PhD Thesis

Accelerated Bayesian Inversion and Calibration for Electrical Tomography

Dipl.-Ing. Markus Neumayer

submitted as thesis to the
Graz University of Technology
Faculty of Electrical and Information Engineering
for the academic degree of
Doctor technicae (Dr. techn.)

Institute of Electrical Measurement and
Measurement Signal Processing
Graz University of Technology

Supervisor:
Univ.-Doz. Dipl.-Ing. Dr.techn. Daniel Watzenig

Graz, October 2011

Abstract

The Bayesian inversion of measurement data for the solution of inverse problems forms an attractive though computational expensive approach to gain statistical knowledge about the quantity of interest given the measured data and the model. This thesis accesses the issue of accelerating computational inference and the topic of Bayesian calibration of computer models for electrical tomography with examples for electrical capacitance tomography (ECT). The first part of the thesis covers the acceleration of Markov Chain Monte Carlo (MCMC) methods by speeding up the forward map, the use of fast approximation techniques and the application of improved sampling techniques. The finite element based forward map is reworked and a framework maintaining Green's functions is developed using mathematical properties of the self adjoint boundary value problem. The framework features several advantageous computational techniques including fast Jacobian and transpose of Jacobian operations as well as exact low-rank updates using the Woodbury identity. The methods are developed to increase the speed of the Metropolis Hastings algorithm by the use of approximations a well as for an efficient implementation of a Gibbs sampler. The second part of the thesis covers the issue of model errors and the effective use of calibration measurements to correct the forward map. Classical calibration schemes are reviewed in the Bayesian sense and two type of auto calibration algorithms are presented. The calibration variables are consequently treated as additional unknowns. Finally a stochastic forward map approach is investigated, which allows the incorporation of knowledge about the error of the forward map to quantify the error of the computer model. Summary statistics and measured data are provided for all experiments to quantify the performance gain of the developed approaches. The approaches to speed up statistical inversion demonstrated the ability to accelerate all methods and increase the statistical efficiency. With the Gibbs sampler independent samples can be drawn with an output frequency similar to nonlinear deterministic methods. The incorporation of calibration knowledge leads to well behaved Markov chains even for the case of strong model errors.

Kurzfassung

Die Bayessche Inversion von Messdaten zur Lösung inverser Probleme stellt einen attraktiven, aber zumeist rechenintensiven Ansatz dar, um für ein gegebenes Modell und gemessene Daten statistisches Wissen über die gesuchte Größe zu erlangen. Diese Arbeit behandelt Ansätze zur Beschleunigung statistischer Inversionsmethoden sowie das Thema der Bayesschen Kalibrierung von elektrischen Tomographiesystemen. Beispiele werden am inverse Problem der elektrischen Kapazitätstomographie (ECT) gezeigt. Der ersten Teil der Arbeit beschäftigt sich mit der Beschleunigung von Markov Chain Monte Carlo (MCMC) Methoden durch schnelle Rechentechniken, Approximationsmethoden und schnelleren Samplingverfahren. Das finite Elemente Gleichungssystem des vorliegenden Problems wird überarbeitet und die Eigenschaften des selbstadjungierten Randwertproblems werden durch einen Lösungsansatz mit Green'schen Funktionen genützt. Die neuen Techniken erlauben schnelle Matrix Operationen auf der Jacobimatrix sowie eine effiziente Benützung der Woodbury Formel. Mit den entwickelten Methoden lässt sich anschließend der Metropolis Hastings Algorithmus beschleunigen sowie ein Gibbs Sampler realisieren. Der zweite Teil der Arbeit behandelt die Einbringung von Kalibrierinformation zur Unterdrückung bzw. Korrektur von Modelfehlern. Standardkalibrierverfahren werden im Bayesschen Rahmen evaluiert und zwei Autokalibrierverfahren werden entwickelt. Zu letzt wird ein stochastisches Computermodel abgeleitet. Dieses nützt bestehendes Wissen von Kalibriermessungen um den aktuellen Modellfehler zu prädizieren und damit eine Unsicherheitsangabe über die Modellausgänge zu generieren. Die Arbeit enthält zusammenfassende Statisitiken und Messdaten über alle Rekonstruktionsexperimente und zeigt deren Verbesserung hinsichtlich Beschleunigung und statistischer Effizienz. Mit dem Gibbs Sampler ist es möglich unabhängige Samples mit der gleichen Frequenz zu generieren, wie typische nichtlineare Verfahren. Die Einbringung von Kalibrierinformation führt zu besseren Verhalten auch im Fall von starken Modellfehlern.

Acknowledgment

To whom it may belong - and perhaps others.

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am ....

(Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Graz, the ....

(Signature)

Contents

List of Figures

1 Introduction

This first chapter provides the introductional part, the problem statement and the outline the thesis.

1.1 A Short Introduction to Inverse Problems

In a number of fields in modern science and the daily work life people are dealing with estimating or finding quantities of interest $\boldsymbol{x} \in \mathbb{R}^N$, which are not directly measurable or observable. Instead they can only take measurements $\tilde{\boldsymbol{d}} \in \mathbb{R}^M$ from some other quantities, which are somehow related to $\boldsymbol{x}$. Such problems are generally referred to as inverse problems or parameter estimation problems [Ram05]. Examples for such problems are process tomographic applications like electrical capacitance tomography (ECT), medical imaging techniques like electrical impedance tomography (EIT) [Hol05], computed tomography or optical diffusion tomography, geophysical problems like reservoir modeling, or quality control applications like nondestructive material testing or remote sensing applications. The underlying physical problem is given by determining $\boldsymbol{x}$ from data $\tilde{\boldsymbol{d}}$. for the given process $P(\cdot)$. An often used model for the measurement process is given by

$$\tilde{\boldsymbol{d}} = P(\boldsymbol{x}) + \boldsymbol{v}, \tag{1.1}$$

where $\boldsymbol{v}$ presents an additive noise term. There also exist a number of other process models like e.g. multiplicative noise models or noise models with a more complex interaction of the noise variable $\boldsymbol{v}$ but in this work the additive noise model will be maintained. For the solution of the given task a (computer) model $F(\boldsymbol{x})$ is constructed, which is referred to as forward map $F : \boldsymbol{x} \mapsto \boldsymbol{y}$. Models or forward maps are the key component of all inverse problems as they build the link between the quantities of interest and the measured quantities. The obvious claim on the forward map is given by

$$F(\boldsymbol{x}) = P(\boldsymbol{x}) \qquad \forall \boldsymbol{x}, \tag{1.2}$$

which means that the model output $\boldsymbol{y}$ equals the noise free data $\boldsymbol{d} = P(\boldsymbol{x})$. Hence, the straightforward approach

$$\boldsymbol{x}^* = F^{-1}(\tilde{\boldsymbol{d}}), \tag{1.3}$$

appears suitable to provide the solution for the inverse problem, which also gives the name to the problem type.

In traditionally deterministic inversion theory, the inversion of $F(\cdot)$ to determine $\boldsymbol{x}$ is mostly treated as an optimization problem [TV82] of form

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} ||F(\boldsymbol{x}) - \tilde{\boldsymbol{d}}||. \tag{1.4}$$

Powerful optimization algorithms are used to provide an efficient manipulation strategy for $\boldsymbol{x}$ in order to minimize the misfit between the model output $\boldsymbol{y} = F(\boldsymbol{x})$ and the measured data $\tilde{\boldsymbol{d}}$ for an appropriate cost function and an appropriate norm [Vog02], [Han98]. However, the pure minimization of the misfit between the model output and the measurements may fail to provide useful inverses for certain problems. This comes due to the inherently ill-posed nature of some of the examples previously mentioned [Bra07], which makes a direct inversion of $F(\cdot)$ unstable. Conditions for ill-posedness were given by Hadamard in 1923 [Had23]. To overcome this problem the original optimization problem is augmented by a stabilization scheme such as Tikhonov regularization [Tar05]. These augmentation schemes are called regularization schemes [Bra07], [EHN02]. Once the setup of the algorithm is completed deterministic methods are able to provide reasonable results within a comparatively small computation time.

A contrasting framework to deterministic inversion theory is statistical inversion theory [KS05]. Within this theory the inverse problem is treated as a Bayesian inference problem. Rather than providing a single result these methods can provide the posterior density function $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ over the solution conditioned on the measured data $\tilde{\boldsymbol{d}}$. Thus, also information about the quality of the result can be obtained. The attractive feature of Bayesian inversion theory is given by the natural way of incorporating existing knowledge about the measurement noise $\boldsymbol{v}$ and other errors as well as prior knowledge for $\boldsymbol{x}$. The drawback is given by the largely increased computational burden and increased modeling requirements, which make the application of statistical methods, especially in large scale inverse problems, often intractable. This especially holds for problems where the underlying forward map is given by a partial differential equation whose solution requires numerical techniques. Adequate computational techniques and schemes have to be deployed, to make statistical inversion methods more suitable for a broader field of applications.

Both types of inversion methods can be applied and once implemented they usually work fine for the task

$$\boldsymbol{x}^* = \mathcal{F}^{-1}(F(\boldsymbol{x}) + \boldsymbol{v}), \tag{1.5}$$

where the calligraphically $\mathcal{F}$ should denote the stabilized inversion overcoming the ill-posedness by using incorporated prior knowledge. This syntectical constellation is referred to as an inverse crime setup [KS07]. It is made when data is generated by the same model, which is used for the reconstruction. Hence, it can only be done in simulation experiments. The use of inverse crime data is sometimes useful to test the functionality of the inversion algorithm. When applying the inversion algorithm to data from $P(\cdot)$, the inversion often shows different effects or even does not work. The culprit for this failure is given by the equal sign in equation (1.2). In fact, equation (1.2) is often more like

$$F(\boldsymbol{x}) \approx P(\boldsymbol{x}). \tag{1.6}$$

More precisely the model inadequacy causes a model error

$$\boldsymbol{e} = P(\boldsymbol{x}) - F(\boldsymbol{x}), \tag{1.7}$$

which can outweight the measurement noise $\boldsymbol{v}$. To overcome this problem a so called model calibration strategy has to be applied [KO00] . This calibration can be performed by several ways either on the so called images space or on the data space and then the calibration information can be applied in either a deterministic of probabilistic way. Using a calibration scheme the problem can be solved using real data or the data from another model, which builds the real test for the ability to solve inverse problems.

1.2 Problem Statement and Motivation

Section 1.1 provided an introduction to inverse problems and the problems one has to face when solving them. This section states more clearly the problems which arise and the motivation to solve them in this thesis. In this work Bayesian methods are used to solve the inverse problem of electrical capacitance tomography (ECT). The two main topics of this work, which also build the title of the thesis, are:

- Accelerating Bayesian methods,
- A Bayesian framework for calibration.

The often mentioned drawback for the application of statistical methods is their comparative slowness with respect to deterministic methods. This comes due to the huge number of forward map evaluations typically necessary, to draw independent samples from the posterior distribution. However, the natural incorporation of any sort of knowledge as well as the availability of uncertainty information about the result are too interesting assets for the application of Bayesian methods. The first part of the work deals with the acceleration of Bayesian methods. This topic covers the items

- Accelerating the forward map.
- Fast computational techniques (Jacobian operations, etc.).

and

- Use of approximations techniques,
- Surrogate models.

The first two points cover numerical techniques, which accelerate the evaluation of the forward map. These developed techniques are numerically exact, which means that they provide the same result up to machine precision with respect to a standard implementation, but in a fraction of the time. The second two points use approximation techniques to further accelerate the evaluation of the forward map but at the cost of approximation errors. Subsequently specialized sampling algorithms will be used to take advantage of the fast computational techniques. This points will be summarized by

- Accelerated sampling using approximations,
- Fast exact sampling methods.

The applicability will then be demonstrated for two types of image representations on test examples for ECT. A main emphasis in the design of the methods lies in a wider usability for a general class of problems. Innovation provided by the use of fast computational methods and surrogate techniques as well as the use of more sophisticated sampling algorithms will be referred

to as "Computational techniques" and "Bayesian algorithms".
The second main part of this thesis considers the problem of model inadequacy and the concerning problems of model errors. As already mentioned in section 1.1, calibration schemes have to be employed as countermeasures to overcome model error problems. Traditional schemes will be reviewed and the possibility of incorporating calibration information in a Bayesian sense will be exploited. The points treated in this part of the thesis are

- Bayesian review of calibration approaches.
- Use of calibration information.
- Uncertainty in the calibration information.
- Calibration in a Bayesian sense.
- Incorporation of uncertainty information to the forward map.

Innovation provided by this points shall be referred to as "Bayesian formulations". Again, the methods are developed to be used for different kind of problems.

Summarizing the problem statement and the motivation provided by this section will also end up in the innovation provided in this thesis which can be summarized by.

- Computational techniques,
- Bayesian algorithms,
- Bayesian formulations.

1.3 Related Work and Scientific Research Fields

Because of the motivation for this thesis and the ambitious aim of the work, the thesis covers a range of several fields of physics, engineering, computational mathematics and science - a matter of fact when dealing with inverse problems. The following list provides a collection related fields to the work.

- Physical modeling.

- Computational physics for linear partial differential equation (PDE) systems.
- Numerical mathematics.
- Approximation techniques.
- Statistical inversion methods.
- Bayesian methods.
- Markov chain Monte Carlo (MCMC) methods.
- Numerical optimization.
- Model errors and calibration techniques.

An important aspect in the development of the algorithms and methods lies in their wider applicability to other inverse problems. This is especially in concern with the computational methods developed for the forward map, as well as with the issue of calibration. The following list collects several fields where the developed computational techniques can be directly applied and thus the full span of methods can be used directly.

- Industrial applications (electrical capacitance tomography).
- Medical applications (electrical impedance tomography).
- Geophysical methods.
- Problems with an underlying resistor network (referred to as $\boldsymbol{WSW}^T$ systems).

1.4 Outline and Novelity

This section provides the outline of the thesis and lists the novelty contributed in each chapter, which are formally given by the terms computational techniques, Bayesian algorithms and Bayesian formulations, in more detail.

Chapter 2 recalls the basics of Bayesian inversion theory and presents the algorithms and formulations used in this thesis. Section 2.2 belong to the topic of Bayesian algorithms and presents the Markov chain Monte Carlo

(MCMC) methods, which are later used in this work to generate samples from the posterior distribution. Section 2.3 is about Bayesian formulations and covers the aspect that the used (computer) model does not fit to the reality and hence a calibration scheme is required to make the model work for real data. In this sense, the major significant points are given by

- Calibration in the Bayesian sense.
- Empirical Bayesian calibration.
- Mutual inference approaches.
- A stochastic forward map.

Chapter 3 deals with computational techniques for linear PDE systems. The focus of this work lies on the forward map for ECT and the efficient solution of the governing boundary value problem. The methods are applicable to all systems where the stiffness matrix is a linear function of the material coefficients and the problem is self adjoint. These systems will be referred to as $\boldsymbol{WSW}^T$ systems. In the first part of the chapter the standard solution framework is replaced by a framework for maintaining Green's functions. The special advantage of the framework is the possibility of directly operating on system specifics like the Jacobian matrix. Compared to the standard framework, where the Jacobian has to be evaluated separately before, the new framework allows the computation of matrix vector products of the Jacobian or its transpose without evaluating the matrix itself. The major points of the rework of the standard solution scheme are,

- Orthogonal decomposition for finite element (FE) systems.
- Integral evaluation by linear maps.
- A Green's functions framework for ECT.
- Fast Jacobian and transpose of Jacobian operations.

Fast low-rank update schemes are developed which become necessary for efficient conditional sampling. These schemes are based on the Woodbury matrix identity and offer the possibility to efficiently solve the model for small changes in the input data. Hereby the increased computational advantages can be achieved with the use of Green's functions. Further, a specialized Woodbury formulation for $\boldsymbol{WSW}^T$ systems is developed. Lastly, domain decomposition techniques are exploited, which decrease the computational costs for the evaluation of the the forward map as only a smaller equation system with respect to the original equation system has to be solved. Finally, a computation time comparison for all methods is presented. Summarizing the new contributions of this second part of the chapter, the novel points are

- Fast low-rank updates.
- A generalized Woodbury formulation for $\boldsymbol{WSW}^T$ systems.
- Domain decomposition techniques.

In the last part of chapter 3 the application of approximation techniques for the forward map is discussed. In these approaches, the forward map is replaced by computational cheap approximation methods. The following points are contributed

- Surrogate approximations.
- Polynomial approximation of the forward map.
- A reduced order physical model.

Chapter 4 is titled "Case studies" and combines the Bayesian algorithms and formulations presented in chapter 2, with computational techniques developed in chapter 3. Necessary details for the implementation of the algorithms are listed and a detailed analysis of the algorithms is presented. The novelty in the results can be summarized by

- A fast deterministic algorithm using Jacobian operations.
- Accelerated inversion with different approximation techniques.
- Gibbs sampling using conditional sampling techniques.
- A Bayesian analysis of the forward map.
- A mutual inference approach.
- A Stochastic forward map.

1.5 Publications

Throughout the genesis of this thesis, the author has published several journal publications and one invited book chapter in the field of inverse problems. The publications are mainly in the field of electrical capacitance tomography (ECT) and capacitive measurement systems but also an eddy current problem is presented. The following list provides a brief summary of each publication

- [NZWF11] This invited book chapter titled "Current reconstruction methods for electrical capacitance tomography" deals with several aspects of ECT including the forward map and reconstruction algorithms. It contains the physics of the ECT problem, and the mathematical tools necessary for solving the forward problem using the finite element method, as well as further methods for deriving gradients with respect to the unknown quantities. The presented survey of current reconstruction algorithms covers the current state of the art techniques like simple back projection algorithms, Landweber iteration methods, offline iteration online reconstruction (OIOR) methods, optimal back projection methods, nonlinear approaches, Bayesian recursive methods and Bayesian inference approaches. For comparison short parts of the book are also repeated in this work, although the developed framework of computational methods for the forward map in this thesis uses a Green's functions approach, which offers advanced possibilities and substantially decreased computation times.

- [NSWZ10] This paper is the journal version of [NSWZ09] and investigates the resolution behavior of nonlinear reconstruction methods on extremely coarse finite element meshes. The paper contains an analysis of the achievable resolution in different parts of the region of interest and presents studies about the applicability of coarse meshes compared to fine meshes.

- [WNF11] This publication is the journal version of [WNF10]. The work is about accelerating MCMC methods for ECT by the use of approximation techniques including surrogate models in combination with the delayed acceptance version of the Metropolis Hastings algorithm. Several parts of the work are also of topic in this thesis and will also appear in this work although this thesis of coarse contains a more advanced treatment of the tasks.

- [NWB11] This work is the journal version of [NWB10] and deals with the inverse problem of stress zone imaging in steel plates of electrical machines. Stress zones occur due to punching or other production processes and cause decreased magnetic properties in the sheet, which result in higher losses. The work describes the eddy current forward problem and the inverse problem and contains exemplary simulation results.
- [NWZ11] This publication is the journal version of [NWZ10] and deals with the robust determination of material parameters in ECT under the impact of model errors. The paper compares the robust H_∞ filter with the Kalman filter for the case, that no calibration is performed.
- [SBN$^+$11] This paper is the journal version of [SBN$^+$10]. The paper presents a sensor fusion concept for capacitance measurements and ultrasound measurements in order to determine the object type and the distance estimation of an object in front of a bumper. The system is thought for roll over protection and parking assistance systems.

1.5.1 Other Publications

Beside the above mentioned journal publications also several other papers related to the thesis were published as output of other work the author was involved during his time as university assistant at the Institute of Electrical Measurement and Measurement Signal Processing at Graz University of Technology. The following lists summarizes these publications including again a brief summary.

- [FZWN08] In this paper a filter algorithm has been developed, which allows edge preserving reconstructions with a 2D ECT system for short 3D objects.
- [NS08] This paper contains parts of the authors masters thesis, where a reconstruction algorithm for the complex conductivity was developed.
- [NS09] This paper contains other parts of the authors masters thesis, where the physical effects inside an ECT sensor are explored and the impact of wave propagation effects due to high excitation frequencies is investigated.
- [NZ09] In this work different magnetic switches using ferromagnetic shielding for a food tablet system with integrated heating were investigated by simulations and measurements.

- [NSWZ09] This paper is the conference version of [NSWZ10].
- [ZN09] In this publication the behavior of ECT sensors with shortened electrodes has been investigated. The work builds a first study for capacitive systems with exposed electrodes for environmental detection tasks. For this class of problems the electrodes are always small compared to the dimensions of the problem domain.
- [NWZ10] This publication is the conference version of [NWZ11].
- [NZ10] In this study the application of ECT algorithms for liquid level measurements in tanks using capacitive sensors is explored.
- [NMGZ10] In this publication a baking dish was equipped with a prototype measurement system for impedance tomography and the electrical properties of dough during a baking process were monitored to provide information during the baking process.
- [NGB$^+$10] The paper describes a simulation and measurement study for robust sensing of human proximity on working tools.
- [SBN$^+$10] This paper is the conference version of [SBN$^+$11].
- [GBN$^+$10] As a demand of an industrial project, a measurement cell for determining the relative permittivity of foam materials was built and presented in the work. The paper includes an uncertainty analysis of the cell.
- [WNF10] This paper is the conference version of [WNF11].
- [NWB10] This paper is the conference version of [NWB11].
- [ZN10] In this work a ratio metric back projection algorithm for ECT is presented. The algorithm is gain invariant, which makes it suitable for low cost hardware systems which offer drift behavior due to temperature variations.
- [NOS$^+$10] This paper summarizes the results of a collaboration with the institute of Numerical Mathematics at TU Graz, where a boundary element method (BEM) level set approach is used for shape determination tasks in ECT.
- [NZ11] This work builds an extension to [NZ10], where different sensors topologies are investigated for the application of liquid level measurements.

2 Bayesian Inversion and Calibration

This chapter of the thesis presents the algorithms for accelerated computational inference for inverse problems and the topic of Bayesian formulations to incorporate calibration information and to calibrate computer models in a Bayesian sense. Hence, the chapter is about methods of statistical inversion theory and some basic knowledge about probability theory for the Bayesian approach is required. The necessary methods will be explained at each step. Detailed information about the required fundamentals can be found e.g. in [Siv96] or [BS 4].

Compared to deterministic inversion methods, which in general provide a point estimate solution, statistical methods offer a different view on the properties of the treated quantities, which can be summarized in the following statements [KS05].

- Every variable is treated as a random variable.
- All knowledge about the quantities is expressed in the form of probability density functions (pdfs).
- Solutions are summary statistics over the posterior density function.

The attractiveness of the Bayesian approach lies in the intuitive way to incorporate existing knowledge about the noise or prior knowledge of the solution into the inversion process and to incorporate uncertainty of models or calibration.

2.1 Bayesian Inversion Theory

As already mentioned in the introductory part of this thesis, an inverse problem is given if one aims on determining a quantity $\boldsymbol{x} \in \mathbb{R}^N$ from information of indirect observations $\tilde{\boldsymbol{d}} \in \mathbb{R}^M$. Formally this measurement process can be written as

$$\tilde{\boldsymbol{d}} = P(\boldsymbol{x}, \boldsymbol{v}), \tag{2.1}$$

where $P(\cdot)$ denotes the physical process that leads to the observations. $\boldsymbol{v}$ denotes a noise process which corrupts the otherwise noise free data $\boldsymbol{d}$ in some way. Often $\boldsymbol{v}$ acts in an additive manner as measurement noise on $\boldsymbol{d}$, but for generality at this stage it should be denoted by this way. Given knowledge about the process $P(\cdot)$ a model

$$\boldsymbol{y} = F(\boldsymbol{x}, \boldsymbol{v}), \tag{2.2}$$

is built, which is referred to as the forward map. Ideally $F(\cdot)$ should behave like the process $P(\cdot)$, but in fact more often the interrelationship $F(\cdot) \approx P(\cdot)$ can be achieved, as every model process is afflicted by either simplifications, or by the fact, that some knowledge about $P(\cdot)$ is not available although one would be able to model it. However, the forward map is assumed to be complete in the sense that $\boldsymbol{y}$ is uniquely defined by $\boldsymbol{x}$ and $\boldsymbol{v}$. So far all elements used in the notation are realizations of random variables, which is necessary due to the physical relationship. This is indicated by the use of small capitals. The corresponding random variables are written in large capitals, i.e. $\boldsymbol{x}$ is a realization of the random variable $\boldsymbol{X}$. Thus, they are defined by their pdfs.

The key element of statistical inversion theory is given by Bayes law [GCSR 5], which is given by

$$\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}}) = \frac{\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})\pi(\boldsymbol{x})}{\pi(\tilde{\boldsymbol{d}})} \propto \pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})\pi(\boldsymbol{x}). \tag{2.3}$$

The terms in equation (2.3) are the prior distribution $\pi(\boldsymbol{x})$, the likelihood $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$ and the evidence $\pi(\tilde{\boldsymbol{d}})$. $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ is the posterior distribution [Gre05]. The likelihood function $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$ is assigned to be a probability density function, which expresses the probability that the measurements $\tilde{\boldsymbol{d}}$ are caused by $\boldsymbol{x}$. In other words, the likelihood answers the question "How likely is it, that $\boldsymbol{x}$ causes the measured data $\tilde{\boldsymbol{d}}$?". Thus, the likelihood quantifies the misfit between the measured data and the output of the forward map using knowledge about the noise process. The prior distribution $\pi(\boldsymbol{x})$ forms a key issue for the successful application of Bayesian inversion theory. Through the prior it is possible to incorporate statistical knowledge about $\boldsymbol{X}$. The most simple prior is given by only evaluating the permissibility of $\boldsymbol{x}$, i.e. $\pi(\boldsymbol{x}) = 0$ for realizations $\boldsymbol{x}$ which are infeasible and $\pi(\boldsymbol{x}) = 1$ for any valid realization. By this, infeasible solutions are effectively neglected. An estimator which maximizes the posterior distribution, is called a maximum *a posteriori* (MAP) estimator. For the simple 0/1 prior the MAP estimator is in essence a maximum likelihood (ML) estimator which avoids infeasible solutions. Using this simple prior the quality of the estimator is already increased. However, by the choice of more specific priors it is possible to build highly effective estimators. The evidence $\pi(\tilde{\boldsymbol{d}})$ is given by

$$\pi(\tilde{\boldsymbol{d}}) = \int_{\mathbb{R}^N} \pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x}, \tag{2.4}$$

which is the marginal density function of the likelihood $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$. The computation of the evidence is in general an infeasible task as the evaluation of the integral becomes untractable for higher dimensional spaces or complex forward maps. However, the evidence has only the role of a normalization constant, which normalizes the posterior distribution. Hence, it can be skipped, which leads to the right hand term in equation (2.3). The distribution $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ is called the posterior distribution. In essential the posterior distribution gives answer to the question "Given the prior, how likely is it, that $\boldsymbol{x}$ is caused by the measured data $\tilde{\boldsymbol{d}}$?". Hence, the posterior distribution provides statistical information about the probability of $\boldsymbol{x}$ given $\tilde{\boldsymbol{d}}$.

The attractiveness of the Bayesian framework for inversion lies in several aspects [Gre05]:

- The approach is applicable to any forward problem.
- No derivatives have to be computed.
- Available prior knowledge can be incorporated in a natural way.
- Information about the measurement noise can be added in the same natural way.
- The result is a probability density function. Complete information about the uncertainty of the result is available.

However, so far Bayes law only provides a way to evaluate the probability of a specific realization of $\boldsymbol{X}$. The following subsections should provide some key issues about the application of statistical inversion and the information one can obtain by applying Bayesian methods.

The Likelihood Function

The likelihood function $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$ is the key element to incorporate knowledge about the measurement noise $\boldsymbol{v}$. The likelihood function is formally given by

$$\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}) = \int_{\mathbb{R}^M} \pi(\tilde{\boldsymbol{d}}, \boldsymbol{v}|\boldsymbol{x})\pi(\boldsymbol{v})d\boldsymbol{v}, \tag{2.5}$$

which means that it is found by marginalization $\boldsymbol{v}$ out of $\pi(\tilde{\boldsymbol{d}}, \boldsymbol{v}|\boldsymbol{x})$. The question is about the distribution of $\pi(\tilde{\boldsymbol{d}}, \boldsymbol{v}|\boldsymbol{x})$. By fixing all variables it becomes clear, that for additive noise the distribution $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}, \boldsymbol{v})$ is given as

$$\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}, \boldsymbol{v}) = \delta(\tilde{\boldsymbol{d}} - F(\boldsymbol{x}) - \boldsymbol{v}), \tag{2.6}$$

where $\delta(\cdot)$ is the Dirac delta function. The integral of equation (2.5) becomes easy to solve, as it is the convolution of a function with a Dirac delta function. Hence, the likelihood function is given as

$$\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}) = \pi_{\boldsymbol{v}}(\tilde{\boldsymbol{d}} - F(\boldsymbol{x})), \tag{2.7}$$

which is the distribution of the measurement noise $\boldsymbol{v}$. For zero mean Gaussian noise the likelihood function is

$$\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{y} - \tilde{\boldsymbol{d}}\right)^T \boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}\left(\boldsymbol{y} - \tilde{\boldsymbol{d}}\right)\right\}, \tag{2.8}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{v}}$ is the covariance matrix of the noise. This derivation assumes additive measurement noise, but the formalism can also be applied to other noise distributions. Examples therefore are given in [KS05].

2.1.1 Estimation of Specific Quantities

The output of Bayesian inversion methods is given by equation (2.3) as the posterior distribution. $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ provides a relative probability about $\boldsymbol{x}$ conditioned on the data $\tilde{\boldsymbol{d}}$. Though $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ contains complete information about the distribution of $\boldsymbol{x}$ conditioned on $\tilde{\boldsymbol{d}}$, the meaning of the information provided by $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ is not that clear at the first.

The representation of the probability of single components of $\boldsymbol{x}$ from $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ is possible, but may not be very meaningful in order to understand the whole information. Hence, representative numbers to represent main issues about $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ are necessary. Two meaningful point estimates are given by the maximum likelihood estimate

$$\boldsymbol{x}_{\mathrm{ML}} = \arg\max_{\boldsymbol{x}} \pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}), \tag{2.9}$$

and the maximum *a posteriori* estimate

$$\boldsymbol{x}_{\mathrm{MAP}} = \arg\max_{\boldsymbol{x}} \pi(\boldsymbol{x}|\tilde{\boldsymbol{d}}). \tag{2.10}$$

The ML estimate presents the result of highest likelihood, which means the result $\boldsymbol{x}$ with the smallest misfit between the model output and the measured data $\tilde{\boldsymbol{d}}$. As no prior knowledge is used, it can generally not be applied to ill-posed inverse problems, which rely on the incorporation of prior knowledge. The MAP estimate uses the incorporated prior knowledge and presents a mode of the posterior distribution. This result is referred to as a point estimate of the posterior distribution.

Looking onto the properties of the likelihood function one can imagine, that this point estimate can be found by solving an optimization problem [KS05]. In fact, this is the practical way to find estimates, as the equivalent optimization problems can be efficiently solved by using methods from numerical optimization. In the case of multi-modal distributions the application of point estimates becomes critical, as the solution dependents on the starting point of the optimization problem. Further the MAP estimate can become less meaningful if the distance of the MAP estimate and the main part of the posteriori distribution is too large. In this case a more representative estimate is the conditional mean (CM)

$$\boldsymbol{x}_{\mathrm{CM}} = \int_{\mathbb{R}^N} \boldsymbol{x}\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})d\boldsymbol{x}, \tag{2.11}$$

which encounters the whole posterior distribution. The uncertainty of the CM estimate can then be stated by the conditional covariance

$$\mathrm{cov}(\boldsymbol{x}, \tilde{\boldsymbol{d}}) = \int_{\mathbb{R}^N} (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{CM}})^T (\boldsymbol{x} - \boldsymbol{x}_{\mathrm{CM}}) \pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})d\boldsymbol{x}. \tag{2.12}$$

The CM estimate is now different to the MAP estimate as it requires the evaluation of an integral which can become inaccessible in the case of complex forward problems or higher dimensional spaces [Liu94]. Numerical integration schemes like quadrature methods can typically not be applied for inverse problems as they rely on knowledge about the support of the function and properties of the function itself. A numerical integration scheme which proofed to converge for such integrals is Monte Carlo integration [MRR+53]. The key idea of Monte Carlo integration is to set up a random number generator which generates samples, or support points, of the target distribution. Thinking about a discrete state space the samples are generated with a frequency proportional to their probability. Hence, the integral reduces to the mean over the samples. A class of algorithms proofed to be effective for sampling pdfs are Markov chain Monte Carlo (MCMC) methods [Has70].

2.2 Markov Chain Monte Carlo Methods

In this section the computational tools for exploring the posterior density function using Markov Chain Monte Carlo (MCMC) methods are explained. Numerical methods for this task generally work on sampling methods which means, that the algorithms aim to produce independent samples of the posterior distribution for the further application of Monte Carlo Integration. These algorithms can generally be seen as random number generators and their efficiency strongly depends on the effective generation of independent samples. Reports on the use of MCMC methods for solving inverse problems have been made for several examples e.g. [MT95], [KKSM00], [WF09].

2.2.1 Markov Chains and Monte Carlo Integration

Markov chains build the key element inside MCMC methods to generate samples from a target distribution $\pi(\cdot)$. Suppose $\mathcal{M} = \{X_n\}_{n=0}^{\infty}$ being a sequence of random variables on the state space $\mathcal{X}$. The process is called a Markov chain if it satisfies the Markov condition

$$Pr(X_{n+1} = x_{n+1} | X_n = x_n, \ldots, X_0 = x_0) = Pr(X_{n+1} = x_{n+1} | X_n = x_n). \tag{2.13}$$

The Markov condition means, that the state of X_{n+1} only depends on the previous state X_n, which is termed a first order Markov chain. The transition probability between two consecutive states is a conditional probability and described by the transition kernel $K(x_n, x_{n+1})$.

Markov chains offer several mathematical properties, which shall shortly be discussed in the following paragraph. Detailed informations about the properties can for example be found in [Tie94] or [GRS96]. The chain $\mathcal{M}$ is called homogeneous if

$$Pr(X_{n+m+1} = x_{n+m+1}|X_{n+m} = x_{n+m}) = Pr(X_{n+1} = x_{n+1}|X_n = x_n) \quad \forall m, \tag{2.14}$$

holds. A more important property is given by the irreducibility of the chain. On discrete state spaces, irreducibility means that all states on $\mathcal{X}$ intercommunicate. Hence, the transition kernel $K(\cdot,\cdot)$ has a non zero transition probability between each combination of the states. On continuous state spaces, irreducibility is defined as the property, that any set of finite probability in the state space can be reached from every starting point. Another important property of Markov chains is called reversibility. This property is given if the so called detailed balance condition

$$\pi(x_{n+1})K(x_n, x_{n+1}) = \pi(x_n)K(x_{n+1}, x_n), \tag{2.15}$$

holds, which is a necessary and sufficient condition. Reversibility is the attribute that for a given sequence $\mathcal{M}$, and for a given proposal kernel, one is not able to determine the direction of simulation.
A Markov chain may have the property of periodicity if the state x_{n+j} can become the state x_n for $j > 2$. For discrete state spaces this property can be given, if the main diagonal of the transition matrix is zero. If the proposal kernel allows that $x_{n+1} = x_n$, the kernel is called aperiodic.

Given this definitions, the sequence $\mathcal{M}$ converges against the target distribution $\pi(\cdot)$ for sufficient large n, if the kernel is irreducible and aperiodic. Thus, samples from the target distribution can be generated and the powerful tool of Monte Carlo integration can be applied [Tie94].

Monte Carlo Integration

The output of the Markov Chain is a set of samples from the probability density function $\pi(\boldsymbol{x})$. A first application, where an integral over the distribution $\pi(\boldsymbol{x})$ has to be evaluated, was presented by the CM estimate (2.11). Monte Carlo integration is a general technique to approximately evaluate integrals of a function $f(\boldsymbol{x})$ using samples $\boldsymbol{x}_i$ of $\pi(\boldsymbol{x})$. For this Monte Carlo integration is given as

$$\int_{\mathbb{R}^N} f(\boldsymbol{x})\pi(\boldsymbol{x})d\boldsymbol{x} \approx \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{x}_i). \tag{2.16}$$

By Monte Carlo integration even complex integrals can be approximated by a mean computation from the output of the chain.

2.2.2 Metropolis Hastings Algorithm

An algorithm to generate samples from a distribution $\pi(\cdot)$ using an underlying Markov chain is the so called Metropolis Hastings (MH) algorithm [MRR$^+$53]. Its generalized form was presented in [Has70]. The algorithm is given by the following:

1. Pick the current state $\boldsymbol{x} = \boldsymbol{X}_n$ of the Markov chain.
2. With proposal density $q(\boldsymbol{x}, \boldsymbol{x}')$ generate a new state $\boldsymbol{x}'$.
3. Compute the acceptance probability $\alpha = \min\left[1, \frac{\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}})q(\boldsymbol{x}',\boldsymbol{x})}{\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})q(\boldsymbol{x},\boldsymbol{x}')}\right]$.
4. With probability α accept $\boldsymbol{x}'$ and $\boldsymbol{X}_{n+1} = \boldsymbol{x}'$, otherwise reject $\boldsymbol{x}'$ and set $\boldsymbol{X}_{n+1} = \boldsymbol{x}$.

Hereby $q(\boldsymbol{x}, \boldsymbol{x}')$ is referred to as proposal density. The so called acceptance probability

$$\alpha = \min\left[1, \frac{\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}})q(\boldsymbol{x}',\boldsymbol{x})}{\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})q(\boldsymbol{x},\boldsymbol{x}')}\right], \tag{2.17}$$

in line 3 represents the computational costly part of the MH algorithm, as the evaluation of $\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}})$ requires the solution of the forward map for the proposal candidate $\boldsymbol{x}'$. A specific of the MH algorithm is the fact, that proposals can become rejected. Hence, an efficient generation of proposal candidates is mandatory. Further, the algorithm has the tendency to produce highly correlated samples of $\boldsymbol{X}$, which means that samples with a low lag between are almost the same. For large lags the samples can be viewed as independent.

An extension to the MH algorithm for scalable state spaces is the so called Metropolis Hastings Green (MHG) or reversible jump MCMC [Gre95] algorithm. The idea behind this extension is to describe the generation of a new candidate $\boldsymbol{x}'$ by a deterministic function $\boldsymbol{x}' = \Psi(\boldsymbol{x}, \gamma)$, where γ is again a random variable. If the function Ψ is differentiable and invertible the MHG algorithm allows to change the dimension of the state space. This writes the acceptance ratio as

$$\alpha = \min \left[1, \frac{\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}}) q(\boldsymbol{x}', \boldsymbol{x})}{\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}}) q(\boldsymbol{x}, \boldsymbol{x}')} |J_m| \right], \tag{2.18}$$

where J_m

$$J_m = \left| \frac{\partial(\boldsymbol{x}', \gamma')}{\partial(\boldsymbol{x}, \gamma)} \right|, \tag{2.19}$$

is the Jacobian for the mapping from composite parameter $(\boldsymbol{x}, \gamma)$ to composite parameter $(\boldsymbol{x}', \gamma')$. Hence, the state space can be accessed and manipulated, which can be a useful feature.

2.2.3 Delayed Acceptance MH Algorithm

The delayed acceptance Metropolis Hastings (DAMH) algorithm was presented in [CF05] and is motivated by the idea to replace the forward map $F(\cdot)$ by a cheap approximation of the forward map $F^*(\cdot)$ in a first step. If a proposal evaluated by $F^*(\cdot)$ is accepted, it is reevaluated using the accurate model $F(\cdot)$. Thus, the DAMH algorithm can be seen like two nested MH algorithms. The algorithm is given by

1. Pick the current state $\boldsymbol{x} = \boldsymbol{X}_n$ from the Markov chain.
2. With proposal density $q(\boldsymbol{x}, \boldsymbol{x}')$ generate a new state $\boldsymbol{x}'$.
3. Compute $\alpha = \min \left[1, \frac{\pi^*_{\boldsymbol{x}}(\boldsymbol{x}'|\tilde{\boldsymbol{d}}) q(\boldsymbol{x}', \boldsymbol{x})}{\pi^*_{\boldsymbol{x}}(\boldsymbol{x}|\tilde{\boldsymbol{d}}) q(\boldsymbol{x}, \boldsymbol{x}')} \right]$.
4. With probability α accept $\boldsymbol{x}'$ to be a proposal for the standard MH algorithm. Otherwise set $\boldsymbol{x}' = \boldsymbol{x}$ and return to 2.
5. Compute $\beta = \min \left[1, \frac{\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}}) q(\boldsymbol{x}', \boldsymbol{x})}{\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}}) q(\boldsymbol{x}, \boldsymbol{x}')} \right]$.
6. With probability β accept $\boldsymbol{x}'$ and $\boldsymbol{X}_{n+1} = \boldsymbol{x}'$, otherwise reject $\boldsymbol{x}'$ and set $\boldsymbol{X}_{n+1} = \boldsymbol{x}$.

It is obvious to observe, that the efficiency of the DAMH algorithm strongly depends on the quality of the approximated forward map $F^*(\cdot)$. The likelihood function $\pi^*_{\boldsymbol{x}}(\boldsymbol{x}'|\tilde{\boldsymbol{d}})$ is used for the evaluation of the approximation. The lower $\boldsymbol{x}$ expresses a possible state dependency of $\pi^*_{\boldsymbol{x}}(\boldsymbol{x}'|\tilde{\boldsymbol{d}})$. Again, the extension towards the reversible jump rule can be applied. Beside this, the DAMH algorithm provides similar overall behavior to the MH algorithm.

2.2.4 Gibbs Sampler

A different algorithm to the class of MH algorithms is the Gibbs sampler which was firstly prominently presented in [GG84]. More detailed information about the algorithm can be found in [CG92]. The algorithm is given by

1. Pick the current state $\boldsymbol{x} = \boldsymbol{X}_n$ from the Markov chain.
2. For every element i of the vector $\boldsymbol{x}$:
    a) Set x_i of $\boldsymbol{x}$ as independent variable, while all other elements are fixed.
    b) Draw $x_i \propto \pi(x_i|x_1, x_,2, \dots, x_{i-1}, x_{i+1}, \dots, x_N)$ from the conditional distribution.
3. Set $\boldsymbol{X}_{n+1} = \boldsymbol{x}$.

The original variant of the Gibbs sampler updates the elements of the vector $\boldsymbol{x}$ componentwise. The Gibbs sampler can be viewed as a MH algorithm that uses the conditional distribution to generate a proposal. In this case α reaches $\alpha = 1$ and every proposal is accepted, which is a main difference between the Gibbs sampler and MH schemes.

To generate a sample from the conditional distribution a one dimensional sampling scheme is required. One possible method is given by inverse transform sampling [Dev86] which computes

$$\Phi_i(t) = \int_{-\infty}^{t} \pi(\boldsymbol{x}) dx_i, \tag{2.20}$$

and then generates a sample by

$$x_i = \Phi_i^{-1}(u). \tag{2.21}$$

where $u \propto \mathcal{U}([0, 1])$. Gibbs sampling has his historic origin for tasks, where the individual (conditional) distributions are known but the joint distribution is unknown. Hence, the joint distribution is found by samples generated from the different conditional distributions.

In the case that only the joint distribution is known, the conditional distribution for an element of $\boldsymbol{x}$ is in principle easy to access, as only the remaining elements of $\boldsymbol{x}$ have to be fixed. Then the conditional distribution can for example be approximated by using some support points. For problems with a complex underlying forward map this requires the repeated evaluation of the forward map. Thus, although each proposal is accepted, the evaluation of the conditional distribution can become computationally costly, which is a main drawback for the application of a Gibbs sampler to complex problems such like inverse problems.

2.2.5 Efficiency of MCMC Methods

A main issue about the application of MCMC methods is their efficiency in terms of producing independent samples form the distribution $\pi(\cdot)$. A measure for the efficiency of MCMC methods is the integrated auto correlation time (IACT) τ_{int}. The meaning of τ_{int} can be explained by the mean estimator

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i, \tag{2.22}$$

for the random variable X. For the case, that the samples of X are independent, the variance of the estimated mean value μ is [Kay93]

$$\sigma_\mu = \frac{\sigma}{N}, \tag{2.23}$$

where σ is the standard deviation of X. This does not hold for correlated samples from X. In this case σ_μ becomes

$$\sigma_\mu = \tau_{\text{int}}\frac{\sigma}{N} \tag{2.24}$$

Thus, τ_{int} presents the number of correlated samples, which have the same variance decreasing behavior as one independent sample [Gey92]. An efficient way to compute τ_{int} is presented in [Wol04]. As a low value of τ_{int} is of high interest, the design of the MH kernel has to be taken with care, as the kernel is responsible for the generation of new proposals. In this sense, the setup of a MH algorithm is always determined by an initial tuning phase of the kernel, which can become an expensive task, if the problem is of high dimension. For the case, that $\boldsymbol{x}$ contains a special data representation, i.e. a shape model, also special moves can be implemented to form the proposal kernel. This moves are mostly motivated by physical considerations and the setup of the kernel may therefore be seen as easier. An example for this is given in [WF09]. However, in the general case the setup of an MH algorithm remains a certain trial and error task. Various schemes for the design of optimal

MH algorithms are presented in [RR01]. The use of adaptive MH algorithms [CMP08] is in general a way to get rid of this problem, but adaptive methods also require a sophisticated setup in the initial phase in order to learn efficiently enough information about the distribution to adapt correctly. A self tuning algorithm which promises some advantages was presented in [CF10].

2.3 Calibration of Computer Models

This section deals with the steps necessary to get a computer model working for dealing with measurement data. So far all methods have been explained for the case, that the model $F(\cdot)$ is accurate enough to describe the behavior of $P(\cdot)$. In this case, the measured data $\tilde{\boldsymbol{d}}$ can be directly used as input data for the inversion algorithm.

The belief of the relation $F(\cdot) = P(\cdot)$ is in fact often more like a wish. The development of computer models and computational methods has seen enormous achievements in the past decades and highly complex large scale simulations like earthquake simulations, weather forecasting, geothermal simulations, etc. are possible today. However, inverse problems and at least the inverse problem presented in this work, often require a higher degree of accuracy. In this context often the two terms

- Absolute imaging,
- Differential imaging,

are used. Absolute imaging refers to the first kind of mentioned problems, where the measurements can be used for the inversion without any further preprocessing [VVSK99]. Differential imaging [AG96] belongs to the second class of described problems. Although this quantification is mainly used for the image reconstruction or the images space itself, it can also be used on the measurements or the data space as well. For the second case it has the meaning that calibration measurements are required. Then the algorithms work on the difference between the actual data and the calibration data. The practical issue of this selection depends on the degree of deviation caused by a change of $\boldsymbol{x}$ compared to the offset value.

Almost every model with the purpose to simulate a physical system will differ from the real system to a certain degree. This differences are caused by

- Neglected physical effects: Often physical effects are neglected in order to keep the forward map computationally simple. The best examples for this are electrical problems, where often Laplacian type partial differential equations (scalar) are used instead of vectorial Helmholtz type partial differential equations, which arise when the full Maxwells equations are solved.

- Incorrect boundary conditions: The assumed boundary conditions do not correspond to the reality. An electrical example for this are Neumann boundary conditions, where i.e. the normal component of the current density is assumed to be constant at the electrode. In fact, the current density at the edges of the electrodes are higher.

- Discretization effects: Every kind of discretization (space and/or time) causes an error. Hence, already the use of the discrete representation for $\boldsymbol{x}$ introduces an error.

- Differences in the geometry: The exact geometry is often not known. The wrong modelling of the domain becomes critical in regions of high sensitivities or wrong boundary conditions.

- Truncated domains: In many cases a 2D model is used, although the real system is always affected by 3D effects. Further, in the case of unbounded domains, the use of the finite element method requires the application of a far boundary, which leads to an unwanted large number of finite elements. Decreasing the far boundary causes larger interactions between the region of interest and the boundary, which lead to different effects.

- Unmodelled behavior of the measurement chain: the behavior of the measurement chain is almost ever neglected. However, every measurement system has at least a feedback effect on the measured object itself.

Hence, the absolute values computed by a model will differ from measured values even in the case, that the model is believed to be representing the exact physical system. If the difference is small enough with respect to the noise absolute imaging can be applied. Otherwise calibration schemes have to be introduced to handle the differences between the model $F(\cdot)$ and the physical process $P(\cdot)$.

As absolute imaging problems are obviously the less difficult case, the following subsections will deal with the handling of problems where differential imaging techniques will have to be applied.

2.3.1 A Standard Framework in Computer Science

As already mentioned with the point "Discretization effects" in the list of the previous section the notations used to introduce the topic are not fully compatible to the notations which are typically found in literature [KO00] on computer model calibration and also do not present all aspects of the problem.

So far all explanations and methods have been demonstrated for the state vector $\boldsymbol{x}$ which was stated as the model input for $F(\cdot)$ and the variables of interest. In fact the process $P(\cdot)$ rather depends on the (continuous) quantity ϕ. The variable $\boldsymbol{x}$ is obtained due to a certain mapping $x : \phi \mapsto \boldsymbol{x}$. An example for such a mapping is given by a finite element discretization, where ϕ is mapped on discrete volumes with piecewise constant properties. Although it is aim to determine ϕ, it is only possible to determine $\boldsymbol{x}$. With this equation (1.1) has in fact to be written as

$$\tilde{\boldsymbol{d}} = P(\phi) + \boldsymbol{v}. \tag{2.25}$$

A basic framework for the calibration of computer models was presented in [KO00]. In their work, the authors start with the assumption, that a computer model

$$\boldsymbol{y} = F(\boldsymbol{x}, \boldsymbol{\zeta}), \tag{2.26}$$

exists, which is adequate to simulate the behavior of the process. The variables $\boldsymbol{\zeta}$ are so called calibration variables. The calibration variables are fixed variables and can be understood as tuning parameters inside the computer model. Examples for such variables are i.e. geometry factors, material specifics, or sensitivities of simulated sensors. In this work we refer to $\boldsymbol{\zeta}$ as image space calibration variables, because it enters $F(\cdot)$ in the same way as the variable $\boldsymbol{x}$. The calibration parameters $\boldsymbol{\zeta}$ are fixed and known for each simulation of the model, but they differ from the true set of calibration parameters $\boldsymbol{\varsigma}$, which would be the ideal set of parameters. $\boldsymbol{x}$ is the vector of interest, which we want to identify from the measurements $\tilde{\boldsymbol{d}}$ after the model is calibrated.

To correct the remaining difference between the model and the true process, a regression parameter $\boldsymbol{\rho}$ and a model inadequacy function $D(\cdot)$ are introduced,

$$P(\phi) \approx \boldsymbol{\rho} F(\boldsymbol{x}, \boldsymbol{\zeta}) + D(\boldsymbol{x}). \tag{2.27}$$

The approach is similar to an offset and gain calibration scheme with $\boldsymbol{\rho}$ and $D(\cdot)$ working in the data space to correct the model $F(\cdot)$. The data to perform the calibration steps is given by calibration measurements $\tilde{\boldsymbol{d}}_c$ or $\boldsymbol{d}_c$, respectively. Hereby $\boldsymbol{d}_c$ denotes a measurement with a better signal to noise ratio (SNR) than $\tilde{\boldsymbol{d}}_c$. This can be achieved by averaging several calibration measurements. These measurements were taken for known values of ϕ_c. In the same, the model is simulated with the corresponding input data $\boldsymbol{x}_c$ in order to obtain the data $\boldsymbol{y}_c$.

The origin of the framework presented in [KO00] is in computer science. Because of this background, the work makes no specifics about the process or the model. Instead, Gaussian process models are used to model the behavior of $F(\cdot,\cdot)$ and $D(\cdot)$. The following subsection will give a brief introduction for.

Gaussian Processes

Gaussian process models are a favorable tool in statistical analysis and computer science [Ras06] to model the behavior between input data $\boldsymbol{x}$ and output data $\boldsymbol{y}$. The approach assumes a Gaussian behavior of form

$$\begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{y}^* \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{x}) & \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{x}^*) \\ \boldsymbol{\Sigma}(\boldsymbol{x}^*, \boldsymbol{x}) & \boldsymbol{\Sigma}(\boldsymbol{x}^*, \boldsymbol{x}^*) \end{bmatrix},\right) \tag{2.28}$$

for the relation between independent (input) data $\boldsymbol{x}$ and dependent output data $\boldsymbol{y}$. As a general approximation approach Gaussian process models are also sometimes applied to model the behavior of systems without any other considerations about the true model which is described by $y_i = F_i(x_i)$. The output $\boldsymbol{y}^*$ for the new input $\boldsymbol{x}^*$ can be predicted by

$$\boldsymbol{y}^* = \boldsymbol{\Sigma}(\boldsymbol{x}^*, \boldsymbol{x}) \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{x})^{-1} \boldsymbol{y}, \tag{2.29}$$

and further, the covariance of $\boldsymbol{y}^*$ given all the data can be quantified by

$$\operatorname{cov}(\boldsymbol{y}^*) = \boldsymbol{\Sigma}(\boldsymbol{x}^*, \boldsymbol{x}^*) - \boldsymbol{\Sigma}(\boldsymbol{x}^*, \boldsymbol{x}) \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{x})^{-1} \boldsymbol{\Sigma}(\boldsymbol{x}, \boldsymbol{x}^*). \tag{2.30}$$

The entries of the matrices $\boldsymbol{\Sigma}(\cdot,\cdot)$ are evaluated using an appropriate covariance function, which relates the input vectors. An often used type of covariance function is the exponential covariance function

$$\boldsymbol{\Sigma}(\boldsymbol{s},\boldsymbol{s}',\boldsymbol{\alpha}) = \lambda \exp\left(-\sum_i \frac{|\boldsymbol{s}-\boldsymbol{s}'|^{p_i}}{\phi_i}\right), \tag{2.31}$$

which belongs to the class of stationary functions, as only the distance between $\boldsymbol{s}$ and $\boldsymbol{s}'$ is evaluated. The vector $\boldsymbol{\alpha}$ is a hyper parameter, which contains all tuning variables of the function. The choice of $\boldsymbol{\alpha}$ is indeed a question on its own. Again, an inferential approach could be made by

$$\pi(\boldsymbol{\alpha}|\boldsymbol{x},\boldsymbol{y}) \propto \pi(\boldsymbol{y}|\boldsymbol{x},\boldsymbol{\alpha})\pi(\boldsymbol{\alpha}), \tag{2.32}$$

to obtain knowledge about the posteriori distribution over $\boldsymbol{\alpha}$. However, this approach is only usable for problems of comparatively low dimension. A more detailed explanation for the determination of $\boldsymbol{\alpha}$ can be found in [KO01].

A Standard Framework in Computer Science - Continued

The application of a Gaussian process model is quite common in computer science and statistical analysis. This is mainly due to the analytical accessibility of the Gaussian functions. However, for large models, the application of this approach requires the nontrivial procedure to find the hyper parameters for the functions $F(\cdot,\cdot)$ and $D(\cdot)$. The whole calibration and the inverse problem can then be treated by

$$\pi(\boldsymbol{x},\boldsymbol{\rho},\boldsymbol{\alpha}_{F,D}|\tilde{\boldsymbol{d}},\boldsymbol{d}_c) \propto \pi(\tilde{\boldsymbol{d}},\boldsymbol{d}_c|\boldsymbol{x},\boldsymbol{x}_c,\boldsymbol{\rho},\boldsymbol{\alpha}_{F,D})\pi(\boldsymbol{x})\pi(\boldsymbol{\alpha}_{F,D}), \tag{2.33}$$

where $\boldsymbol{\alpha}_{F,D}$ contains the hyper parameters for the Gaussian process models replacing $F(\cdot,\cdot)$ and $D(\cdot)$. Because of dimensionality reasons the direct approach in (2.33) is in general not applicable.

2.3.2 The Enhanced Error Model

In contrast to the previously discussed calibration methods in this section, the enhanced error model [KS05] is not primarily motivated by the problem of model inadequacy, which is a main part in [KO00]. Instead, the enhanced error model provides a suitable scheme to replace the accurate but computational expensive simulator $F(\cdot)$ by a reduced simulator $F^*(\cdot)$ without large losses in the quality in the result. The following assumptions are made for the development of the enhanced error model

1. There exists an exact physical model for the process $P(\cdot)$.
2. There exists a computer model $F(\cdot)$ which represents the physical process $P(\cdot)$ with enough accuracy, but the model may not be usable for the inversion as it is computationally to expensive.
3. There exists a computer model $F^*(\cdot)$, which is usable for the inversion.

Hence, the enhanced error model requires an absolute imaging problem. The enhanced error model starts with the expression

$$\boldsymbol{y} = F^*(\boldsymbol{x}^*) + (F(\boldsymbol{x}) - F^*(\boldsymbol{x}^*)). \tag{2.34}$$

One can see that $\boldsymbol{y}$ equals the output of $F(\cdot)$, but it is evaluated by using $F^*(\boldsymbol{x}^*)$. For this, the second part has to be seen as a correction term. It might also be the case, that $F^*(\cdot)$ operates on $\boldsymbol{x}$, but for using the same notation as in [KS05], $\boldsymbol{x}^*$ is used. When using the reduced model $F^*(\cdot)$, the correction term

$$\boldsymbol{e}_n = F(\boldsymbol{x}_n) - F^*(\boldsymbol{x}_n^*), \tag{2.35}$$

presents the deterministic model error. The idea of the enhanced error model is know to apply a stochastic model for the correction term. Samples using any available prior knowledge are drawn and computed on both models. This does not mean that only samples from $\boldsymbol{x}$ are drawn, but also samples from $\boldsymbol{\zeta}$ can be included to incorporate knowledge about the uncertainty of model parameters. Then, Gaussian distributions are computed to describe the model error by

$$\boldsymbol{e} \propto \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{e}}, \boldsymbol{\Sigma}_{\boldsymbol{e}}), \tag{2.36}$$

and the prior knowledge by

$$\boldsymbol{x} \propto \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{x}}, \boldsymbol{\Sigma}_{\boldsymbol{x}}), \tag{2.37}$$

where $\boldsymbol{\mu}_{(\cdot)}$ is the mean and $\boldsymbol{\Sigma}_{(\cdot)}$ is the corresponding covariance matrix. Hence, the step equals a marginalization of $\boldsymbol{e}$ from $\boldsymbol{x}$ as well as the uncertain calibration parameters $\boldsymbol{\zeta}$ [KS05]. Subsequently, either a sampling algorithm or, because of the analytical expressions, a MAP based optimization can be used to solve the inverse problem. The prior based determination of $\boldsymbol{\mu}_{\boldsymbol{e}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{e}}$ leads to a global description of the model error. Local behavior of the error gets therefore lost and also the fact that the marginal error over the posterior distribution may be quite different to the marginal error over the prior makes the global treatment of $\boldsymbol{e}_n$ not be an optimal choice.

It should be mentioned that the enhanced error model can also be used to get rid of uncertain calibration parameters $\boldsymbol{\zeta}$. In this case, $\boldsymbol{\zeta}$ is treated as random variable during the sampling procedure to build the enhanced error model. After the reconstruction it is even possible to determine the real value of $\boldsymbol{\zeta}$.

Posterior Based Update of the Enhanced Error Model

A suitable way to update the mean and covariance structure of the enhanced error model appears within the DAMH structure, as the acceptance of a proposal generated by $F^*(\cdot)$ leads to an exact evaluation of the accurate model $F(\cdot)$. The following update equations can be used [Cui10], to build an adaptive version of the enhanced error model [CFO$^+$10], [FCN11]

$$\begin{aligned} \boldsymbol{\mu}_{\boldsymbol{e},n} &= \frac{1}{n}\left((n-1)\boldsymbol{\mu}_{\boldsymbol{e},n-1} + \boldsymbol{e}_n\right), && (2.38)\\ \boldsymbol{C}_{\boldsymbol{e},n} &= \boldsymbol{C}_{\boldsymbol{e},n-1} + \boldsymbol{e}_n\boldsymbol{e}_n^T, && (2.39)\\ \boldsymbol{\Sigma}_{\boldsymbol{e},n} &= \frac{1}{n-1}\left((n-1)\boldsymbol{C}_{\boldsymbol{e},n} - n\boldsymbol{\mu}_{\boldsymbol{e},n}\boldsymbol{\mu}_{\boldsymbol{e},n}^T\right). && (2.40) \end{aligned}$$

The enhanced error model learns automatically about the local properties of the error and the likelihood function is given by

$$\pi_{\boldsymbol{x}^*}(\boldsymbol{x}|\tilde{\boldsymbol{d}}) = \exp\left\{-\frac{1}{2}\left(\boldsymbol{y}^* + \boldsymbol{\mu}_{\boldsymbol{e},n} - \tilde{\boldsymbol{d}}\right)^T (\boldsymbol{\Sigma}_{\boldsymbol{v}} + \boldsymbol{\Sigma}_{\boldsymbol{e},n})^{-1}\left(\boldsymbol{y}^* + \boldsymbol{\mu}_{\boldsymbol{e},n} - \tilde{\boldsymbol{d}}\right)\right\}. \tag{2.41}$$

Hence, an automatic error learning within the DAMH algorithm can be established.

2.3.3 A More Detailed Look on Calibration

At least two aspects in metrological and model based inverse problems lead to a difficulty with the calibration framework presented in section 2.3.1.

- Good knowledge about the computer model $F(\cdot,\cdot)$ is present.
- The inadequacy function $D(\cdot)$ is not known.

The first point addresses the fact that at least good knowledge about the physical behavior is available, and that the implemented computer model is sufficient to solve the equations in a correct manner. This means that the model is at least able to provide correct trends, although its not correct in its absolute values. The replacement of $F(\cdot,\cdot)$ by a Gaussian process seems less adequate. The second point means, that knowledge about the additive model inadequacy is typically not given. The only thing which one could do, is to compare the model against a model which is believed to be more accurate. Based on the difference an inadequacy function $D(\cdot)$ could be designed using a Gaussian regression model. However, even the more accurate model will not meet the real process for differential imaging problems and hence the inadequacy function is in this case useless.

The calibration strategy presented in [KO00] aims more on the determination of the $\boldsymbol{\zeta}$. A detail about the calibration strategy for $\boldsymbol{\zeta}$ is the general allowance of even physically impossible values for $\boldsymbol{\zeta}$. This seems to be an odd fact. Certain elements of $\boldsymbol{\zeta}$ might have larger or lower values compared to their nominal value in order to compensate other effects, but the elements should not reach impossible values, even if they would lead too a perfect fitted model. An impossible or infeasible value for $\boldsymbol{\zeta}$ should more be an indicator, that a certain effect is neglected. In this work $\boldsymbol{\zeta}$ should be set to a fixed value although the actual realization may differ from the true value. Consequently the calibration is performed in the sense of the corrections by $\boldsymbol{\rho}$ and $D(\cdot)$ for the given model as given in equation (2.27).

Because of the problem of designing an inadequacy function equation (2.27) will be simplified to

$$P(\boldsymbol{x}) \approx \boldsymbol{\rho} F(\boldsymbol{x}, \boldsymbol{\zeta}) + \boldsymbol{c}, \tag{2.42}$$

which is again an offset gain calibration, but with the constant $\boldsymbol{c}$ replacing the inadequacy function. The affine approach in equation (2.42) means, that we assume that $F(\cdot,\cdot)$ gives the correct trends, but has a wrong gain and is shifted by an offset. In this sense the affine approach is also a quality measure for the model, as a nonlinear calibration function would mean, that the model does not even provide the correct trends.

Retaining the calibration scheme (2.42), the question is about the determination of $\boldsymbol{\rho}$ and $\boldsymbol{c}$ out of the calibration data $\boldsymbol{d}_c$ and the meaning of this procedure. Collecting the parameters $\boldsymbol{\rho}$ and $\boldsymbol{c}$ with the hyper parameter

$\boldsymbol{\xi} = \{\boldsymbol{\rho}, \boldsymbol{c}\}$, the whole Bayesian inferencial problem can be written as

$$\pi(\boldsymbol{x}, \boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c) \propto \pi(\tilde{\boldsymbol{d}}, \boldsymbol{d_c}|\boldsymbol{x}, \boldsymbol{\xi})\pi(\boldsymbol{x})\pi(\boldsymbol{\xi}). \tag{2.43}$$

Using the product rule, the left hand term of equation (2.43) can be expressed by

$$\pi(\boldsymbol{x}, \boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c) \propto \pi(\boldsymbol{x}|\boldsymbol{\xi}, \tilde{\boldsymbol{d}}, \boldsymbol{d}_c)\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c). \tag{2.44}$$

The term $\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$ offers the dependency of the calibration data $\boldsymbol{\xi}$ to the available information $\tilde{\boldsymbol{d}}$ and $\boldsymbol{d}_c$. It seems odd, that the calibration vector $\boldsymbol{\xi}$ itself relies on the data $\tilde{\boldsymbol{d}}$. However, this expresses the fact, that a parameter set $\boldsymbol{\xi}$ can be found, which provides the best fit of the model output for the measured data. The dependency of $\boldsymbol{\xi}$ on the data $\tilde{\boldsymbol{d}}$ seems clear, but due to the structure of the calibration scheme (2.42) it remains unfavorable, as the fitting of the model towards the data can also be done only by $\boldsymbol{\xi}$.

2.3.4 An Empirical Bayes Approach

Knowing that $\boldsymbol{\xi}$ also depends on the data $\tilde{\boldsymbol{d}}$ and that $\boldsymbol{\xi}$ is a random variable described by the distribution $\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$, it is in practice often the case that one just uses a meaningful fixed value for $\boldsymbol{\xi}$. It is also a computationally cheaper way to treat the problem as no computational inference about $\boldsymbol{\xi}$ is required.

This means, that a sample from the distribution $\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$ has to be generated. As this can only be done with the calibration data $\boldsymbol{d}_c$, the data $\tilde{\boldsymbol{d}}$ is neglected. For the determination of the elements of $\boldsymbol{\xi}$, the affine calibration approach requires at least two calibration measurements $\boldsymbol{d}_c$, to obtain the components of $\boldsymbol{\rho}$ and $\boldsymbol{c}$ for each output of the model $F(\cdot)$.

For the case of several calibration measurements, a typical and obvious approach to find $\boldsymbol{\xi}$ is given by a least squares fit of form

$$\boldsymbol{\xi}_{\mathrm{LSQ}} = \arg \min_{\boldsymbol{\xi}=\{\boldsymbol{\rho},\boldsymbol{c}\}} ||\boldsymbol{\rho} F(\boldsymbol{x}_c) + \boldsymbol{c} - \boldsymbol{d}_c||_2^2. \tag{2.45}$$

For the case of only two measurements, the solution of (2.45) is directly given by

$$\begin{aligned} \boldsymbol{d} &= \frac{\boldsymbol{d}_{c,H} - \boldsymbol{d}_{c,L}}{\boldsymbol{y}_{c,H} - \boldsymbol{y}_{c,L}} \left(F(\boldsymbol{x}) - \boldsymbol{y}_{c,L}\right) + \boldsymbol{d}_{c,L}, && (2.46) \\ &= \underbrace{\frac{\boldsymbol{d}_{c,H} - \boldsymbol{d}_{c,L}}{\boldsymbol{y}_{c,H} - \boldsymbol{y}_{c,L}}}_{\boldsymbol{\rho}} F(\boldsymbol{x}) + \underbrace{\boldsymbol{d}_{c,L} \left(1 - \frac{\boldsymbol{d}_{c,H} - \boldsymbol{d}_{c,L}}{\boldsymbol{y}_{c,H} - \boldsymbol{y}_{c,L}}\right)}_{\boldsymbol{c}}, && (2.47) \end{aligned}$$

where $\boldsymbol{d}_{c,L}$ and $\boldsymbol{d}_{c,H}$ are the two calibration measurements and the subscript L (low) and H (high) should denote two calibration points. Looking, at equation (2.46), the approach fixes the model on the end points. This is quite an obvious decision for the choice of $\boldsymbol{\xi}$ and can be referred to as a two point calibration.

In Bayesian terms, this calibration strategy follows the rules of empirical Bayes approaches [Cas85]. Empirical Bayes approaches are characterized by the approach to describe a distribution $\pi(\cdot)$ by knowledge from existing samples from the distribution. One can also say that one is conditioning on the best estimate of some nuisance parameters such as the calibration parameters. As the measurements are actually not used, equation (2.44) reduces to

$$\pi(\boldsymbol{x}, \boldsymbol{\xi} | \tilde{\boldsymbol{d}}, \boldsymbol{d}_c) \approx \pi(\boldsymbol{x} | \boldsymbol{\xi}, \tilde{\boldsymbol{d}}, \boldsymbol{d}_c) \pi(\boldsymbol{\xi} | \boldsymbol{d}_c), \tag{2.48}$$

and by the knowledge, that a least squares fit equals the ML estimate in the case of Gaussian noise, the two point calibration provides the ML estimate for $\boldsymbol{\xi}$ of the distribution $\pi(\boldsymbol{\xi} | \boldsymbol{d}_c)$.

2.3.5 A Full Bayesian Approach for Mutual Inference

The calibration based on the empirical Bayes approach is in many cases a sufficient strategy to solve the given inverse problem with the given computer code.

The approach of fixing the parameter $\boldsymbol{\xi}$ from calibration data $\boldsymbol{d}_c$ makes the model in some sense only optimal for the calibration data. In the case that $\boldsymbol{\xi}$ also depends on the data $\tilde{\boldsymbol{d}}$, a calibration in the empirical Bayesian sense could lead to the behavior, that the determination of an unknown $\boldsymbol{x}$ out of data $\tilde{\boldsymbol{d}}$ is only possible, if $\boldsymbol{x}$ is somehow close to $\boldsymbol{x}_c$. In this case, $\boldsymbol{\xi}$ has to be allowed to change its value, which means, that it has to become part of the estimation process. The extended state vector to be estimated is given by $\boldsymbol{x}_e^T = \left[\begin{array}{cc} \boldsymbol{x}^T & \boldsymbol{\xi}^T \end{array}\right]$, but for clarity in the following parts, $\boldsymbol{x}$ and $\boldsymbol{\xi}$ will be treated separately. For this mutual inference approach equation (2.43) has to be rewritten in the form

$$\pi(\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N_c)} | \tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N_c)}) \propto \pi(\tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N_c)} | \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N_c)}) \pi(\boldsymbol{x}) \pi(\boldsymbol{\xi}) \pi(\boldsymbol{x}_c^{(N_c)}), \quad (2.49)$$

where $\boldsymbol{d}_c^{(N_c)}$ collects the N_c calibration measurements and $\boldsymbol{x}_c^{(N_c)}$ collects the corresponding calibration inputs. As the calibration measurements have equal prior probability $\pi(\boldsymbol{x}_c^{(N)})$ is constant and can be skipped. The same holds in principle for $\pi(\boldsymbol{\xi})$, but this is because of the lack of knowledge about $\boldsymbol{\xi}$. To ensure a proper behavior of the chain, a uniform distribution with some respectful range centered around the least squares estimate can be applied.

For the likelihood function $\pi(\tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N_c)} | \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N_c)})$, an approach similar to applications in radio carbon dating can be used [CCL95]. Hereby, the likelihood function is expressed by the product

$$\pi(\tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N_c)} | \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N_c)}) \propto \pi(\tilde{\boldsymbol{d}} | \boldsymbol{x}, \boldsymbol{\xi}) \prod_{i=1}^{N_c} \omega_i \pi(\boldsymbol{d}_c^i | \boldsymbol{\xi}, \boldsymbol{x}_c^i), \quad (2.50)$$

which means that calibration measurements and the current measurement are independent. ω_i is a weighting factor for the calibration data and presents the trust in the calibration measurements so ω_i corresponds to $\pi(\boldsymbol{x}_c^i)$. The approach itself is computational inexpensive in the sense, that the forward map has only to be evaluated for a new proposal of $\boldsymbol{x}$ but not for $\boldsymbol{\xi}$. However, the number of variables for determining the elements of $\boldsymbol{\xi}$ is still twice the number of measurements. Hence, an MCMC algorithm for determining both, $\boldsymbol{x}$ and $\boldsymbol{\xi}$, will require a sophisticated setup to be efficient.

2.3.6 Sampling with a Distribution $\pi(\boldsymbol{\xi}|\cdot)$

The empirical Bayes approach and the full Bayesian calibration approach for mutual inference are a logical consequence of the likelihood function $\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$ of equation (2.44). The empirical Bayes approach gets rid of it by fixing $\boldsymbol{\xi}$. This step requires a certain accuracy, as a wrong choice of $\boldsymbol{\xi}$ will bias the solution. The full Bayesian approach takes the dependency of $\boldsymbol{\xi}$ on the data into care, but this times $\boldsymbol{\xi}$ becomes an unknown variable to estimate itself. This is not to worry about, but the dependency on the data remains in some sense unfavorable, as it could happen, that the calibration just fits the model output towards the data $\tilde{\boldsymbol{d}}$. A more preferable posterior distribution would be of kind

$$\pi(\boldsymbol{x}, \boldsymbol{\xi} | \tilde{\boldsymbol{d}}, \tilde{\boldsymbol{d}}_c) \propto \pi(\tilde{\boldsymbol{d}}, \boldsymbol{\xi} | \boldsymbol{x}) \pi(\boldsymbol{x}) \pi(\boldsymbol{\xi} | \boldsymbol{d}_c). \tag{2.51}$$

The meaning of the scheme of this posterior distribution is, that a prior distribution of $\boldsymbol{\xi}$ is generated based on calibration data $\boldsymbol{d}_c$. Then the sampling is performed in the way, that in a first samples of $\boldsymbol{x}$ are generated for a fixed $\boldsymbol{\xi}$. In a second step samples from $\boldsymbol{\xi}$ are generated from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$.

2.3.7 A Stochastic Forward Map Approach

So far, the calibration data has been used to design likelihood functions $\pi(\tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N_c)} | \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N_c)})$ and taking $\boldsymbol{\xi}$ as part of the sampling problem, which is at least one way to incorporate calibration information.

A different approach is the idea of using calibration measurements to form a so called stochastic forward map. The known error obtained by the calibration measurements is used to form an error or uncertainty estimate of the model output for the current state. Hence, the otherwise deterministic forward map becomes a statistical model itself, which assigns a quality measure to its output variables. Subsequently, the deterministic output $\boldsymbol{y}$ should be replaced by $\boldsymbol{\mathcal{Y}}$. An approach for such a model is given by

$$\boldsymbol{\mathcal{Y}} = \tilde{F}(\boldsymbol{x}) + \mathcal{D}(\cdot). \tag{2.52}$$

This structure is in principal the same as the structure of (2.27) presented in [KO00]. The function $\tilde{F}(\boldsymbol{x})$ should present the best available deterministic model, which could include a calibration like

$$\tilde{F}(\boldsymbol{x}) = \boldsymbol{\rho} F(\boldsymbol{x}) + \boldsymbol{c}. \tag{2.53}$$

The function $\mathcal{D}(\cdot)$ is again a model inadequacy function. It can compensate the model error, but its main sense is to incorporate statistical knowledge about $\tilde{F}(\boldsymbol{x})$. I.e. if it is known from calibration data, that some model outputs suffer from a certain set of data, the model inadequacy function should present this information in form of an uncertainty statement about the model, if $\boldsymbol{x}$ gets close to the region where the calibration measurement is done. Then the likelihood is given by

$$\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mathcal{Y}}-\tilde{\boldsymbol{d}}\right)^T (\boldsymbol{\Sigma}_{\boldsymbol{v}}+\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}})^{-1}\left(\boldsymbol{\mathcal{Y}}-\tilde{\boldsymbol{d}}\right)\right\}, \tag{2.54}$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\mathcal{Y}}}$ presents the uncertainty of $\tilde{F}(\boldsymbol{x})$ in the sense of a covariance matrix. Although the idea of such a function $\mathcal{D}(\cdot)$ is clear, the design becomes a demanding and problem dependent task. As a last statement, it should be stated, that the way to design a function $\mathcal{D}(\cdot)$, or algorithms in general, can be separated into

- Image space based methods: $\mathcal{D} = \mathcal{D}(\boldsymbol{x}, \cdot)$.
- Data space based methods: $\mathcal{D} = \mathcal{D}(\tilde{\boldsymbol{d}}, \cdot)$.

2.4 Summary

In this chapter the basics of Bayesian inversion theory and inference approaches were introduced. After discussing the general idea of using a Bayesian formulation the powerful tool of Monte Carlo integration is explained and the issue of statistical algorithms is treated. Two different sampling algorithms, the MH algorithm and the Gibbs sampler were presented and their behavior in concern of sampling from the posterior distribution from inverse problems is discussed. For the MH algorithm also an accelerated variant called the DAMH algorithm is presented and a framework for adaptive error learning is introduced in order to improve the behavior of the algorithm.

Then the topic of model errors and Bayesian formulations for differential imaging problems is discussed. Standard approaches from computer science are reviewed for their appropriateness to the solution of inverse problems and appropriate versions are derived. The aspects about this calibration schemes are explained in a Bayesian sense and two type of auto calibration algorithms are suggested and discussed. Finally the idea of a stochastic forward map approach is introduced. This approach uses calibration information to quantify the quality of the output of the forward map.

Throughout the chapter a general notation for all quantities is used allowing its application to a broader field of problems. After introducing computational techniques like or approximated forward map $F^*(\cdot)$ in the next chapter the introduced algorithms and strategies are tested in chapter 4.

3 Computational Techniques for Linear PDE Systems

In this chapter, computational methods for the ECT forward map are developed. After reviewing the set of standard techniques to solve the forward problem and build the forward map the first part of this chapter presents a computational framework using a Green's functions approach which offers enormous computational advantages. Although the presented methods have a special focus on ECT, they are in general applicable to problems with an underlying linear (partial) differential equation in a self adjoint problem. For this class of systems, the stiffness matrix of the finite element system is a linear function of the unknown coefficients. In this sense the system matrix can be expresses in form of the matrix product $\boldsymbol{WSW}^T$.

The second part of this chapter deals with approximation techniques to replace the computational expensive but accurate model $F(\cdot)$ by a computationally cheaper model $F^*(\cdot)$, or even by a so called surrogate model.

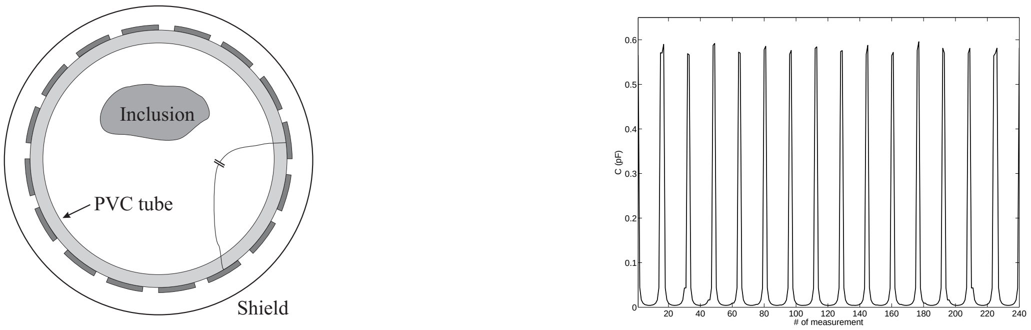
3.1 Electrical Capacitance Tomography

This section provides a short introduction to the inverse problem of electrical capacitance tomography (ECT). As the focus of the chapter lies on the ECT forward map and computational methods for it, the topic of process tomography, which is one major application of ECT will only be touched in a slender way. The following list should collect some topics in the field and the application of ECT for process tomography, together with some major references.

- ECT sensors: [HDX$^+$88], [Yan10], [JB00].
- ECT hardware: [Yan96], [WHL09], [WFHK05], [WFHK08].
- Process tomography: [Gla97], [BHW03].
- Process control: [SM05], [BWLW02].

- General reconstruction methods: [YP03], [LFY$^+$04] ,[Isa96], [NZWF11].
- Statistical approaches: [WBS07], [WF08], [SWF08], [WF09].
- Sensor Fusion: [SWW05], [DSSW08].
- Applications: [Whi02], [WHWB96], [FZL07], [HXW$^+$94], [CP05], [HL10].
- Others: [WSF$^+$07], [FZWN08], [SW08], [NSWZ10].

ECT is an inverse problem, which uses capacitance measurements for determining information about a certain region of interest Ω_{ROI}. Figure 3.1(a) depicts a typical scheme for a 2D ECT system suitable for process tomography, where it is well suited due to the good contrast in terms of the permittivity of typical industrial processes like multiphase flows. A number of N_{elec} electrodes are mounted on the exterior of a process pipe (PVC tube). By measurements of the inter-electrode capacitances the spatial dielectric permittivity distribution in Ω_{ROI} can be determined. To take measurement of the capacitances typically an AC-voltage is applied to one of the electrodes and the displacement currents on the other electrodes are measured. Figure 3.1(b) depicts a typical measurement pattern.



(a) Scheme of an ECT sensor.

(b) Typical measurement pattern.

Figure 3.1: Scheme of an ECT sensor and typical measurement pattern.

Following [BHW03], the electrical effects inside an ECT sensor are described by the potential equation

$$\nabla \cdot (\varepsilon_0 \varepsilon_r \nabla V) = 0, \tag{3.1}$$

where V is the electric scalar potential and $\varepsilon_0 = 8.854 \times 10^{-12}\,\mathrm{AsV^{-1}m^{-1}}$ is the absolute permittivity. ε_r denotes the dimensionless relative permittivity. For the explained measurement modi the boundary conditions are given by

$$\begin{aligned} V_{\partial\Omega} &= 0, && (3.2)\\ V_{\Gamma_j} &= V_0, && (3.3)\\ V_{\Gamma_i} &= 0 \qquad \forall j \neq i, && (3.4) \end{aligned}$$

where $\partial\Omega$ denotes the boundary of the problem domain and Γ_j denotes the surface of an individual electrode. After solving (3.1) for the specified boundary conditions, the capacitances $C_{i,j}$ between electrode i and electrode j can be computed by

$$C_{i,j} = -\frac{1}{V_0}\oint_{\Gamma_i} \varepsilon_0 \varepsilon_r \nabla V_j \cdot \vec{\mathbf{n}} d\Gamma \qquad i \neq j. \tag{3.5}$$

The computation of the capacitances given the material distribution is referred to as forward map $F : \varepsilon \mapsto \boldsymbol{C}$. This model does not include any other effects in the real process $P : \varepsilon \mapsto \boldsymbol{C}$, like 3D effects, specifics of measurement electronics, etc.. The effect of this physical approximation has to be removed by calibration. As an analytic solution is typically not available, numerical schemes like the finite element method (FEM) or the boundary element method (BEM) have to be used to build a numerical evaluated forward map $F : \boldsymbol{\varepsilon} \mapsto \boldsymbol{C}$. In the case of the FEM it is common to assign the permittivity value of each finite element in Ω_{ROI} to the entries of the vector $\boldsymbol{x} = \boldsymbol{\varepsilon}$. In the case of the BEM typically shape models are used to describe the contour of an inclusion. Hence, the parameters of the shape model build the entries of $\boldsymbol{x}$.

In terms of absolute and differential imaging, ECT generally belongs to the class of differential imaging problems. Typically offset capacitances between the electrodes are in the range of pF, whereas the changes caused by inclusions are in the range of fF. As ECT belongs to the class of problems where the signal deviation is only a fraction of the offset value the application of calibration measurements becomes necessary and hence ECT is truly a differential imaging problem.

3.1.1 Data Representation

An issue which comes to hand in any higher dimensional problem is about the representation of the results. Typically, a classification is made between

- Low level representation.
- Mid level representation.
- High level representation.

Low level and mid level representations build representations about the material distribution. Pixel like image representations are a classical example for low level representations. For ECT this kind of representation is typically associated with the material values of the finite elements. This scheme can be applied to any problem but is disadvantageous for further processing and the image quality of such schemes is generally low. Mid level schemes typically utilize a generic model to describe the data in a more specific way. A typical example for imaging problems is i.e. the use of some kind of shape model to describe the boundary of an inclusion. Hence mid level representations typically allow a more informative representation of the data and also permit a suitable access for further information extraction (i.e. the evaluations of areas or volumes). High level representations feature more specific issues about the data and are suitable for answering more general questions like "How many inclusions or objects are inside the domain Ω_{ROI}?".

3.2 Standard Solution Techniques for ECT

This section briefly explains the standard computation steps for the ECT forward map using the finite element method [Pol06].

3.2.1 Finite Element Forward Model

Solving the partial differential equations (3.1) for the boundary conditions (3.2) to (3.4) by a numerical scheme like the finite element method or the boundary element method ends up in a linear equation system of the form

$$\hat{\boldsymbol{K}}\boldsymbol{v} = \boldsymbol{r}. \tag{3.6}$$

In the case of the FEM, the scalar potential V from equation (3.1) is approximated by

$$V \approx \sum_{i=1}^{p+1} v_i N_i, \tag{3.7}$$

in the domain Ω_e, which is the domain of the finite element. v_i denotes the weights of linear independent basis functions N_i. The stiffness matrix $\boldsymbol{K}$ is assembled out of the N_e element matrices $\boldsymbol{K}_e$ by

$$\boldsymbol{K} = \sum_{i=1}^{N_e} \varepsilon_i \boldsymbol{K}_{e,i}, \tag{3.8}$$

where ε_i is the permittivity of the finite element. For this the element matrices are computed by evaluating the coefficients using the Ritz-Galerkin equation

$$k_{i,j} = \int_{\Omega_e} \left(\frac{\partial N_i}{\partial x} \frac{\partial N_j}{\partial x} + \frac{\partial N_i}{\partial y} \frac{\partial N_j}{\partial y} \right) d\Omega. \tag{3.9}$$

Then the matrix $\hat{\boldsymbol{K}}$ and the right hand side vector $\boldsymbol{v}$ are formed by applying the Dirchlet type boundary conditions to $\boldsymbol{K}$ and $\boldsymbol{v}$ by matrix manipulations.

3.2.2 Charge Computation

The evaluation of equation (3.5) is referred to as charge computation. To avoid computing ∇V the charge method [YSW98] can be applied. The computation is given by

$$Q_{\text{elec}} = \sum_{n_{\text{elec}}} (\boldsymbol{K}\boldsymbol{v})_{n_{\text{elec}}}, \tag{3.10}$$

where n_{elec} refers to the nodes of the finite element mesh, which are situated on the boundary of the electrode.

3.2.3 Jacobian Computation

An important property of systems is the Jacobian matrix $\boldsymbol{J}$. For the described ECT forward map the Jacobian is defined by

$$\boldsymbol{J} = \left[\begin{array}{ccccc} \nabla C_1 & \nabla C_2 & C_3 & \dots & \nabla C_N \end{array} \right]^T. \tag{3.11}$$

The Jacobian contains the derivatives of the output variables with respect to the input variables. An existing efficient method to compute the Jacobian is the adjoint variable method (AVM) [Bra03]. The elements of $\boldsymbol{J}$ can be computed by

$$dC_{i,j} = \boldsymbol{\gamma}_i^T \left[\left[\frac{\partial \boldsymbol{r}}{\partial \varepsilon_k} \right] - \left[\frac{\partial \hat{\boldsymbol{K}}}{\partial \varepsilon_k} \right] \boldsymbol{v}_j \right] d\varepsilon_k, \tag{3.12}$$

where j denotes the number of the active electrode, i denotes the number of the receiver electrode and k is the number of the finite element. $\boldsymbol{\gamma}$ is the solution of the adjoint problem

$$\hat{\boldsymbol{K}} \boldsymbol{\gamma}_i = \left[\frac{\partial Q_{i,j}}{\partial \boldsymbol{v}_j} \right]. \tag{3.13}$$

The adjoint variable method suggests the computation of the Jacobian at the cost of an additional forward problem and can be seen as a quite efficient and straight forward method to obtain derivatives.

3.3 A Green's Function Approach

The standard computation steps seem obvious as they correspond to the governing physical laws. Also the AMV method to compute the Jacobian seems efficient, as only one additional forward problem has to be solved and by maintaining efficient coding techniques it seems obvious that a fast performing computer code can be built.

However, in fact these computation steps offer some drawbacks. In this subsection numerical methods based on the use of Green's functions are introduced, which provide computational advantages, i.e. the approach allows operations on the Jacobian matrix without solving another forward problem. This means that one is able to compute matrix vector multiplications of the Jacobian or its transpose without the explicit evaluation of the Jacobian. Also an exact fast low-rank update scheme based on the Woodbury formula is developed. By this, a line search can be performed at decreased computational costs. The first two subsections provide an efficient way to assembly the stiffness matrix and to compute the charges. These methods provide the basis for some later introduced computations. Some of the techniques will also be used in subsection 3.5.2 to build a reduced forward model. Further, the forward map $F : \boldsymbol{\varepsilon} \mapsto \boldsymbol{C}$ is replaced by $F : \boldsymbol{\varepsilon} \mapsto \boldsymbol{Q}$ of convenience reasons.

3.3.1 Fast Stiffness Update

To update the stiffness matrix after a material update requires a reevaluation of equation (3.8). To keep the number of operations low typically a stiffness matrix $\boldsymbol{K}_{\text{ini}}$ is formed out of the non effected finite elements. Thus, the sum in equation (3.8) is reduced to N_{ROI}, which is the number of finite elements in Ω_{ROI}. However, the update still requires a loop operation, which is known to be slow in languages like MATLAB.

A significantly faster method based on an orthogonal decomposition is mentioned in [Fox05]. For any square matrix $\boldsymbol{A}$ an eigenvector decomposition of form

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^{-1}, \tag{3.14}$$

can be made, where the columns of $\boldsymbol{Q}$ are the eigenvectors and $\boldsymbol{D}$ is a diagonal matrix of the eigenvalues. For $\boldsymbol{A}$ being a symmetric matrix, which is the case for the element matrices, the inverse in equation (3.14) can be replaced by its transpose. By assembling the sparse matrices $\boldsymbol{W}_l$ out of the eigenvectors of the element matrices multiplied by the square root of the corresponding eigenvalue, it is possible the compute the material update by

$$\boldsymbol{K} = \boldsymbol{K}_{\text{ini}} + \sum_{l=1}^{p} \boldsymbol{W}_l \boldsymbol{\mathcal{E}} \boldsymbol{W}_l^T, \tag{3.15}$$

where p is the rank of the element matrix which is the order of the finite element minus one. $\boldsymbol{\mathcal{E}}$ is a (sparse) diagonal matrix of permittivity values. As only sparse matrix operations are required, this form of the material update is exceptional fast when using linear algebra packages and in the case of linear triangular finite elements even only $p = 2$ sums are necessary.

Due to the specifics of the geometry of the ECT sensor a further advantage can be used. As the interior of the pipe and the electrodes are separated by the tube wall, the boundary conditions do not affect the interior. Thus, the initial stiffness matrix $\boldsymbol{K}_{\text{ini}}$ can already contain the Dirichlet type boundary conditions and after the material update $\boldsymbol{K}$ equals $\hat{\boldsymbol{K}}$.

3.3.2 Charge Map

The charge map is referred to as an approach to compute the charges on the electrodes as a function of the potential distribution on $\partial\Omega_{\text{ROI}}$, which is the interior boundary of the tube. To be more precisely this approach of computing the charges splits the charges into a constant part and a part which depends on $V_{\partial\Omega_{\mathrm{ROI}}}$. For this the linear map $Q_c : V_{\partial\Omega_{\mathrm{ROI}}} \mapsto \Delta\boldsymbol{Q}$ is introduced. The idea of this concept may appear to be without any benefit, as the charge method presented in 3.2.2 is already highly efficient. However, there are two major points:

- In section 3.5.2 a matrix manipulation will be used to reduce the dimension of the stiffness matrix in order to solve the problem only for the potential in the domain Ω_{ROI}. Then, the charge map is used to compute the charges out of the potential distribution on the interior boundary of the tube.
- In the following an approach using Green's functions is explained. In this concept the use of the charge map appears in a natural way and offers several advantages (for example see 3.3.4 or 3.3.5).

Further, the charge method requires the stiffness matrix $\boldsymbol{K}$, which is not necessary when using the charge map.

The application and computation of the charge map is motivated by thinking about the problem in terms of an electrode arrangement. From linear network theory it is known that the relation between the potentials on the electrodes and the charges on the electrodes are related by the capacitance matrix $\boldsymbol{C}$. To determine the elements of $\boldsymbol{C}$ a voltage is applied to one electrode, while all other electrodes are grounded. Then the charges on the electrodes are determined and the elements of the capacitance matrix can be found using the relation $q = Cu$. For the computation of the charge map this means solving (3.1) in the domain $\Omega \setminus \Omega_{\mathrm{ROI}}$, where the boundary conditions are given by

$$\begin{aligned} V_{\partial\Omega} &= 0, && (3.16)\\ V_{\partial\Omega_{\mathrm{ROI}}} &= \delta(\boldsymbol{z} - \boldsymbol{z}_i) \qquad \forall i && (3.17)\\ V_{\Gamma_j} &= 0 \qquad \forall j, && (3.18) \end{aligned}$$

for $\boldsymbol{z} \in \partial\Omega_{ROI}$. In the case of the finite element implementation the coordinates $\boldsymbol{z}_i$ are given by the coordinates of the nodes on Ω_{ROI}. Then the discrete version of Q_c becomes the $N_{\mathrm{elec}} \times N_{\partial\Omega_{\mathrm{ROI}}}$ matrix $\boldsymbol{Q}_c : \boldsymbol{v}_{\partial\Omega_{\mathrm{ROI}}} \mapsto \Delta\boldsymbol{Q}$, where $N_{\partial\Omega_{ROI}}$ is the number of boundary nodes. To compute the total charges on the electrodes, an affine part given by the $N_{\mathrm{elec}} \times N_{\mathrm{elec}}$ matrix $\boldsymbol{Q}_a$ has to be computed, which presents the back side capacitance between the electrodes.

For this $v_{\partial\Omega_{\mathrm{ROI}}}$ is set to zero and the problem is solved for each electrode. The charge are finally computed by

$$\boldsymbol{Q} = \boldsymbol{Q}_a + \boldsymbol{Q}_c \boldsymbol{V}_{\partial\Omega_{\mathrm{ROI}}}, \tag{3.19}$$

where $\boldsymbol{V}_{\partial\Omega_{\mathrm{ROI}}}$ contains the finite element solutions (column vectors) on $\partial\Omega_{\mathrm{ROI}}$. In this representation, every column of $\boldsymbol{Q}$ corresponds to the measurements of the corresponding transmitter electrode.

3.3.3 Adjoint Problem and Green's Functions

The fast stiffness matrix update and the charge map approach can be applied to the standard solution framework and already lead to an accelerated forward map. An approach for solving a PDE using Green's functions is now presented. Its quite different approach and is motivated by considerations from functional analysis. Although the use of the new framework requires the same number of matrix inversions for the forward problem itself, it permits the use of other fast operations, which beat all standard methods in terms of computation time.

In the following, the theory behind the approach is presented. Consider a PDE of form

$$Lu = f, \tag{3.20}$$

in the domain Ω with the boundary conditions B

$$Bu = c, \tag{3.21}$$

on $\partial\Omega$, where L denotes a linear differential operator of order p in n variables. Using a multi-index notation $k = (k_1, k_2, \ldots, k_n)$ and defining the length $|k|$ of the multi-index as

$$|k| = k_1 + k_2 + \ldots + k_n, \tag{3.22}$$

and the differential operator D^k as

$$D^k = \frac{\partial^{|k|}}{\partial x_1^{k_1} \partial x_2^{k_2} \ldots \partial x_n^{k_n}}, \tag{3.23}$$

the operator L of equation (3.20) can be formally written as

$$L = \sum_{|k| \leq p} a_k(x) D^k. \tag{3.24}$$

A Green's function g is defined as the fundamental solution to the equation

$$Lg(x,\xi) = \delta(x-\xi), \tag{3.25}$$

that satisfies the boundary conditions

$$Bg = 0, \tag{3.26}$$

where $\delta(x-\xi)$ is the Dirac delta function at the source position ξ. By the definition of the inner product

$$\langle f, h\rangle = \int_{\Omega} f(x)\overline{h}(x)dx, \tag{3.27}$$

of functions f and h in the domain Ω, with $f, h : \Omega \mapsto \mathbb{C}$, the inner product of a function with the Dirac delta function gives

$$\langle u, \delta(x-\xi)\rangle = u(\xi). \tag{3.28}$$

Evaluating the inner product $\langle Lu, w\rangle$, where w is an arbitrary but differentiable function, turns to an equation of form

$$\langle Lu, w\rangle = \langle u, L^{*}w\rangle + \text{remaining part}, \tag{3.29}$$

where the operator L^{*} is termed the formal adjoint operator. Maintaining the multi index scheme, L^{*} can be expressed by

$$L^{*} = \sum_{|k|\leq p} (-1)^{|k|} D^{k} a_{k}(x). \tag{3.30}$$

To give a more precise statement about the "remaining part" in equation (3.29), the Lagrange's identity can be used to write

$$(Lu)w - u(L^{*}w) = \nabla \cdot J(u,w), \tag{3.31}$$

where J is a vectorial bilinear form of u, w and their derivatives to the order $p-1$. Taking the domain integral of (3.31) ends up in Green's 2^{nd} theorem

$$\langle Lu, w\rangle \;=\; \langle u, L^{*}w\rangle + \int_{\partial\Omega} \vec{\mathbf{n}} \cdot J(u,w)ds, \tag{3.32}$$

where the domain integral of $\nabla \cdot J(u,w)$ becomes a boundary integral due to Gauss' divergence theorem. This boundary integral is termed the bilinear concomitant or bilinear confluent. Because of the appearance of the bilinear confluent the boundary value problem is only called formally adjoint. If the bilinear confluent vanishes and $L = L^*$ holds the problem is termed to be self adjoint.

Due to the linearity of the gradient and the inner product of the potential equation (3.1), the differential operator for the ECT problem is a linear operator and by applying the definition (3.30) to equation (3.1) in order to obtain the adjoint operator, it can be seen that $L = L^*$. Thus, the operator is formally self adjoint. Working out the bilinear concomitant for equation (3.1) gives Green's formula for the ECT problem

$$\langle Lu, w\rangle - \langle u, L^* w\rangle = \int_{\partial\Omega} \varepsilon_0 \varepsilon_r \frac{\partial u}{\partial \vec{\boldsymbol{n}}} w - \varepsilon_0 \varepsilon_r \frac{\partial w}{\partial \vec{\boldsymbol{n}}} u ds. \tag{3.33}$$

The objective now is to find a set of boundary conditions in order to achieve that the bilinear confluent vanishes. If this can be achieved, the problem offers some advantageous properties, which will be used later. For Dirchlet type boundary conditions this task is simple, as u and w only have to be restricted to functions such that the original boundary conditions $Bu = 0$ and the adjoint boundary conditions $B^* u = 0$ are zero. Then the bilinear confluent vanishes. In this case the problem is self adjoint. The property of self adjoint problems is the most important property for the further steps. Considering h to be the Green's function of the adjoint problem,

$$L^* h(x, \eta) = \delta(x - \eta), \tag{3.34}$$

such that $h(x, \eta)$ fulfills the homogeneous adjoint boundary conditions. As the bilinear confluent vanishes for self adjoint problems one can write

$$\begin{aligned} \langle Lg(x,\xi), h(x,\eta)\rangle &= \langle g(x,\xi), L^* h(x,\eta)\rangle, && (3.35) \\ \langle \delta(x-\xi), h(x,\eta)\rangle &= \langle g(x,\xi), \delta(x-\eta)\rangle, && (3.36) \\ \bar{h}(\xi,\eta) &= g(\xi,\eta). && (3.37) \end{aligned}$$

The conjugate adjoint Green's function equals the original Green's function with reversed arguments. This is called the reciprocity principle. In the case of a self adjoint boundary value problem also

$$g(x,\xi) = \bar{h}(x,\xi) = \bar{g}(\xi, x), \tag{3.38}$$

holds. The reciprocity principle states, that the relation between an excitation source and a sink remains unchanged if the two positions of source and sink are switched. This relation can be used for solving (3.20) for non zero boundary conditions $Bu = c$.

Replacing w in equation (3.32) by h gives

$$\begin{aligned}\langle Lu, h\rangle &= \langle u, L^*h\rangle + \int_{\partial\Omega} \vec{\mathbf{n}} \cdot J(u,h)ds, &(3.39)\\ \langle f(x), h(x,\xi)\rangle &= \underbrace{\langle u, \delta(x-\xi)\rangle}_{u(\xi)} + \int_{\partial\Omega} \vec{\mathbf{n}} \cdot J(u,h)ds. &(3.40)\end{aligned}$$

Thus

$$u(\xi) = \langle f(x), h(x,\xi)\rangle - \int_{\partial\Omega} \vec{\mathbf{n}} \cdot J(u,h)ds, \qquad (3.41)$$

holds, but as $h = \bar{g}$ one can write

$$\begin{aligned}u(\xi) &= \langle f(x), \bar{g}(x,\xi)\rangle - \int_{\partial\Omega} \vec{\mathbf{n}} \cdot J(u,\bar{g})ds, &(3.42)\\ &= \langle f(x), \bar{g}(x,\xi)\rangle - \int_{\partial\Omega} \vec{\mathbf{n}} \cdot J(c,\bar{g})ds. &(3.43)\end{aligned}$$

This offers the possibility to find the solution u for the problem $Lu = f$ with the boundary conditions $Bu = c$ by applying Green's function on the right hand side term. In a short notation, this can be written as [You88]

$$u = L^{-1}f = Gf, \qquad (3.44)$$

where G is an integral operator whose kernel g is the Greens's function. In this way, it can be seen that G acts in an inverse sense to L.

Green's Functions for ECT

This inside view on the problem is based on methods from functional analysis but gives a deeper understanding about the applicability of Green's functions. The same theory can be applied to linear equation systems where it turns out that Hermitian matrices offer the property of self adjointness [You88]. As the finite element stiffness matrix of the electrostatic field problem has the property of being a Hermitian matrix, the problem is self adjoint and thus Green's functions can be used.

Instead of an evaluation of the standard forward problem (3.6) now

$$\hat{\boldsymbol{K}}\boldsymbol{g}_k = \boldsymbol{e}_k, \tag{3.45}$$

is solved, where $\boldsymbol{e}_k$ is the k^{th} unit vector and k belongs to the nodes on $\partial\Omega_{\mathrm{ROI}}$. Using a matrix notation for the Green's functions and the identity vectors

$$\hat{\boldsymbol{K}}\boldsymbol{G} = \boldsymbol{E}_{\partial\Omega_{\mathrm{ROI}}}, \tag{3.46}$$

the solutions $\boldsymbol{V}_{\partial\Omega_{\mathrm{ROI}}}$ can be computed by

$$\boldsymbol{V}_{\partial\Omega_{\mathrm{ROI}}} = \boldsymbol{G}^T\boldsymbol{R}, \tag{3.47}$$

where $\boldsymbol{R}$ contains the right hand side vectors of (3.6). Again the Green's functions have an inverse behavior to $\hat{\boldsymbol{K}}$. Similar, the charges on the electrodes can be computed by

$$\boldsymbol{Q} = \boldsymbol{Q}_a + \boldsymbol{Q}_c\boldsymbol{G}^T\boldsymbol{R}. \tag{3.48}$$

However, this approach has the immense drawback, that $N_{\partial\Omega_{\mathrm{ROI}}}$ Green's functions have to be determined. To remember, for the standard framework only N_{elec} matrix inversions were necessary. From the computations of the charges one can see that a left side multiplication by $\boldsymbol{Q}_c$ has to be done. This can be used to manipulate (3.46) by

$$\begin{aligned}
\hat{\boldsymbol{K}}\boldsymbol{G}\boldsymbol{Q}_c^T &= \boldsymbol{E}_k\boldsymbol{Q}_c^T, && (3.49)\\
\hat{\boldsymbol{K}}\boldsymbol{G}_Q &= \boldsymbol{E}_k\boldsymbol{Q}_c^T, && (3.50)\\
\hat{\boldsymbol{K}}\boldsymbol{G}_Q &= \boldsymbol{R}_Q. && (3.51)
\end{aligned}$$

Now again only N_{elec} problems have to be solved and the charges can be computed by

$$\boldsymbol{Q} = \boldsymbol{Q}_a + \boldsymbol{G}_Q^T\boldsymbol{R}. \tag{3.52}$$

By this approach, the forward map comes so far at the same cost in terms of matrix inversions as the standard framework. In the following, several computational techniques will be developed, which take extensive use of $\boldsymbol{G}_Q$, which can be referred to as Green's functions for the charges.

3.3.4 Jacobian Operations

It was already mentioned, that matrix vector multiplications of type $\boldsymbol{J\varepsilon}$ and $\boldsymbol{J}^T\boldsymbol{q}$ are needed in several algorithms. Hereby $\boldsymbol{q}$ is the vectorial form of $\boldsymbol{Q}$. With the standard computation schemes, this means that the Jacobian $\boldsymbol{J}$ is evaluated to perform the multiplications. Although for e.g. the AVM offers an efficient scheme to compute $\boldsymbol{J}$, it requires a further evaluation of an adjoint problem, which comes at the same cost as the forward map. Further, the storage of the Jacobian becomes costly as the Jacobian is a dense matrix and the Jacobian itself is not of interest. Only matrix vector products with the Jacobian are required. A method to by-pass the expensive evaluation of the Jacobian are Jacobian operations [Fox05]. Operating on the Jacobian means that the matrix vector multiplications are performed using mathematical relationships and properties of the problem, but without an explicit evaluation of the Jacobian.

In the following, the scheme will be explained for $J : \boldsymbol{\varepsilon} \mapsto \boldsymbol{v}$. The derivation starts with the extension of the forward problem (3.6) by

$$(\hat{\boldsymbol{K}} + d\hat{\boldsymbol{K}})(\boldsymbol{v} + d\boldsymbol{v}) = \boldsymbol{r}, \tag{3.53}$$

which can be rearranged to

$$\hat{\boldsymbol{K}} d\boldsymbol{v} = -d\hat{\boldsymbol{K}}(\boldsymbol{v} + d\boldsymbol{v}). \tag{3.54}$$

By this, a derivative with respect to a component of $\boldsymbol{\varepsilon}$ can be expressed as

$$\frac{d\boldsymbol{v}}{d\varepsilon_j} = -\hat{\boldsymbol{K}}^{-1}\frac{d\hat{\boldsymbol{K}}}{d\varepsilon_j}\boldsymbol{v}. \tag{3.55}$$

An extension to all elements of $\boldsymbol{\varepsilon}$ can be done by using the chain rule

$$\begin{aligned} d\boldsymbol{v} = \boldsymbol{J} d\boldsymbol{\varepsilon} &= -\sum_j -\hat{\boldsymbol{K}}^{-1}\frac{d\hat{\boldsymbol{K}}}{d\varepsilon_j}\boldsymbol{v}\, d\varepsilon_j && (3.56)\\ &= -\hat{\boldsymbol{K}}^{-1}\left[\sum_j \frac{d\hat{\boldsymbol{K}}}{d\varepsilon_j} d\varepsilon_j\right]\boldsymbol{v} && (3.57)\\ &= -\hat{\boldsymbol{K}}^{-1}\hat{\boldsymbol{K}}_{d\boldsymbol{\varepsilon}}\boldsymbol{v}. && (3.58)\end{aligned}$$

So far this scheme has no advantage compared to the AVM method, as still an inversion of $\hat{\boldsymbol{K}}$ is necessary. Now, as the problem is self adjoint, the inverse behavior of Green's functions presented in equation (3.47) can be used, to express equation (3.58) by

$$d\boldsymbol{v} = -\boldsymbol{G}^T \hat{\boldsymbol{K}}_{d\boldsymbol{\varepsilon}} \boldsymbol{G}. \tag{3.59}$$

The result is remarkable, as due to the use of the Green's functions no further forward problem has to be solved. The evaluation of $\boldsymbol{J}\boldsymbol{\varepsilon}$ comes at the cost of some matrix multiplications.

To operate on the Jacobian for the charges, $\boldsymbol{G}$ can be replaced by $\boldsymbol{G}_Q$ due to the linearity of the charge map. So the Jacobian operation $J : \boldsymbol{\varepsilon} \mapsto \boldsymbol{Q}$ is given by

$$d\boldsymbol{Q} = -\boldsymbol{G}_Q^T \hat{\boldsymbol{K}}_{d\boldsymbol{\varepsilon}} \boldsymbol{G}_Q. \tag{3.60}$$

In combination with the orthogonal decomposition of the stiffness matrix $d\boldsymbol{Q}$ becomes

$$d\boldsymbol{Q} = -\boldsymbol{G}_Q^T \left[\sum_l^p \boldsymbol{W}_l d\boldsymbol{\mathcal{E}} \boldsymbol{W}_l^T \right] \boldsymbol{G}_Q. \tag{3.61}$$

In the same way the transpose of Jacobian operation $J^T : \boldsymbol{Q} \mapsto \boldsymbol{\varepsilon}$ can be derived as [Fox05]

$$\boldsymbol{J}^T \boldsymbol{q} = -\sum_{l=1}^p \left(\boldsymbol{G}_Q^T \boldsymbol{W}_l \right)^T \boldsymbol{Q} \left(\boldsymbol{G}_Q^T \boldsymbol{W}_l \right), \tag{3.62}$$

where $\boldsymbol{Q}$ is the matrix form of the vector $\boldsymbol{q}$. It turns out, that due to the matrix structure of $\boldsymbol{Q}$ the result is given as the main diagonal of (3.62). The result is given by

$$\boldsymbol{J}^T \boldsymbol{q} = -\mathrm{diag} \left(\sum_{l=1}^p \left(\boldsymbol{G}_Q^T \boldsymbol{W}_l \right)^T \boldsymbol{Q} \left(\boldsymbol{G}_Q^T \boldsymbol{W}_l \right) \right), \tag{3.63}$$

and in order to avoid the computation of the off-diagonal elements, this can be written as

$$\boldsymbol{J}^T \boldsymbol{q} = -\left(\sum_{l=1}^p \left(\boldsymbol{G}_Q^T \boldsymbol{W}_l \right)^T \otimes \left(\boldsymbol{Q} \boldsymbol{G}_Q^T \boldsymbol{W}_l \right) \right), \tag{3.64}$$

where $\otimes$ expresses the column and row wise multiplication.

3.3.5 Exact Fast Low-Rank Updates

During the solution process of the inverse problem one is often faced by the task to determine the solution for a marginally changed material distribution. This happens when the material value of one pixel (one finite element) changes. The change in the charges could be approximated by the Jacobian approximation. A method to compute the exact solution without evaluating all forward problems again is given by the Woodbury matrix identity [Hag89]

$$(\boldsymbol{A}+\boldsymbol{L}\boldsymbol{U})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{L}\left(\boldsymbol{I}+\boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{U}\boldsymbol{A}^{-1}, \tag{3.65}$$

where the matrix $\left(\boldsymbol{I}+\boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{L}\right)$ is referred to as the Woodbury matrix. The Woodbury identity provides an efficient method to compute the inverse of the matrix $\boldsymbol{B} = (\boldsymbol{A}+\boldsymbol{L}\boldsymbol{U})$ for a low-rank update of the matrix $\boldsymbol{A}$, when $\boldsymbol{A}^{-1}$ is known. For the ECT problem, the matrix $\boldsymbol{B}$ is given by

$$\boldsymbol{B} = \hat{\boldsymbol{K}}_{\text{new}} = \hat{\boldsymbol{K}}_{\text{old}} + \gamma\sum_{l=1}^{p}\boldsymbol{W}_l\Delta\boldsymbol{\mathcal{E}}\boldsymbol{W}_l^T, \tag{3.66}$$

where the second term expresses the change in the permittivity values and γ denotes a scaling variable of the update expression. Several methods are possible to perform the decomposition

$$\gamma\boldsymbol{L}\boldsymbol{U} = \gamma\sum_{l=1}^{p}\boldsymbol{W}_l\Delta\boldsymbol{\mathcal{E}}\boldsymbol{W}_l^T. \tag{3.67}$$

Again an eigenvector decomposition could be used to take advantage of the rank deficiency of the element matrices. This further decreases the dimension of the inverse matrix in the Woodbury formula and in the case of single element update an analytic evaluation becomes possible by the use of the the adjoint extension to express the inverse of the 2×2 Woodbury matrix $\left(\boldsymbol{I}+\boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{L}\right)$. However, for larger dimensional updates the eigenvector decomposition becomes comparatively slow. A Cholesky decomposition with $\boldsymbol{L} = \boldsymbol{C}$ and $\boldsymbol{U} = \boldsymbol{C}^T$ cannot be used due to the rank deficiency. Hence, the LU-decomposition becomes a suitable decomposition. For a single finite element update, the matrix $\boldsymbol{L}$ becomes a $n_{\text{node}} \times 3$ matrix with entries in the rows C, which correspond to the nodes of the effected finite element.

The efficiency of the Woodbury identity is introduced by the use of Green's functions to replace $\boldsymbol{A}^{-1} = \hat{\boldsymbol{K}}_{\text{old}}^{-1}$ on the left and on the right side of the

Woodbury matrix. Following the developments in [FN97] only the Green's functions for the nodes C have to be evaluated and the effect of a low-rank update can be evaluated by

$$\Delta\boldsymbol{Q} = -\gamma \boldsymbol{G}_Q^T \boldsymbol{L} \left(\boldsymbol{I} + \gamma \boldsymbol{U}_{:,C} \boldsymbol{G}_{C,C} \boldsymbol{L}_{C,:}\right)^{-1} \boldsymbol{U} \boldsymbol{G}_Q, \tag{3.68}$$

where a MATLAB like notation is used and $\boldsymbol{G}_{C,C}$ denotes the set of Green's functions for the nodes C. Thus, for the application of the Woodbury identity the forward problem has to be solved for C identity vectors. Then, an arbitrary scaling of $\Delta\boldsymbol{\mathcal{E}}$ by the factor γ can be computed at the cost of inverting the Woodbury matrix, which is a 3×3 matrix for the case, that only one (linear and triangular) finite element is updated.

Although the scheme is efficient the additional evaluation of $\boldsymbol{G}_{C,C}$ is still a painful bottleneck in the case of higher dimensional updates. The application of the Woodbury identity has to be done carefully with respect to the number of elements which are updated. In gradient-based optimization schemes the number of updates is typically high and efficient efficient line search schemes only require a low number of additional evaluations of the forward map. In this case the number of computations to obtain the needed Green's functions $\boldsymbol{G}_{C,C}$ is likely to increase the number of forward computations in the normal lines search scheme. However, for low dimensional updates the application of the Woodbury formula offers an enormous computational advantage.

3.3.6 Fast Exact Update for $\boldsymbol{WSW}^T$ Systems

The necessary evaluation of additional Green's functions in the Woodbury formula (3.68) still requires evaluations of the forward map to apply the scheme. Considering, that the $m \times m$ matrix $\boldsymbol{A}$ is of form

$$\boldsymbol{A} = \boldsymbol{W}_1 \boldsymbol{S}_0 \boldsymbol{W}_1^T, \tag{3.69}$$

where $\boldsymbol{S}_0$ is a $n \times n$ diagonal matrix containing the material values and $\boldsymbol{W}_1$ is a constant $m \times n$ projection matrix to build the system matrix $\boldsymbol{A}$ with $n > m$. The update of the matrix $\boldsymbol{A}$ is of form

$$\boldsymbol{B} = \boldsymbol{A} + \boldsymbol{W}_1 \boldsymbol{S} \boldsymbol{W}_1^T, \tag{3.70}$$

where $\boldsymbol{S}$ is again a diagonal matrix. Using the decomposition

$$\begin{aligned} \boldsymbol{L} &= \boldsymbol{W}_1, && (3.71)\\ \boldsymbol{U} &= \boldsymbol{S}\boldsymbol{W}_1^T, && (3.72) \end{aligned}$$

the Woodbury identity writes as

$$\begin{aligned} \boldsymbol{B}^{-1} &= \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{L}\left(\boldsymbol{I} + \boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{L}\right)^{-1}\boldsymbol{U}\boldsymbol{A}^{-1}, && (3.73)\\ &= \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{W}_1\left(\boldsymbol{I} + \boldsymbol{S}\boldsymbol{W}_1^T\boldsymbol{A}^{-1}\boldsymbol{W}_1\right)^{-1}\boldsymbol{S}\boldsymbol{W}_1^T\boldsymbol{A}^{-1}. && (3.74) \end{aligned}$$

Again the inverse of $\boldsymbol{A}$ is necessary. Recall that $\boldsymbol{A}^{-1}$ can be expressed by

$$\boldsymbol{A}^{-1} = \boldsymbol{W}_2\boldsymbol{S}_0^{-1}\boldsymbol{W}_3^T. \tag{3.75}$$

As $\boldsymbol{S}_0$ is a diagonal matrix, the inversion of $\boldsymbol{S}_0$ comes at the cost of computing the reciprocal values in the main diagonal, which is the cheapest way of inverting the matrix. To evaluate $\boldsymbol{W}_2$ and $\boldsymbol{W}_3$ the relation $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{I}$ is used to analyze the problem. This appears as

$$\boldsymbol{W}_1\boldsymbol{S}_0\boldsymbol{W}_1^T\boldsymbol{W}_2\boldsymbol{S}_0^{-1}\boldsymbol{W}_3^T = \boldsymbol{I}. \tag{3.76}$$

As $\boldsymbol{S}_0\boldsymbol{S}_0^{-1} = \boldsymbol{I}$ is an obvious fact, the equation suggests the mathematical relations

$$\boldsymbol{W}_1^T\boldsymbol{W}_2 = \boldsymbol{I}, \tag{3.77}$$

and

$$\boldsymbol{W}_1\boldsymbol{W}_3^T = \boldsymbol{I}. \tag{3.78}$$

Equation (3.78) builds an over-determined equation system and $\boldsymbol{W}_3$ is given by

$$\boldsymbol{W}_3 = \left(\boldsymbol{W}_1\boldsymbol{W}_1^T\right)^{-1}\boldsymbol{W}_1, \tag{3.79}$$

which is the Moore Penrose inverse or pseudo inverse of $\boldsymbol{W}_1^T$. Thus, $\boldsymbol{W}_3$ can be evaluated directly from $\boldsymbol{W}_1$. This computation has to be done only for once in the initial phase. Equation (3.77) is an undetermined problem. A matrix $\boldsymbol{W}_2$ fulfilling (3.77) cannot be found. However, the existence of $\boldsymbol{W}_3$ satisfying equation (3.78) and the structure in equation (3.76)

$$\boldsymbol{W}_1\underbrace{\boldsymbol{S}_0\boldsymbol{W}_1^T\boldsymbol{W}_2\boldsymbol{S}_0^{-1}}_{\boldsymbol{I}}\boldsymbol{W}_3^T = \boldsymbol{W}_1\boldsymbol{W}_3^T = \boldsymbol{I}, \tag{3.80}$$

suggests the correctness of equation (3.77). Hence, equation (3.75) can be substituted into the Woodbury matrix although knowing, that $\boldsymbol{W}_2$ cannot be found. One can obtain

$$\begin{aligned}\left(\boldsymbol{I}+\boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{L}\right)^{-1} &= \left(\boldsymbol{I}+\boldsymbol{S}\boldsymbol{W}_1^T\boldsymbol{A}^{-1}\boldsymbol{W}_1\right)^{-1}, &(3.81)\\ &= \left(\boldsymbol{I}+\boldsymbol{S}\underbrace{\boldsymbol{W}_1^T\boldsymbol{W}_2}_{\boldsymbol{I}}\boldsymbol{S}_0^{-1}\boldsymbol{W}_3^T\boldsymbol{W}_1\right)^{-1}, &(3.82)\\ &= \left(\boldsymbol{I}+\boldsymbol{S}\boldsymbol{S}_0^{-1}\boldsymbol{W}_3^T\boldsymbol{W}_1\right)^{-1}, &(3.83)\end{aligned}$$

where $\boldsymbol{W}_2$ cancels out. The result in equation (3.83) is remarkable, as the scheme only requires the inversion of the diagonal matrix $\boldsymbol{S}_0$ during the runtime and the evaluation of $\boldsymbol{W}_3$ is only required for one time during the initialization. Although $\boldsymbol{W}_3^T\boldsymbol{W}_1$ is a full matrix, the Woodbury matrix is diagonal dominant, as $\boldsymbol{S}$ has only nonzero entries in the main diagonal where an update should be evaluated. If only the diagonal elements C of $\boldsymbol{S}$ are nonzero equation (3.83) can be further reduced to

$$\left(\boldsymbol{I}+\boldsymbol{U}\boldsymbol{A}^{-1}\boldsymbol{L}\right)^{-1} = \left(\boldsymbol{I}+\boldsymbol{S}_C\boldsymbol{S}_{0,C}^{-1}\boldsymbol{W}_{3,:,C}^T\boldsymbol{W}_{1,:,C}\right)^{-1}, \tag{3.84}$$

where $\boldsymbol{S}_C$ and $\boldsymbol{S}_{0,C}$ are diagonal matrices of dimension $C \times C$.

The scheme presented offers a remarkable advance compared to the scheme of equation (3.68) but it requires a decomposition of the equation system into the form of equation (3.69). In fact this decomposition can be done for any problem which is similar to a resistor network [Str86]. This means that an underlying graph exists as it is the case for finite element methods. Hence, the matrix $\boldsymbol{S}$ belongs to the edges of the updated finite element mesh and $\boldsymbol{W}_1$ can be found as decomposition of the Laplacian matrix by

$$(\boldsymbol{W}_1\boldsymbol{W}_1^T)_{l,m} = \begin{cases}\deg(v_l) & \text{if } l = m,\\ -1 & \text{if } l \neq m \text{ and } v_i \text{ is adjacent to } v_j,\\ 0 & \text{else.}\end{cases} \tag{3.85}$$

The matrix contains the connectivity information between the nodes and the edges. The Laplacian equals the admittance matrix $\boldsymbol{Y}$ of nodal voltage analysis $\boldsymbol{Y}\boldsymbol{v} = \boldsymbol{i}$ for a network with all resistors having the value one. However, for the solution it is necessary to remove at least one node which gives the reference potential. The matrix $\boldsymbol{W}_1$ can even be found in a more simple way, as its columns only contain a one in the row which becomes projected into the main diagonal and a minus in the rows which build the off diagonal entries in the Laplacian matrix. In this sense also the computation of $\boldsymbol{W}_3$ is not critical, as the inversion of $\boldsymbol{W}_1\boldsymbol{W}_1^T$ in equation (3.79) comes to the cost of the forward problem. Further, only the columns of $\boldsymbol{W}_3$ corresponding to the edges which are effected by an update have to be computed. This evaluation can further be done by using Green's functions.

To compute $\boldsymbol{S}$, a map $\boldsymbol{P}_S : \boldsymbol{\mathcal{E}} \mapsto \boldsymbol{S}$ has to be formed. This is possible by decomposing the element matrices of each finite element into a 3×3 resistor network on the edges [Bra03]. On the neighboring edge of two finite elements, the resistors are in parallel and $\boldsymbol{P}_S$ becomes a linear map of form

$$\boldsymbol{S} = \mathrm{diag}(\boldsymbol{P}_S\boldsymbol{\mathcal{E}}_{\mathrm{diag}}), \tag{3.86}$$

where $\boldsymbol{\mathcal{E}}_{\mathrm{diag}}$ denotes the vector of the main diagonal of the matrix $\boldsymbol{\mathcal{E}}$. Given this decomposition, for the finite element system, the change $\Delta\boldsymbol{Q}$ can be computed by

$$\Delta\boldsymbol{Q} = -\boldsymbol{G}_{Q,C_2,:}^T\boldsymbol{W}_{1,:,C}\left(\boldsymbol{I} + \boldsymbol{S}_C\boldsymbol{S}_{0,C}^{-1}\boldsymbol{W}_{3,:,C}^T\boldsymbol{W}_{1,:,C}\right)^{-1}\boldsymbol{S}\boldsymbol{W}_{1,:,C}^T\boldsymbol{G}_{Q,C_2,:}, \tag{3.87}$$

where C again denotes the effected edges of the finite element and C_2 denote the nodes not on the boundary $\partial\Omega$. C_2 becomes necessary, as the scheme works on the reduced matrix.

Given this way to build $\boldsymbol{S}$ and $\boldsymbol{W}_1$, the decomposition is somehow remarkable, as the physical law and the geometry are presented in the diagonal of $\boldsymbol{S}$, whereas $\boldsymbol{W}_1$ only contains the connectivity information from the edges in form of integer values. Also the fact, that $\boldsymbol{W}_3$ an thus the term $\boldsymbol{W}_3^T\boldsymbol{W}_1$ in equation (3.83) only depends on the connectivity reveals a deeper structure inside the used terms as no material values are included. In fact, the Laplacian presents a discrete version of the Laplace operator. A slight drawback is the fact that this scheme operates on the edges of the finite elements, which means that the dimension of the Woodbury matrix has the tendency to become larger than in the previous applications of the Woodbury identity. However, for a single material finite element update, the Woodbury matrix (3.84) is of dimension 3×3, which is the same as presented in equation (3.68) and this new approach does not require the evaluation of further Green's functions.

An arguable point in the generality of the approach lies in the fact, that $\boldsymbol{W}_1$

has to be the reduced adjacency matrix in order to incorporate the Dirichlet type boundary conditions. This means that the rows of $\boldsymbol{W}_1$ which correspond to nodes on Dirichlet type boundaries are removed. Without this, the Laplacian matrix is a singular matrix.

Efficient Solution of Matrix Pencils

The evaluation of a line search or a conditional sampling requires multiple inversions of the Woodbury matrix and an efficient scheme for a further improvement seems suitable. The structure of the Woodbury matrix as the sum of two matrices forming an equation system

$$(\boldsymbol{A} + \gamma \boldsymbol{B})\, \boldsymbol{x} = \boldsymbol{c}, \tag{3.88}$$

is referred to as matrix pencil of degree two or as a linear matrix pencil. For the Woodbury matrix, the fact that $\boldsymbol{A} = \boldsymbol{I}$ forms a further advantage. A modified version of an algorithm presented in [ES78], which already takes care to the Woodbury structure is given by the following computations.

In a first step, a Hessenberg decomposition of the matrix $\boldsymbol{B}$ is performed by

$$\boldsymbol{B} = \boldsymbol{L}\boldsymbol{H}\boldsymbol{L}^{-1}. \tag{3.89}$$

Then the equation system

$$\boldsymbol{L}\boldsymbol{z}_1 = \boldsymbol{c}, \tag{3.90}$$

is solved, using the property $\boldsymbol{L}^H = \boldsymbol{L}^{-1}$. Now, the equation system

$$\left(\gamma^{-1}\boldsymbol{I} + \boldsymbol{H}\right) \boldsymbol{y} = \gamma^{-1}\boldsymbol{z}_1, \tag{3.91}$$

has to be solved. Subsequently $\boldsymbol{x}$ can be found as

$$\boldsymbol{x} = \boldsymbol{L}\boldsymbol{y}. \tag{3.92}$$

The algorithm provides a reduction of about 25 % of the computation time compared to the repeated solution of (3.88). However, the necessity to solve (3.91) every time is still a drawback.

Another way to solve (3.88) is given by using an eigenvector decomposition of the matrix $\boldsymbol{B}$. The inverse of the Woodbury matrix is then given by

$$(\boldsymbol{I}+\gamma\boldsymbol{B})^{-1}=\boldsymbol{Q}_B\,(\boldsymbol{I}+\gamma\boldsymbol{D}_B)^{-1}\,\boldsymbol{Q}_B^{-1}, \tag{3.93}$$

where $\boldsymbol{Q}_B$ contains the eigenvectors of $\boldsymbol{B}$ and $\boldsymbol{D}_B$ is a diagonal matrix of the eigenvalues of $\boldsymbol{B}$. The decomposition and the computation of the inverse of $\boldsymbol{Q}_B$ have only to be done for once and due to the diagonal structure of $(\boldsymbol{I}+\gamma\boldsymbol{D}_B)$, the evaluation of $(\boldsymbol{I}+\gamma\boldsymbol{D}_B)^{-1}$ comes at the cost of inverting the elements in the main diagonal.

This is an attractive approach except that the eigenvector decomposition is comparatively costly. The application of the method becomes a trade off between cost of the eigenvector decomposition and the number of low-rank evaluations. For larger updates, where the eigenvector decomposition becomes even more expensive, the approach might even get insufficient. A further, problem is given by the fact, that the matrices $\boldsymbol{Q}_B$ and its inverse $\boldsymbol{Q}_B^{-1}$ are dense matrices. Hence also the multiplication of the right hand side terms in (3.93) becomes expensive.

3.3.7 Domain Decomposition Techniques

The size of the equation system (3.6) to solve is essentially affected by the degree of discretization. It was reported in [KS07] and [KS05] that the error through the discretization should be smaller than the measurement noise. In cite [WNF10] an unstructured finite element mesh with about 6000 elements was presented as the coarsest mesh to meet these requirements for the used ECT sensor. This number is mainly caused by the discretization in the region of the electrodes, as for an accurate flux integration a finer discretization in this region is required. The main number of the entries of $\hat{\boldsymbol{K}}$ represents the domain $\Omega\setminus\Omega_{\mathrm{ROI}}$, which stays constant. Only the elements of $\hat{\boldsymbol{K}}$ corresponding to Ω_{ROI} change during a material update.

A decomposition seems suitable in order to invert only the part of $\hat{\boldsymbol{K}}$ which is affected by a material update. The charge map already presents a method where the charges can be computed using the potential distribution on $\partial\Omega_{\mathrm{ROI}}$. By rearranging the elements of the vector $\boldsymbol{v}$, equation (3.6) can be presented as

$$\begin{bmatrix} \hat{\boldsymbol{K}}_{\mathrm{ROI}} & \hat{\boldsymbol{K}}_{\mathrm{Diag}} \\ \hat{\boldsymbol{K}}_{\mathrm{Diag}}^T & \hat{\boldsymbol{K}}_{\mathrm{Rest}} \end{bmatrix} \begin{bmatrix} \boldsymbol{v}_{\mathrm{ROI}} \\ \boldsymbol{v}_{\mathrm{Rest}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{r}_1 \\ \boldsymbol{r}_2 \end{bmatrix}, \tag{3.94}$$

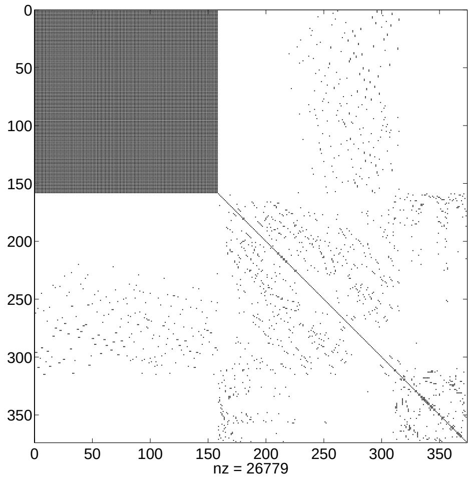


Figure 3.2: Typical spy plot of the stiffness matrix after using the Schur complement (no permutation scheme has been applied). The block structure is introduced by the exterior part.

where $\boldsymbol{v}_{\text{ROI}}$ collects the nodal potentials in Ω_{ROI} and on its boundary $\partial\Omega_{\text{ROI}}$. Now the Schur complement [ZF05] can be used to set up the smaller equation system

$$\left(\hat{\boldsymbol{K}}_{\text{ROI}} - \hat{\boldsymbol{K}}_{\text{Diag}}\hat{\boldsymbol{K}}^{-1}_{\text{Rest}}\hat{\boldsymbol{K}}^{T}_{\text{Diag}}\right)\boldsymbol{v}_{\text{ROI}} = \boldsymbol{r}_1 - \hat{\boldsymbol{K}}_{\text{Diag}}\hat{\boldsymbol{K}}^{-1}_{\text{Rest}}\boldsymbol{r}_2. \tag{3.95}$$

As only the matrix $\hat{\boldsymbol{K}}_{\text{ROI}}$

$$\hat{\boldsymbol{K}}_{\text{ROI}} = \boldsymbol{K}_{\text{ini,ROI}} + \sum_{l=1}^{p} \boldsymbol{W}_{l,\text{ROI}}\boldsymbol{\mathcal{E}}_{\text{ROI}}\boldsymbol{W}^{T}_{l,\text{ROI}}, \tag{3.96}$$

is changed, the inversion of the larger matrix $\hat{\boldsymbol{K}}_{\text{Rest}}$ can be done in a pre computation step and during the runtime only a matrix problem of the dimension of the matrix $\hat{\boldsymbol{K}}_{\text{ROI}}$ has to be solved. However, this reduction comes at the cost of introducing a block matrix structure into the new stiffness matrix. Figure 3.2 depicts a typical spy plot of the reduced stiffness matrix. Due to the block structure the bandwidth of the matrix increases and thus the performance of direct solvers decreases drastically. Hence, the application of Cholesky decomposition becomes necessary.

A further improvement can be made by again using the Woodbury formula (3.65) in order to pre evaluate the inverse of the constant parts of the reduced matrix (3.95). For this,

$$\boldsymbol{A} = \boldsymbol{K}_{\text{ini,ROI}} - \hat{\boldsymbol{K}}_{\text{Diag}} \hat{\boldsymbol{K}}_{\text{Rest}}^{-1} \hat{\boldsymbol{K}}_{\text{Diag}}^{T}, \tag{3.97}$$

has to be used, as the matrix $\hat{\boldsymbol{K}}_{\text{Diag}} \hat{\boldsymbol{K}}_{\text{Rest}}^{-1} \hat{\boldsymbol{K}}_{\text{Diag}}^{T}$ cannot be inverted itself as it is rank deficient. $\boldsymbol{K}_{\text{ini,ROI}}$ has to be of full rank, which can be ensured by setting up $\boldsymbol{K}_{\text{ini}}$ of equation (3.15) for $\varepsilon_r = 1$ in Ω_{ROI}. Then the susceptibility $\chi = \varepsilon_r - 1$ has to be used for the material update and the update term of the Woodbury identity is given by

$$\boldsymbol{LU} = \sum_{l=1}^{p} \boldsymbol{W}_{l,\text{ROI}} \boldsymbol{\chi}_{\text{ROI}} \boldsymbol{W}_{l,\text{ROI}}^{T}. \tag{3.98}$$

Again a Cholesky decomposition can be applied to invert the Woodbury matrix.

3.3.8 Performance Estimation and Discussion

The previous two sections of this chapter provided information about the standard solution techniques and techniques using Green's functions. The first section contains the solution of the forward problem and the evaluation of the Jacobian. The use of Green's functions enables further computational abilities, such as fast low-rank updates to the solution of the stiffness matrix. Table 3.1 represents a list of the computation times for the methods. The times were determined on an Intel Core 2 Duo CPU with 2.10 GHz clock frequency on a 64 bit operating system with 4 GB RAM for two finite element meshes. Mesh 1 is a mesh with about 3100 elements in total, which results in about 1950 nodes. This mesh has about 580 degrees of freedom. Meshes of this size are mostly used on ECT systems, as the computational costs vs. the spatial accuracy are in a good trade of for use with deterministic methods. Mesh 2 is a mesh with about 12200 finite elements in total and about 2650 elements in the interior of the pipe, resulting in an equation system with about 7200 unknowns. The table summarizes the most important computer times. All times of the new methods have to be seen with respect to the times of a normal evaluation of the forward map or the Jacobian, respectively. For the block concerning the domain decomposition techniques not all possible varieties are included which means, that. not all possible combinations of the Schur complement and the Woodbury identity are stated. However, the listed times give a statement on the usefulness of some combinations. It also

Nr.	Operation/Method		t_{mesh1} ms	t_{mesh2} ms
Forward map:				
1	Forward problem standard	$F : \boldsymbol{\varepsilon} \mapsto \boldsymbol{C}$	96	640
2	Forward problem new	$F : \boldsymbol{\varepsilon} \mapsto \boldsymbol{Q}$	4.8	26
3	Standard material update	(3.8)	35.5	230
4	Fast material update	(3.15)	0.039	2
5	Matrix inv. $\hat{\boldsymbol{K}}\boldsymbol{v} = \boldsymbol{r}$ $(\times N_{\mathrm{elec}})$	(3.6)	3.3	19
6	Cholesky d. of $\hat{\boldsymbol{K}}$	$\hat{\boldsymbol{K}} = \boldsymbol{C}\boldsymbol{C}^T$	1.4	9
7	Matrix inv. with Chol.		2.2	12
Jacobian operations:				
8	Jacobian by AVM	(3.12)	360	> 2000
9	Jacobian op. $J : \boldsymbol{\varepsilon} \mapsto \boldsymbol{Q}$	(3.61)	0.48	3.9
10	Transp. of Jacobian op. $J^T : \boldsymbol{Q} \mapsto \boldsymbol{\varepsilon}$	(3.64)	3.3	15.5
Low-rank updates:				
11	Exact low-rank update (1 elem.)	(3.68)	3.5	16.3
12	$\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^T$ Woodbury (1 elem.)	(3.87)	0.66	2.1
13	Exact low-rank update (20 elem.)	(3.68)	11.6	55.4
14	$\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^T$ Woodbury (20 elem.)	(3.87)	2.7	14.1
15	Exact update 1 elem × 50	(3.68)	7.3	26.3
16	$\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^T$ Woodbury 1 elem × 50	(3.87)	2.4	3.9
Domain Decomposition:				
17	Domain d. by Schur c.		23.8	66
18	Chol. d. for Schur c.		29	75
19	Schur c. with Cho.		2.9	8
20	LU d. for Wood. identity		6.1	514
21	Woodb. matrix for Schur c.		104	> 4600
22	Woodb. for Schur c. with Chol.		0.895	17.1

Table 3.1: Computation times of the methods presented in section 3.2 and section 3.3.

has to be mentioned that a poor comparison of the computation times is not the fairest way to compare the two implementations. In fact a comparison maintaining a Landau notation [Bla11] should be used. However, the times give at least an inside view about the gain of a fast implementation. For the further comparisons the use of the time table is a fair approach, as the new methods build an extension of the new forward map.

The first block in table 3.1 presents times regarding the evaluation of the forward map. Line 1 presents the computation times for the forward problem when implemented in the standard way. Line 2 presents the times with the new approaches using the fast material update and the charge map based on $\boldsymbol{G}_Q$. The performance gain is evident. Line 3 and 4 give an insight about the efficiency of the eigenvector decomposition to assembly the stiffness matrix. Another computational issue which becomes time consuming is the introduction of the Dirichlet type boundary conditions to form $\hat{\boldsymbol{K}}$ out of $\boldsymbol{K}$. This operation is not necessary in the new method, as the charges are computed by the charge map and the boundary conditions are introduced by $\boldsymbol{K}_{\mathrm{ini}}$. Line 5 presents the time amount for solving the 16 forward problems. The times for solving the equation system using a Cholesky decomposition and the Cholesky decomposition itself are presented in line 6 and 7. The overall time when using Cholesky decomposition is in this case higher. However, this results were obtained for an air filled pipe. In this case the stiffness matrix has the lowest condition number. For the case of material inclusions the condition number increases and in this case the direct inversion takes longer, while the inversion using Cholesky decomposition takes the same time. Hence, it is always preferable to use the Cholesky decomposition. This holds especially for larger equation systems.

The second block contains the computation times in for the with Jacobian operations. Line 8 presents the times for the evaluation of the Jacobian using the adjoint variable method. It is obvious, that the AVM becomes the bottleneck when used. Line 9 contains the times for the Jacobian operations. It can be seen, that the Jacobian operation outperforms the AVM by a factor of 750 for mesh 1 and > 130 for mesh 2. The transpose of Jacobian operations require more computation time compared to the Jacobian operations, as they require reasonable more multiplications. However, the operation is still several times faster than the evaluation of the Jacobian.

The third block contains the performance of low-rank updates. The computation times of this methods have to be viewed in contrast to block 1, as their aim is to provide the solution for a small update without computing the full forward problem again. The lines 11 and 12 contain the evaluation times when a single finite element is updated. In the first case, the evaluation of (3.68) requires the computation of the three Green's functions which correspond to the finite element. Hence, the computation times in line 11 are only slightly lower than the computation times in line 2. In the case of the generalized Woodbury matrix, the performance gain is obvious. The lines 13 and 14 present the evaluation time of 20 elements. The elements were not connected, which presents the worst case for the evaluation, as this requires the computation of 60 Green's functions or leads to 60 independent edges in the generalized form. This fact appears in line 13 for the standard Woodbury scheme. The times in line 14 are only slightly lower than the times in line 2. It turned out, that the multiplication in equation (3.83) become more expensive and causes the bottleneck in this computation. For repeated evaluations of the Woodbury formula it becomes mandatory to evaluate the right hand term $\boldsymbol{G}^T_{Q,C_2,:}\boldsymbol{W}_{1,:,C}$ and the left hand term $\boldsymbol{S}\boldsymbol{W}^T_{1,:,C}\boldsymbol{G}_{Q,C_2,:}$ of the Woodbury matrix separately to save computation time. The lines 15 and 16 present the times for the repeated update of the same single finite element. Again the Woodbury formulation provides an enormous speed update.

The last block represents the computation times when domain decomposition methods are applied. As can be seen in line 17, the direct application of the Schur complement using the presented domain decomposition ends up in a significantly increased computation time. This is due to the introduced block structure of the Schur complement. Solving the same problem using a Cholesky decomposition leads to a remarkable reduction of the computation time for mesh 2, as can be seen in line 19. However, the Cholesky decomposition itself is expensive, as can be seen in line 18. To use the Woodbury matrix and LU decomposition of the update term has to be evaluated. The computation times are comparatively slow, but this is more likely to be caused by the Matlab implementation. As the update is a sparse matrix, a direct update of the LU decomposition seems possible, to reduce this evaluation time. In line 21 the inverse of the Woodbury matrix is evaluated using the sum of the block matrix of the Schur complement and the initial stiffness matrix as constant matrix with known inverse using the LU decomposition to describe the update. However, with a further Cholseky decomposition a remarkable performance gain can be found for mesh 1, as line 22 indicates.

3.4 Overview of Approximation Techniques

In the following several common approximation techniques for the forward map are presented in short, which will be used for within the DAMH algorithm. Detailed information can be found in [FSK08].

3.4.1 Local Approximations

An obvious method to estimate the behavior of the function $F(\cdot)$ in the region of the point $\boldsymbol{x}$ is by applying a Taylor series expansion about the point $\boldsymbol{x}$. The model output is locally approximated by

$$F(\boldsymbol{x}+\Delta\boldsymbol{x}) \approx F^{*}(\boldsymbol{x}+\Delta\boldsymbol{x}) = F(\boldsymbol{x}) + \left.\frac{\partial F(\boldsymbol{x})}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}} \Delta\boldsymbol{x}, \tag{3.99}$$

where the derivative of $F(\cdot)$ with respect to $\boldsymbol{x}$ is called the Jacobian. The quality of this type of approximations either depends on the nonlinearity of the function $F(\cdot)$, as well as on the displacement $\Delta\boldsymbol{x}$. A major aspect about the application of this approximation is the efficient computation of the Jacobian matrix. With the availability of efficient schemes for the given application the Jacobian based approximation forms a suitable tool within the DAMH algorithm.

3.4.2 Polynomial Approximations

Polynomial approximations try to approximate $F(\cdot)$ by a matrix vector product given by

$$F^{*}(\boldsymbol{x}) = \boldsymbol{P}\tilde{\boldsymbol{x}}, \tag{3.100}$$

where $\tilde{\boldsymbol{x}}$ is an augmented input vector. Hence, each row of the matrix $\boldsymbol{P}$ builds an approximation for each output quantity. The augmentation of $\boldsymbol{x}$ up to a polynomial approximation for $F(\cdot)$ by degree M is given by $\tilde{\boldsymbol{x}} = \begin{bmatrix} 1 & \boldsymbol{x}^T & \dots & (\boldsymbol{x}^T)^M \end{bmatrix}^T$. Only using the initializing 1 in $\tilde{\boldsymbol{x}}$ an affine approximation is built. To obtain the rows of the matrix $\boldsymbol{P}$, the Vandermonde matrix has to be built out of sampled data. Then each row of $\boldsymbol{P}$ is given by applying the Moore Penrose pseudo inverse. It has been demonstrated in [FSK08], that a polynomial approximation of order M is in essential a Taylor series approximation of order $M+1$. Hence, the use of a higher order polynomials would theoretically increase the approximation quality but too high dimensional polynomial approximations are likely to lead to overfitted approximations. Also effects like saturation can not be handled with polynomial approximations.

Within the DAMH algorithm also an adaptive form of the polynomial approximation can be applied by using the result of the second step to update the matrix $\boldsymbol{P}$. The update can for example be done by a recursive least squares method (RLS). A computational less expensive update method is given by the least mean squares (LMS) algorithm [Hay01]

$$\boldsymbol{p}_{k+1} = \boldsymbol{p}_{k+1} + \mu(F(\boldsymbol{x}) - \boldsymbol{p}_k^T \tilde{\boldsymbol{x}})\tilde{\boldsymbol{x}}, \tag{3.101}$$

where μ denotes the adaptation constant.

3.4.3 Other Techniques

Radial Basis Functions

In a number of engineering disciplines it is common to describe the behavior of systems or signals in form of a weighted sum of basis functions which offer a well known behavior. A prominent example for periodic functions is Fourier analysis. For approximating smooth and continuous functions the use of symmetric basis functions, so called radial basis functions (RBF), has turned out to be a more efficient approach. Every output of $F(\cdot)$ is approximated by

$$F_i^*(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{\psi}(\boldsymbol{x}) = \sum_{j=0}^{n_c} w_j \psi(||\boldsymbol{x} - \boldsymbol{c}_j||), \tag{3.102}$$

where $\psi(\cdot)$ denotes the radial basis function centered at $\boldsymbol{c}$ and w_j denotes the individual weight. Thus, the value of $\psi(\cdot)$ depends only on the Euclidian distance between $\boldsymbol{x}$ and the center $\boldsymbol{c}$. Beside the choice of the center location, the choice of the RBF type $\psi(r)$ itself is of essential concern when using this approximation type. Prominent examples for simple basis functions are the linear function $\psi(r) = r$, the cubic function $\psi(r) = r^3$ or the thin plate spline $\psi(r) = r\log(r)$. Examples for parametric basis functions are the Gaussian function $\psi(r) = e^{-r^2/\sigma^2}$ or the multiquadratic RBF $\psi(r) = \sqrt{r^2 + \sigma^2}$. For fixed parameters and center points the weights $\boldsymbol{w}$ can again be found by a pseudo inverse approach on sampled data. The design matrix for this case is referred to as Gram matrix.

Kriging and LOLIMOT

Kriging is another approximation technique, which was originally invented for geostatical applications [Kri51]. The main characteristics of Kriging techniques is the use of basis functions of form [FSK08]

$$\psi^i = \exp\left(-\sum_{j=1}^{k} \theta_j |x_j^i - x_i|^{p_j}\right). \tag{3.103}$$

Hence, the basis functions equals a Gaussian kernel of $p_j = 2, \forall j$ and $\theta_j = 1/\sigma^2, \forall j$. However, due to these parameters the basis function (3.103) allows to be more specific. Based on this, one can see that the basis function for Kriging is in essential a covariance function. In fact, the approach for Kriging is then the same as for the presented Gaussian Process regression presented in subsection 2.3.1.

Local linear model tree (LOLIMOT) techniques [Nel01] are comparatively new techniques for approximation although their origin dates back to [TS85]. This technique basically works on the principle of neuronal networks, which makes it in principal similar to the previous presented techniques. LOLIMOT strategies divide the the complex model into several smaller subproblems. In the ideal case this subproblems are independent such that each subproblem can be solved independently. Then simple general approaches like linear models are used to describe the behavior of each subproblem.

3.5 Two Surrogate Approximations for ECT

In this section two approximations for ECT are implemented. One surrogate model is based on the polynomial approximation techniques discussed in the previous section. The second is a physically reduced model.

3.5.1 A Polynomial Surrogate Model

This section presents the design of a polynomial surrogate for ECT. The approach was presented by the author in [WNF10], where the polynomial surrogate was directly used within the MH algorithm. As explained in 3.4.2, the determination of the coefficients of the matrix $\boldsymbol{P}$ requires samples of the permittivity distribution $\boldsymbol{x}$ and the capacitances to build the Vandermonde matrix. In principle such samples can be drawn by exploring the prior distribution $\pi(\boldsymbol{x})$ using a MCMC scheme. However, a prior distribution $\pi(\boldsymbol{x})$ which is suitable for the solution of the inverse problem, may be insufficient to draw representative samples from $\boldsymbol{x}$, as $\pi(\boldsymbol{x})$ is mainly designed to incorporate knowledge about the solution. Instead it may appear more simple and efficient to directly draw patterns of the material distribution from the imagination about concrete realizations.

As a typical scenario for ECT often circular inclusions are given. Hence, an intuitive choice for the generation of samples is to place cylindric objects in an arbitrary sense inside the pipe. The following procedure presents a way to generate such patterns

1. Draw $\#_{\text{incl}} \sim \lfloor \mathcal{U}(1, \#_{\text{max incl}}) \rfloor$.
2. For the selected number of inclusions perform the following three steps:
    a) Draw $[x_{\text{center}}, y_{\text{center}}] \sim \mathcal{U}(\text{in } \Omega_{\text{ROI}})$.
    b) Draw $\varepsilon_r \sim \mathcal{U}(\varepsilon_{r,\text{min}}, \varepsilon_{r,\text{max}})$.
    c) Draw $r_{\text{incl}} \sim \mathcal{U}(r_{\text{min}}, r_{\text{max}})$.
3. Place the circular inclusion in the pipe Ω_{ROI}.
4. Compute the capacitances for the given pattern.

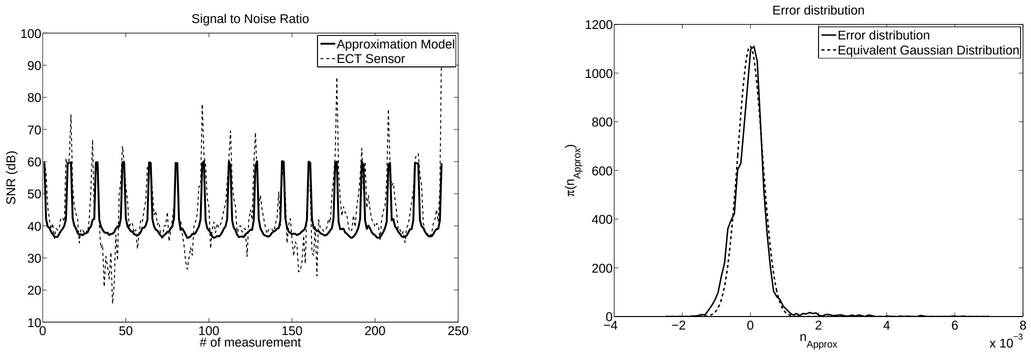
Alternatively, the material distribution can also be assumed to be Gaussian bumps. Hence, some smoothness assumption about the material distribution is incorporated. Similar to the choice of the radius when rods are used as prior samples, the sampled radius can be used as full width half maximum value (FWHM) to characterize such a Gaussian bump. Figure 3.3 depicts two typical samples of the two prior distributions. Like in [WNF10] an affine approximation of the forward map was built for low contrast material distributions (PVC rods) to replace a finite element solver with about 560 finite elements in the interior of the pipe. Figure 3.4 depicts the behavior of the approximation error for low permittivities. The achieved SNR of this approximation is depicted in Figure 3.4(a) together with a SNR measurement of the ECT system under empty conditions. One can see that the approximation has almost the same quality as the measurements. Figure 3.4(b) depicts a specific distribution of the approximation error. One can see that the distribution is nonsymmetric and has a tail towards positive values. However, like in [WNF10] the error was approximated as a Gaussian function as the main bulk of the distribution looks similar to one. The likelihood function

(a) Typical rod like pattern.

(b) Typical Gaussian bump like pattern

Figure 3.3: Typical material patterns obtained by the presented procedure to generate samples for building up an approximated forward map.

in the MCMC algorithm remains a Gaussian function, and the total covariance matrix is the sum of the measurement noise covariance matrix and the approximation error covariance matrix.



(a) Signal to noise ratio of the affine approximation a the ECT sensor.

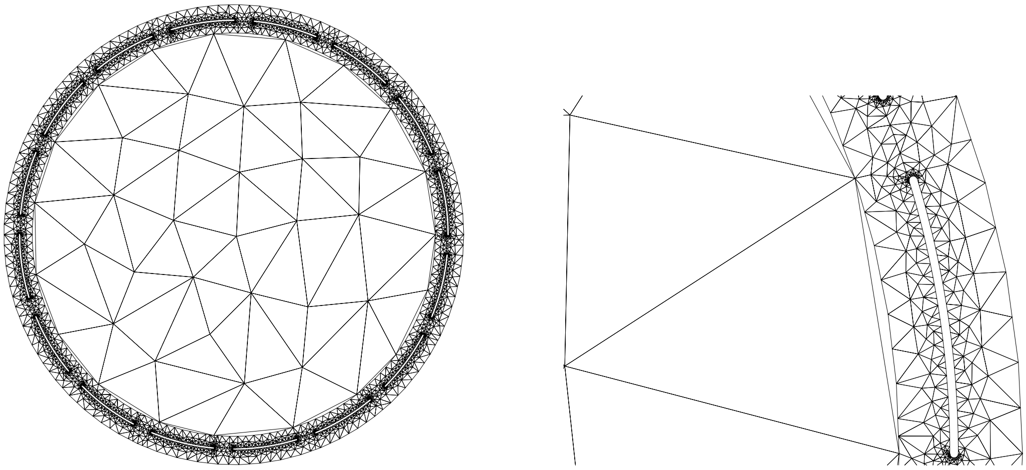
(b) Typical distribution of the approximation error.

Figure 3.4: Behavior of the affine approximation compared to ECT measurements.

3.5.2 A Reduced Physical Model - Mesh Fusion

The presented polynomial surrogate model is a local, non-physical approximation of the ECT forward map. This localness is a drawback, as it limits its use to certain material scenarios i.e. low permittivities. In this subsection a physically based surrogate is built. The reduction of the computational load is based on the use of a coarse mesh in the domain Ω_{ROI}.

The fine discretization in the region near the electrodes typically effects the element size even in some distant region from the electrode to be small and thus the number of finite elements to be high. This is because of the fact, that the expansion from small finite elements to larger ones requires a certain distance from the electrode. However, a fine discretization in the center region of the tube is not mandatorily required, as the sensitivity in the interior in the center region or in regions with some distance to the electrodes is low and thus in generally only a lower resolution can be achieved. This was explored in [SWZ$^+$07] and [NSWZ10]. It seems useful to replace the finite element discretization in the interior of the pipe by a coarser discretization. Figure 3.5(a) depicts a coarse mesh, which was inserted into the interior of the tube. For the domain $\Omega \setminus \Omega_{\mathrm{ROI}}$ the fine mesh remains. Figure 3.5(b) gives a more detailed view of the mesh fusion. As the internal mesh forms a polygon with less corners than the internal boundary of the tube, a gap between the both meshes comes up. It would be possible to form a curved finite element to close the gap. However, it was decided to keep the gap. Without any arrangements in the stiffness matrix the boundary of the gap would act


(a) Finite element mesh with coarse discretization in Ω_{ROI}.

(b) Detailed view on the tube internal boundary and the gap caused by the coarse mesh.

Figure 3.5: Accurate finite element discretization with coarse mesh in Ω_{ROI}.

like a homogeneous Neumann boundary condition, which introduces a wrong physical behavior. To avoid this, the potential of the boundary nodes of the fine mesh is approximated using linear combinations of the potential of the boundary nodes of the coarse mesh. The weights of the linear combination were determined using the ratios of the arc lengths on the interior of the tube. The number of unknowns of the equation system is decreased because the nodes on the interior of the tube can be eliminated after introducing the linear relationships to $\hat{\boldsymbol{K}}$, but this comes at the cost of an increased condition number of the matrix. In the further, the Schur complement (3.95) can be applied, as the ratio of unknowns in Ω_{ROI} compared to the remaining number of unknowns in $\Omega \setminus \Omega_{\mathrm{ROI}}$ suggests the use of a domain decomposition approach. The reduced stiffness matrix again has a block structure. However, due to the highly decreased number of unknowns, this equation system can be solved faster, even without further decompositions. Figure 3.6 finally depicts a potential plot of the electric scalar potential on the reduced mesh. The potential on the boundaries on the gaps is approximated by the corresponding edge points of the coarse mesh. Although the potential plot looks quite good, the approximation leads to large errors in regions of high field gradients. Hence, the approximation error can be expected to be large on electrodes close to the active electrode. A drawback of this approximation is the loss of the self adjoint property of the forward problem. This could be worked out by finding the boundary conditions which make the problem

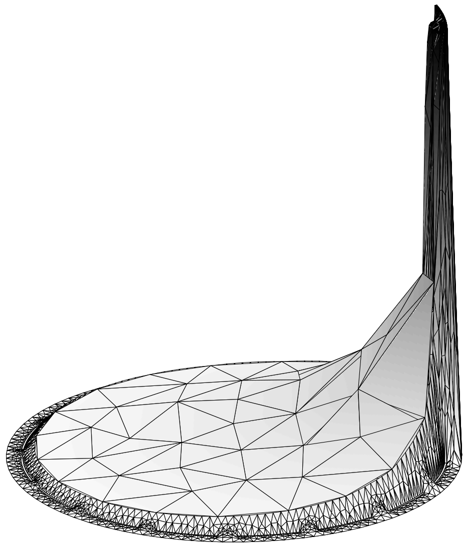


Figure 3.6: Potential distribution for the reduced physical model.

again self adjoint. The computation time of the reduced forward map was measured with about 1.3 ms. This is so far not a big gain compared to the other approximation schemes. However, the reduced physical model works over the whole permittivity range which is a valuable property.

3.6 Summary

In this chapter fast computational methods and techniques maintaining Green's functions as well as a specific model order reduction and approximation approaches were developed. The presented methods in this chapter lead to a variety of different reconstruction schemes in combination with the statistical inversion methods presented in chapter 2. Even the speed of the standard MH algorithm can be improved, as the speed comparison of subsection 3.3.8 suggests the use of the new approach compared to the standard approach presented in section 3.2. This can also be used to increase the speed of deterministic methods. The following list provides several combinations of the methods and algorithms which appear as interesting combinations. Lines with a star at the end have the main purpose to accelerate the inversion.

- A fast deterministic scheme*.
- Standard MH using the standard models.
- Standard MH using an approximation*.
- DAMH algorithm using an approximation*.
- Gibbs sampling using low-rank updates*.

The first entry is stated due to the availability of the Jacobian transpose operation which forms an efficient way to implement optimization-based inversion algorithms. The second point is the standard MH algorithm and is stated for completeness. However, this algorithm will be used to provide reference solutions compared to the accelerated schemes. The third point suggests the use of an approximated forward model directly within the standard MH method. This can be done by modifying the likelihood function to take care to the approximation error, which has already been presented in [WNF10], where the forward map has been replaced by an affine approximation. The DAMH algorithm in line four offers several different forms for combining computational methods and approximations together, with the main purpose to speed up Bayesian inversion methods. Most of this items allow at least several different forms or extensions like the application of the enhanced error model, an adaptive form of the polynomial approximation or the use of additional calibration information. The application of Gibbs sampling is motivated by the ability of computing fast low-rank updates. I.e. an arbitrary update of a single finite element can be done at the cost of a repeated inversion of the 3×3 matrix pencil in the Woodbury formula. Hence, it would be even possible to evaluate the conditional probabilities in an analytic way, once the Woodbury matrix is given. Thus, the application of Gibbs sampling seems a reasonable way, as the costly evaluation of the full forward map can be avoided during each conditional sampling step.

4 Case Studies

This chapter contains case studies using the discussed approaches and the developed methods in the two previous chapters. The first section gives an overview about the two types of data representation, the prior models, and describes the simulation setup.

Sections 4.2 a fast deterministic inversion scheme is presented which uses the BFGS (Broyden-Fletcher-Goldfarb-Shanno) algorithm and takes advantage of the Jacobian transposed operation. Section 4.3 contains the result of the DAMH reconstructions using the three presented approximation types and their variants, in section 4.4 a Gibbs sampler is presented. Finally section 4.5 contains the results from the Bayesian calibration studies where the different Bayesian formulations and the stochastic forward map are investigated and the topic of model errors and calibration strategies is treated.

All sections include the necessary details about the specific implementation and the results of the different algorithms and methods are presented and analyzed.

4.1 Reconstruction Framework and Simulation Setup

4.1.1 Volumetric Parameter Description

The volumetric parameter description is referred to as a low level representation. In this case, the elements of the parameter vector $\boldsymbol{x}$ correspond to the elements of the material vector $\boldsymbol{\varepsilon}$ or $\boldsymbol{\varepsilon}_r$ of the finite element forward map. Hence, $\boldsymbol{x} = \boldsymbol{\varepsilon}$ holds. The posterior probability density function is given by

$$\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}}) \propto \exp\left(-\frac{1}{2}\boldsymbol{e}^T\boldsymbol{\Sigma}_{\boldsymbol{v}}^{-1}\boldsymbol{e}\right)\exp\left(-\frac{1}{2}\alpha\boldsymbol{x}^T\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{x}\right)I(\boldsymbol{x}). \tag{4.1}$$

The first term in the exponential function builds the likelihood function for additive measurement noise, where $\boldsymbol{e}$ denotes the misfit $\boldsymbol{e} = F(\boldsymbol{x}) - \tilde{\boldsymbol{d}}$. The second term belongs to the prior distribution, where $\boldsymbol{L}$ incorporates knowledge about the solution. For most experiments a discrete second order operator has been used for $\boldsymbol{L}$, which incorporates a smoothness assumption for $\boldsymbol{x}$ [BHW03]. Hence, the reconstruction of sharp material boundaries is generally not possible due to this type of prior. α in equation (4.1) is a weight factor for the prior. The function $I(\boldsymbol{x})$ is an indicator function, which has the value one for allowed realizations of $\boldsymbol{X}$ and zero for infeasible realizations of $\boldsymbol{X}$. $I(\boldsymbol{x})$ is used to incorporate the constraints about the relative permittivity. The logarithm of the posterior distribution in equation (4.1) is in essential the same, as a least squares cost function in deterministic inversion schemes. Hence, the results obtainable by this prior and the used representation will in general be similar to results obtained by deterministic algorithms. For some illustrations of the result an adaptive threshold operation is applied in order to present a filtered result. The threshold level is defined by

$$\mathrm{th} = ((\varepsilon_{\mathrm{max}} - \varepsilon_{\mathrm{min}})/2 + 1), \tag{4.2}$$

which is the half peak value of the permittivity distribution.

The volumetric description will be used for deterministic reconstruction approach and for Gibbs sampling. A proposal generation scheme is not necessary, as samples are generated directly from the conditional density. For the application of standard MH algorithms, typical proposal generation scheme for the MH algorithm can be found in [FN97].

4.1.2 Contour Description

In contour reconstruction tasks it is aim to determine the shape of a contour in an otherwise uniform background material. The permittivity values are assumed to be known. Hence, an efficient way to present close boundaries is required. Such a representation is referred to as mid level representation. Especially in concern with a finite element forward map, a representation which features a simple way to map the shape of the contour on the permittivity vector $\boldsymbol{\varepsilon}$ is required. A very efficient scheme in combination with the finite element forward map are radial basis functions (RBFs) [BI98], [UPV04].

With the RBF approach an object is represented in implicit form given by $f(\boldsymbol{z}) = 0$, where $\boldsymbol{z}$ represents the Cartesian coordinates of a point on the boundary of the shape. The function $f(\cdot)$ is of form

$$f(\boldsymbol{z}) = \sum_{i=1}^{N} \lambda_i \phi(||\boldsymbol{z} - \boldsymbol{c}_i^N||) \tag{4.3}$$

where $\boldsymbol{c}_i^N$ represent a set of N given locations on the boundary of the object and the function ϕ is the used RBF. The support points $\boldsymbol{c}_i^N$ build the state vector $\boldsymbol{x}$. For the representation of closed contours in $2D$ the thin plate RBF

$$\phi(r) = r^2 \cdot \log(r), \tag{4.4}$$

is used, where r is the Euclidean distance $||\boldsymbol{z} - \boldsymbol{c}||$. The weights λ_i are determined by solving the equation system

$$\boldsymbol{A\lambda} = \boldsymbol{h}, \tag{4.5}$$

using the constraint $f(\boldsymbol{z} = \boldsymbol{c}_i) = h_i = 0$. For the thin plate spline, the set of basis functions needs to be augmented by the linear function $P(\boldsymbol{x}) = az_x + bz_y + c$, wherewith equation (4.3) becomes

$$f(\boldsymbol{z}) = \sum_{i=1}^{N} \lambda_i \phi(||\boldsymbol{z} - \boldsymbol{c}_i||) + P(\boldsymbol{z}). \tag{4.6}$$

Consequently equation (4.5) is extended to

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{Q} \\ \boldsymbol{Q}^T & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \boldsymbol{c}_p \end{bmatrix} = \begin{bmatrix} \boldsymbol{h} \\ \boldsymbol{0} \end{bmatrix}, \tag{4.7}$$

where $\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{x}_p & 1 \end{bmatrix}$ and $\boldsymbol{c}_p = \begin{bmatrix} a & b & c \end{bmatrix}^T$. To use the RBF scheme in an efficient way within a finite element scheme, the set $\boldsymbol{c}_i^N$ has to be extended by a further point, which lies inside the contour. For this point a negative value for h_i is assigned. Then, to determine if a point lies inside or outside the object, only the sign of the function f has to be evaluated. This procedure is referred to as mapping. Figure 4.1 depicts an exemplary RBF to illustrate this principle. For an accurate mapping onto a finite element mesh a grid of evaluation points has to be placed onto the finite elements. Then, the mapping becomes a counting task with respect to the points inside and outside the contour. As prior distribution $\pi(\boldsymbol{x})$ the following equation

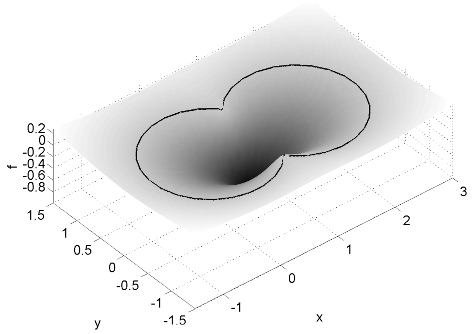


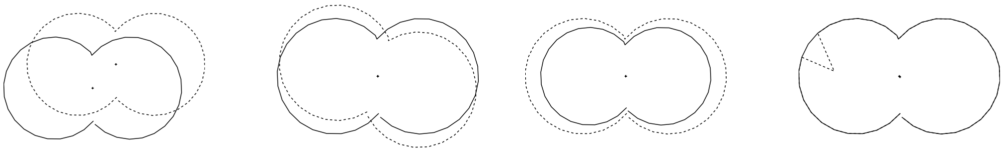
Figure 4.1: Exemple of a RBF shape description.

is used

$$\pi(\boldsymbol{x}) = \exp\left(-\frac{1}{2\sigma_{\mathrm{pr}}^2}\frac{c(\boldsymbol{x})}{\sqrt{\pi\Gamma(\boldsymbol{x})}}\right) I(\boldsymbol{x}), \tag{4.8}$$

where $c(\boldsymbol{x})$ denotes the circumference of the shape described by $\boldsymbol{x}$ and $\Gamma(\boldsymbol{x})$ denotes the area of the object [WF09]. Hence, the prior (4.8) evaluates the deviation of the shape to that of a circle. The degree of deviation is controlled by the variable σ_{pr}^2. $I(\boldsymbol{x})$ is again an indicator function. For the shape reconstruction, the indicator function proofs the contour for self intersection, correct interior angles, zig zag behavior, etc..

The proposal generation is done by using four basic moves to manipulate the shape. This moves are referred to as translation, rotation, scale and corner move [WF09] and are depicted in figure 4.2.

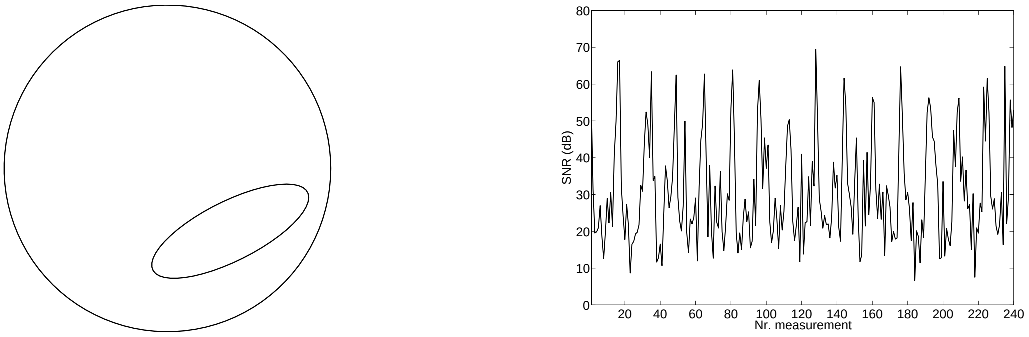


(a) Translation. (b) Rotation. (c) Scale. (d) Corner move.

Figure 4.2: The four moves for the proposal generation used for the shape reconstruction tasks.

An important point is the correct evaluation of the prior and the postprocessing of the data. The state vector $\boldsymbol{x}$ only contains the support points of the shape. The shape is determined by $f(\boldsymbol{z}) = 0$. An evaluation of the prior has to be done for the contour defined by this root finding problem. As the RBF description supports curved connection lines between the support points an evaluation of the direct connection line between two support points can lead to strong errors. To describe the uncertainty of the boundary estimation so called scatter plots can be used [WF09], which depict points of the posteriori distribution of the contour, i.e. uniform sampled points along the surface of the contour. Only plotting the support points does not present the contour and due to a certain clustering behavior this plots are less meaningful.

4.1.3 Simulation Setup



(a) True inclusion.

(b) Typical SNR.

Figure 4.3: True position of the material inclusion for the shape reconstruction tasks and typical SNR of the measurements.

For the simulation studies, the data $\boldsymbol{d} = P(\boldsymbol{x}_{\text{true}})$ for the true material distributions $\boldsymbol{x}_{\text{true}}$ was generated on the fine mesh which was used for the computation time studies in the previous chapter. Figure 4.3(a) depicts the true inclusion, which is used for the shape reconstruction tasks. The permittivity of the inclusion is assumed to be known and has a value of either $\varepsilon_r = 3.5$ (PVC) for the low permittivity case or $\varepsilon_r = 80$ (water) for the high permittivity case, respectively. The parameter σ_{pr} of the prior distribution (4.8) has been set to $\sigma_{\text{pr}} = 1$ for all shape reconstructions. For the finite element based reconstruction tasks a different distribution is reconstructed.

In all simulations the input data $\tilde{\boldsymbol{d}}$ was generated by corrupting $\boldsymbol{d}$ with additive, zero-mean Gaussian noise with a variance of $\sigma^2 = 5 \times 10^{-6}$, which was stated to be a typical noise level [WSF+07]. Figure 4.3(b) depicts the signal to noise ratio (SNR) for a realization of the input signal. Compared to the SNR level depicted in figure 3.4(a) the SNR for the reconstructions is considerable lower. However, this level will be used, as the obtained results depict the behavior of the algorithms under harder conditions.

To provide a fair comparison between the different MH and DAMH variants for shape reconstructions the same proposal kernel has been used for all simulations. The probability of the corner move has been set to 3 % while the remaining moves had equal probability. The same kernel is also used for the later reconstructions, where calibration becomes an issue.

For the experiments aiming on the acceleration of statistical inversion methods the calibration was performed using the empirical Bayes approach based on the two point calibration. For this, calibration data was generated for $\boldsymbol{x}_L = 1$, $\boldsymbol{x}_{H_1} = 3.5$ and $\boldsymbol{x}_{H_2} = 80$, which means that the domain Ω_{ROI} was completely filled with a medium of the corresponding relative permittivity (air, PVC, water).

4.2 An Accelerated Deterministic Scheme

Although this thesis is focused on using of Bayesian inversion methods, the developed techniques in section 3.3 motivate the application of an accelerated deterministic inversion algorithm for element-based reconstruction. This section will briefly demonstrate an accelerated deterministic method. The motivation for this lies much less in the application of deterministic algorithms, but to use the output of a deterministic method to initialize a statistical inversion algorithm. This is motivated by the idea to decrease the duration of the burn in period by providing an initial mean estimate for $\boldsymbol{x}$.

Deterministic algorithms are often treated as optimization problems which minimize an appropriate cost functional. An often used approach for ECT is given by [BHW03]

$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \left\| F(\boldsymbol{x}) - \tilde{\boldsymbol{d}} \right\|_2^2 + \alpha \boldsymbol{x}^T \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{x}, \tag{4.9}$$

$$\text{s. t. } I(\boldsymbol{x}) \tag{4.10}$$

which is exactly the logarithm of the posterior distribution (4.1). However, the prior terms are named regularization terms and α denotes the regularization parameter. For fast convergence mostly second order optimization schemes are used to solve (4.9). This leads to the scheme

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - s\left(\mathbf{J}^T\mathbf{J} + \alpha\boldsymbol{L}^T\boldsymbol{L}\right)^{-1}\left(\boldsymbol{J}^T\boldsymbol{r}_k + \alpha\boldsymbol{L}^T\boldsymbol{L}\boldsymbol{x}_k\right), \tag{4.11}$$

where $\boldsymbol{J}$ is the Jacobian and $\boldsymbol{r}$ is the residual vector containing the difference between the model output $F(\boldsymbol{x})$ and the measured data $\tilde{\boldsymbol{d}}$. The parameter s denotes a step-size or line search parameter. The fundamentals of (4.11) are given by the Gauss-Newton method, where the Hessian matrix is approximated by

$$\boldsymbol{H} = 2\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{J}, \tag{4.12}$$

and the gradient is given by

$$\boldsymbol{g} = 2\boldsymbol{J}^{\boldsymbol{T}}\boldsymbol{r}. \tag{4.13}$$

Then, for a positive definite Hessian matrix a descent direction for (4.9) is given by

$$\boldsymbol{s} = -\boldsymbol{H}^{-1}\boldsymbol{g}, \tag{4.14}$$

which is referred to as Newton direction [Fle87]. To incorporate the indicator function $I(\boldsymbol{x})$, the scheme typically requires methods used in constrained optimization like an active set method [Fle87], to ensure the avoidance of infeasible values of $\boldsymbol{x}$.

4.2.1 BFGS based Hessian Update

A major concern about the Gauss Newton scheme (4.11) is the requirement of the Jacobian $\boldsymbol{J}$ to approximate the Hessian matrix $\boldsymbol{H}$. Although the Jacobian can be reconstructed by Jacobian operations, it seems more suitable to use a modified scheme which directly operates on the Hessian. An efficient update scheme which directly works on the inverse of the Hessian is the BFGS (Broyden-Fletcher-Goldfarb-Shanno) scheme [Fle87]. The iteration steps are given by

1. Evaluate the Newton direction $\boldsymbol{p}_k = -\boldsymbol{H}_k^{-1}\boldsymbol{g}_k$.
2. Find s to set $\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + s\boldsymbol{p}_k$ and set $\boldsymbol{s}_k = s\boldsymbol{p}_k$.
3. Compute $\boldsymbol{z}_k = \boldsymbol{g}(\boldsymbol{x}_{k+1}) - \boldsymbol{g}(\boldsymbol{x}_k)$.
4. Evaluate $\boldsymbol{H}_{k+1}^{-1} = \boldsymbol{H}_k^{-1} + \frac{\boldsymbol{s}_k^T\boldsymbol{z}_k + \boldsymbol{z}_k^T\boldsymbol{H}_k^{-1}\boldsymbol{z}_k}{(\boldsymbol{s}_k^T\boldsymbol{z}_k)^2}\boldsymbol{s}_k\boldsymbol{s}_k^T - \frac{\boldsymbol{H}_k^{-1}\boldsymbol{z}_k\boldsymbol{s}_k^T + \boldsymbol{s}_k\boldsymbol{z}_k^T\boldsymbol{H}_k^{-1}}{(\boldsymbol{s}_k^T\boldsymbol{z}_k)}$.

As the gradient $\boldsymbol{g}_k$ can be evaluated by the Jacobian transpose operation, the scheme fits suitable to the developed framework. The efficiency of the BFGS algorithm is based on the direct update of $\boldsymbol{H}_k^{-1}$ in line 4. Hence, the scheme is in fact nothing else than a version of the Woodbury identity. Although it is not the most genteel way, the value of $\boldsymbol{x}_{k+1}$ in step 2 should be corrected for feasibility by clipping $\boldsymbol{x}_{k+1}$, in order to obtain physical meaningful results.

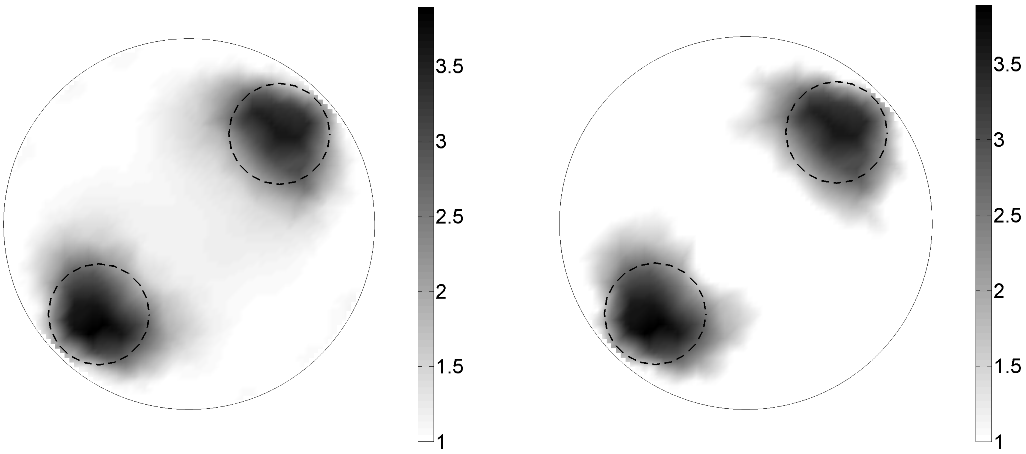
4.2.2 Deterministic Results with Different Regularization Terms

Figure 4.4(a) depicts the result obtained by the deterministic algorithm using the smoothness prior. The regularization parameter has been set to $\alpha = 10^{-7}$. This value was found by experiment. A systematic to find an appropriate regularization parameter is given by the L-curve criterion [Han98] but also adaptive methods exist [WBH04]. The two black circles mark the true position of the inclusions. The true relative permittivity was set to $\varepsilon_r = 3.5$. For the reconstruction, the line search parameter s has been set to $s = 1$ and a fixed number of 50 iterations was evaluated. The result depicted in figure 4.4 is a typical result for deterministic algorithms. Because of the low level representation, the image quality is in general comparatively low.

To demonstrate the ability of an equivalent Gaussian prior distribution, a covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{x}}$ has been computed from samples generated by the method described in section 3.5.1. Then the regularization matrix $\boldsymbol{L}$ for equation (4.9) is given by the Cholesky decomposition $\boldsymbol{L}^T\boldsymbol{L} = \boldsymbol{\Sigma}_{\boldsymbol{x}}$. Figure 4.5 depicts the results using the Gaussian prior. The algorithm settings have been the same as for the reconstruction using the smoothness prior. Compared to the results depicted in figure 4.5, the results are less blurred and offer a better better spatial resolution with respect to the contours of the inclusion. Remembering that the Gaussian prior was formed by just using samples from imaginary material distributions the approach is at least an effective way to find suitable regularization terms.

4.3 MH Algorithm with Approximations

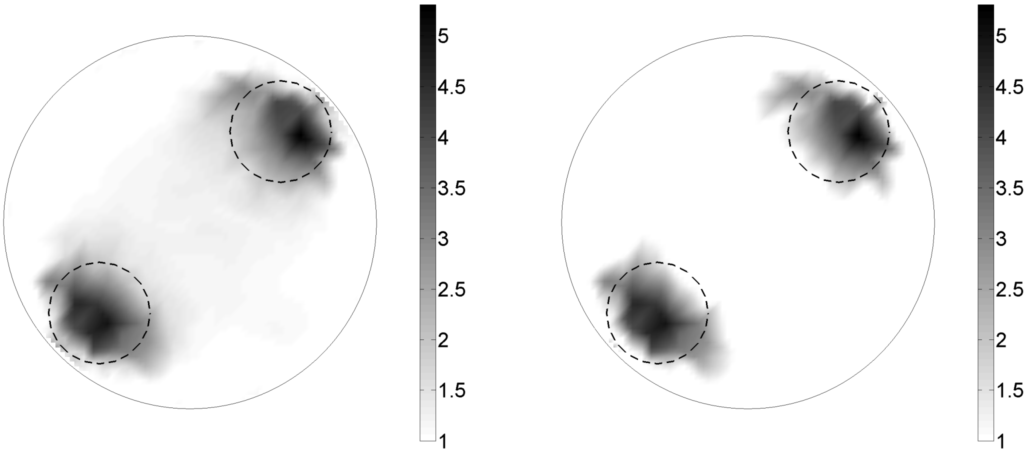
In this section, the DAMH is used to perform the reconstruction of the inclusion depicted in figure 4.3(a) using different approximation techniques. For comparison subsection 4.3.1 provides a reference solution obtained by the standard MH algorithm. For all shape reconstruction experiments the a



(a) Deterministic result.

(b) Deterministic result (threshold applied).

Figure 4.4: Result obtained by the accelerated deterministic scheme using the BFGS update.



(a) Deterministic result.

(b) Deterministic result (threshold applied).

Figure 4.5: Result obtained by the accelerated deterministic scheme using the BFGS update with a sample based Gaussian prior.

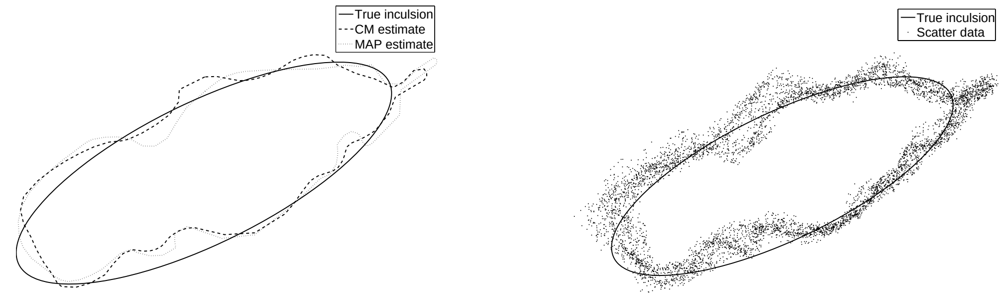
circular inclusion placed in the center of the pipe has been used to initialize the algorithm. For the evaluation the first 100000 samples were removed from the chain in order to be not effected by the the burn in phase.

4.3.1 Standard MH Reference Solution

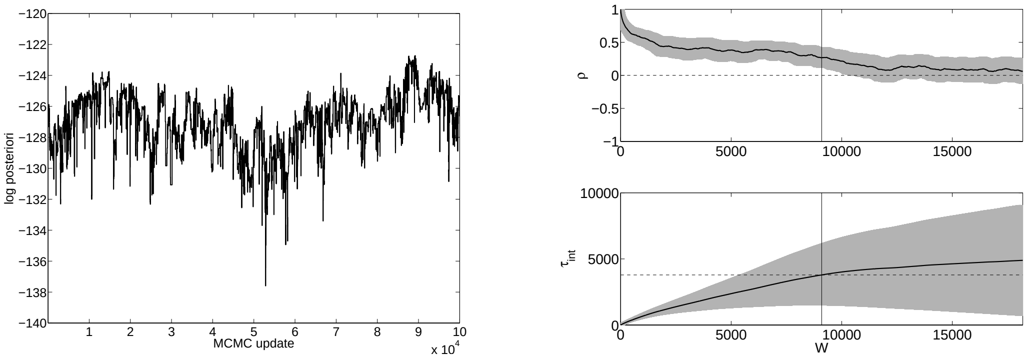
Figure 4.6 and 4.7 depict the results obtained with the standard MH algorithm. A summary of the behavior of the algorithm is also included in the first block of table 4.1, where Ac_α is the acceptance ratio for the proposal. Line 1 and 2 of table 4.1 list the results when inverse crime data is used. For all other reconstructions the appropriate two point calibration scheme was applied.

The results for the reconstruction of the low permittivity case (PVC) depicted in figure 4.6 are acceptable. The conditional mean estimate an the ML estimate offer a small part sticking out of the main volume of the inclusion. Beside this the remaining parts of the inclusion are well estimated. The oscillations along the boundary and the part sticking out indicate that the used prior is comparatively moderate. A much stricter prior would smooth the result and the behavior should vanish. Because of the decreased sensitivity in the interior of the pipe, the determination of the boundary part of the contour facing towards the center of the pipe shows an increased variance. This can be oberved in the scatter plot depicted in figure 4.6(b). Although the result is acceptable, the chain is yet not in equilibrium as can be seen by the output trace depicted in figure 4.6(c). This can also be seen by the autocovariance function and the IACT τ_{int} depicted in figure 4.6(d) and the large large uncertainty illustrated by the gray shaded regions. Thus the chain would have to run longer but for the purpose of this section, to demonstrate the accelerated behavior of the chains using the approximations, the simulation was stopped. Although its large uncertainty the IACT was evaluated with about 4000 samples which is considerable long. The result with inverse crime data looks almost the same but it has a lower IACT and the chain has a better behavior. This indicates that the calibration scheme works but already introduces an error.

In contrast, the results for the high permittivity case depicted in figure 4.7 offer an unsatisfactory behavior, as the determination of the shape offers a visible biased behavior. The low variance in the scatter plot in figure 4.7(b) indicate that the chain almost stucks. This bad behavior of the chain can also be observed by the output trace depicted in figure 4.7(c) and 4.7(d). For

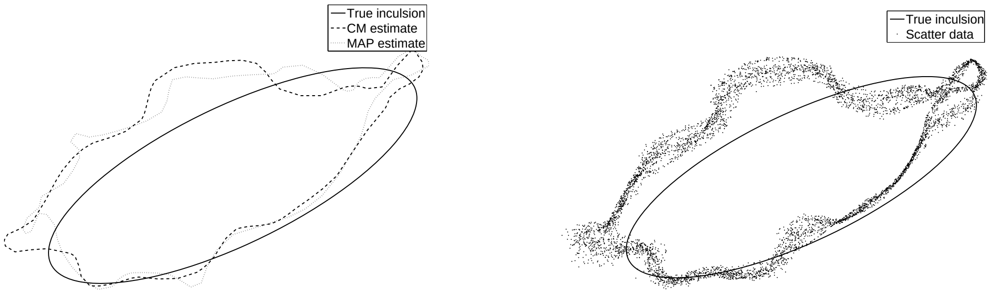


(a) Mean estimate. (b) Scatter plot.

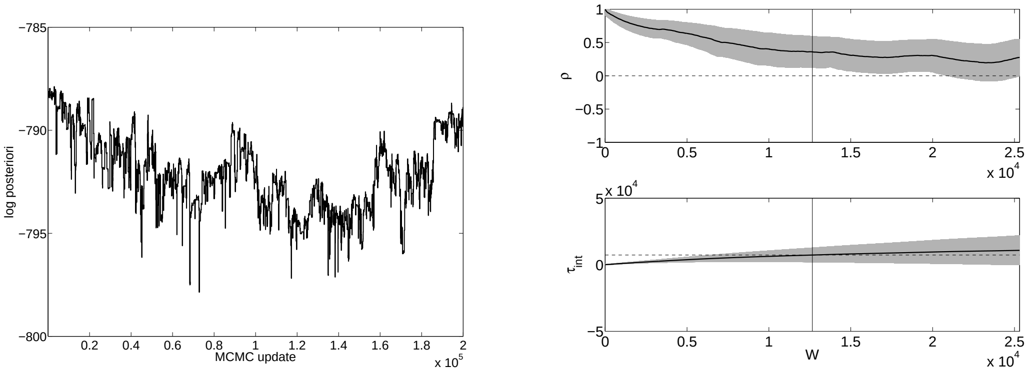


(c) MCMC output. (d) IACT.

Figure 4.6: MH reference results for the low permittivity case (PVC).



(a) Mean estimate. (b) Scatter plot.



(c) MCMC output. (d) IACT.

Figure 4.7: MH reference results for the high permittivity case (water).

inverse crime data the reconstruction leads to a proper result. The IACT for the inverse crime case is increased with respect to the low permittivity case. This can be explained by a better fitting of the proposal kernel compared to the low permittivity case, which is effected due to the nonlinear behavior of the forward map. However, the bad behavior of the chain and the large value of τ_{int} in line 4 have a different origin. Compared to the inverse crime case this behavior has its origin in an improperness in the input data for the MCMC algorithm, which is caused by the calibration scheme. The calibration for high permittivities fails. This will be of concern in the later sections. By the large IACT it can be stated, that for high permittivities the Markov chain almost gets stuck due to improper input data.

Although this first results indicate the problem of model errors and that the used calibration scheme has problems for inclusions with high permittivities it will be used in the next subsection to compare the performance of the different approximation types.

4.3.2 DAMH Setups for the Different Approximations

This subsection describes the setups for the DAMH algorithm for the three types of used approximations. In the following subsections the term enhanced error model [KS05] will be used to represent the adaptive form of it, which is based on the posterior distribution, as it was presented in subsection 2.3.2.

First Order Approximation

The setup for the DAMH algorithm using a first order approximation is considerable simple, as the approximation is given by the Jacobian operation. No further adjustments on the approximation itself have to be made. For the evaluation of the likelihood of the proposal the full set of data $\tilde{\boldsymbol{d}}$ is used. For both types of inclusions the following experiments were performed:

- Pure DAMH.
- DAMH with adaptive enhanced error model.

Polynomial Approximation

For the DAMH algorithm using a polynomial approximation, in a first step two affine approximations were built as described in subsection 3.5.1. To evaluate the likelihood of the proposal, a reduced data vector $\tilde{\boldsymbol{d}}_r$ is used. In this case, the data of the electrodes next to the transmitter electrode has been skipped. The following experiments were performed for both types of inclusions:

- Pure DAMH with static approximation.
- DAMH with adaptive enhanced error model and static approximation.
- DAMH with adaptive enhanced error model and adaptive approximation.

The adaptive version of the affine approximation is based on the LMS update presented in equation (3.101). For this case, the evaluation of $\boldsymbol{e}_n$ for the adaptive enhanced error model requires a re-evaluation of the model output on the updated forward map.

Reduced Physical Model

Compared to the other approximations, the application of the reduced physical model within the DAMH algorithm requires a more sensitive setup of the algorithm. Although the potential plot in figure 3.6 looks quite reasonable, the approximation of the boundary potential by the nodal potentials of the coarse mesh, is a major error source. This especially holds for the charges on the electrodes close to the active electrode, as the field gradient in this region has its maximum. Again a reduced data vector $\tilde{\boldsymbol{d}}_r$ is used for the evaluation of the likelihood of the proposal. A further essential detail about the application of the posterior based enhanced error model, is given by the computation of the error $\boldsymbol{e}_n = F(\boldsymbol{x}_n) - F^*(\boldsymbol{x}_n^*)$. As the model $F(\cdot)$ has a different gain and a different offset compared to the approximation $F^*(\cdot)$, the output $\boldsymbol{y}$ of $F(\cdot)$ has to be mapped into the range of the approximation $F^*(\cdot)$ to build the enhanced error model. This is done by a linear interpolation in form of an offset gain correction similar to the two point calibration given by formula (2.46).

4.3.3 Summary of DAMH Results

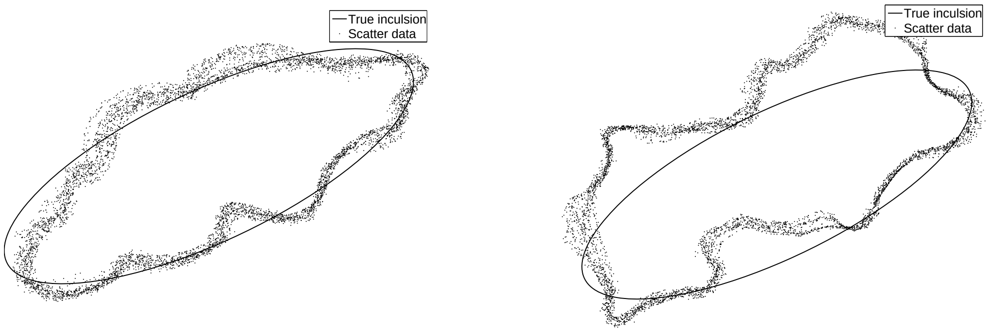
Table 4.1 collects an analysis about the reconstruction tasks using the MH algorithm and the DAMH algorithm with different approximations and for different setups. Some selected reconstruction results are depicted in figure 4.8. For the DAMH algorithm, Ac_α in table 4.1 is the acceptance ratio for the proposal when the approximation is used. $Ac_{\beta|\alpha}$ then is the acceptance rate for the second step. Hence, $Ac_{\beta|\alpha}$ can be understood as a quality measure for the approximation. Ac_{β} is the overall acceptance rate of the DAMH algorithm.

The second block in table 4.1 contains the results using the DAMH algorithm in combination with the Jacobian approximation. Some results of this experiments are depicted in Figure 4.8(a), 4.8(b) and 4.8(c). For the low permittivity case the application of the Jacobian approximation introduces a large improvement as can be seen by figure 4.8(a) and 4.8(c). The algorithm behaves almost like the MH algorithm. The almost same value of $Ac_{\beta|\alpha}$ indicates that the the approximation error is almost negligible as can be seen in line 7 and 8. The expected behavior for the Jacobian approximation of being a good approximation for small changes is fulfilled. Figure 4.8(b) depicts the result for the high permittivity case without using the enhanced error model. The result from the scatter plot is of equal quality as the result obtained by the MH algorithm but from the statistics provided in table 4.1 one can see that the chains for this case almost stick. Applying the enhanced error model for the high permittivity case introduces an interesting behavior as can be seen in line 8. Surprisingly, the acceptance rate Ac_{α} and Ac_{β} drop and $Ac_{\beta|\alpha}$ reaches 100%. This is explainable, by the large approximation error for high permittivities. As the enhanced error model adopts, it learns to this large approximation error and leads to a significant low acceptance rate Ac_{α}. Those proposals are then accepted at almost 100% rate leading to $Ac_{\alpha} \approx Ac_{\beta}$. However, the overall acceptance rate is too low to make the approach usable. The result could be improved by adopting the moves in order to obtain smaller differences between the actual state and the proposal leading to a smaller approximation error. However, for large differences between the current state and the proposal the use of the Jacobian approximation is critical.

In the third block of table 4.1 the results using the affine approximation are listed. Figure 4.8(d), 4.8(e), and 4.8(f) depict the results for the low permittivity case including the static approximation without enhanced error model, the static approximation with the enhanced error model and the adaptive version of the affine approximation. The affine approximation did not work in the high permittivity case for the same region as the Jacobian approximations and thus no results are depicted. This failure for high permittivities is caused by the increased model error of the affine map for high permittivities in the static case. In the adaptive case, it appears that the forward map becomes to local and thus an exploration of the full state space

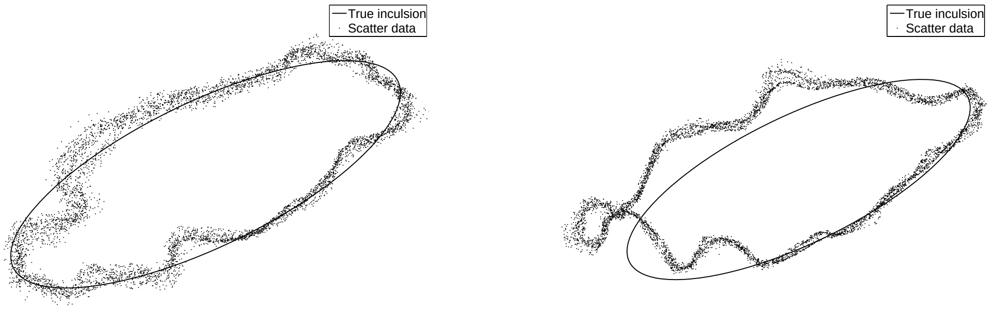
Nr.	Simulation	ε_r	Conv.	Ac_{α} %	$Ac_{\beta\|\alpha}$ %	Ac_{β} %	τ_{int}
Standard MCMC:							
1	Inverse crime	3.5	y	6.09	X	X	1832
2	Inverse crime	80	y	3.73	X	X	3965
3	Normal	3.5	y	4.31	X	X	3788
4	Normal	80	y	2.41	X	X	7286
DAMH with first order approximation:							
5	Normal	3.5	y	3.97	84.29	3.35	1949
6	Normal	80	y	2.56	92.7	2.37	10049
7	With EEM	3.5	y	4.75	84.24	4.00	3828
8	With EEM	80	n	0.004	100	0.004	10209
DAMH with affine approximation:							
9	Normal	3.5	y	2.91	60.54	1.76	3854
10	Normal	80	n	0.35	78.99	0.28	5624
11	With EEM	3.5	y	4.50	73.02	3.28	1199
12	With EEM	80	n	1.03	58.14	0.60	1144
13	Adap. appr. with EEM	3.5	y	5.68	53.4	3.03	4584
14	Adap. appr. with EEM	80	n	0.21	77.46	0.16	11169
DAMH with reduced model:							
15	With EEM	3.5	y	6.49	53.71	3.49	1945
16	With EEM	80	y	9.79	10.90	1.06	4584

Table 4.1: Results from the reconstructions using the MH and the DAMH algorithm in combination with different approximations.



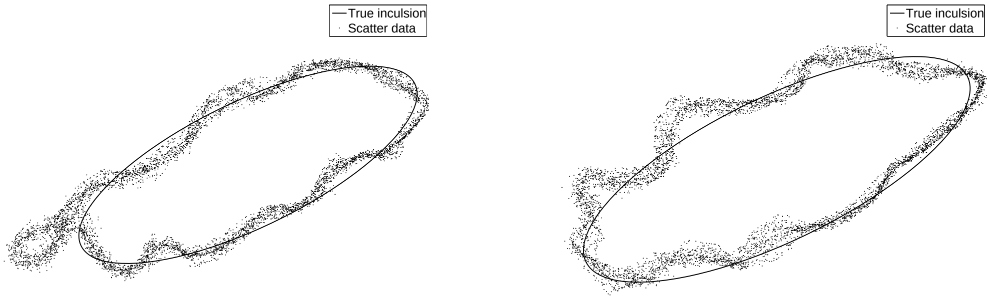
(a) Line 5, Jacobian, no EEM, $\varepsilon_r = 3.5$.

(b) Line 8, Jacobian, $\varepsilon_r = 80$.



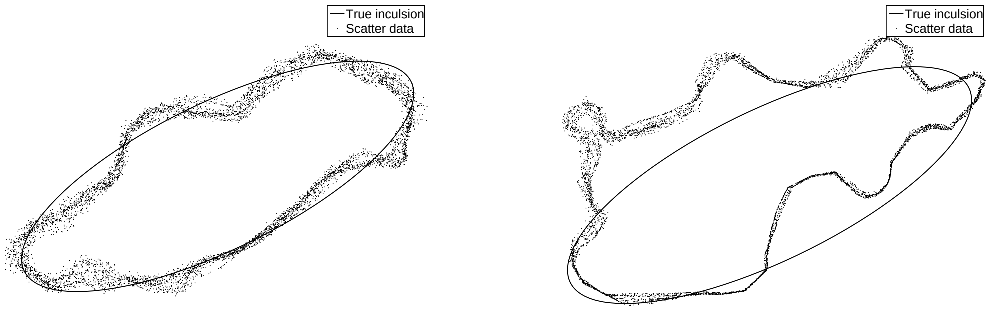
(c) Line 7, Jacobian, $\varepsilon_r = 3.5$

(d) Line 9, affine map, no EEM, $\varepsilon_r = 3.5$.



(e) Line 11, affine map, $\varepsilon_r = 3.5$.

(f) Line 13, adapt. affine map, $\varepsilon_r = 3.5$.



(g) Line 15, red. model, $\varepsilon_r = 3.5$.

(h) Line 16, red. model, $\varepsilon_r = 80$.

Figure 4.8: Some results obtained by the DAMH algorithm in combination with different approximations. The subtitles at the figures correspond to the lines of table 4.1.

becomes impossible once the approximation has adopted to a too local area. Hence, also the adaptive version of the affine approximation fails for the high permittivity case. Looking at the results from figure 4.8(d) to figure 4.8(f) and keeping in mind the completely artificial nature of the approximation the results are outstanding. This especially holds for the last two cases. Also the increased values of the IACT are remarkable. The acceptance rate $Ac_{\beta|\alpha}$ is lower than for the Jacobian approximation. Comparing line 9 with line 11, the effect of the enhanced error model becomes observable by the increased acceptance rate and the decreased value of τ_{int}. In this case the error between the forward map and the approximation is small enough compared to the case for the Jacobian approximation in line 8. Thus, the enhanced error model is able to effectively compensate the error. In the adaptive case, it appears that the forward map becomes to local and thus an exploration of the full state space becomes impossible once the approximation has adopted to a too local area.

The fourth block contains the DAMH results using the reduced model. The two reconstruction results are depicted in figure 4.8(g) and 4.8(h). Both results are equal to the results obtained by the standard MH algorithm. For the low permittivity case depicted in figure 4.8(g) the result even appears as one of the best results of the series of scatter plots. However, the low acceptance rate $Ac_{\beta|\alpha}$ indicates, that the used approximation is almost at the limit of its meaningful usability as only 50% of the accepted. This value even drops for the high permittivity case as can be seen in line 16. The overall behavior of the DAMH using the reduced model is equal to the standard MH algorithm for the low permittivity case. The approximation fulfills its purpose but it should be stated as a reminder that approximations like the presented reduced physical model have to be handled with care. As this reductions directly play with physical properties of the system the introduced error can become considerable large. Model reduction based on physical reductions should be handled with care. Nevertheless, the results also presents the powerfulness of the enhanced error model.

4.4 Gibbs Sampling for ECT

In this section Gibbs sampling on the finite element representation is presented. For the application of a Gibbs sampler the computational scheme in concern with the computational methods is always given by

1. Draw a conditional sample.

2. Compute the new forward map to obtain the new Green's functions $\boldsymbol{G}_Q$ for the cheap evaluation of the next conditional sample.

The scheme requires the continued evaluation of the forward map after each conditional sampling step in order to obtain the new Green's functions $\boldsymbol{G}_Q$. This appears as computationally expensive. However, the applicability of the Gibbs sampler is motivated by the ability to performing the conditional sampling with low computational costs due to the availability of low-rank updates.

When using the standard Woodbury scheme, the time amount of the reevaluation of the new set of Green's functions $\boldsymbol{G}_Q$ for the next element can be reduced, by solving the update of the forward map using the extended right hand side matrix $\left[\begin{array}{cc}\boldsymbol{R}_Q & \boldsymbol{E}_C\end{array}\right]$. $\boldsymbol{E}_C$ includes the identity vectors for the nodes C, which correspond to the next finite element in the update chain of the Gibbs sampler. This advantage is only because of the special implementation of the matrix inversion scheme in MATLAB and does not hold for the general case. For the Woodbury formula using the $\boldsymbol{WSW}^T$ formulation this extension of the right hand side is not necessary. However, to keep the Woodbury matrix a 3×3 matrix, the update is mandatory.

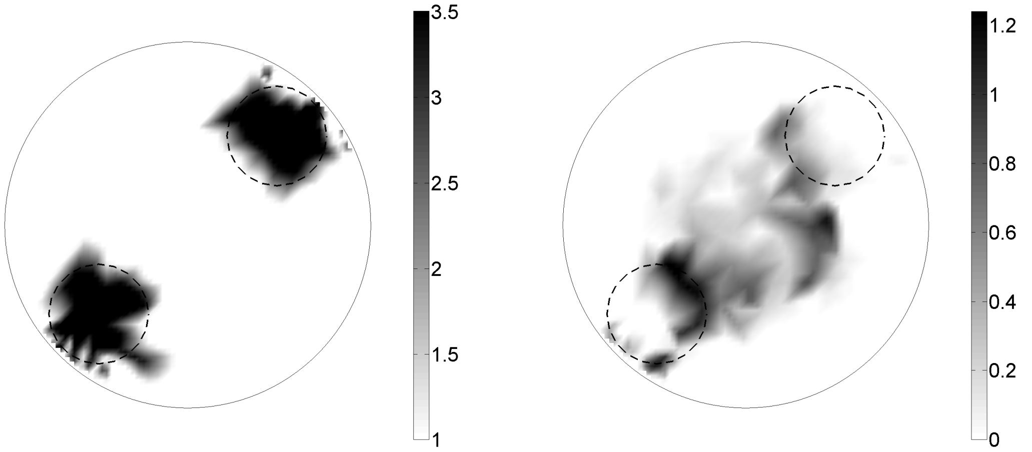
It could be argued, that reevaluations of the forward map could be saved by using larger low-rank updates as the inversion of even a larger Woodbury matrix seems cheaper in a numerical sense than the solution of the forward map. However, this scheme was were not implemented.

4.4.1 Gibbs Sampling for Bimodal Material Distributions

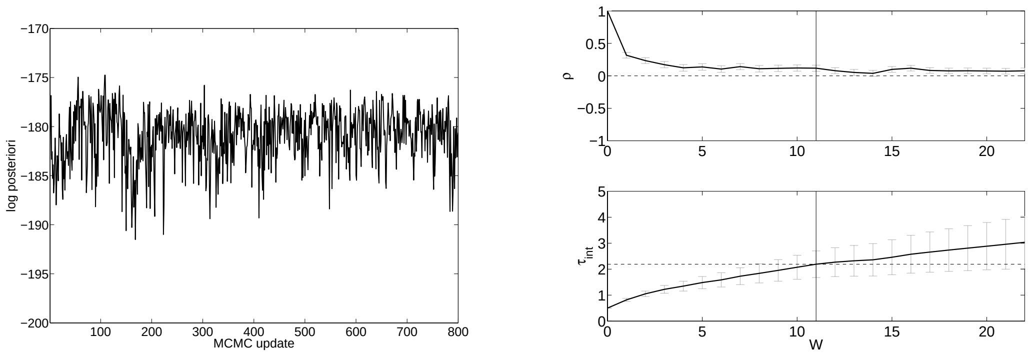
The simplest application of Gibbs sampling is given for bimodal material distributions. In this case, the update scheme is given as

1. Flip one element of $\boldsymbol{x}$ to generate the proposal $\boldsymbol{x}'$.
2. Compute $\alpha = \min\left[1, \frac{\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}})}{\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})}\right]$.
3. Accept $\boldsymbol{x}'$ with probability α.

The selection of the finite element to be toggled can be done by several ways. In the most simple case, only a loop is applied to update all elements consecutively during a sweep. The probability of the opposite state is evaluated



(a) Mean estimate (threshold applied). (b) Standard deviation.



(c) MCMC output. (d) IACT.

Figure 4.9: Results from Gibbs sampling for bimodal material distribution.

using a low-rank update scheme. For small changes, i.e. the difference between the permittivity values is small even the Jacobian operation can be used, to approximate the probability of the opposite state. To initialize the algorithm, the output of the deterministic algorithm can be used after applying a threshold operation with respect to the bimodal distribution. Figure 4.9 depicts the results obtained from bimodal Gibbs sampling. The algorithm was initialized with the result of the deterministic algorithm. As the difference between the the two permittivity states is low, the probability of the opposite state was evaluated using the Jacobian operation. The regularization parameter α has been set to $\alpha = 10^{-7}$.

Compared to the deterministic result, the mean estimate depicted in figure 4.9(a) offers an excellent result. The standard deviation shown in figure 4.9(b) reflects the low sensitivity in the center region of the pipe. On a part of the boundary of the left inclusion the variance is considerably higher, which is an expectable behavior and already known from the shape reconstructions. However, the interpretation of the standard deviation for low level representation has a slightly unclear character which comes due to the inherently low quality of this representation. An interesting effect about the efficiency of the algorithm is depicted in figure 4.9(c) and figure 4.9(d). The IACT of $\tau_{\text{int}} = 2$ is considerable low, which means that the bimodal Gibbs sampler, once in equilibrium, produces an independent sample at every second iteration. As the use of the Jacobian operation provide an enormous speed up, this algorithm offers the potential of real time application.

4.4.2 Gibbs Sampling for Arbitrary Material Distributions

The more demanding case for Gibbs sampling is given for the case of unknown material values. For this samples from the full conditional posterior distribution have to be generated. The straight forward way for providing conditional samples is given by a numerical implementation of inverse transform sampling. The conditional distribution becomes a function of the scaling variable γ in equation (3.66) and the support of γ is determined by the indicator function $I(\boldsymbol{x})$. For a numerical implementation the conditional distribution is approximated by some support points. This is a possible option, although it requires a larger number of low-rank evaluations, in order to provide an accurate representation of the conditional density. However, the scheme proved to have some numerical difficulties with fast declining densities, as it operates directly on the posterior distribution and not on the logarithm of it. Hence, a different algorithm has to be used.

Conditional Sampling of the Posterior Distribution

A scheme to avoid inverse transform sampling is rejection sampling [CG92]. To take samples from a target density $g(X)$, rejection sampling requires a second density function $g_u(X)$, with $g_u(X) \geq g(X)\ \forall X$. If $g(X)$ is easy to sample from, samples of $g(X)$ can be effectively generated by

1. Sample a point x from $g_u(X)$ and $u \propto \mathcal{U}([0,1])$.
2. If $u \leq \frac{g(x)}{g_u(x)}$ accept x, otherwise repeat.

Also a version using a lower bound $g_l(x) \leq g(x) \leq g_u(x)$ exists, where the sampling is done by

1. Sample a point x from $g_u(X)$ and $u \propto \mathcal{U}([0,1])$.
2. If $u \leq \frac{g_l(x)}{g_u(x)}$ accept x.
3. If not, accept x in the case of $u \leq \frac{g(x)}{g_u(x)}$. Repeat otherwise.

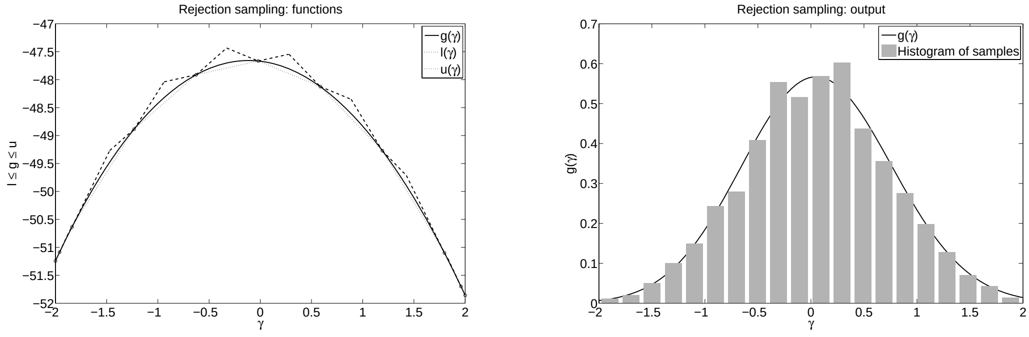
Due to the constraints given by the indicator function $I(\boldsymbol{x})$ the support will become bounded. In this case, the reconstruction of the functions $g_l(X)$ and $g_u(X)$ is further simplified [GBT95].

For the case that $g(x)$ offers log concave behavior [GBT95], another version of rejection sampling can be implemented, which allows an even simpler construction of the upper and the lower function. In this case the rejection sampler also works on the logarithm of $g(x)$, by using $h(x) = \ln(g(x))$. The two bounding functions $l(x)$ and $h(x)$ have to fulfill $l(x) \leq h(x) \leq u(x)$. As the scheme works on the logarithm, the functions $l(x)$ and $h(x)$ can be designed using simple piecewise linear functions. Then, a sample can be generated by the following

1. Draw a sample x from $u(x)$ and $w \propto \mathcal{U}([0,1])$.
2. If $w \leq \exp(l(x) - u(x))$ accept x.
3. If not, accept x in the case of $w \leq \exp(h(x) - u(x))$. Repeat otherwise.

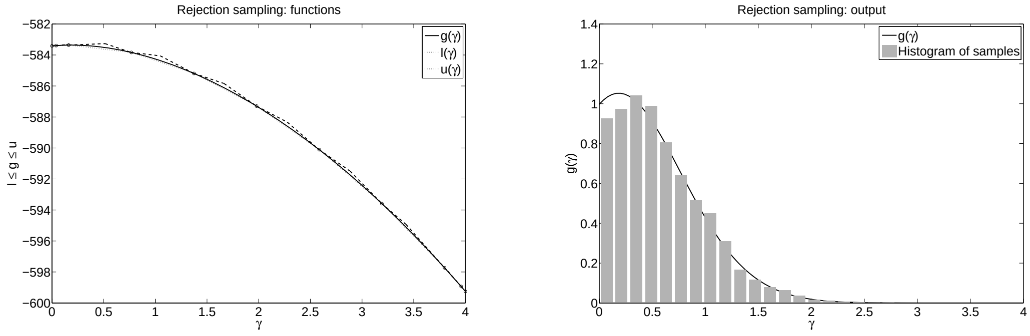
A simple method to sample from $u(x)$ is presented in [Lux78]. An extension to this scheme is adaptive rejection sampling, where the rejected point of the evaluation of $g(x)$ is used to update the upper and the lower bounding functions.

It is in some sense a controversial issue that schemes like rejection sampling were designed to generate a huge number of independent random variables from a known distribution. This is motivated by the application of Monte Carlo integration for the numerical evaluation of integrals over complex functions. A Gibbs sampler for inverse problems on the other hand requires only one sample of the conditional distribution but the conditional distribution itself is not given directly. Only support points are available but this points are in general expensive to evaluate as the full forward map has to be evaluated. For this problem the Woodbury approach is used for a cheap evaluation of the conditional distribution. Only by this the Gibbs sampler becomes usable within an expectable time. As a further reevaluation of the forward map is also not desirable only the two last schemes of rejection sampling become suitable, but with the simplification that the third step is neglected to avoid a reevaluation of $g(x)$. The sampler only uses the upper and the lower function. Hence, the third version is used, as it provides the possibility to work on the logarithm of the posterior distribution, which offers numerical advantages. Figure 4.10 depicts the output of the sampling scheme, for two exemplary conditional distributions.



(a) The functions $l(x) \leq h(x) \leq u(x)$.

(b) The distribution g(X).



(c) The functions $l(x) \leq h(x) \leq u(x)$.

(d) The distribution g(X).

Figure 4.10: Two examples for rejection sampling on conditional distributions.

Although figure 4.10 suggests that the implemented rejection sampler scheme works the algorithm is not proved for correct convergence. The problem can be seen as less critical, as the conditional sampling is performed repeatedly for each element. Hence, a not completely correct taken sample is seen as less critical. The more concerning point is the the assumption of the log concave behavior for the conditional density of (4.1).

Global non log concave behavior has so far not been observed in the data. This may be more an effect of the smoothness assumption in the prior of the conditional posterior density (4.1). However, as figure 4.11 depicts, local non log concave behavior can be observed which suggests that the conditional posterior density may not be a simple function. In this case, the permittivity of an element, with a true permittivity value of one was increased starting from one. For very small values of γ, the conditional density increases showing non log concave behavior. As the conditional distribution changes its trend into the direction of lower probabilities, which is the correct trend the function offers log concave behavior. However, as a global log concave behavior could be observed within the test runs, a probably non log concave behavior of a local distribution can mostly be neglected. In the case that the non log concave method becomes a problem also appropriate rejection sampling schemes for this case can be applied [MCP08].

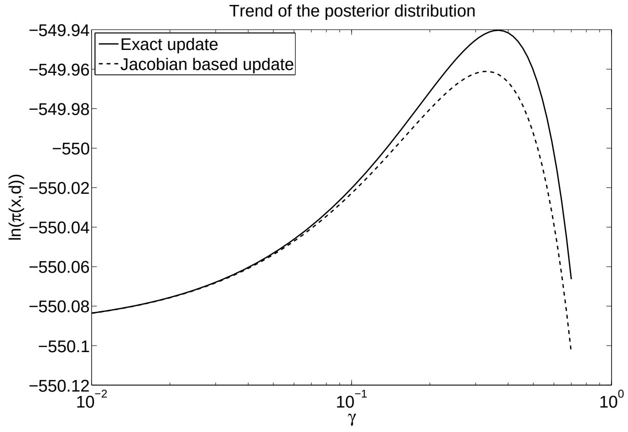
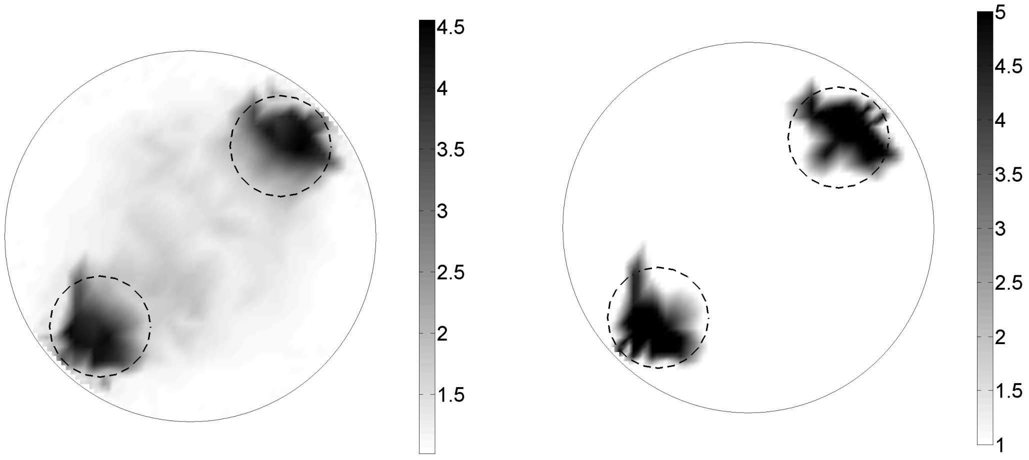


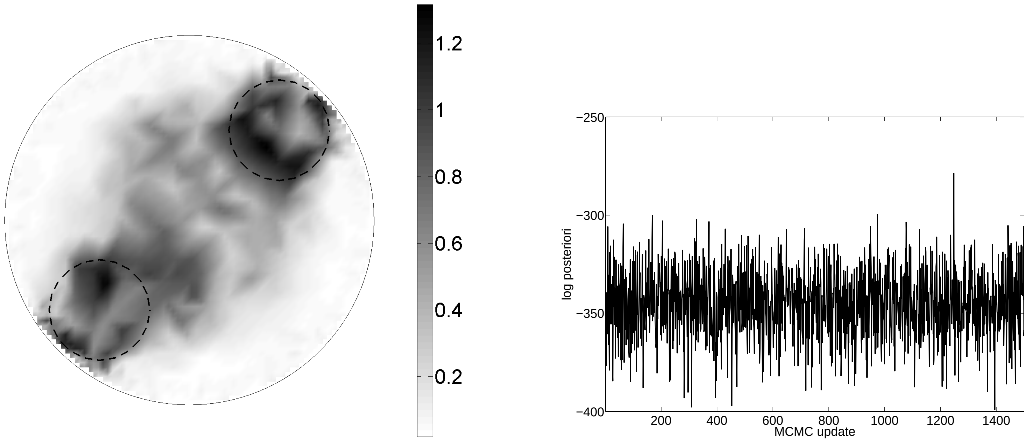
Figure 4.11: Example for non log concave behavior of the conditional posterior density.

Gibbs Sampling for Arbitrary Material Distributions - Continued

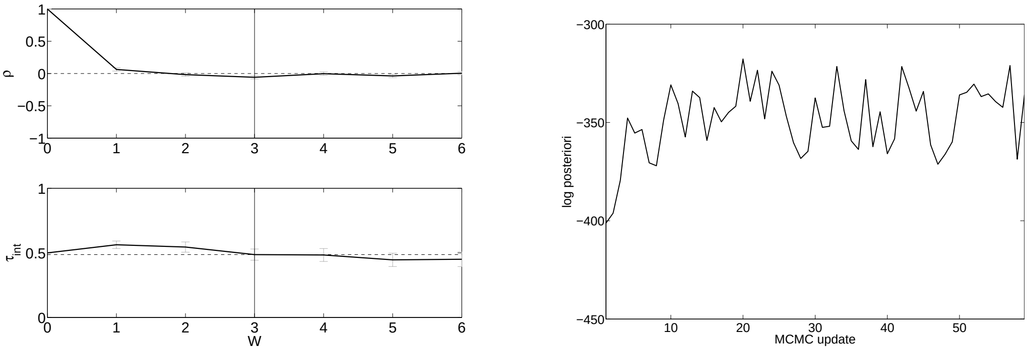
With the scheme for drawing conditional samples from the posterior distribution the Gibbs sampler can be implemented. Again, the result of the deterministic algorithm can be used to initialize the algorithm. Figure 4.12 summarizes the results. The parameter α was again set to $\alpha = 10^{-7}$. By the mean estimate illustrated in figure 4.12(a), one can see that the estimated



(a) Mean estimate. (b) Mean estimate (threshold applied).



(c) Standard deviation. (d) MCMC output.



(e) IACT. (f) Burn in phase.

Figure 4.12: Mean estimate with applied threshold threshold operation for figure 4.12(a) and trend of the posterior probability in the initial phase.

inclusions are too small. This becomes especially obvious on the result after a threshold operation has been applied as shown in figure 4.12(b). The underestimation can be observed on those parts of the inclusion facing the center of Ω_{ROI}. One can see by the standard deviation in figure 4.12(c), that the uncertainty in this region is comparatively high. Hence, the standard deviation suggests, that the right inclusions are larger. Figure 4.12(d) depicts the posteriori probability in equilibrium and figure 4.12 the IACT τ_{int}. The samples obtained after one sweep over all variables are independent! Once the chain is in equilibrium, each sweep produces an independent sample of the material distribution. Figure 4.12(f) shows the posterior distribution in the burn in phase. It takes only a comparatively low number of about 25 samples, before the chain reaches the equilibrium.

A sweep over all variables takes only about some seconds which is approximately same amount of time as the presented deterministic scheme takes. This is remarkable, as the sampling algorithm is able to deliver an independent sample in almost the same time, as a deterministic algorithm delivers a result. This could further be improved by using a different update schemes [LYHN05], [RC06], [Fis96]. Although the image quality itself is low, the availability of samples allows a statement about the quality of the result, whereas a deterministic algorithm only provides the same result. Hence, the possibilities provided by the Gibbs sampler are extremely valuable.

4.5 Bayesian Calibration

For the reconstruction tasks of the previous sections the empirical Bayes approach was used for calibration, which has been presented in subsection 2.3.4. For the low permittivity case the results are acceptable but for the high permittivity case they are biased (i.e. compare figure 4.7) and almost not usable. In fact, simulations on inverse crime data showed almost perfect behavior concerning the estimation of the boundary of the ellipsoidal object used in the investigations. This was stated in table 4.1.

In this section, the issue of calibration is treated and the Bayesian formulations are tested. Subsequently the calibration parameter $\boldsymbol{\xi}$ becomes part of the estimation problem or the likelihood function is adopted for the Bayesian forward map approach, respectively.

4.5.1 Impact of Neglected Calibration

First, the impact of neglecting any calibration strategy should be reviewed. For this, the simulated data of the fine model is used and the reconstruction is performed on the coarse model.

Figure 4.13(a) depicts the deterministic result when the raw data of the fine mesh is directly used for the reconstruction. One can see that the reconstruction result is a blurred image with large artefacts. Using a threshold operator, the result looks almost as in the calibrated case, but this is out of interest.

Figure 4.13(b) shows the result when the enhanced error model is applied. In this case the optimization problem is given by

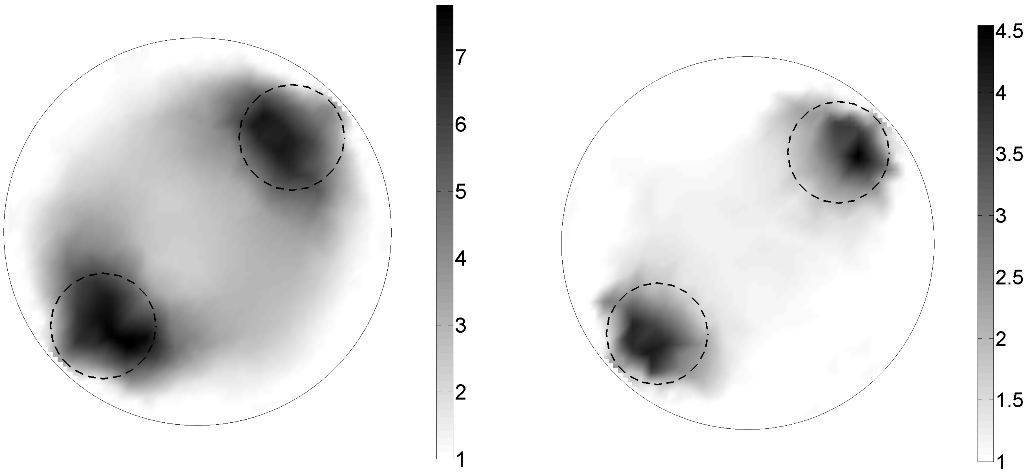
$$\boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \left\| \boldsymbol{C_w} \left(F^*(\boldsymbol{x}) + \boldsymbol{\mu_e} - \tilde{\boldsymbol{d}} \right) \right\|_2^2 + \alpha \boldsymbol{x}^T \boldsymbol{L}^T \boldsymbol{L} \boldsymbol{x}, \tag{4.15}$$

$$\text{s. t. } I(\boldsymbol{x}), \tag{4.16}$$

where $\boldsymbol{\mu_e}$ is the mean of the model error $\boldsymbol{e}_n = F(\boldsymbol{x}_n) - F^*(\boldsymbol{x}_n^*)$ and $\boldsymbol{C_w}$ is given by $\boldsymbol{C_w^T C_w} = (\boldsymbol{\Sigma_e} + \boldsymbol{\Sigma_v})^{-1}$, which is the Cholesky decomposition of the inverse total covariance matrix . For the practical implementation it turned out to be useful to normalize $\boldsymbol{C_w}$ by its maximum value. Again, the enhanced error model has been built by samples drawn with the method described in subsection 3.5.1. It has to be mentioned that this approach cannot be used for ECT, as the enhanced error model requires an exact forward map which is not the case for differential imaging problems. However, for academic reasons the enhanced error model will be applied to study its effect.

The result obtained with the enhanced error model is shown in figure 4.13(b) and is of the same quality as the result using the normal calibration scheme. Figure 4.14 illustrates the elements of the main diagonal of $\boldsymbol{C_w}$, which presents the weighting of the residuals. One can see, that measurements of adjacent electrodes with respect to the driving electrode have full weight, whereas the residuals of measurements from electrodes close to the active electrode are less weighted. It turned out that the choice $\boldsymbol{C_w} = \boldsymbol{I}$ has no large influence on the result. This, however, is also due to the low image quality of low level representations, which covers this effect. The effect of calibration is mostly to compensate the offset error.

The results of this subsection point out the need of calibration schemes or



(a) Deterministic result.

(b) Deterministic result with enhanced error model.

Figure 4.13: Result obtained by the accelerated deterministic scheme without calibration and with the enhanced error model.

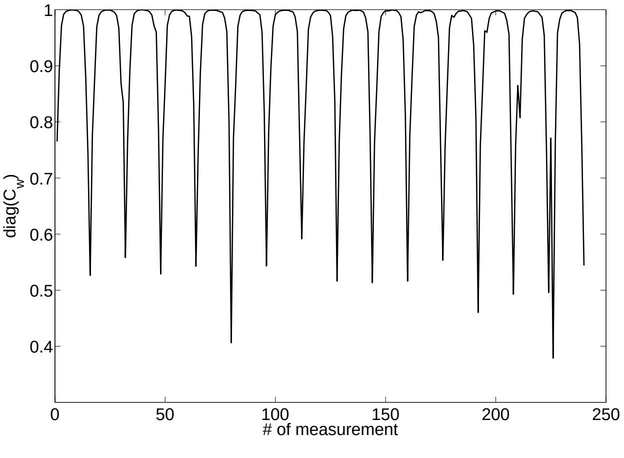


Figure 4.14: Main diagonal of the matrix $\boldsymbol{C_w}$.

model correction schemes like the enhanced error model to deal with data. In [NWZ10], the same problem was treated using a robust H_∞ approach [Sim06] showing the ability to outperform Kalman filter approaches [GA93]. However, as the enhanced error model can only be applied if a model $F(\cdot)$ exists, which is accurate enough, differential imaging problems always rely on calibration schemes.

4.5.2 The Problem of Empirical Bayesian Calibration

This subsection should give a first introduction about the concern of treating $\boldsymbol{\xi}$ as part of the estimation problem. For the previous used empirical Bayes calibration scheme, $\boldsymbol{\xi}$ was determined out of a least squares fit from calibration data $\boldsymbol{d}_c$, which was obtained from simulations of homogeneous filled pipes. Calibration data $\boldsymbol{d}_{c,L}$ and $\boldsymbol{d}_{c,H}$ was generated using material values corresponding to the permittivity of the inclusion. $\boldsymbol{d}_{c,L}$ was always generated for an air filled pipe, while $\boldsymbol{d}_{c,H}$ was generated for either PVC or water. Figure 4.16 shows the trends of the ML estimates for $\boldsymbol{\rho}$ and $\boldsymbol{c}$ for several calibration schemes. The trend "air - rods -pvc" comprises a third set of calibration measurements, where PVC rods were placed on several positions inside the pipe. This positions are depicted in figure 4.15. The trends in figure 4.16(a) display a non negligible difference in the behavior of the gain correction $\boldsymbol{\rho}$. Especially the difference to calibrations where water is used, is significantly. Basically, the calibration using water means a larger system deviation compared to a calibration measurement using PVC. As the flux depends on the gradient of the electric field, a finer discretization is required. This mainly causes the difference between the coarse and the fine model. For the reconstruction scenario of the elliptic object only a small number of electrodes has a close distance to the inclusion and hence require a larger gain. Thus, the gain at the remaining electrodes is too large and the calibration is essentially not suitable for this scenario. For the case that one tries to reconstruct an inclusion in water, the air-water calibration is the calibration to use. The least squares estimate of $\boldsymbol{c}$, which is depicted in figure 4.16(b), offers appropriate behavior on electrodes with a certain distance to the transmitter electrode, but shows strong deviations on electrodes close to the transmitter electrode. Especially for the calibration using water, $\boldsymbol{c}$ has also a different sign for electrodes close to the transmitter electrode. Thus, the calibration causes a too large misfit, which can be seen as the major reason for the stucking behavior of the Markov chains.

Using this knowledge it becomes clear, that the reconstruction of the elliptic

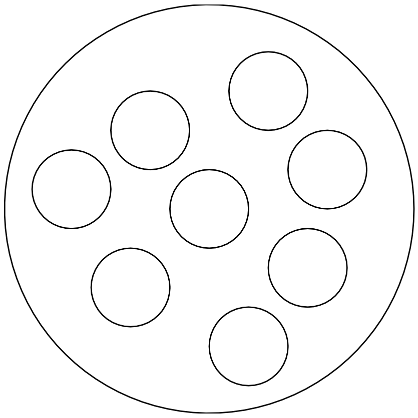
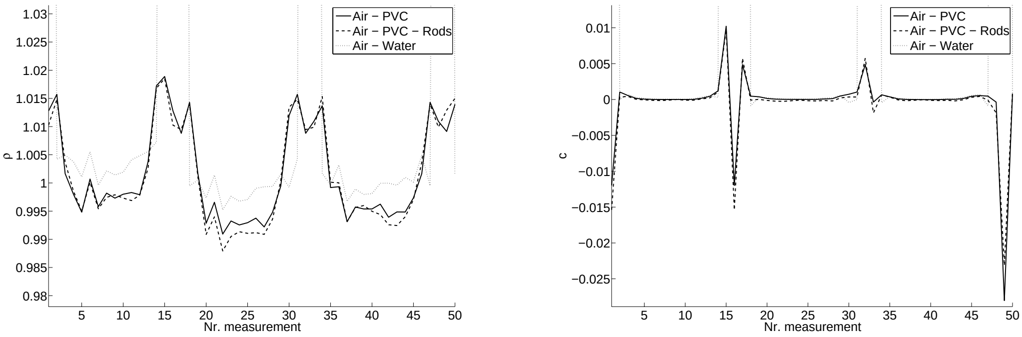


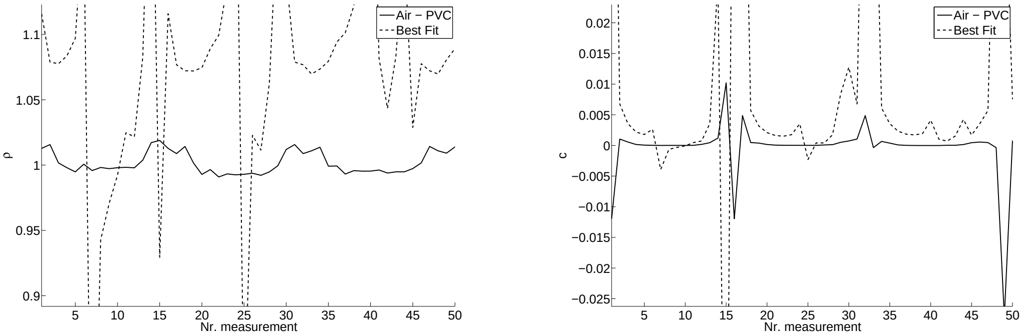
Figure 4.15: Positions of several rods used for providing calibration data.



(a) Trend of $\boldsymbol{\rho}$.

(b) Trend of $\boldsymbol{c}$.

Figure 4.16: Example of trends for the calibration parameters $\boldsymbol{\rho}$ and $\boldsymbol{c}$ for different empirical Bayes calibrations.



(a) Trend of $\boldsymbol{\rho}$.

(b) Trend of $\boldsymbol{c}$.

Figure 4.17: Example of trends for the calibration parameters $\boldsymbol{\rho}$ and $\boldsymbol{c}$ for a ML calibration and for a best fit approach, where the model is fitted to the data for a known state $\boldsymbol{x}$.

inclusion used in this work is much more demanding for the high permittivity case than for the low permittivity case. The parts of the inclusion close to the pipe cause a large deviation of the electric fields in this region, which results in a larger calibration error for electrodes close to the object. The fixed calibration cannot compensate this and the error can show full impact. It would be simpler if the inclusion is placed close to the center of Ω_{ROI} or with some distance to the electrodes. In this case the deviation of the electric field is smaller and the error has smaller impact. This is a property of soft field tomography systems, where the objects have an influence on the direction of the used sensing quantity.

Another extreme case which can be thought for calibration, is the case where the model is fitted to the data for a known realization of $\boldsymbol{X}$. The trends of $\boldsymbol{\rho}$ and $\boldsymbol{c}$ for such an approach are depicted in figure 4.17, where the coarse model was fitted to the accurate model using the two point calibration scheme for air and the test distribution illustrated in figure 4.3(a). One can see, that in this case, the trends of $\boldsymbol{\rho}$ and $\boldsymbol{c}$ differ completely from their trends for homogeneous fillings, which are also depicted in figure 4.17. This difference is because of the appearance of small differences in the denominator of fraction terms for $\boldsymbol{\rho}$ and $\boldsymbol{\xi}$ in equation (2.47). As the inclusion used for the experiment is relatively small and near to the pipe wall, the change in the charges on the electrodes on the opposite site is small and hence, the small differences appear. In this sense it is clear, that such a two point calibration is a bad approach, as a certain signal deviation is required. This also imposes that $\boldsymbol{x}_c$ has to differ from $\boldsymbol{x}$ certainly enough. The homogeneous filled pipes are an example for $\boldsymbol{x}_c$, which differ enough from $\boldsymbol{x}$, because as the results already showed, this calibration works.

4.5.3 Information from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$

Fixing the value of $\boldsymbol{\xi}$ has already been investigated as a problem, as $\boldsymbol{\xi}$ offers a large variability with respect to the used calibration scheme and the inclusion itself. Hence, it is of interest to obtain more information about $\boldsymbol{\xi}$. This can be done by drawing samples from the distribution $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$, which is given as

$$\pi(\boldsymbol{\xi}|\boldsymbol{d}_c) \propto \prod_{i=1}^{N_c} \pi(\boldsymbol{d}_c|\boldsymbol{x}_c, \boldsymbol{\xi}). \tag{4.17}$$

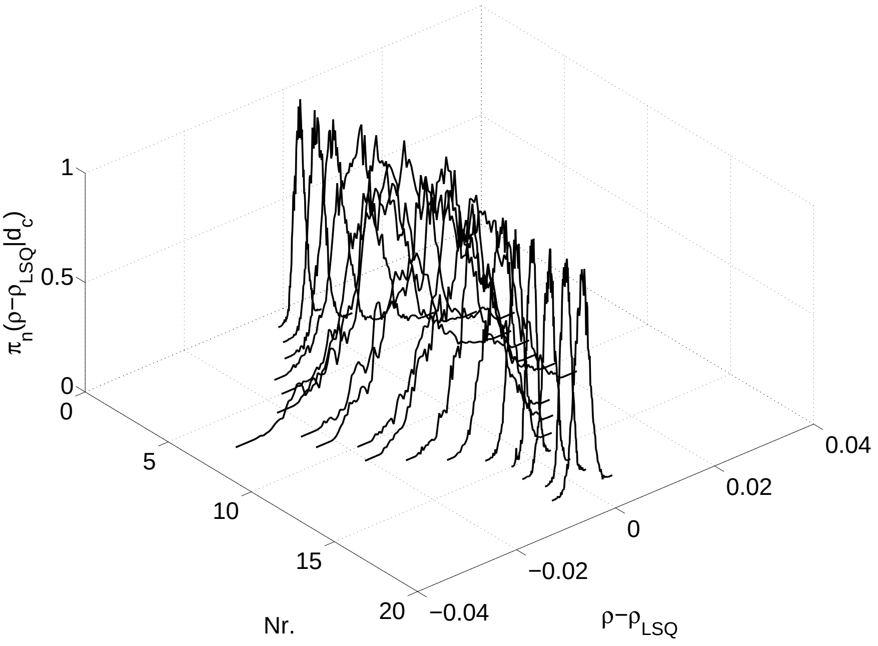
Since the individual calibration measurements are independent, the density function can be written as

$$\pi(\boldsymbol{\xi}|\boldsymbol{d}_c^{(N)}) \propto \exp\left\{ -\frac{1}{2} \sum_{i=1}^{N_c} \left(\boldsymbol{\rho} \boldsymbol{y}_{c,i} + \boldsymbol{c} - \boldsymbol{d}_{c,i}\right)^T \boldsymbol{\Sigma}_c^{-1} \left(\boldsymbol{\rho} \boldsymbol{y}_{c,i} + \boldsymbol{c} - \boldsymbol{d}_{c,i}\right) \right\}. \tag{4.18}$$

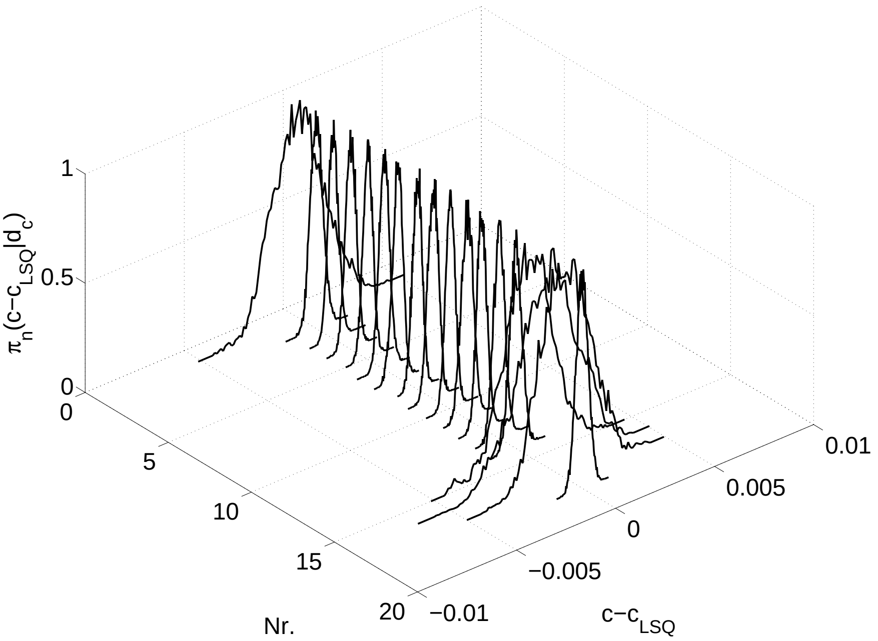
To take samples from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$, a Gibbs sampler can be used, which means that a sweep over the variables is performed. Unfortunately it is not possible to recast (4.18) in the form of a known probability density function. Because of the nonlinear forward map it was not mentioned in the previous parts, but if the conditional distribution is a known standard distribution, a standard random number generator could be used within the Gibbs sampler. As this is not possible, again a rejection sampling scheme has to be used to draw samples from the parameter vector $\boldsymbol{\xi}$. As the domain of $\boldsymbol{\xi}$ is now unbounded, which means that the support of $g(X)$ is infinite, a slightly different reconstruction scheme for $g_l(X)$ and $g_u(X)$ has to be implemented [GBT95].

Figure 4.18 and figure 4.19 depict some normalized probability density functions of elements of the vector $\boldsymbol{\xi} - \boldsymbol{\xi}_{\mathrm{LSQ}}$ after sampling from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c^{(N)})$ using homogeneous fillings for calibration. This plots provide information about the distribution of $\boldsymbol{\xi}$ and especially their deviation from the least squares estimate. For the case of low permittivities, the results for $\boldsymbol{\rho}$ show at least a good agreement with the used two point scheme. Electrodes with a certain distance to the transmitter electrode offer a higher variance for the gain. The offset correction offers a larger variance for electrodes close to the transmitter electrode. For high permittivities, one can see a very proper behavior of $\boldsymbol{\rho}$ and $\boldsymbol{c}$ for electrodes which are not directly next to the transmitter electrode. For electrodes directly next to the transmitter electrode, the elements of $\boldsymbol{\xi}$ offer a large uncertainty in both, $\boldsymbol{\rho}$ and $\boldsymbol{c}$.

The results correspond to the experience, that for ECT systems or similar soft field systems like EIT, the measurements from electrodes close to the active electrode are often not usable. This especially holds for high contrast problems because of the already mentioned reason. The weight of the residuals is often less weighted or the measurements are completely neglected. This is exactly the behavior which can be seen in the diagonal of the covariance matrix of the enhanced error model. Neglecting this data is a way to get rid of the problem, but it means also the abandonment of measurement data. However, the justification of this experience by samples from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ is in

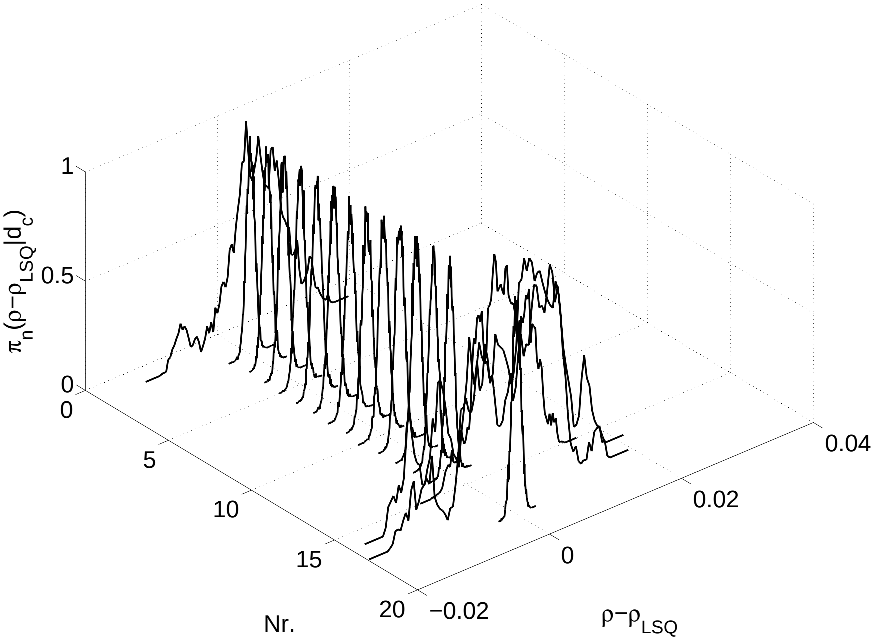


(a) Gain correction $\boldsymbol{\rho}$.

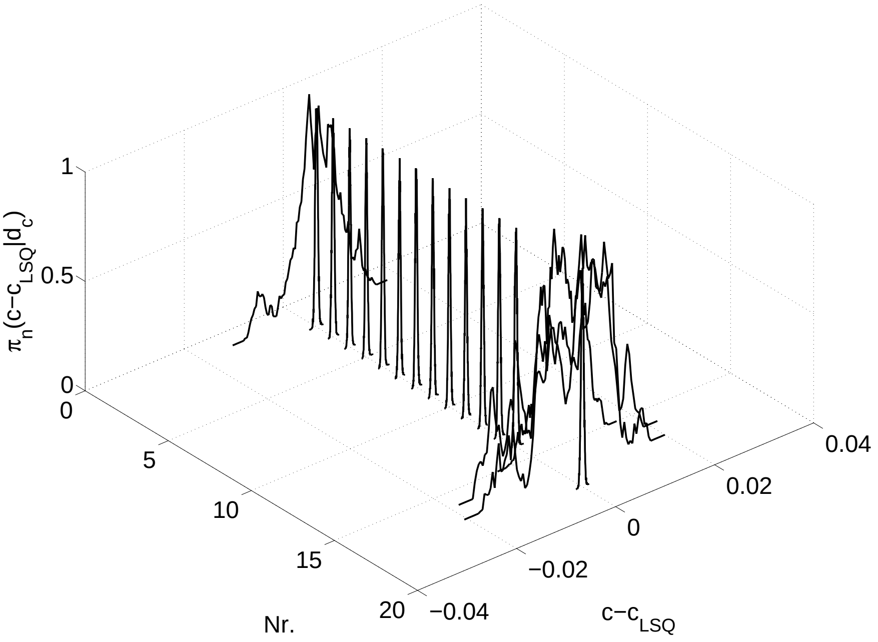


(b) Offset correction $\boldsymbol{c}$.

Figure 4.18: Normalized probability density functions $\pi(\xi_i - \xi_{i,\mathrm{LSQ}}|\boldsymbol{d}_c)$ for several electrodes as result from sampling from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ for low permittivities.



(a) Gain correction $\boldsymbol{\rho}$.



(b) Offset correction $\boldsymbol{c}$.

Figure 4.19: Normalized probability density functions $\pi(\xi_i - \xi_{i,\mathrm{LSQ}}|\boldsymbol{d}_c)$ for several electrodes as result from sampling from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ for high permittivities.

fact a very tough result, as the strong variation of $\boldsymbol{\xi}$ for the close electrodes implicates, that the model is out of range for high permittivities. Thus, not the measurements are wrong, but the model is unusable. The ECT forward map is in fact not able to describe the process $P(\cdot)$ over the whole range, which is a serious fault.

Summarizing this information one can interpret them also in the sense that in fact the likelihood function uses a wrong noise model. So far the likelihood function is designed for zero mean Gaussian noise. With model errors introducing a mean error and a gain error this noise model is not correct. The enhanced error model incorporates (global) knowledge about model errors by approximating them as a Gaussian distribution. However, from the results of the previous subsection it was seen that model errors in soft field tomography have local behavior. In the following two subsections two different approaches are presented to overcome the problem of model errors.

4.5.4 Mutual Inference for $\boldsymbol{x}$ and $\boldsymbol{\xi}$

This subsection covers the approaches introduced in the subsections 2.3.5 and 2.3.6. All results of this subsection have to be compared with the results of the standard MH algorithm (i.e. with figure 4.6 and 4.7) presented in section 4.3.1. The idea of full Bayesian calibration is to treat the calibration variables $\boldsymbol{\xi}$ as part of the estimation problem. For this, to different approaches are made:

- $\boldsymbol{\xi}$ depends on the data.
- $\boldsymbol{\xi}$ is independent from the data.

The likelihood function for sampling from the posterior distribution (2.50) is given by

$$\pi(\tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N_c)} | \boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N)}) = \pi(\tilde{\boldsymbol{d}} | \boldsymbol{x}, \boldsymbol{\xi}) \prod_{i=1}^{N} \pi(\boldsymbol{d}_c^i | \boldsymbol{\xi}, \boldsymbol{x}_c^i), \tag{4.19}$$

where $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}, \boldsymbol{\xi})$ is the likelihood function for $\tilde{\boldsymbol{d}}$ and $\pi(\boldsymbol{d}_c^i|\boldsymbol{\xi}, \boldsymbol{x}_c^i)$ are the likelihood functions of the calibration measurements. All calibration information, which comprises the homogeneous fillings and the data for each rod are used to form the product term. The weights ω_i in equation (2.50) have been set to one, which means that all calibration measurements have the same uncertainty. In more detail, the likelihood function is given by

$$\begin{aligned}\pi(\tilde{\boldsymbol{d}}, \boldsymbol{d}_c^{(N)}|\boldsymbol{x}, \boldsymbol{\xi}, \boldsymbol{x}_c^{(N)}) \quad \propto \quad & \exp\left\{-\frac{1}{2}\left(\boldsymbol{\rho y} + \boldsymbol{c} - \tilde{\boldsymbol{d}}\right)^T \boldsymbol{\Sigma}_v^{-1} \left(\boldsymbol{\rho y} + \boldsymbol{c} - \tilde{\boldsymbol{d}}\right)\right. \\ & \left. -\frac{1}{2}\sum_{i=1}^{N_c} \left(\boldsymbol{\rho y}_{c,i} + \boldsymbol{c} - \boldsymbol{d}_{c,i}\right)^T \boldsymbol{\Sigma}_c^{-1} \left(\boldsymbol{\rho y}_{c,i} + \boldsymbol{c} - \boldsymbol{d}_{c,i}\right)\right\}.\end{aligned} \tag{4.20}$$

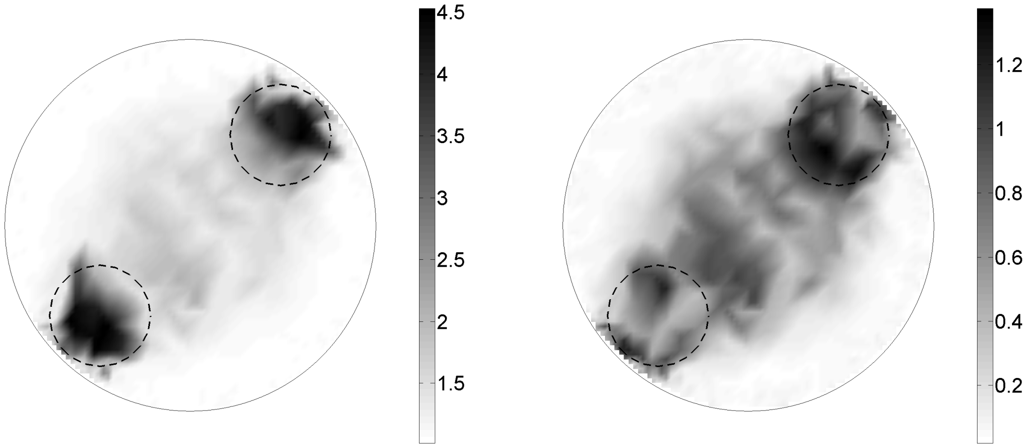
For sampling from the posterior distribution, in principal every algorithm of the presented MCMC algorithms in this work can be used. For the MH algorithm this also means that an appropriate proposal kernel for $\boldsymbol{\xi}$ has to be designed. Because of the large dimension of $\boldsymbol{\xi}$ compared to the number of unknowns for boundary reconstructions this is a critical issue. for the shape reconstructions, where 30 boundary points were used, the additional estimation of the $2 \times N_{\text{meas}} = 480$ elements of $\boldsymbol{\xi}$ increases the problem dimensionality by a factor of 17. Hence, the same Gibbs sampler presented in 4.5.3 will be used for generating samples from $\boldsymbol{\xi}$. The algorithm structure for both inference approaches about $\boldsymbol{x}$ and $\boldsymbol{\xi}$ is given by

1. Draw samples from $\boldsymbol{x}$ using the likelihood (4.20).
2. Draw samples form $\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$ or $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$, respectively.
3. When sampling $\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ update the posterior probability of the new state with the new calibration variables $\boldsymbol{\xi}$.

For the Gibbs sampler, the most obvious way is to run the sampling procedure for $\boldsymbol{\xi}$ after one sweep over the material parameters. For the MH scheme, a certain number of standard MH steps is performed and then a sweep over $\boldsymbol{\xi}$ is done. Because of this, the Gibbs sampler for $\boldsymbol{\xi}$ can be seen as part of the proposal kernel.

For sampling from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$, the probability density function is given by the product

$$\pi(\boldsymbol{\xi}|\boldsymbol{d}_c) \propto \prod_{i=1}^{N_c} \pi(\boldsymbol{d}_c|\boldsymbol{x}_c, \boldsymbol{\xi}), \tag{4.21}$$



(a) Mean estimate.

(b) Standard deviation.

Figure 4.20: Mean estimate and standard deviation for Gibbs sampling with full Bayesian calibration.

where again all calibration measurements have equal weight. In detail, $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ is implemented as

$$\pi(\boldsymbol{\xi}|\boldsymbol{d}_c^{(N)}) \quad \propto \quad \exp\left\{-\frac{1}{2}\sum_{i=1}^{N_c}\left(\boldsymbol{\rho}\boldsymbol{y}_{c,i}+\boldsymbol{c}-\boldsymbol{d}_{c,i}\right)^T\boldsymbol{\Sigma}_c^{-1}\left(\boldsymbol{\rho}\boldsymbol{y}_{c,i}+\boldsymbol{c}-\boldsymbol{d}_{c,i}\right)\right\}, \tag{4.22}$$

which is exactly the right part of $\pi(\tilde{\boldsymbol{d}},\boldsymbol{d}_c^{(N)}|\boldsymbol{x},\boldsymbol{\xi},\boldsymbol{x}_c^{(N)})$.

Results for the Low Level Representation

Figure 4.20 depicts the mean estimate and the standard deviation for $\boldsymbol{x}$ obtained by the Gibbs sampler for the full Bayesian approach. The result is in principle the same, as in the case when the pure two point calibration is used. It is a matter of fact, that the quality of the low level representation covers the effects of a non appropriate choice of $\boldsymbol{\xi}$. This could i.e. also be observed when using a two point calibration with air and water for PVC reconstructions. Although the image quality is in general low, the effect that low level representations and algorithms are insensitive with respect to the calibration makes this kind of representation more suitable for tasks, where a lack of calibration measurements is present.

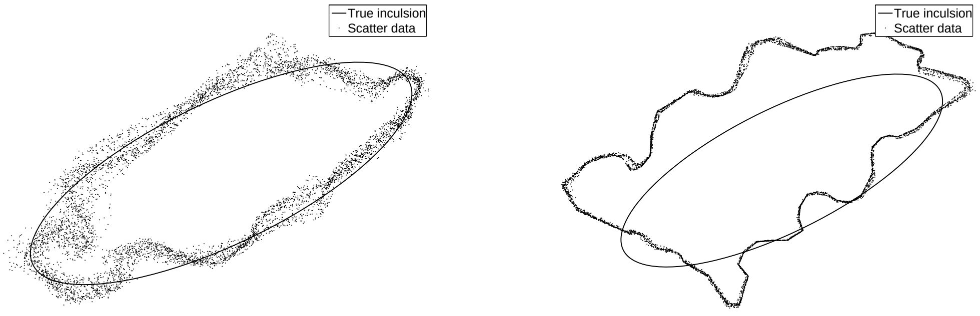
Results for Mid Level Representation

Table 4.2 lists the results for shape the reconstruction tasks using the two Bayesian calibration approaches. Line 1 and 2 summarize the behavior when using the full Bayesian approach, where samples from $\boldsymbol{\xi}$ are drawn from $\pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$. The results for low and high permittivities are depicted in figure 4.21(a) and 4.21(b). Comparing the results of this mutual inference approach with the results of the standard approach presented in subsection 4.3.1, the results are almost equal to the case using the fixed calibration. Looking on the IACTs of the full Bayesian approach a far lower value for τ_{int} appears for both reconstructions. Yet this chains reached the equilibrium. For the reconstruction of the inclusion with the high permittivity, the IACT is again large. This is an indicator, that the calibration works well for small permittivities, but suffers for large ones. Line 3 and 4 of table 4.2 list the results when samples from $\boldsymbol{\xi}$ are drawn from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$. The corresponding results are depicted in figure 4.22(a) and 4.22(b). The figures offer a high variance in the posterior distribution as can be seen by the scatter plots. The estimated shape of the inclusion with the high permittivity is again biased but in this case the result is quite better compared to all other experiments so far. An interesting fact is the highly decreased IACT for both reconstructions. For the low permittivity the result is acceptable although. For the high permittivity case the result is biased but the clustering has vanished. Hence the approach for sampling $\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ shows some benefit for the behavior of the chain.

Nr.	Simulation	ε_r	Conv	Ac_α	τ_{int}
1	$\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}\|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$	3.5	y	4.51	569
2	$\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}\|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$	80	y	0.43	104
3	$\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}\|\boldsymbol{d}_c)$	3.5	y	3.84	76
4	$\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}\|\boldsymbol{d}_c)$	80	y	2.63	101

Table 4.2: Results from Bayesian calibration.

Summarizing, it can be stated, that the mutual inference approach increases the quality and the behavior when reconstructing low permittivity distributions. For inclusions with a high permittivity the dependence of $\boldsymbol{\xi}$ on the data showed to be unsuccessful. Sampling $\boldsymbol{\xi}$ from $\pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$ decreased the bias but leads to a higher variance in the posterior distribution.



(a) $\varepsilon_r = 3.5$.

(b) $\varepsilon_r = 80$.

Figure 4.21: Result from the mutual inference approach with $\boldsymbol{\xi} \propto \pi(\boldsymbol{\xi}|\tilde{\boldsymbol{d}}, \boldsymbol{d}_c)$.

(a) $\varepsilon_r = 3.5$.

(b) $\varepsilon_r = 80$.

Figure 4.22: Result from the mutual inference approach with $\boldsymbol{\xi} \propto \pi(\boldsymbol{\xi}|\boldsymbol{d}_c)$.

4.5.5 A Stochastic Forward Map for ECT

This subsection deals with the issue discussed in subsection 4.5.3 in order to design a stochastic forward map using calibration information to quantify the uncertainty of the forward map. This subsection deals with the aspect of designing a model inadequacy function $\mathcal{D}(\cdot)$ from calibration data and consequently manipulate the likelihood function to incorporate this knowledge. As the design of a specific function for the model inadequacy is in general an infeasible task, the function $\mathcal{D}(\cdot)$ has to be designed from a known class of functions. Because of the universal applicability of Gaussian processes, the inadequacy function $\mathcal{D}(\cdot)$ will be designed as a Gaussian process and hence, the stochastic forward map becomes

$$\boldsymbol{\mathcal{Y}} = \tilde{F}(\boldsymbol{x}) + \mathcal{GP}_{\mathcal{D}_e}(\cdot). \tag{4.23}$$

The Gaussian process $\mathcal{GP}_{\mathcal{D}_e}(\cdot)$ is of form

$$\begin{bmatrix} \boldsymbol{e}_c \\ \boldsymbol{e} \end{bmatrix} = \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{\Sigma}(\boldsymbol{x}_{c,r}, \boldsymbol{x}_{c,r}) & \boldsymbol{\Sigma}(\boldsymbol{x}_{c,r}, \boldsymbol{x}_r) \\ \boldsymbol{\Sigma}(\boldsymbol{x}_r, \boldsymbol{x}_{c,r}) & \boldsymbol{\Sigma}(\boldsymbol{x}_r, \boldsymbol{x}_r) \end{bmatrix}\right), \tag{4.24}$$

where $\boldsymbol{e}_c$ is the known error from calibration measurements, which is defined by

$$\boldsymbol{e}_c = \boldsymbol{d}_c - \tilde{F}(\boldsymbol{x}_c). \tag{4.25}$$

The error $\boldsymbol{e}$ is the error to estimate for the state $\boldsymbol{x}_r$. The lower r for $\boldsymbol{x}_r$ in the Gaussian process (4.24) expresses, that a reduced version of the state is used. The meaning of this will be explained in the next subsection. The last subsection presents results where the stochastic forward map is used for reconstructions.

Design of an Inadequacy Function $\mathcal{D}_e$

As already mentioned in sub-section 2.3.7, the inadequacy function $\mathcal{GP}_{\mathcal{D}_e}(\cdot)$ can generally be designed as a data space function or as an image space function. However, because of the nonlinear forward map an image space based inadequacy function is at least preferable. The inadequacy function has the obvious form $\mathcal{GP}_{\mathcal{D}}(\boldsymbol{x})$, which is a treatable approach, as long as $\boldsymbol{x}$ is of comparatively low dimension. Thus, an appropriate reduction approach $R : \mathbb{R}^N \mapsto \mathbb{R}^{N_r}$, $N_r \ll N$, has to be developed, to design the inadequacy function $\mathcal{GP}_{\mathcal{D}}(\boldsymbol{x}_r)$ for a reduced vector $\boldsymbol{x}_r$, which depends on $\boldsymbol{x}$.

The design of a reduction $R : \boldsymbol{x} \mapsto \boldsymbol{x}_r$ is in principle of heuristic nature.

However, as the Gaussian process uses correlation information between the current (reduced) state and the reduced states of the calibration measurements, the reduction $R_{\boldsymbol{x}}$ has to emphasis specifics about $\boldsymbol{x}$ in order to obtain a good correlation. For this, knowledge about the behavior of the error can be used. This knowledge is given as

- The forward map works for inclusions of low permittivity.
- For high permittivities, the charges on the electrodes next to the transmitter electrodes suffer from large errors.

A further point which was not mentioned yet is about the type of input for the inadequacy function. So far it was assigned that the inadequacy function $\mathcal{D}(\cdot)$ is a function of $\boldsymbol{x}$. This is not useful for the shape description because the support points of the shape model can be moved on the boundary without changing the shape. Hence, a more suitable approach is to used fixed domains as input for $\mathcal{D}(\cdot)$. Thus, the inadequacy will be designed using the permittivity vector $\boldsymbol{\varepsilon}$ of the finite element forward map.

Based on this, a possible formulation for a reduction is given by

$$x_{\mathrm{r},i} = \int_{\Omega_{\mathrm{ROI}}} \frac{(\varepsilon_r(\boldsymbol{z}) - 1)^{p_{\boldsymbol{x}}}}{|\boldsymbol{z} - \boldsymbol{z}_{\mathrm{ref},i}|^{p_z}} d\Omega, \tag{4.26}$$

where $\boldsymbol{z}$ denotes the Cartesian coordinate. The point $\boldsymbol{z}_{\mathrm{ref},i}$ are reference points to incorporate geometrical knowledge. The exponentials $p_{\boldsymbol{x}}$ and p_z are used to control the behavior of the function. The approach incorporates the value of $\boldsymbol{x}$ together with spatial information. The choice of $(\varepsilon_r(\boldsymbol{z}) - 1)$ is because of the lower bound of the relative permittivity. As the model error is known to be correlated with the distance to the electrodes a possible choice for $\boldsymbol{z}_{\mathrm{ref}}$ is given by the corners of the electrodes, as the local flux density has its highest value in this region. This leads to $N_r = 2 \cdot N_{\mathrm{elec}}$ elements for the vector $\boldsymbol{x}_r$. For the numerical implementation of (4.26) the integral can be evaluated previously and hence the reduction can be written as

$$\boldsymbol{x}_{\mathrm{r}}^T = \left((\boldsymbol{x} - 1)^{p_{\boldsymbol{x}}}\right)^T \boldsymbol{P}. \tag{4.27}$$

To build up individual correlation matrices $\boldsymbol{\Sigma}(\boldsymbol{x}_{r,i}, \boldsymbol{x}_{r,j})$ only the elements of $\boldsymbol{x}_r$ corresponding to the transmitter electrode and the receiver electrode are used. The same holds for $\boldsymbol{\Sigma}(\boldsymbol{x}_r, \boldsymbol{x}_r^*)$. Figure 4.23 exemplary depicts the behavior of the Gaussian surrogate for the material distribution depicted in figure 4.3(a), which was used for the shape reconstructions. One can see, that the estimated variance is high for measurements where the error becomes large. Thus, the inadequacy function provides the wanted behavior.

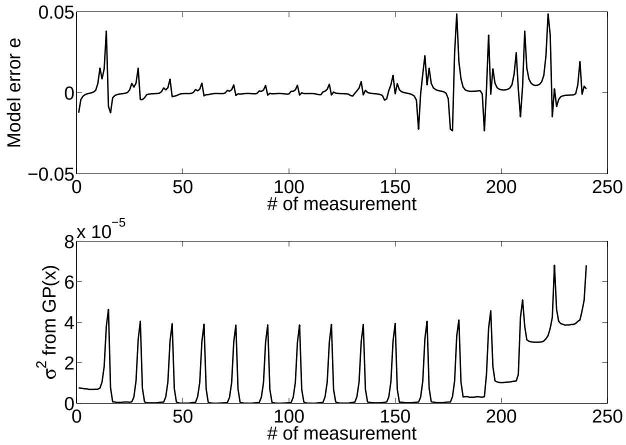
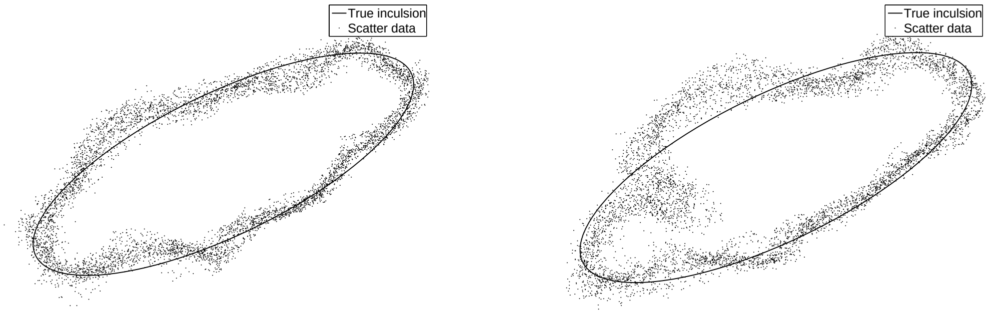


Figure 4.23: Example for the behavior of $\mathcal{GP}_\mathcal{D}(\boldsymbol{x}_r)$. The upper plot shows the absolute error of the model for the material distribution depicted in figure 4.3(a). The lower plot contains the variance of the predicted model error.

Shape Determination with the Stochastic Forward Map

To investigate the behavior of the stochastic forward map, the shape determination is performed for the low and the high permittivity case. For the experiments, different calibration types (air-PVC or air-water) were used, in combination with different ways to incorporate the information from the stochastic forward map. These were either just to take the covariance information from $\mathcal{GP}_\mathcal{D}(\boldsymbol{x}_r)$ or to take both, the estimated mean and the covariance from $\mathcal{GP}_\mathcal{D}(\boldsymbol{x}_r)$ to incorporate this knowledge into the likelihood function. The inadequacy function includes calibration measurements of the rod data of figure 4.15 for low and high permittivities and homogeneous fillings. Only the error $\boldsymbol{e}_c$ of equation (4.24) has to be adapted according to the calibration type for $\tilde{F}(\boldsymbol{x})$.

Table 4.3 collects the results of the experiments. All results have to be compared with the results of the MH reference solution presented in section 4.3.1, where a fixed calibration is used. Especially the case for the high permittivity is of interest.



(a) Line 1. (b) Line 2.

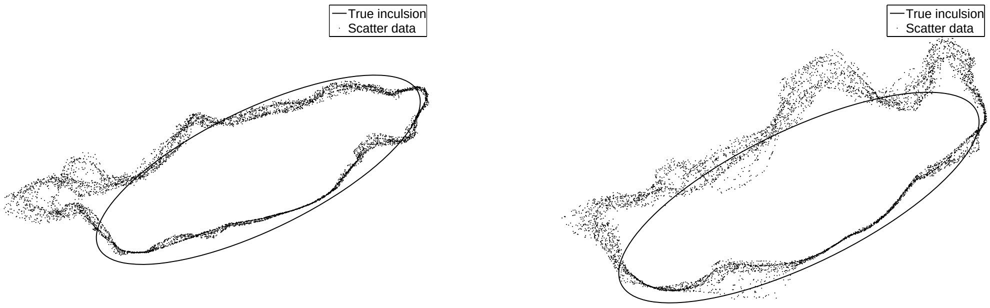
Figure 4.24: Results obtained by using the Bayesian forward map for the low permittivity case (to compare with figure 4.6).

Line 1 and 2 contain the results for the low permittivity case. The corresponding scatter plots are depicted in figure 4.24. This results have to be compared with the results depicted in figure 4.6(b). Especially the case when only the covariance information is used to manipulate the likelihood function provides an improvement to the reference solution of figure 4.6(b). The scatter plot when using also the mean is depicted in figure 4.24(b). In this case the quality of the reconstruction is decreased but the prediction of the error is even more complicate and a wrong prediction of μ can lead to a serious fault. From the IACT values in line 1 and 2 of table 4.3 one can see decreased values of τ_{int}. The chains for this case behaved well.

Nr.	Simulation	ε_r	Conv.	Calib.	Ac_α	τ_{int}
1	Only σ	3.5	y	Air-PVC	4.32	1717
2	With μ and σ	3.5	y	Air-PVC	4.78	2388
3	No Bayesian forward map	80	y	Air-PVC	1.81	4722
4	Only σ	80	y	Air-PVC	1.71	3525
5	With μ and σ	80	y	Air-PVC	1.52	10417
6	Only σ	80	y	Air-Water	2.39	6382
7	With μ and σ	80	n	Air-Water	2.30	4201

Table 4.3: Results from Bayesian calibration.

The more interesting case is given for the high permittivity reconstructions.



(a) Line 3.

(b) Line 4.

(c) Line 5.

(d) Line 6.

Figure 4.25: Results obtained by using the Bayesian forward map for the high permittivity case (to compare with figure 4.7).

They are documented in line 3 to 7 of table 4.3 and figure 4.25. Line 3 and figure 4.25(a) present the result when the standard forward map and an air-PVC calibration is used. The result looks quite good with respect to figure 4.7(b) although it is effected by an artefact at the left side of the inclusion. As can be seen by the high value of τ_{int} in table 4.3, the chain has again a bad convergence behavior, which is similar to the standard MH reference solution of subsection 4.3.1. However, the result looks better than the reference result using the air-water calibration, which is depicted in figure 4.7. The calibration using water is of more critical nature for the reason explained in subsection 4.5.3.

Line 4 and 5 contain the results when the stochastic forward map is used for the same calibration type. The two scatter plots are depicted in figure 4.25(b) and 4.25(c). For the first experiment only the covariance information is used. For the second experiment both, the covariance and the expected mean, are used to manipulate the likelihood. The results are yet not perfect with respect to the low permittivity case but especially the first result depicted in figure 4.25(b) provides a clear improvement compared to figure 4.7(b). While the IACT for the first experiment is slightly decreased the IACT for the second experiment is high suggesting that the chain has stalled. Again the estimation of the error by the Gaussian process in order to correct the likelihood is the more demanding task this behavior can be expected. The result of the first experiment is quite remarkable as the artefact are successfully suppressed. Beside the stalling behavior of the chain also the result of the second experiment can be accepted. Line 6 and 7 list the result for the Bayesian forward map when the air-water calibration is used. This results have to be compared with the standard MH result in subsection 4.3.1, where the same calibration has been used. Using only the covariance information to correct the likelihood function results again in a decreased IACT. This result is illustrated in figure 4.25(d). Although the chain behaves better in terms of the IACT the result is too much biased. For the case that also the mean of $\mathcal{GP}_{\mathcal{D}}(\boldsymbol{x}_r)$ is used, the reconstruction failed. The estimated model error to correct the forward map is to large and hence, the reconstruction is not possible. It turned out, that the setup of $\mathcal{GP}_{\mathcal{D}}(\boldsymbol{x}_r)$ for the air-water calibration is more complicate compared to the air-PVC case.

Summarizing the results of the Bayesian forward map compared to the results only using the standard forward map with the standard calibration it can be stated, that the results can outperform the previous approaches and the Markov chains show much better behavior in terms of the IACTs.

5 Summary and Conclusion

In this thesis schemes for accelerated Bayesian inversion and calibration approaches are presented and demonstrated for the inverse problem of ECT. The two main topics of this thesis, which also build the title of the work, are the acceleration of Bayesian inversion methods and the issue of model errors and how to overcome them by means of calibration. The acceleration of Bayesian inversion methods covers a large range starting from the development of efficient computation schemes for the forward map, the design of approximation techniques and the application of fast sampling techniques. The topic of model errors and calibration accesses schemes to make computational inverse problems working with real data. Summarizing these points the thesis provided contributions for

- Computational techniques,
- Bayesian algorithms,
- Bayesian formulations.

The fundamentals of Bayesian inversion theory, which build the backbone of this work are presented in chapter 2. In the Bayesian inversion framework every variable is treated as random variable leading the inversion problem to become a Bayesian inferential problem. The attractiveness of the approach lies in the natural incorporation about both, measurement errors and prior information about the quantities of interest. Several Markov Chain Monte Carlo algorithms are presented, which provide an acceleration with respect to the standard MH algorithm. These statistical algorithms are used in the later part of the thesis in combination with the developed computational methods.

As major new part of chapter 2 the general calibration of computer models is treated and the issue of model errors is discussed. Model errors build a fundamental problem to overcome when solving inverse problems for measurement data. This is especially the case for differential imaging problems. The nature of traditional calibration schemes is reviewed in a Bayesian sense and the use of further calibration information is investigated for two mutual inference approaches. With these Bayesian formulations two auto calibrating algorithms were built, which allow a natural incorporation of further calibration measurements and solve the inverse problem in one step without any further preprocessing steps. Last, a different approach called stochastic forward map is developed. In this approach calibration information is used to design a forward map, which quantifies the uncertainty of its output. Thus, the forward map does not only provide a number but also provides a statement about the quality or the confidence in this number. These Bayesian formulations are tested in the last part of the chapter "Case studies" and improvements in terms of statistical efficiency and behavior could be achieved with respect to the standard solution.

Chapter 3 presents computational techniques for the forward map of ECT using the finite element method. The concrete methods are developed with focus on ECT but can be applied to all problems where the stiffness matrix can be decomposed into $\boldsymbol{W}\boldsymbol{S}\boldsymbol{W}^T$ systems, which holds for all problems which are similar to a resistor network. Several aspects of the forward map are reworked in order to obtain a forward map which only contains fast matrix vector operations. Then a computational framework maintaining Green's functions is developed to take use of the self adjointness of the forward problem. Beside improvements in computation time the use of Green's functions permits to several other time saving assets like Jacobian operations and exact low-rank updates using the Woodbury identity. Jacobian operations allow a fast and efficient numerical access to the Jacobian without a direct evaluation of the Jacobian. Thus, fast local approximations can be built or gradient information can be obtained using the current solution. With the Woodbury formula an efficient and exact evaluation of low-rank updates can be done. By this the influence of small changes in the input data can be evaluated without an evaluation of the full forward map. This mathematical tools then built the computational backbone within the used MCMC algorithms. Further some aspects about the use of domain decomposition techniques are discussed. The second point in chapter 3 is about approximation techniques or so called surrogate models. Two surrogate models, a polynomial approximation and a reduced physical model are discussed.

Chapter 4 combines the approaches of chapter 2 with the computational techniques developed in chapter 3. Two types of data representations, a low level representation and a mid level representation using a shape model, are used to perform different reconstruction tasks.

As first reconstruction algorithm, a fast deterministic second order scheme using a BFGS update to approximate the inverse Hessian is demonstrated. The scheme uses several computational tricks of the developed computational framework, making it a least an efficient and fast deterministic scheme.

The third part of the chapter deals with the acceleration of MH methods using different approximation techniques within the DAMH algorithm. The three presented approximation techniques are used for a shape reconstruction task and implemented in several versions to compare either the benefit of using the enhanced error model on the posterior distribution or to work out another specific variant of the approximation like the adaptive version of the polynomial approximation. Beside the polynomial approximation for high permittivities all approximations could be used for accelerating the algorithm reaching the same result as obtained with the standard MH reference solution. However, the adjustment of a proper DAMH algorithm requires some experience to find a working setup. This especially holds for the task to neglect certain measurements in the first step of the DAMH where approximations were used and is at least the case for the reduced physical model, as it is close to the limit. To override physical laws by approximating the potential on the gaps between the two meshes is critical, as the gaps are close to the electrodes. Hence, such tricks are critical in regions of high sensitivity with respect to the model output. However, in regions of low sensitivity such methods may be well applied.

The fourth block of chapter 4 is about the application of a Gibbs sampler for ECT using a low level representation. The application of a Gibbs sampler strongly relies on the efficient computation of fast low-rank updates which is possible due to the use of the Woodbury formula. Two versions of the Gibbs sampler, one for bimodal material distributions and one for arbitrary material distributions, are implemented and show excellent results compared to the deterministic results. A remarkable aspect about the application of the Gibbs sampler are the low integrated auto correlation times. Due to the efficient numerical evaluation the presented Gibbs sampling schemes produce an independent sample in nearly the same time as a nonlinear deterministic algorithm. Hence, this schemes come close to the demands of real time application.

The last part of chapter 4 covers the topic of calibration and model errors. Model errors are a serious problem when solving inverse problems with measurement data. First, the typical calibration approaches are reviewed and the variability in the calibration parameters is demonstrated, to analyze to origin of the model errors. Two different mutual inference approaches are implemented for inference about both, the data and the calibration variables. In the first approach the calibration was dependent on the data. For the reconstruction of the inclusion with the low permittivity, the approach showed a good behavior and decreased the integrated auto correlation time. In the high permittivity case the approach has the same behavior as when using fixed calibration parameters. For the approach, that the calibration only depends on the calibration measurements, the results offered a drastically decreased integrated auto correlation time. However, the results are also effected by a high variance in the posterior distribution. Thus, the improvement of the mutual inference approach is in general low. As last step a stochastic forward map approach is tested by using calibration information to design an inadequacy function, which provides information about the quality of the result. With this approach, the integrated auto correlation time could be drastically decreased leading to statistical efficient Markov chains even for the more demanding high permittivity case.

5.1 Outlook an Future Work

Although the thesis contributes several new aspects to the field also several new and further points appear as future work.

The Green's functions approach provides several improvements for the forward map. Remaining tasks for further investigations can be summarized by

- Direct influence of the material update on the Cholesky decomposition,
- A better workout of the domain decomposition techniques,
- A fully analytic low-rank scheme.

The first point means, that the material update should directly result in the Cholesky matrix. By this one could save further time as no Cholesky decomposition is necessary. The second point covers the use of domain decomposition techniques. Last, a full analytic version of the Woodbury identity would be fruitful for further speeding up of the conditional sampling. This will be explained with the next points.

With respect to the statistical inversion methods the following points seem as useful future tasks:

- Gibbs sampling using different update schemes.
- Automatic prior adjustment.
- Gibbs sampling for mid level representations.
- Inverse transform sampling.

This points mainly address the application of the Gibbs sampler. Although the presented MH algorithms can be accelerated by using approximations, the advantages of the Gibbs sampler remain an attractive issue for further work.

The most simple improvement for the speed of the Gibbs sampler is to apply a different update scheme for the sweeps. The second point addresses the choice of the prior parameter α. So far the prior parameter α has been fixed similar as it is done for deterministic methods. However, also the prior parameter could become a variable to estimate. By this an automatic adjusting could be implemented and one could get rid of the choice of the regularization parameter. Further, this parameter could also be used to improve the results of deterministic methods.

A more demanding task is given by applying the Gibbs sampler for shape determination tasks. In this case, one has to work out the influence of a move onto the finite elements and then use the Woodbury identity to sample from the conditional distribution. Because of the good behavior the Gibbs sampler showed for the low level representation this approach seems extremely useful to accelerate sampling for mid level representations.

Inverse transform sampling is in concern with the mentioned full analytic low-rank update. If it would be possible to derive an analytic expression for the low-rank update, one could think about describing the conditional distribution as an analytical term and hence the idea of inverse transform sampling appears possible. Thus, the now used rejection sampling schemes could be replaced, which again results in an enormous increased computation time. This presents the highest end of possible methods.

Another issue for accelerating the inversion methods is given by

- Importance sampling using polynomial acceleration schemes.

Importance sampling [Gey92] is a technique for Monte Carlo integration where samples from different distribution are generated. Then the importance weight is evaluated for each of the samples by computing the posterior probability the target distribution. This importance factor is used as a weight for the Monte Carlo integration. If the first distribution is easy to sample from and the difference between the two distributions is not too large, the method provides a fast way to perform Monte Carlo integration. Hence, the idea of importance sampling is in principle simple, but the application of this scheme to inverse problems is a demanding task, as it means, that the posterior distribution has to be replaced by a distribution where an efficient generation of independent samples is possible. Polynomial acceleration is an efficient scheme to accelerate sampling from even high dimensional Gaussian distributions. It seems preferable to built a similar Gaussian distribution for the posterior distribution. For the mean in principle the output of a deterministic algorithm can be used. The more complex case is given by generation of a representative covariance matrix.

A further point covers the issue of calibration and model errors. An interesting point in this concern is given by

- A more formal design for inadequacy functions to form stochastic forward maps.

The two mutual inference approaches showed so far less success in the case of larger model errors as it is the case for the reconstruction of inclusions with a high permittivity. The far more interesting point is given by the stochastic forward map. The design of the inadequacy functions in this work is explained and done in a heuristic way, which requires a lot of expert knowledge. A more general way to design inadequacy functions for different problems seems preferable. In this work a Gaussian process using an exponential covariance function has been used. Also this seems to be a point for improvements, as more richer classes of functions are available.

Although, this work was done using standard personal computers for all simulations, a final statement about the implementation of MCMC algorithms and forward map calculations using field programmable gate arrays (FPGAs) should be made. These devices allow a specific implementation of algorithms in hardware, which allows an enormous speed up or even parallelization. In combination with statistically efficient sampling algorithms like the Gibbs sampler the use of such hardware could enable real time statistical inversion for inverse problems.

Bibliography

[AG96] A. Adler and R. Guardo. Electrical impedance tomography: regularized imaging and contrast detection. *IEEE Transactions on Medical Imaging*, 15(2):170 –179, apr 1996.

[BHW03] B. Brandstätter, G. Holler, and D. Watzenig. Reconstruction of inhomogeneities in fluids by means of capacitance tomography. *The International Journal for Computation and Mathematics in Electrical Engineering (COMPEL)*, 22(12):508–519, 2003.

[BI98] A. Blake and M. Isard. *Active Contours: The Application of Techniques from Graphics,Vision,Control Theory and Statistics to Visual Tracking of Shapes in Motion*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1st edition, 1998.

[Bla11] P. E. Black. Big-O Notation, 2011. NIST U.S. National Institute of Standards and Technology, This is an electronic document. Date retrieved: January 14, 2011.

[Bra03] B. Brandstätter. Jacobian calculation for electrical impedance tomography based on the reciprocity principle. *IEEE Transactions on Magnetics*, 39(3):1309 – 1312, May 2003.

[Bra07] B. Brandstätter. Inverse problems, Ill-posedness and regularization - an illustrative example. *e&i Elektrotechnik und Informationstechnik*, 124:224–231, 2007. 10.1007/s00502-007-0447-2.

[BS 4] J. M. Bernardo and A. F. M. Smith. *Bayesian Theory*. John Wiley & Sons, New York, 1994 (ISBN: 0-471-92416-4).

[BWLW02] M. A. Bennett, R. M. West, S. P. Luke, and R. A. Williams. The investigation of bubble column and foam processes using electrical capacitance tomography. *Minerals Engineering*, 15(4):225 – 234, 2002.

[Cas85] G. Casella. An Introduction to Empirical Bayes Data Analysis. *The American Statistician*, 39(2):pp. 83–87, 1985.

[CCL95] J. A. Christen, R. S. Clymo, and C. D. Litton. Bayesian approach to the use of C^{14} dates in the estimation of the age of peat. *Radiocarbon*, (37):431–442, 1995.

[CF05] J. A. Christen and C. Fox. Markov chain Monte Carlo Using an Approximation. *Journal of Computational and Graphical Statistics*, 14(4):795–810, 2005.

[CF10] J. Andrés Christen and Colin Fox. A General Purpose Sampling Algorithm forContinuous Distributions (the t-walk). *Bayesian Analysis*, 5(2):263–282, 2010.

[CFO$^+$10] T. Cui, C. Fox, M.J. O'Sullivan, University of Auckland. Dept. of Engineering Science, and University of Auckland. Faculty of Engineering. *Adaptive error modelling in MCMC sampling for large scale inverse problems.* Report (University of Auckland. Faculty of Engineering). Engineering Science, School of Engineering, University of Auckland, 2010.

[CG92] G. Casella and E. I. George. Explaining the Gibbs Sampler. *The American Statistician*, 46(3):pp. 167–174, 1992.

[CMP08] B. Cai, R. Meyer, and F. Perron. Metropolis-Hastings algorithms with adaptive proposals. *Statistics and Computing*, 18:421–433, December 2008.

[CP05] G. Chaplin and T. Pugsley. Application of electrical capacitance tomography to the fluidized bed drying of pharmaceutical granule. *Chemical Engineering Science*, 60(24):7022 – 7033, 2005.

[Cui10] T Cui. *Bayesian Calibration of Geothermal Reservoir Models via Markov Chain Monte Carlo.* PhD thesis, University of Auckland, 2010, 2010.

[Dev86] L. Devroye. Non-uniform random variate generation, 1986.

[DSSW08] C. Deinhammer, G. Steiner, M. Sommer, and D. Watzenig. Electromechanical flow imaging using ultrasound and electrical capacitance data. In *Sensors, 2008 IEEE*, pages 160 –163, October 2008.

[EHN02] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems.* Kluwer Verlag, Dordrecht, 2002.

[ES78] W. H. Enright and Steven M. Serbin. A note on the efficient solution of matrix pencil systems. *BIT Numerical Mathematics*, 18:276–281, 1978. 10.1007/BF01930897.

[FCN11] C. Fox, T. Cui, and M. Neumayer. Adaptive approximations and error models for mcmc applied to inverse problems. In *Proceedings of the 7th International Congress on Industrial and Applied Mathematics (ICIAM)*, 2011.

[Fis96] G. S. Fishman. Coordinate Selection Rules for Gibbs Sampling. *The Annals of Applied Probability*, 6(2):pp. 444–465, 1996.

[Fle87] R. Fletcher. *Practical methods of optimization; (2nd ed.)*. Wiley-Interscience, New York, NY, USA, 1987.

[FN97] C. Fox and G. Nicholls. Sampling Conductivity Images via MCMC. In *University of Leeds*, pages 91–100, 1997.

[Fox05] C. Fox. Fast Jacobian and Transpose of Jacobian Operation for EIT (and other inverse problems), 2005. This is an electronic document. Date retrieved: January 14, 2011.

[FSK08] A. Forrester, A. Sobester, and A. Keane. *Engineering Design Via Surrogate Modelling: A Practical Guide*. Wiley, 2008.

[FZL07] W. Fu, J. Zhao, and J. Lei. ECT Technique and Its Application to Gas/Solid Flows Concentration Measurement. In *Electronic Measurement and Instruments, 2007. ICEMI '07. 8th International Conference on*, pages 4–274 –4–276, July 16-18 2007.

[FZWN08] A. Fuchs, H. Zangl, D. Watzenig, and M. Neumayer. Investigations on inverse filtering to sharpen electrical capacitance tomography images. In *13th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, Graz, Austria, September 21-24 2008.

[GA93] Mohinder S. Grewal and Angus P. Andrews. *Kalman filtering: theory and practice*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[GBN$^+$10] B. George, Th. Bretterklieber, M. Neumayer, H. Zangl, and G. Brasseur. Precision Measurement of Humidity Dependent Permittivity of Foams. In *IEEE Sensors Conference 2010*, 2010.

[GBT95] W. R. Gilks, N. G. Best, and K. K. C. Tan. Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 44(4):pp. 455–472, 1995.

[GCSR 5] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis.* Chapman & Hall, London, 1995 (ISBN 0-412-03991-5).

[Gey92] C. J. Geyer. Practical Markov Chain Monte Carlo. *Statistical Science*, 7(4):pp. 473–483, 1992.

[GG84] S. Geman and D. Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-6(6):721 –741, 1984.

[Gla97] L. F. Gladden. Process tomography: Principles, techniques and applications. *Measurement Science and Technology*, 8(4), 1997.

[Gre95] P. J. Green. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732, 1995.

[Gre05] P. C. Gregory. *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica Support.* Cambridge, Cambridge, 2005.

[GRS96] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. *Markov chain Monte Carlo in practice.* Chapman & Hall/CRC, 1996.

[Had23] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations.* Yale University Press, New Haven, 1923.

[Hag89] W. W. Hager. Updating the inverse of a matrix. *SIAM Review*, 31(2):pp. 221–239, 1989.

[Han98] P. C. Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, USA, 1998.

[Has70] W. Hastings. Monte Carlo sampling using Markov chains and their applications. *Biometrica*, 57(1):pp. 97–109, 1970.

[Hay01] S. Haykin. *Adaptive Filter Theory (4th Edition).* Prentice Hall, September 2001.

[HDX$^+$88] S. M. Huang, T. Dyakowski, C. G. Xie, A. B. Plaskowski, L. A. Xu, and M. S. Beck. A tomographic flow imaging system based on capacitance measuring techniques. In *Pattern Recognition, 1988., 9th International Conference on*, pages 570 –572 vol.1, nov 1988.

[HL10] M. Hamidipour and F. Larachi. Characterizing the liquid dynamics in cocurrent gas-liquid flows in porous media using twin-plane electrical capacitance tomography. *Chemical Engineering Journal*, 165(1):310 – 323, 2010.

[Hol05] D. S. Holder. *Electrical Impedance Tomography: Methods, History and Applications.* Institute of Physics Publishing, Philadelphia, USA, 2005.

[HXW$^+$94] R. He, C. G. Xie, R. C. Waterfall, M. S. Beck, and C. M. Beck. Engine flame imaging using electrical capacitance tomography. *Electronics Letters*, 30(7):559 –560, mar 1994.

[Isa96] O. Isaksen. A review of reconstruction techniques for capacitance tomography. *Measurement Science and Technology*, 7(3):325, 1996.

[JB00] A. J. Jaworski and G. T. Bolton. The design of an electrical capacitance tomography sensor for use with media of high dielectric permittivity. *Measurement Science and Technology*, 11(6):743, 2000.

[Kay93] S. M. Kay. *Fundamentals of statistical signal processing: estimation theory.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

[KKSM00] J. Kaipio, V. Kolehmainen, E. Somersalo, and Vauhkonen M. Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems*, 16(5):1487, 2000.

[KO00] M. C. Kennedy and A. O'Hagan. Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society, Series B, Methodological*, 63:425–464, 2000.

[KO01] M. C. Kennedy and A. O'Hagan. Supplementary details on Bayesian Calibration of Computer Models, 2001.

[Kri51] D. G. Krige. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, December 1951.

[KS05] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer, 2005.

[KS07] 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. Applied Computational Inverse Problems.

[LFY+04] S. Liu, L. Fu, W.Q. Yang, H.G. Wang, and F. Jiang. Prior-online iteration for image reconstruction with electrical capacitance tomography. *Science, Measurement and Technology, IEE Proceedings* -, 151(3):195 – 200, May 2004.

[Liu94] J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, New York, 1994.

[Lux78] I. Lux. A special method to sample some probability density functions. *Computing*, 20:183–188, 1978. 10.1007/BF02252347.

[LYHN05] R. Levine, Z. Yu, W. Hanley, and J. Nitao. Implementing random scan Gibbs samplers. *Computational Statistics*, 20:177–196, 2005. 10.1007/BF02736129.

[MCP08] R. Meyer, B. Cai, and F. Perron. Adaptive rejection Metropolis sampling using Lagrange interpolation polynomials of degree 2. *Computational Statistics & Data Analysis*, 52(7):3408 – 3423, 2008.

[MRR+53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[MT95] K. Mosegaard and A. Tarantola. Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research*, 100(B7), 1995.

[Nel01] O. Nelles. *Nonlinear System Identification.* Applied Mathematical Sciences. Springer, 2001.

[NGB$^+$10] M. Neumayer, B. George, Th. Bretterklieber, H. Zangl, and G. Brasseur. Robust sensing of human proximity for safety applications. In *IEEE Instrumentation and Measurement Technology Conference 2010*, 2010.

[NMGZ10] M. Neumayer, M. Moser, B. George, and H. Zangl. Optimization of Baking Processes by Means of Electrical Capacitance Tomography. In *Proceedings of the 6th World Congress of Industrial Process Tomography*, 2010.

[NOS$^+$10] M. Neumayer, G. Of, A. Schwaigkofler, O. Steinbach, G. Steiner, and D. Watzenig. Shape Determination in ECT using an Energy Norm Formulation. In *14th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, 2010.

[NS08] M. Neumayer and G. Steiner. Industrial Process Tomography for Complex Impedance Reconstruction. In *13th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, Graz, Austria, September 21-24 2008.

[NS09] M. Neumayer and G. Steiner. Impact of Wave Propagation Effects in Electrical Tomography. In *Compumag 2009*, Florianopolis, Brazil, November 22-26 2009.

[NSWZ09] M. Neumayer, G. Steiner, D. Watzenig, and H. Zangl. Spatial resolution analysis for real time applications in electrical tomography. In *W3MDM*, Leeds, UK, June 23-24 2009.

[NSWZ10] M. Neumayer, G. Steiner, D. Watzenig, and H. Zangl. Spatial resolution analysis for real time applications in electrical capacitance tomography. *Nuclear Engineering and Design*, In Press, Corrected Proof:–, 2010.

[NWB10] M. Neumayer, D. Watzenig, and B. Brandstätter. Stress Zone Imaging in Steel Plates of Electrical Machines. In *14th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, 2010.

[NWB11] M. Neumayer, D. Watzenig, and B. Brandstätter. Stress Zone Imaging in Steel Plates of Electrical Machines. *Accepted for The International Journal for Computation and Mathematics in Electrical Engineering (COMPEL)*, 30(6), 2011.

[NWZ10] M. Neumayer, D. Watzenig, and H. Zangl. A H_∞ Approach for Robust Estimation of Material Parameters in ECT. In *Proceedings of the 6th World Congress of Industrial Process Tomography*, 2010.

[NWZ11] M. Neumayer, D. Watzenig, and H. Zangl. A H_∞ Approach for Robust Estimation of Material Parameters in ECT. *Accepted for Measurement Science and Technology*, 2011.

[NZ09] M. Neumayer and H. Zangl. High Ferrous Shielding Ratio for Magnetic Proximity Switch Applications. In *IEEE Sensors Conference*, Christchurch, New Zealand, October 25-28 2009.

[NZ10] M. Neumayer and H. Zangl. Modeling of a Tank System for Liquid Level Measurement by Means of Electrical Tomography. In *Proceedings of the 6th World Congress of Industrial Process Tomography*, 2010.

[NZ11] M. Neumayer and H. Zangl. Electrical Capacitance Tomography for Level Measurements of Separated Liquid Stacks. In *Sensors + Test*, Nürnberg, Germany, June 7-9 2011.

[NZWF11] M. Neumayer, H. Zangl, D Watzenig, and A. Fuchs. *New Developments and Applications in Sensing Technology*, volume 83 of *Lecture Notes in Electrical Engineering*, chapter Current Reconstruction Algorithms in Electrical Capacitance Tomography, page 337. Springer, 2011.

[Pol06] A.C. Polycarpou. Introduction to the finite element method in electromagnetics. *Synthesis Lectures on Computational Electromagnetics*, 4:1–126, 2006.

[Ram05] A. G. Ramm. *Inverse Problems - Mathematical and Analytical Techniques with Applications to Engineering.* Springer Science, Boston, 2005.

[Ras06] C. E. Rasmussen. Gaussian processes for machine learning. MIT Press, 2006.

[RC06] A. L. Richard and G. Casella. Optimizing random scan Gibbs samplers. *Journal of Multivariate Analysis*, 97(10):2071 – 2100, 2006.

[RR01] G. O. Roberts and J. S. Rosenthal. Optimal Scaling for Various Metropolis-Hastings Algorithms. *Statistical Science*, 16(4):pp. 351–367, 2001.

[SBN$^+$10] Th. Schlegl, Th. Bretterklieber, M. Neumayer, H. Zangl, and G. Brasseur. A Novel Sensor Fusion Concept for Distance Measurement in Automotive Applications. In *IEEE Sensors Conference 2010*, 2010.

[SBN$^+$11] Th. Schlegl, Th. Bretterklieber, M. Neumayer, H. Zangl, and G. Brasseur. A Novel Sensor Fusion Concept for Distance Measurement in Automotive Applications. *IEEE Sensors Journal*, 2011.

[Sim06] D. Simon. *Optimal State Estimation: Kalman, H_∞, and Nonlinear Approaches.* Wiley & Sons, 1. auflage edition, August 2006.

[Siv96] D. S. Sivia. *Data Analysis: A Bayesian Tutorial (Oxford Science Publications).* Oxford University Press, July 1996.

[SM05] D. M. Scott and H. McCann. *Process Imaging For Automatic Control (Electrical and Computer Enginee).* CRC Press, Inc., Boca Raton, FL, USA, 2005.

[Str86] G. Strang. *Introduction to Applied Mathematics.* Wellesley-Cambridge Press, Box 157, Wellesley, Massachusets 02181, 1986.

[SW08] G. Steiner and D. Watzenig. Electrical Capacitance Tomography with physical bound constraints. In *SICE Annual Conference, 2008*, pages 1100 –1105, aug. 2008.

[SWF08] C. Schwarzl, D. Watzenig, and C. Fox. Estimation of contour parameter uncertainties in permittivity imaging using mcmc sampling. In *Sensor Array and Multichannel Signal Processing Workshop, 2008. SAM 2008. 5th IEEE*, pages 446 –450, July 2008.

[SWW05] G. Steiner, H. Wegleiter, and D. Watzenig. A Dual Mode Ultrasound and Electrical Capacitance Process Tomography Sensor. In *Sensors, 2005 IEEE*, pages 696 –699, November 2005.

[SWZ+07] G. Steiner, D. Watzenig, H. Zangl, H. Wegleiter, and A. Fuchs. Impact of the reconstruction method on the point spread function in electrical tomography. In Ratko Magjarevic, Hermann Scharfetter, and Robert Merwa, editors, *13th International Conference on Electrical Bioimpedance and the 8th Conference on Electrical Impedance Tomography*, volume 17 of *IFMBE Proceedings*, pages 428–431. Springer Berlin Heidelberg, 2007.

[Tar05] A. Tarantola. *Inverse Problem Theory and Methods for Model Parameter Estimation.* Society for Industrial and Applied Mathematics (SIAM), Boston, 2005.

[Tie94] L. Tierney. Markov chains for exploring posterior distributions. *Annals of Statistics*, 22:1701–1762, 1994.

[TS85] T. Takagi and M. Sugeno. Fuzzy identification of systems and its applications to modeling and control. *IEEE Tran. Systems, Man and Cybernetics*, 15:116–132, 1985.

[TV82] A. Tarantola and B. Valette. Generalized Nonlinear Inverse Problems Solved Using the Least Squares Criterion. *Review of Geophysics and Space Physics*, 20:219–232+, 1982.

[UPV04] K. Uhlíř, J. Patera, and S. Václav. Radial basis function method for iso-line extraction. In *Electronic computers and informatics*, pages 439–444, Košice, 2004. VIENALA Press.

[Vog02] C. R. Vogel. *Computational Methods for Inverse Problemss.* Society for Industrial and Applied Mathematics (SIAM), Philadelphia, USA, 2002.

[VVSK99] 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, September 1999.

[WBH04] D. Watzenig, B. Brandstätter, and G. Holler. Adaptive regularization parameter adjustment for reconstruction problems.

IEEE Transactions on Magnetics, 40(2):1116 – 1119, march 2004.

[WBS07] D. Watzenig, M. Brandner, and G. Steiner. A particle filter approach for tomographic imaging based on different state-space representations. *Measurement Science and Technology*, 18(1):30, 2007.

[WF08] D. Watzenig and C. Fox. Posterior variability of inclusion shape based on tomographic measurement data. *Journal of Physics: Conference Series*, 135(1):012102, 2008.

[WF09] D. Watzenig and C. Fox. A review of statistical modelling and inference for electrical capacitance tomography. *Measurement Science and Technology*, 20(5):052002, 2009.

[WFHK05] H. Wegleiter, A. Fuchs, G. Holler, and B. Kortschak. Analysis of hardware concepts for electrical capacitance tomography applications. In *Sensors, 2005 IEEE*, page 4 pp., November 2005.

[WFHK08] H. Wegleiter, A. Fuchs, G. Holler, and B. Kortschak. Development of a displacement current-based sensor for electrical capacitance tomography applications. *Flow Measurement and Instrumentation*, 19(5):241 – 250, 2008.

[Whi02] R. B. White. Using electrical capacitance tomography to monitor gas voids in a packed bed of solids. *Measurement Science and Technology*, 13(12):1842, 2002.

[WHL09] B. Wang, Z. Huang, and H. Li. Design of high-speed ECT and ERT system. *Journal of Physics: Conference Series*, 147(1):012035, 2009.

[WHWB96] R. C. Waterfall, R. He, N. B. White, and C. M. Beck. Combustion imaging from electrical impedance measurements. *Measurement Science and Technology*, 7(3):369, 1996.

[WNF10] D. Watzenig, M. Neumayer, and C. Fox. Accelerated Markov Chain Monte Carlo Sampling in Electrical Capacitance Tomography. In *14th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, 2010.

[WNF11] D. Watzenig, M. Neumayer, and C. Fox. Accelerated Markov Chain Monte Carlo Sampling in Electrical Capacitance Tomography. *Accepted for The International Journal for Computation and Mathematics in Electrical Engineering (COMPEL)*, 30(6), 2011.

[Wol04] U. Wolff. Monte Carlo errors with less errors. *Computer Physics Communications*, 156(2):143 – 153, 2004.

[WSF$^+$07] D. Watzenig, G. Steiner, A. Fuchs, H. Zangl, and B. Brandstätter. Influence of the discretization error on the reconstruction accuracy in electrical capacitance tomography. *The International Journal for Computation and Mathematics in Electrical Engineering (COMPEL)*, 26:661–676(16), 19 June 2007.

[Yan96] W. Q. Yang. Hardware design of electrical capacitance tomography systems. *Measurement Science and Technology*, 7(3):225, 1996.

[Yan10] W. Yang. Design of electrical capacitance tomography sensors. *Measurement Science and Technology*, 21(4):042001, 2010.

[You88] N. Young. *An introduction to Hilbert space*. Cambridge University Press, New York, NY, USA, 1988.

[YP03] W. Q. Yang and L. Peng. Image reconstruction algorithms for electrical capacitance tomography. *Measurement Science and Technology*, 14(1):R1, 2003.

[YSW98] H. Yan, F.Q. Shao, and S. Wang. Fast calculation of sensitivity distributions in capacitance tomography sensors. *Electronics Letters*, 34(20):1936 –1937, October 1998.

[ZF05] Zhang and Fuzhen. *The Schur Complement and Its Applications*. Springer, New Haven, 2005.

[ZN09] H. Zangl and M. Neumayer. Impact of leakage effects in electrical capacitance tomography due to short electrode length. In *W3MDM*, Leeds, UK, June 23-24 2009.

[ZN10] H. Zangl and M. Neumayer. A Fast Gain Invariant Reconstruction Method for Electrical Tomography. In *14th International IGTE Symposium on Numerical Field Calculation in Electrical Engineering*, 2010.