$$\left(A^{\mathrm{T}}A + a\,G_c\right)\boldsymbol{c} = A^{\mathrm{T}}L_\eta \boldsymbol{y}^{(l)}, \qquad A = L_\eta J^{(l)}, \tag{2.25}$$

where, as previously, $\boldsymbol{y}^{(l)} = \boldsymbol{f} - \mathcal{F}(\boldsymbol{c}^{(l)}) + J^{(l)}\boldsymbol{c}^{(l)}$, G_c is defined by (2.17) and $L_\eta^{\mathrm{T}}L_\eta = \Gamma_\eta^{-1}$ is a Cholesky factorization. Using this formulation, we can introduce an algorithm to solve the nonlinear parameter reconstruction problem with an edge-enhancing prior.

Algorithm 4. Assume *that the to-be-minimized functional is of form* (2.14) *with* R *being a discretized version of the regularizer* (2.12)*. Let* a, Γ_η^{-1} *and the function* r *related to* R *together with its free parameter(s) be fixed.* Initialize *by calculating a Cholesky factorization* $L_\eta^{\mathrm{T}}L_\eta = \Gamma_\eta^{-1}$ *and by using Algorithm 1 (with a suitable initial guess* $c^{(0)} \in \mathbb{R}$*) to solve*

$$c_0 = \arg\min_{c\in\mathbb{R}} \frac{1}{2}(\boldsymbol{f} - \mathcal{F}(c\mathbf{1}))^{\mathrm{T}}\Gamma_\eta^{-1}(\boldsymbol{f} - \mathcal{F}(c\mathbf{1})),$$

where $\mathbf{1} = [1,\ldots,1]^{\mathrm{T}} \in \mathbb{R}^n$*. Set* $\boldsymbol{c}^{(0)} = c_0\mathbf{1}$.

Iterate *for* $l = 0, 1, 2, \ldots$, *while* $\mathcal{C}(\boldsymbol{c}^{(l)}) \neq$ TRUE*:*
1. evaluate $\mathcal{F}(\boldsymbol{c}^{(l)})$ *and* $J^{(l)} = J_c^{\mathcal{F}}(\boldsymbol{c}^{(l)})$*,*
2. build the matrix G_c *according to* (2.17) *with* $\hat{c} = \hat{c}^{(l)} = \sum_{j=1}^n \boldsymbol{c}_j^{(l)}\phi_j$*,*
3. use the preconditioned LSQR with $\boldsymbol{c}_0 = \mathbf{0}$ *to solve* $\boldsymbol{c}^{(l+1)}$ *from* (2.25)*.*

In the initialization step, only a *homogeneous estimate* of the parameter is needed, which can be approximated using the basic Gauss–Newton algorithm. Notice also that in steps 2–3 we have suppressed one level of iterations by taking only one step of the lagged diffusivity iteration per each linearization of the forward model. That is, in principle these steps could be replaced by:

2. set $\boldsymbol{e}^{(0)} = \boldsymbol{c}^{(l)}$*,*

Iterate *for* $j = 0, 1, 2, \ldots$, *while* $\mathcal{C}(\boldsymbol{e}^{(j)}) \neq$ TRUE*:*
i. build the matrix G_c *according to* (2.17) *with* $\hat{c} = \hat{e}^{(j)} = \sum_{i=1}^n \boldsymbol{e}_i^{(j)}\phi_i$*,*
ii. use the preconditioned LSQR to solve $\boldsymbol{e}^{(j+1)}$ *from* (2.25)*,*

3. set $\boldsymbol{c}^{(l+1)} = \boldsymbol{e}^{(j)}$.

However, since each step of the lagged diffusivity iteration requires the solution of a normal equation (with another iterative method), this approach would considerably increase the computational cost of the algorithm. We have also observed in practice that taking several lagged diffusivity steps per linearization does not seem to improve reconstructions.

Finally, we note that, even though not written out explicitly, the whole algorithm consists of two nested iterations, which may seem inefficient. However, using the priorconditioning approach, the number of inner LSQR iterations is typically not very high. Moreover, since the information corresponding to the solution from the previous iteration round is included in the prior matrix G_c, it is usually enough to make only a few linearizations (outer iterations) before the algorithm converges.

Stopping conditions

So far we have presented all the algorithms with a general stopping condition $\mathcal{C} : \boldsymbol{c}^{(l)} \mapsto \{\text{TRUE}, \text{FALSE}\}$. The choice of the used condition(s) is, however, a delicate issue, which is particularly important when working with inverse problems: if the iterative algorithm minimizing a cost functional of form (1.1) is run too long, the reconstruction starts to fit to noise. In consequence, we want to stop the iteration before actually reaching a critical point. A widely used approach, originally presented in [77], is to terminate the iteration, when the discrepancy in the minimized functional reaches the noise level.

Condition 1 (Morozov discrepancy principle)**.** *Assume that the to-be-minimized functional is of the form* $\Phi(\boldsymbol{c}) = \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)^{\mathrm{T}} \Gamma_\eta^{-1} \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)$*, with the data vector* $\boldsymbol{f}$ *being a noisy version of the exact measurement* $\boldsymbol{f}^*$ *satisfying* $\big(\boldsymbol{f} - \boldsymbol{f}^*\big)^{\mathrm{T}} \Gamma_\eta^{-1} \big(\boldsymbol{f} - \boldsymbol{f}^*\big) \leq \varepsilon$*. Then, with a fixed (small)* $\tau > 1$*,*

$$\mathcal{C}(\boldsymbol{c}^{(l)}) := \Big(\Phi(\boldsymbol{c}^{(l)}) \leq \tau \varepsilon \Big) .$$

Under the assumption (2.7) of an additive noise model for $\boldsymbol{f} \in \mathbb{R}^k$, we can approximate the noise level by

$$\varepsilon \approx \mathbb{E}\left[\big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big)^{\mathrm{T}} \Gamma_\eta^{-1} \big(\boldsymbol{f} - \mathcal{F}(\boldsymbol{c})\big) \right] = \mathbb{E}\left[\boldsymbol{\eta}^{\mathrm{T}} \Gamma_\eta^{-1} \boldsymbol{\eta} \right] = k.$$

The early stopping by using Condition 1 combined with the conjugate gradient iteration can in fact be considered as a regularization method [45].

When the cost functional includes a regularization term as in (2.14), it may be reasonable to let the iteration converge to a critical point even in

the context of inverse problems. In this case, the most obvious condition $\|\nabla\Phi(\boldsymbol{c}^{(l)})\| \leq \delta$ is often considered inadequate, since the measured quantity — and hence a suitable $\delta > 0$ — is strongly case-dependent [32]. A modified approach is to compare the change in the target functional to that of the parameter for example as follows.

Condition 2. *Denote the to-be-minimized functional by $\Phi(\boldsymbol{c})$. Then, with a user-specified observation length $h \in \mathbb{N}$ and a tolerance $\delta > 0$,*

$$\mathcal{C}(\boldsymbol{c}^{(l)}) := \left(\frac{|\Phi(\boldsymbol{c}^{(l-h)}) - \Phi(\boldsymbol{c}^{(l)})|}{\|\boldsymbol{c}^{(l-h)} - \boldsymbol{c}^{(l)}\|} \leq \delta \right), \qquad l > h.$$

Finally, it is possible for an iterative method to stagnate before reaching the noise level or a critical point. Therefore it may also be reasonable to consider stopping conditions that check if the reconstruction is still improving. In practice this is done by measuring the change in the parameter, which can be implemented for example as follows.

Condition 3. *Denote the to-be-minimized functional by $\Phi(\boldsymbol{c})$. Then, with a user-specified observation length $h \in \mathbb{N}$ and a tolerance $\delta > 0$,*

$$\mathcal{C}(\boldsymbol{c}^{(l)}) := \left(\|\boldsymbol{c}^{(l)} - \boldsymbol{c}^{(l-h)}\| \leq \delta \right), \qquad l > h.$$

For detailed considerations of stopping conditions in general, we refer to [32]. In addition, suitable stopping criteria are considered in the method-specific references, such as [45, 46, 62, 86, 112].

3. Tomographic imaging

In this thesis we consider the parameter reconstruction problems related to three tomographic imaging modalities, whose forward operators can be described with the help of boundary value problems for elliptic PDEs. We model the imaged body as a bounded domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, which is assumed to have a connected complement and a Lipschitz boundary $\partial\Omega$. In addition, the to-be-reconstructed parameters live in the set

$$L^\infty_+(\Omega) = L^\infty_+(\Omega, \mathbb{R}) := \{v \in L^\infty(\Omega, \mathbb{R}) \,|\, \operatorname{ess\,inf} v > 0\}\,.$$

In the following sections we describe the forward models for electrical impedance tomography, diffuse optical tomography and quantitative photoacoustic tomography.

3.1 Electrical impedance tomography

Electrical impedance tomography (EIT) is an imaging modality that employs electrical signals to recover the internal distribution of electrical properties in the imaged object. More precisely, the principal parameter of interest is the *conductivity* distribution and the measurements are performed by controlling the electrical *current* and measuring the corresponding *potential* on the object boundary. Since the electrical properties of air, different tissues and blood differ greatly [8, 13, 54], EIT can be useful in medical applications such as monitoring lung problems or heart function and blood flow, imaging brain function, and screening for breast cancer. The industrial applications include process tomography and nondestructive testing of materials. Consult the review articles [13, 24, 108] or the books [54, 78] and references therein for more information about the theoretical background, numerical implementation and possible applications of EIT.

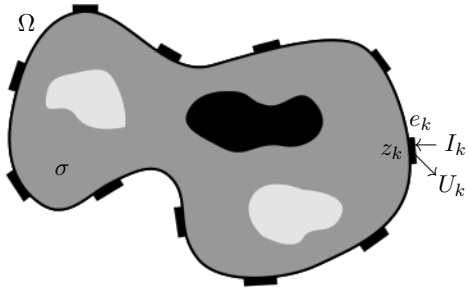

Figure 3.1. Illustration of the parameters of the CEM for EIT: on each electrode e_k we control the net current I_k and measure the potential U_k, which depends on the internal conductivity σ and the contact resistances z_k, $k = 1, \ldots, K$.

Complete electrode model

There are a number of forward models that have been considered for EIT (see e.g. [78]). Here we introduce the complete electrode model (CEM) [25, 98] which is the most accurate model for the real-life EIT measurements used in Publications I and IV. Consider the domain Ω illustrated in Figure 3.1. Attached to the boundary, there are $K \geq 2$ pairwise disjoint, open and connected electrodes, $e_k \subset \partial\Omega$, $k = 1, \ldots, K$. We control the net current I_k through each electrode and, corresponding to a particular current pattern $I = [I_k]_{k=1}^K$, we measure the constant potentials $U = [U_k]_{k=1}^K$ on the electrodes. Due to the principle of charge conservation and by choosing the ground level of potential appropriately, it holds that $I, U \in \mathbb{R}_\diamond^K := \{V \in \mathbb{R}^K | \sum_{k=1}^K V_k = 0\}$. The parameter of interest is the internal conductivity distribution $\sigma \in L_+^\infty(\Omega)$. However, to account for the imperfectly conducting layers between the electrodes and the domain, the CEM introduces contact resistances, $z = [z_k]_{k=1}^K \in \mathbb{R}_+^K$, which typically also need to be reconstructed from the EIT measurements. With a system of K electrodes, there are $K - 1$ linearly independent current patterns $I^{(1)}, \ldots, I^{(K-1)}$ for which the measurements are produced and hence the full forward operator is of form $\mathcal{F}(\sigma, z) = [U^{(1)}, \ldots, U^{(K-1)}] \in \mathbb{R}^{K \times (K-1)}$, where the dependence on the current patterns is assumed implicitly.

According to the CEM, the state of the EIT system corresponding to a current pattern $I \in \mathbb{R}_\diamond^K$ is characterized by the internal electromagnetic potential $u \in H^1(\Omega)$ together with the electrode potentials $U \in \mathbb{R}_\diamond^K$. The pair $(u, U) \in H^1(\Omega) \oplus \mathbb{R}_\diamond^K =: \mathcal{H}^1$ is the unique solution of the state equation

described by the elliptic boundary value problem

$$\begin{aligned} \nabla \cdot \sigma \nabla u &= 0 \quad && \text{in } \Omega \\ \nu \cdot \sigma \nabla u &= 0 \quad && \text{on } \partial\Omega \setminus \cup_{k=1}^{K} \overline{e}_k \\ u + z_k \nu \cdot \sigma \nabla u &= U_k \quad && \text{on } e_k, \quad k = 1, \ldots, K, \\ \int_{e_k} \nu \cdot \sigma \nabla u \, dS &= I_k, \quad && k = 1, \ldots, K, \end{aligned} \tag{3.1}$$

where $\nu : \partial\Omega \to \mathbb{R}^d$ denotes the exterior unit normal of $\partial\Omega$.

The variational formulation corresponding to (3.1) is to find $(u, U) \in \mathcal{H}^1$ that satisfies [98]

$$B\big((u,U),(v,V)\big) = \sum_{k=1}^{K} I_k V_k \quad \text{for all } (v,V) \in \mathcal{H}^1, \tag{3.2}$$

where the bounded, coercive bilinear form $B : \mathcal{H}^1 \times \mathcal{H}^1 \to \mathbb{R}$ is defined by

$$B\big((u,U),(v,V)\big) := \int_\Omega \sigma \nabla u \cdot \nabla v \, dx + \sum_{k=1}^{K} \frac{1}{z_k} \int_{e_k} (u - U_k)(v - V_k) \, dS.$$

Following the reasoning of Section 2.1 (see e.g. [110] for details about the FE implementation), after the discretization of σ and u, the (approximate) state (u, U) can be solved from (3.2) for all $K - 1$ current patterns. By applying the measurement operator $\mathcal{M} : (u, U) \mapsto U$ to all of the resulting states, we then get the measurement $\mathcal{F}(\sigma, z) = [U^{(k)}]_{k=1}^{K-1} \in \mathbb{R}^{K \times (K-1)}$.

To write down the Fréchet derivative of the forward operator, consider first the derivative corresponding to the solution of the state equation (3.1). For a fixed current pattern I, the solution itself is given by the operator

$$L_+^\infty(\Omega) \times \mathbb{R}_+^K \ni (\sigma, z) \mapsto (u, U) \in \mathcal{H}^1$$

and its Fréchet derivative at $(\sigma, z) \in L_+^\infty(\Omega) \times \mathbb{R}_+^K$ is given by the linear map

$$L^\infty(\Omega) \times \mathbb{R}^K \ni (\theta, \xi) \mapsto (u', U') \in \mathcal{H}^1,$$

where $(u', U') = \big((u', U')(\sigma, z)\big)(\theta, \xi)$ is the unique solution of the variational problem [60, 68, 111]

$$B\left((u',U'),(v,V)\right) = -\int_\Omega \theta \nabla u \cdot \nabla v \, dx + \sum_{k=1}^{K} \frac{\xi_k}{z_k^2} \int_{e_k} (u - U_k)(v - V_k) \, dS$$

for all $(v, V) \in \mathcal{H}^1$. Combining this with the measurement operator then gives the derivative of the forward operator

$$\mathcal{F}'(\sigma, z) \, : \, (\theta, \xi) \mapsto [U'^{(k)}]_{k=1}^{K-1}.$$

In Publication IV we also need the Fréchet derivatives with respect to the electrode sizes and locations. Their computation is based on [29, 30]; the details are given in Section 2.2 of Publication IV.

3.2 Diffuse optical tomography

In optical tomography (OT) the imaged object is illuminated with near-infrared (NIR) light and the aim is to determine its internal optical properties. To be more precise, typically there are two parameters of interest: the *absorption* and the (reduced) *scattering* coefficient. The measurement is the flux of photons coming out of the object corresponding to a specific illumination pattern. The most important potential applications of OT, which include detecting breast cancer and imaging the brain, are based on the different absorption spectra of some physiologically interesting molecules such as hemoglobin and melanin (see e.g. [2, 92]). For details about the theory of light propagation in tissue, medical applications and instrumentation for OT, we refer to the review articles [2, 4, 12, 40, 48, 81] and references therein.

Diffusion approximation

The widely accepted model for light propagation in tissue is the radiative transfer equation (RTE), see e.g. [60, 104] for details and implementation. However, because the RTE is computationally rather expensive, in practical applications it is commonly approximated by simpler models. Here we focus on the diffusion approximation (DA) of the RTE [2, 49], which is the model considered in Publications II and III. The formulation of OT with the DA as the forward model is also known as *diffuse* optical tomography (DOT). In a strongly scattering medium, such as most tissues, the DA is a fairly accurate forward model. As a result, DOT is the typical imaging approach in medical applications [2]. Notice, however, that the DA has its limitations. In particular, it is somewhat inaccurate close to light sources and boundaries (see e.g. [69] and references therein).

We present the DA in the frequency domain denoting the fixed harmonic frequency, with which the illumination is modulated, by $\omega \in \mathbb{R}$. In practice the illuminations are usually time-modulated [48], but any such measurement can equivalently be expressed in the frequency domain (see e.g. [2, 4]). Consider the domain Ω illustrated in Figure 3.2. There are two types of optical fibers attached to the boundary: photon sources $s_k \subset \partial\Omega$, $k = 1, \ldots, K$, and measurement sensors $m_l \subset \partial\Omega$, $l = 1, \ldots, L$. The sources and sensors are modeled as pairwise disjoint, open and connected subsets of the boundary. The object is illuminated through each source s_k, one at a time, and the corresponding outward *photon flux* $Y_{lk} \in \mathbb{C}$

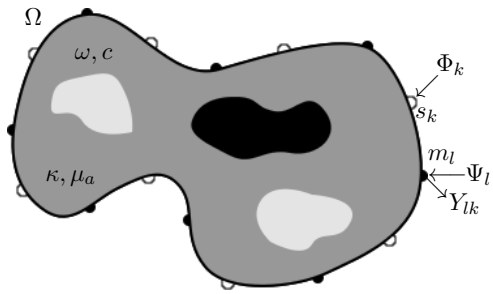

Figure 3.2. Illustration of the parameters of the DA of the RTE for OT: we control the input amplitude Φ_k on each photon source s_k and account for the 'device function' Ψ_l of each measurement sensor m_l to measure the outward photon flux Y_{lk}, which depends on the diffusivity κ and the absorption μ_a as well as on the (known) harmonic frequency ω and the speed of light c in Ω.

is measured at each sensor m_l resulting altogether in the measurement $Y = [Y_{lk}]_{l,k=1}^{L,K} \in \mathbb{C}^{L\times K}$. The two internal parameters in the DA are the *diffusion* and absorption coefficients, which are here assumed to be isotropic and denoted by $\kappa, \mu_a \in L^\infty_+(\Omega, \mathbb{R})$, respectively. The scattering coefficient μ_s is related to these parameters through

$$\kappa = \frac{1}{3(\mu_a + \mu_s')},$$

where $\mu_s' := (1-g)\mu_s$ is the reduced scattering coefficient, and $g \in (0,1)$ is a free parameter accounting for the fact that scattering in biological tissues is in reality anisotropic. Hence, the forward model is of form $\mathcal{F}(\kappa, \mu_a) = Y \in \mathbb{C}^{L\times K}$, where the dependence on the frequency ω is implicit. Obviously, instead of the diffusivity κ, we could as well choose to use μ_s' or μ_s as the model parameter accompanying μ_a.

The illuminations through the sources $s_1, \dots, s_K$ are characterized by the input amplitudes $\Phi^{(1)}, \dots, \Phi^{(K)} \in L^2(\partial\Omega, \mathbb{R})$, which are modeled by suitable functions supported on the corresponding sources. The state of the DOT system corresponding to an input amplitude (or inward photon flux) Φ is the internal *photon density* (or *fluence*) $\varphi \in H^1(\Omega) = H^1(\Omega, \mathbb{C})$. Ignoring the boundary reflections for simplicity, φ is the unique solution of the elliptic boundary value problem [42]

$$\begin{aligned} -\nabla\cdot\kappa\nabla\varphi + \Big(\mu_a + \frac{\omega}{c}\mathrm{i}\Big)\varphi = 0 \quad &\text{in } \Omega, \\ \gamma_d\varphi + \frac{1}{2}\nu\cdot\kappa\nabla\varphi = \Phi \quad &\text{on } \partial\Omega, \end{aligned} \tag{3.3}$$

where $\mathrm{i} = \sqrt{-1}$, c is the constant speed of light in Ω, and γ_d is a dimension-dependent constant with $\gamma_2 = 1/\pi$ and $\gamma_3 = 1/4$. The variational formula-

tion corresponding to (3.3) is to find $\varphi \in H^1(\Omega)$ that satisfies

$$B(\varphi, v) = 2\int_{\partial\Omega} \Phi\bar{v}\, dS \quad \text{for all } v \in H^1(\Omega). \tag{3.4}$$

Here the bounded and coercive sesquilinear form $B : H^1(\Omega) \times H^1(\Omega) \to \mathbb{C}$ is given by

$$B(\varphi, v) := \int_\Omega \left(\kappa\nabla\varphi\cdot\nabla\bar{v} + \left(\mu_a + \frac{\omega}{c}\mathrm{i}\right)\varphi\bar{v}\right) dx + 2\gamma_d\int_{\partial\Omega}\varphi\bar{v}\, dS.$$

After discretizing κ, μ_a and $\Phi^{(k)}$ for all $k = 1, \ldots, K$, the FEM approximations of $\varphi^{(1)}, \ldots, \varphi^{(K)}$ can be solved from (3.4) as described in Section 2.1. In the framework of the DA, the boundary measurement at the lth sensor corresponding to the state $\varphi^{(k)}$ can be written as [49]

$$Y_{lk} \;=\; \int_{\partial\Omega} \Psi^{(l)}\left(\gamma_d\varphi^{(k)} - \frac{1}{2}\nu\cdot\kappa\nabla\varphi^{(k)}\right) dS \;=\; 2\gamma_d\int_{\partial\Omega}\Psi^{(l)}\varphi^{(k)}\, dS, \tag{3.5}$$

where $\Psi^{(l)} \in L^2(\partial\Omega, \mathbb{R})$ is a suitable 'device function' supported on m_l. The second equality in (3.5) follows from the boundary condition in (3.3) since the supports of $\Phi = \Phi^{(k)}$ and $\Psi^{(l)}$ are disjoint. Applying the measurement operator $\mathcal{M}(\varphi^{(k)}) = [Y_{lk}]_{l=1}^L \in \mathbb{C}^L$ to all K states then gives the full measurement $\mathcal{F}(\kappa, \mu_a) = [\mathcal{M}(\varphi^{(k)})]_{k=1}^K = Y \in \mathbb{C}^{L\times K}$.

Consider the operator corresponding to solving the state equation (3.3) with the input amplitude Φ,

$$[L^\infty_+(\Omega)]^2 \ni (\kappa, \mu_a) \mapsto \varphi \in H^1(\Omega).$$

Its Fréchet derivative at $(\kappa, \mu_a) \in [L^\infty_+(\Omega)]^2$ is given by the linear map

$$[L^\infty(\Omega)]^2 \ni (\theta, \xi) \mapsto \varphi' \in H^1(\Omega),$$

where $\varphi' = \big(\varphi'(\kappa, \mu_a)\big)(\theta, \xi)$ is the unique solution of the variational problem (see e.g. [33])

$$B(\varphi', v) = -\int_\Omega \left(\theta\nabla\varphi\cdot\nabla\bar{v} + \xi\varphi\bar{v}\right) dx \quad \text{for all } v \in H^1(\Omega). \tag{3.6}$$

The derivative of the forward operator is then given by

$$\mathcal{F}'(\kappa, \mu_a) : (\theta, \xi) \mapsto [Y'_{lk}]_{l,k=1}^{L,K},$$

where

$$\big(Y'_{lk}(\kappa, \mu_a)\big)(\theta, \xi) \;=\; 2\gamma_d\int_{\partial\Omega}\Psi^{(l)}\varphi'^{(k)}\, dS. \tag{3.7}$$

Instead of a direct computation based on (3.7), the Fréchet derivative can also be efficiently sampled (see e.g. [2]) using the 'dual problem'. The

variational formulation of this problem is to find the unique $\psi^{(l)} \in H^1(\Omega)$ for which

$$B(\psi^{(l)}, v) = 2\int_{\partial\Omega} \Psi^{(l)}\bar{v}\, dS \quad \text{for all } v \in H^1(\Omega). \tag{3.8}$$

Using (3.8) and (3.6), we may then write (3.7) as

$$\big(Y'_{lk}(\kappa, \mu_a)\big)(\theta, \xi) \;=\; -\gamma_d \int_\Omega \Big(\theta \nabla\varphi^{(k)} \cdot \nabla\psi^{(l)} + \xi\varphi^{(k)}\psi^{(l)}\Big)\, dx.$$

3.3 Quantitative photoacoustic tomography

Photoacoustic tomography (PAT) is a hybrid imaging modality for which the imaging signal comes from an optical source, but the measurements are based on acoustic waves. The aim is, as in OT, to reconstruct the optical properties of the imaged target. More precisely, the examined object is illuminated by an external light pulse (using e.g. NIR or visible light), which propagates inside the target. When the light is absorbed, the released energy generates local heating, which causes an increase of pressure in the material. Assuming that the underlying medium is elastic in nature (such as most tissues), the pressure relaxation acts as a source for acoustic signals that propagate through the material as ultrasonic waves. The receivers at the object boundary then record these waves as functions of time. The images of PAT are based on the *initial pressure distribution*, which can be formed as in conventional (pulse-echo) ultrasound imaging. The modality aims to combine the desirable qualities of OT and ultrasound imaging to create a high resolution image with good contrast. The main potential applications are biomedical; PAT has been used for example to visualize human blood vessels and to detect tumors as well as to create whole-body images of small animals. For further reading about the underlying physics, practical implementation and imaging capabilities of PAT, we recommend [10, 71, 113, 114] and references therein.

The initial pressure distribution provided by PAT describes the qualitative behavior of internal optical parameters, but does not reveal their actual values. This information, however, is essential e.g. for accurate functional and molecular imaging, where the absolute concentrations of light-absorbing molecules are required. In *quantitative* photoacoustic tomography (QPAT) the aim is to use the image of PAT to go a step further and estimate the absolute distribution of optical parameters. Hence, it corresponds to solving two inverse problems: the inverse initial value problem

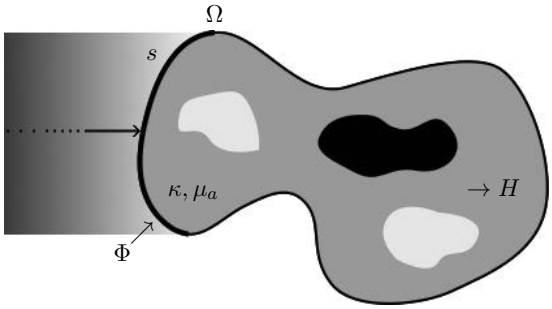

Figure 3.3. Illustration of the parameters of the DA of the RTE for QPAT: we control the input photon flux Φ on s and get the absorbed energy density H as an 'internal measurement' that depends on the diffusivity and absorption distributions κ, μ_a in Ω.

in acoustics to reconstruct the initial pressure distribution from the ultrasound measurements, and the inverse parameter estimation problem to reconstruct the optical parameters from the solution of the first problem. In addition to the ill-posed inversion, QPAT provides numerical challenges, since the data for the latter inverse problem are high-resolution three-dimensional images of considerable size, and the resolution is expected to be carried over to the reconstruction. For more information about QPAT see for example [6, 27, 39, 79, 103].

Optical forward problem

As in OT, the light propagation in QPAT can be modeled by the RTE (see e.g. [44, 75, 90, 93, 102]) or by its approximations. Here we consider the DA, but this time without harmonic modulation i.e. with $\omega = 0$. The set-up is illustrated in Figure 3.3. The domain Ω is illuminated by a light pulse (or possibly several) that propagates into the domain through a part of the boundary $s \subset \partial\Omega$. The output (or 'measurement') of the model is the corresponding *absorbed energy density* $H \in L^2(\Omega, \mathbb{R})$, which represents the amount of energy generated when the light is absorbed inside the domain. If the to-be-reconstructed optical parameters are again the diffusion and absorption coefficients, $\kappa, \mu_a \in L^\infty_+(\Omega, \mathbb{R})$, the optical forward operator is given by $\mathcal{F}(\kappa, \mu_a) = H \in L^2(\Omega, \mathbb{R})$.

The light pulse is characterized by the input amplitude $\Phi \in L^2(\partial\Omega, \mathbb{R})$ supported on s and the corresponding state of the optical forward problem is given by the photon density $\varphi \in H^1(\Omega, \mathbb{R})$, which is the unique solution of (3.3) (with $\omega = 0$). Given the state, the measurement operator produces the absorbed energy density via $\mathcal{M}(\varphi) = \mu_a\varphi = H \in L^2(\Omega, \mathbb{R})$. After the

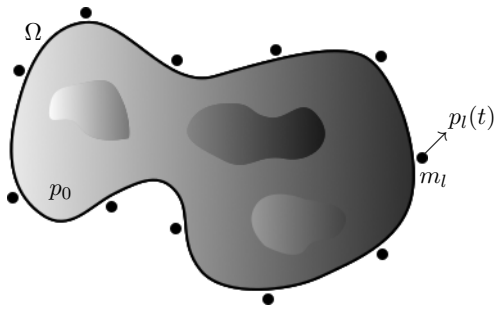

Figure 3.4. Illustration of the parameters of the acoustic forward model in QPAT: the initial pressure distribution p_0 propagates through the domain and the resulting pressure $p_l(t)$ is measured at the sensors m_l for $t \in [0, T]$.

discretization of κ, μ_a and Φ, the discretized $\boldsymbol{\varphi} \in \mathbb{R}^n$ can be solved using the variational formulation (3.4). The approximate model for the measurement is then $\boldsymbol{H} = [(\boldsymbol{\mu}_a)_j \boldsymbol{\varphi}_j]_{j=1}^n \in \mathbb{R}^n$.

The Fréchet derivative of the forward operator $\mathcal{F}$ at $(\kappa, \mu_a) \in [L_+^\infty(\Omega)]^2$ to the direction $(\theta, \xi) \in [L^\infty(\Omega)]^2$ is given by

$$\mathcal{F}'(\kappa, \mu_a) : (\theta, \xi) \mapsto \mu_a \varphi' + \xi\varphi \in L^2(\Omega),$$

where $\varphi' = \big(\varphi'(\kappa, \mu_a)\big)(\theta, \xi) \in H^1(\Omega)$ is the unique solution of (3.6).

Acoustic forward problem

The absorbed energy density H generates an initial acoustic pressure distribution $p_0 = \hat{\Gamma} H \in L^2(\Omega)$, where the coefficient $\hat{\Gamma} \in L_+^\infty(\Omega)$ represents the *photoacoustic efficiency* (Grüneisen parameter for an absorbing fluid) i.e. the efficiency of the conversion of heat to pressure. Starting from the given initial value $p(x, 0) = p_0(x)$, the pressure distribution $p(x, t)$ describes how the acoustic wave propagates through the space. The pressure $p_l(t)$ is then recorded for $t \in [0, T]$ at a number of sensors m_l, $l = 1, \ldots, L$, around the domain Ω. The propagation of the pressure wave is described by a wave equation [28], whose detailed description is beyond the scope of this thesis, but the measurement configuration is sketched in Figure 3.4. The corresponding forward model is of form $\mathcal{F}(p_0) = [p_l(t)]_{l=1}^L = P(t)$, $t \in [0, T]$. The inverse problem of estimating p_0 based on the measurements $P(t)$ is considered for example in [55, 66, 67, 107, 113, 114]. See also [34, 38, 44, 88, 97, 99] for approaches where the two inverse problems are combined to estimate the optical parameters directly from $P(t)$. Furthermore, e.g. [1, 7, 79, 89, 96] provide discussion about estimation of the Grüneisen parameter together with the optical properties.

4. Summary of results

I: Edge-enhancing reconstruction algorithm for three-dimensional electrical impedance tomography

Consider the nonlinear parameter estimation problem of EIT described in Section 3.1 and assume that the to-be-reconstructed conductivity distribution has an approximately constant background value in which there are distinct inclusions. In Publication I this problem is tackled by assuming an edge-enhancing prior and utilizing Algorithm 4 to compute the MAP estimate for the conductivity. In particular, the algorithm is modified to include also the estimation of contact resistances, which are given an *uninformative* prior, i.e. all realizations are considered equally likely. In fact, the equation corresponding to (2.25) can then be restructured to separate the reconstruction of the conductivity from that of the contact resistances, see Section 4.2 of Publication I for details.

The functionality of the method is tested with three-dimensional numerical experiments using simulated data in two different geometrical settings and real-life measurements from a water tank. Two different priors, the total variation and the Perona–Malik, are employed for comparison. Despite a slight imperfection in the reconstruction of the contact resistances (see Figure 1 in Publication I), the results demonstrate that the method produces conductivity reconstructions that are well in accordance with the prior information. Furthermore, the reconstructions are produced in just a few minutes on dense unstructured FE meshes (around 10 000 – 30 000 nodes) and with relatively large data vectors (240 – 2 256 entries).

II: Edge-promoting reconstruction of absorption and diffusivity in optical tomography

In Publication II we consider the application of Algorithm 4 to the setting of DOT (see Section 3.2), where two internal parameter distributions are to be estimated from complex-valued measurements. In this case it proves to be useful to consider the (componentwise) transformation $\tilde{\boldsymbol{c}} = \log(\boldsymbol{c}/c_0)$ for the discretized parameters of interest, where c_0 is the homogeneous estimate from the initialization phase of Algorithm 4. The transformation ensures the positivity and 'normalizes' the parameters, producing better results when considering simultaneous reconstruction (see e.g. [95, 104] for similar approaches). The transformed versions of both parameters are given edge-enhancing priors of form (2.8) with different free parameters $a, b > 0$. Then, assuming that the to-be-reconstructed parameters are independent, their MAP estimate corresponds to the solution of the equation (2.25) with a block diagonal matrix G_c (see Section 4.1 of Publication II). The two blocks in G_c correspond to the priors given for the two parameters of interest. The relative strength of the priors is controlled by the ratio b/a, introducing an additional free parameter into the algorithm.

In the simulated numerical experiments of Publication II, we observe that if one of the parameters is assumed to be known, the reconstruction of the other is comparable to those presented in Publication I. This observation is based on reconstructions corresponding to unmodulated measurements (i.e. $\omega = 0$ in (3.3)). However, as demonstrated in the other examples, to successfully reconstruct the two parameters simultaneously, frequency modulated measurements are required to prevent 'cross-talk' in the reconstructions (see [94]). These examples also demonstrate the computational capabilities of the algorithm when the number of the parameter values or the size of the data vector increases. The reconstructions on dense three-dimensional FE meshes (around 50 000 nodes) with a data vector of fixed length (496 entries) were performed in about five minutes for both the separate (one parameter) and the simultaneous case (two parameters). However, when the length of the data vector was doubled (to 992 entries), the computation took also approximately twice the time.

III: Efficient inclusion of total variation type priors in quantitative photoacoustic tomography

As explained in Section 3.3, the optical inverse problem of QPAT provides a computationally challenging setting. Both the measurements and the parameters of interest are inherently infinite-dimensional and thus require high-dimensional discretizations. Therefore any practical method for solving this inverse problem must perform efficiently for systems of considerable size. In Publication III we propose to solve the parameter estimation problem of QPAT by assuming edge-enhancing priors for the (transformed) optical parameters and applying Algorithm 4 together with a matrix-free approach for calculating the needed Jacobians (see the Appendix of Publication III). The reconstruction algorithm is a modification of the one presented in Publication II. In particular, the use of homogeneous estimates of the parameters for initialization leads in this case to a slow convergence in the first iteration round. However, utilizing the specific structure of the measurement of QPAT, we are able to significantly improve the initialization phase of the algorithm (details are given in Section 5 of Publication III when describing the first numerical experiment).

The two simulated numerical experiments demonstrate that when two or more (well positioned) illuminations are used, we are able to produce accurate reconstructions of both optical parameters in elaborate targets. Moreover, the matrix-free approximation of the Jacobians reduces the computational cost so that we are able to produce reconstructions with huge data vectors (about 100 000 – 200 000 entries) on dense unstructured FE meshes with computational times comparable to those reported in Publication II.

IV: Compensation for geometric modeling errors by electrode movement in electrical impedance tomography

In Publication IV we consider the modeling errors resulting from an inaccurately known domain shape in the context of EIT (see Section 3.1). It is well known that mismodeling of the geometric measurement setting can easily ruin the conductivity reconstruction if not taken into account [9, 15, 65]. Instead of the full reconstruction of the domain boundary (see [30, 31] for this approach), we propose that the incorrect shape of the reconstruction domain can be compensated by letting the electrode loca-

tions and sizes alter. To theoretically support this claim, we prove that in two dimensions the CEM is approximately conformally invariant.

In the numerical studies a basic adaptation of the Gauss–Newton iteration (Algorithm 3) is employed with Gaussian priors for all the parameters (the conductivity and contact resistances as well as the electrode positions and shapes). We demonstrate that in two dimensions the approach functions as expected with both simulated and real-life data. However, we also observe that a similar technique does not, in general, give good results in three dimensions.

References

[1] G. S. Alberti and H. Ammari. Disjoint sparsity for signal separation and applications to hybrid inverse problems in medical imaging. *Applied and Computational Harmonic Analysis*, 2015. In Press.

[2] S. R. Arridge. Optical tomography in medical imaging. *Inverse Problems*, 15(2):R41–R93, 1999.

[3] S. R. Arridge, M. M. Betcke, and L. Harhanen. Iterated preconditioned LSQR method for inverse problems on unstructured grids. *Inverse Problems*, 30(7):075009, 2014.

[4] S. R. Arridge and J. C. Hebden. Optical imaging in medicine: II. Modelling and reconstruction. *Physics in Medicine and Biology*, 42(5):841–853, 1997.

[5] S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen. Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems*, 22(1):175–195, 2006.

[6] G. Bal and K. Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Problems*, 27(7):075003, 2011.

[7] G. Bal and K. Ren. On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inverse Problems*, 28(2):025010, 2012.

[8] D. C. Barber and B. H. Brown. Applied potential tomography. *Journal of Physics E: Scientific Instruments*, 17(9):723–733, 1984.

[9] D. C. Barber and B. H. Brown. Errors in reconstruction of resistivity images using a linear reconstruction technique. *Clinical Physics and Physiological Measurement*, 9(A):101–104, 1988.

[10] P. Beard. Biomedical photoacoustic imaging. *Interface Focus*, 1(4):602–631, 2011.

[11] Å. Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[12] 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(6):57–75, 2001.

[13] L. Borcea. Electrical impedance tomography. *Inverse Problems*, 18(6):R99–R136, 2002.

[14] A. Boyle, A. Adler, and W. R. B. Lionheart. Shape deformation in two-dimensional electrical impedance tomography. *IEEE Transactions on Medical Imaging*, 31(12):2185–2193, 2012.

[15] W. R. Breckon and M. K. Pidcock. Data errors and reconstruction algorithms in electrical impedance tomography. *Clinical Physics and Physiological Measurement*, 9(A):105–109, 1988.

[16] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*. Springer, 3rd edition, 2008.

[17] W. L. Briggs and S. F. Henson, V. E. McCormick. *A Multigrid Tutorial*. SIAM, 2nd edition, 2000.

[18] D. Calvetti. Preconditioned iterative methods for linear discrete ill-posed problems from a Bayesian inversion perspective. *Journal of Computational and Applied Mathematics*, 198(2):378–395, 2007.

[19] D. Calvetti, D. McGivney, and E. Somersalo. Left and right preconditioning for electrical impedance tomography with structural information. *Inverse Problems*, 28(5):055015, 2012.

[20] D. Calvetti and E. Somersalo. Priorconditioners for linear systems. *Inverse Problems*, 21(4):1397–1418, 2005.

[21] D. Calvetti and E. Somersalo. *Introduction to Bayesian Scientific Computing: Ten Lectures on Subjective Computing*. Springer, 2007.

[22] D. Calvetti and E. Somersalo. Hypermodels in the Bayesian imaging framework. *Inverse Problems*, 24(3):034013, 2008.

[23] G. Chavent. *Nonlinear Least Squares for Inverse Problems: Theoretical Foundations and Step-by-Step Guide for Applications*. Springer, 2010.

[24] M. Cheney, D. Isaacson, and J. C. Newell. Electrical impedance tomography. *SIAM Review*, 41(1):85–101, 1999.

[25] K.-S. Cheng, D. Isaacson, J. C. Newell, and D. G. Gisser. Electrode models for electric current computed tomography. *IEEE Transactions on Biomedical Engineering*, 36(9):918–924, 1989.

[26] P. G. Ciarlet. *The Finite Element Method for Elliptic Problems*. North–Holland, 1978.

[27] B. Cox, J. G. Laufer, S. R. Arridge, and P. C. Beard. Quantitative spectroscopic photoacoustic imaging: a review. *Journal of Biomedical Optics*, 17(6):061202, 2012.

[28] B. T. Cox and P. C. Beard. Fast calculation of pulsed photoacoustic fields in fluids using k-space methods. *The Journal of the Acoustical Society of America*, 117(6):3616–3627, 2005.

[29] J. Dardé, H. Hakula, N. Hyvönen, S. Staboulis, and E. Somersalo. Fine-tuning electrode information in electrical impedance tomography. *Inverse Problems and Imaging*, 6(3):399–421, 2012.

[30] J. Dardé, N. Hyvönen, A. Seppänen, and S. Staboulis. Simultaneous reconstruction of outer boundary shape and admittivity distribution in electrical impedance tomography. *SIAM Journal on Imaging Sciences*, 6(1):176–198, 2013.

[31] J. Dardé, N. Hyvönen, A. Seppänen, and S. Staboulis. Simultaneous recovery of admittivity and body shape in electrical impedance tomography: an experimental evaluation. *Inverse Problems*, 29(8):085004, 2013.

[32] J. E. Dennis, Jr and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice–Hall, Inc., 1983.

[33] T. Dierkes, O. Dorn, F. Natterer, V. Palamodov, and H. Sielschott. Fréchet derivatives for some bilinear inverse problems. *SIAM Journal on Applied Mathematics*, 62(6):2092–2113, 2002.

[34] T. Ding, K. Ren, and S. Vallélian. A one-step reconstruction algorithm for quantitative photoacoustic imaging. *Inverse Problems*, 31(9):095005, 2015.

[35] L. Eldén. A weighted pseudoinverse, generalized singular values, and constrained least squares problems. *BIT Numerical Mathematics*, 22(4):487–502, 1982.

[36] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, 1996.

[37] R. Gagliardi and B. L. McClennan, editors. *A History of the Radiological Sciences: Diagnosis*. Radiology Centennial, Inc., 1996.

[38] H. Gao, J. Feng, and L. Song. Limited-view multi-source quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065004, 2015.

[39] H. Gao, S. Osher, and H. Zhao. *Mathematical Modeling in Biomedical Imaging II*, chapter Quantitative Photoacoustic Tomography, pages 131–158. Springer, 2012.

[40] A. P. Gibson, J. C. Hebden, and S. R. Arridge. Recent advances in diffuse optical imaging. *Physics in Medicine and Biology*, 50(4):R1–R43, 2005.

[41] G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, 3rd edition, 1996.

[42] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Pitman, 1985.

[43] J. Hadamard. *Lectures on the Cauchy Problem in Linear Partial Differential Equations*. Yale University Press, 1923.

[44] M. Haltmeier, L. Neumann, and S. Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 2015.

[45] M. Hanke. *Conjugate gradient type methods for ill-posed problems*. Longman Scientific & Technical, 1995.

[46] M. Hanke. Regularizing properties of a truncated Newton-CG algorithm for nonlinear inverse problems. *Numerical Functional Analysis and Optimization*, 18(9–10):971–993, 1997.

[47] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, 1998.

[48] J. C. Hebden, S. R. Arridge, and D. T. Delpy. Optical imaging in medicine: I. Experimental techniques. *Physics in Medicine and Biology*, 42(5):825–840, 1997.

[49] J. Heino and E. Somersalo. Estimation of optical absorption in anisotropic background. *Inverse Problems*, 18(3):559–573, 2002.

[50] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, 49(6):409–436, 1952.

[51] J. W. Hilgers. On the equivalence of regularization and certain reproducing kernel Hilbert space approaches for solving first kind problems. *SIAM Journal on Numerical Analysis*, 13(2):172–184, 1976.

[52] M. Hintermüller and T. Wu. Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization-based superlinearly convergent solver. *SIAM Journal on Imaging Sciences*, 6:1385–1415, 2013.

[53] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms I: Fundamentals*. Springer–Verlag, 1996.

[54] D. S. Holder, editor. *Electrical impedance tomography: Methods, History and Applications*. IOP Publishing, 2005.

[55] Y. Hristova, P. Kuchment, and L. Nguyen. Reconstruction and time reversal in thermoacoustic tomography in acoustically homogeneous and inhomogeneous media. *Inverse Problems*, 24(5):055006, 2008.

[56] M. Jacobsen, P. C. Hansen, and M. A. Saunders. Subspace preconditioned LSQR for discrete ill-posed problems. *BIT Numerical Mathematics*, 43(5):975–989, 2003.

[57] M. Jehl, J. Avery, E. Malone, D. Holder, and T. Betcke. Correcting electrode modelling errors in EIT on realistic 3D head models. *Physiological Measurement*, 36(12):2423–2442, 2015.

[58] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2005.

[59] J. Kaipio and E. Somersalo. Statistical inverse problems: discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504, 2007.

[60] J. P. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen. Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems*, 16(5):1487–1522, 2000.

[61] M. E. Kilmer, P. C. Hansen, and M. I. Espanol. A projection-based approach to general-form Tikhonov regularization. *SIAM Journal on Scientific Computing*, 29(1):315–330, 2007.

[62] A. Kirsch. *An Introduction to the Mathematical Theory of Inverse Problems*. Springer, 2nd edition, 2011.

[63] V. Kolehmainen, M. Lassas, and P. Ola. The inverse conductivity problem with an imperfectly known boundary. *SIAM Journal on Applied Mathematics*, 66(2):365–383, 2005.

[64] V. Kolehmainen, M. Lassas, and P. Ola. The inverse conductivity problem with an imperfectly known boundary in three dimensions. *SIAM Journal on Applied Mathematics*, 67(5):1440–1452, 2007.

[65] V. Kolehmainen, M. Vauhkonen, P. A. Karjalainen, and J. P. Kaipio. Assessment of errors in static electrical impedance tomography with adjacent and trigonometric current patterns. *Physiological Measurement*, 18(4):289–303, 1997.

[66] R. Kowar and O. Scherzer. Photoacoustic imaging taking into account attenuation. *arXiv:1009.4350*, 2010.

[67] P. Kuchment and L. Kunyansky. Mathematics of thermoacoustic tomography. *European Journal of Applied Mathematics*, 19(02):191–224, 2008.

[68] A. Lechleiter and A. Rieder. Newton regularizations for impedance tomography: a numerical study. *Inverse Problems*, 22(6):1967–1987, 2006.

[69] O. Lehtikangas, T. Tarvainen, and A. D. Kim. Modeling boundary measurements of scattered light using the corrected diffusion approximation. *Biomedical Optics Express*, 3(3):552–571, 2012.

[70] M. S. Lehtinen, L. Päivärinta, and E. Somersalo. Linear inverse problems for generalised random variables. *Inverse Problems*, 5(4):599, 1989.

[71] C. Li and L. V. Wang. Photoacoustic tomography and sensing in biomedicine. *Physics in Medicine and Biology*, 54(19):R59–R97, 2009.

[72] C. Lieberman, K. Willcox, and O. Ghattas. Parameter and state model reduction for large-scale statistical inverse problems. *SIAM Journal on Scientific Computing*, 32(5):2523–2542, 2010.

[73] D. Liu, V. Kolehmainen, S. Siltanen, A.-M. Laukkanen, and A. Seppänen. Estimation of conductivity changes in a region of interest with electrical impedance tomography. *Inverse Problems and Imaging*, 9(1):211–229, 2015.

[74] D. Liu, V. Kolehmainen, S. Siltanen, and A. Seppänen. A nonlinear approach to difference imaging in EIT; assessment of the robustness in the presence of modelling errors. *Inverse Problems*, 31(3):035012, 2015.

[75] A. V. Mamonov and K. Ren. Quantitative photoacoustic imaging in the radiative transport regime. *Communications in Mathematical Sciences*, 12(2):201–234, 2012.

[76] A. Mandelbaum. Linear estimators and measurable linear transformations on a Hilbert space. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 65(3):385–397, 1984.

[77] V. A. Morozov. On the solution of functional equations by the method of regularization. In *Soviet Mathematics Doklady*, volume 7, pages 414–417, 1966.

[78] J. L. Müller and S. Siltanen. *Linear and nonlinear inverse problems with practical applications*. SIAM, 2012.

[79] W. Naetar and O. Scherzer. Quantitative photoacoustic tomography with piecewise constant material parameters. *SIAM Journal on Imaging Sciences*, 7(3):1755–1774, 2014.

[80] Y. Nesterov. *Introductory Lectures on Convex Optimization: A Basic Course*. Springer, 2004.

[81] I. Nissilä, T. Noponen, K. Kotilahti, T. Katila, L. Lipiäinen, T. Tarvainen, M. Schweiger, and S. Arridge. Instrumentation and calibration methods for the multichannel measurement of phase and amplitude in optical tomography. *Review of Scientific Instruments*, 76(4):044302, 2005.

[82] A. Nissinen, L. M. Heikkinen, and J. P. Kaipio. The Bayesian approximation error approach for electrical impedance tomography — experimental results. *Measurement Science and Technology*, 19(1):015501, 2008.

[83] A. Nissinen, V. Kolehmainen, and J. P. Kaipio. Reconstruction of domain boundary and conductivity in electrical impedance tomography using the approximation error approach. *International Journal for Uncertainty Quantification*, 1(3):203–222, 2011.

[84] A. Nissinen, V. P. Kolehmainen, and J. P. Kaipio. Compensation of modelling errors due to unknown domain boundary in electrical impedance tomography. *IEEE Transactions on Medical Imaging*, 30(2):231–242, 2011.

[85] C. C. Paige and M. A. Saunders. Algorithm 583: LSQR: Sparse linear equations and least squares problems. *ACM Transactions on Mathematical Software*, 8(2):195–209, 1982.

[86] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, 8(1):43–71, 1982.

[87] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.

[88] A. Pulkkinen, B. T. Cox, S. R. Arridge, H. Goh, J. P. Kaipio, and T. Tarvainen. Direct estimation of optical parameters from photoacoustic time series in quantitative photoacoustic tomography. Submitted, 2016.

[89] A. Pulkkinen, B. T. Cox, S. R. Arridge, J. P. Kaipio, and T. Tarvainen. A Bayesian approach to spectral quantitative photoacoustic tomography. *Inverse Problems*, 30(6):065012, 2014.

[90] A. Pulkkinen, B. T. Cox, S. R. Arridge, J. P. Kaipio, and T. Tarvainen. Quantitative photoacoustic tomography using illuminations from a single direction. *Journal of Biomedical Optics*, 20(3):036015, 2015.

[91] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[92] J. L. Sandell and T. C. Zhu. A review of *in-vivo* optical properties of human tissues and its impact on PDT. *Journal of Biophotonics*, 4(11–12):773–787, 2011.

[93] T. Saratoon, T. Tarvainen, B. T. Cox, and S. R. Arridge. A gradient-based method for quantitative photoacoustic tomography using the radiative transfer equation. *Inverse Problems*, 29(7):075006, 2013.

[94] M. Schweiger and S. R. Arridge. Application of temporal filters to time resolved data in optical tomography. *Physics in Medicine and Biology*, 44(7):1699–1717, 1999.

[95] M. Schweiger, S. R. Arridge, and I. Nissilä. Gauss–Newton method for image reconstruction in diffuse optical tomography. *Physics in Medicine and Biology*, 50(10):2365–2386, 2005.

[96] P. Shao, B. Cox, and R. J. Zemp. Estimating optical absorption, scattering, and Grueneisen distributions with multiple-illumination photoacoustic tomography. *Applied Optics*, 50(19):3145–3154, 2011.

[97] P. Shao, T. Harrison, and R. J. Zemp. Iterative algorithm for multiple illumination photoacoustic tomography (MIPAT) using ultrasound channel data. *Biomedical Optics Express*, 3(12):3240–3249, 2012.

[98] E. Somersalo, M. Cheney, and D. Isaacson. Existence and uniqueness for electrode models for electric current computed tomography. *SIAM Journal on Applied Mathematics*, 52(4):1023–1040, 1992.

[99] N. Song, C. Deumié, and A. Da Silva. Considering sources and detectors distributions for quantitative photoacoustic tomography. *Biomedical Optics Express*, 5(11):3960–3974, 2014.

[100] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.

[101] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, 2005.

[102] T. Tarvainen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography. *Inverse Problems*, 28(8):084009, 2012.

[103] T. Tarvainen, A. Pulkkinen, B. T. Cox, J. P. Kaipio, and S. R. Arridge. Bayesian image reconstruction in quantitative photoacoustic tomography. *IEEE Transactions on Medical Imaging*, 32(12):2287–2298, 2013.

[104] T. Tarvainen, M. Vauhkonen, and S. R. Arridge. Gauss–Newton reconstruction method for optical tomography using the finite element solution of the radiative transfer equation. *Journal of Quantitative Spectroscopy & Radiative Transfer*, 109(17–18):2767–2778, 2008.

[105] L. Tierney. Markov chains for exploring posterior distributions. *The Annals of Statistics*, 22(4):1701–1728, 1994.

[106] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. In *Soviet Mathematics Doklady*, volume 4, pages 1035–1038, 1963.

[107] B. E. Treeby, E. Z. Zhang, and B. T. Cox. Photoacoustic tomography in absorbing acoustic media using time reversal. *Inverse Problems*, 26(11):115003, 2010.

[108] G. Uhlmann. Electrical impedance tomography and Calderón's problem. *Inverse Problems*, 25(12):123011, 2009.

[109] P. J. Vauhkonen, M. Vauhkonen, and J. P. Kaipio. Errors due to the truncation of the computational domain in static three-dimensional electrical impedance tomography. *Physiological Measurement*, 21(1):125–135, 2000.

[110] P. J. Vauhkonen, M. Vauhkonen, T. Savolainen, and J. P. Kaipio. Three-dimensional electrical impedance tomography based on the complete electrode model. *IEEE Transactions on Biomedical Engineering*, 46(9):1150–1160, 1999.

[111] T. Vilhunen, J. P. Kaipio, P. J. Vauhkonen, T. Savolainen, and M. Vauhkonen. Simultaneous reconstruction of electrode contact impedances and internal electrical properties: I. Theory. *Measurement Science and Technology*, 13(12):1848–1854, 2002.

[112] C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*, 17(1):227–238, 1996.

[113] L. V. Wang, editor. *Photoacoustic Imaging and Spectroscopy*. CRC press, 2009.

[114] M. Xu and L. V. Wang. Photoacoustic imaging in biomedicine. *Review of Scientific Instruments*, 77(4):041101, 2006.

[115] K. Yosida. *Functional Analysis*. Springer–Verlag, 6th edition, 1980.

Errata

Publication I

On page B63 after the equation (3.4) the domain of the function r should be $\mathbb{R}_+$.

Publication II

On page 6 after the equation (3.6) the domain of the function r should be $\mathbb{R}_+$.

Aalto-DD 181/2016

ISBN 978-952-60-7007-0 (printed)
ISBN 978-952-60-7006-3 (pdf)
ISSN-L 1799-4934
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

Aalto University
School of Science
Department of Mathematics and Systems Analysis
www.aalto.fi

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL DISSERTATIONS