DISSERTATION

Optimal Bayesian Inversion in Computational Uncertainty Quantification for PDE Models

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen wissenschaften unter der Leitung von

Associate Prof. Dipl.-Ing. Dr. techn. Clemens Heitzinger
E101 – Institut für Analysis und Scientific Computing, TU Wien

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

Ahmad Karimi
Matrikelnummer: 11941366

Diese Dissertation haben begutachtet:

1. **Prof. Dr. Clemens Heitzinger**
   Institut für Analysis und Scientific Computing, TU Wien
2. **Prof. Dr. Roger Georges Ghanem**
   University of Southern California
3. **Prof. Dr. Christian Ringhofer**
   Arizona State University

Wien, Juli 2024

Kurzfassung

Die Bayes'sche Analyse bei der inversen Modellierung zielt darauf ab, die Wahrscheinlichkeitsverteilung von unsicheren Parametern zu berechnen, die durch Vorwärtsmodelle (PDE) eingeschränkt werden, wobei ein Vorwissen über den unbekannten Parameter und verrauschte Beobachtungen vorausgesetzt werden. Prominente Beispiele sind Bayes'sche inverse Probleme in der Medizin, Biologie, Bildgebung und unterirdischen Strömung sowie Klima- und Wettervorhersagen. Eine der Hauptfragen, die sich hier stellt, ist, wie viel und wie optimal man durch einen Bayes'schen Ansatz aus gegebenen Messdaten in einem Experiment Informationen extrahieren kann. In dieser Dissertation beantworten wir diese Frage, indem wir Methoden zur optimalen Versuchsplanung (OED) für inverse Bayes'sche Probleme entwickeln. Zu diesem Zweck wird der erwartete Informationsgewinn (Expected Information Gain, EIG) als Optimalitätskriterium berechnet, und zwar als erwartetes logarithmisches Verhältnis zwischen der Posterior- und der Prior-Verteilung. Die Optimierung dieses Verhältnisses führt zu optimalen Versuchsplänen mit den informativsten Daten. Dies führt zu einer Verringerung der Unsicherheit bei den Bayes'schen Parametern, die im mathematischen Modell von Interesse sind.

Der erwartete Informationsgewinn umfasst verschachtelte Integrale und lässt sich in der Regel nicht in einer geschlossenen Form berechnen. In dieser Dissertation werden Double-Loop-Monte-Carlo-Sampling-Methoden entwickelt, um den EIG in realen Anwendungen wie dem inversen Problem der elektrischen Impedanztomographie (EIT) in der medizinischen Bildgebung zu schätzen, bei dem es darum geht, die optimale Elektrodenkonfiguration zu finden. Für dieses inverse Problem wird eine nichtlineare elliptische partielle Differentialgleichung, nämlich die Poisson-Boltzmann-Gleichung, als das entsprechende Vorwärts-PDE-Modell verwendet. Da Double-Loop-Sampling-Methoden inhärent rechenintensiv sind, wird eine Laplace-Approximationstechnik entwickelt, um dieses Problem zu überwinden und das EIG schneller und effizienter zu berechnen. Darüber hinaus wird ein Ansatz zur Varianzreduzierung entwickelt, nämlich die mehrstufige Quasi-Monte-Carlo-Methode zur effizienten Approximation des EIG.

Die Bayes'sche Inversion und die optimale Versuchsplanung haben verschiedene Anwendungen in den Computerwissenschaften und -techniken, der Medizin und der Biologie. In dieser Dissertation wird als weitere Anwendung die Bayes'sche inverse Modellierung von epidemiologischen Krankheiten erforscht. Um die Ausbreitung von Infektionskrankheiten zu modellieren, wird ein SIR (susceptible-infectedremoved) System gewöhnlicher Differentialgleichungen entwickelt und Bayes'sche Inversionstechniken werden für die robuste und zuverlässige Schätzung der unbekannten Parameter im epidemiologischen Modell eingesetzt.

Abstract

Bayesian analysis in inverse modeling aims to compute the probability distribution of uncertain parameters, constrained by forward (PDE) models, given a prior knowledge on the unknown parameter, and noisy observations. Prominent examples include Bayesian inverse problems in medicine, biology, imaging, and subsurface flow as well as climate and weather forecasts. One of the main questions arises here is that how much and how optimal information one can extract through a Bayesian approach from given measurement data in an experiment. In this dissertation, we answer this question by developing optimal experimental design (OED) methodologies for Bayesian inverse problems. To this end, the expected information gain (EIG) is calculated as an optimality criterion which specifically is the expected logarithmic ratio between the posterior and prior distributions. Optimizing this ratio leads to optimal designs of experiments with the most informative data. This results in the reduction of uncertainty in the Bayesian parameter of interest in the mathematical model.

The expected information gain includes nested integrals and usually does not have a closed form to compute. In this dissertation, double loop Monte Carlo sampling methods are developed to estimate the EIG in real world applications such as electrical impedance tomography (EIT) inverse problem in medical imaging, where the goal is to find the optimal electrode configuration. For this inverse problem, a nonlinear elliptic partial differential equation, namely, Poisson-Boltzmann equation is used as the corresponding forward PDE model. Moreover, as double-loop sampling methods are inherently computationally expensive, a Laplace approximation technique is developed to overcome this issue and to compute the EIG in a faster and more efficient way. Furthermore, a variance reduction approach namely the multilevel quasi Monte Carlo method for the efficient approximation of the EIG is developed.

Bayesian inversion and optimal experimental design have various applications in computational science and engineering, medicine and biology. In this dissertation, as another application, the Bayesian inverse modeling of epidemiological diseases is explored. In order to model the spread of infectious diseases, a SIR (susceptible-infected-removed) system of ordinary differential equations is developed and Bayesian inversion techniques are deployed for the robust and reliable estimation of the unknown parameters in the epidemiological model.

Acknowledgements

I would like to express my deepest gratitude to my PhD advisor, Prof. Dr. Clemens Heitzinger, for his unwavering guidance, encouragement, and invaluable insights throughout the entire research process. His expertise and commitment to excellence have been instrumental in shaping the trajectory of this dissertation.

I am grateful to the members of my dissertation committee, Prof. Dr. Roger Georges Ghanem and Prof. Dr. Christian Ringhofer, for their constructive feedback and thoughtful suggestions that significantly enriched the quality of this work.

I want to express my appreciation to the Institute of Analysis and Scientific Computing, where I had the privilege of being part of a vibrant and intellectually stimulating academic community.

I would like to acknowledge the financial support from the Austrian Science Fund (FWF) during this research.

I am indebted to my family for their infinite support, patience, and encouragement throughout this demanding journey.

Vienna, July 2024 Ahmad Karimi

Contents

1 Introduction

1.1 Bayesian Inversion and Optimal Experimental Design

Various phenomena around us can be modeled by mathematical models, often with the means of differential equations or systems. The coefficients and parameters used in the mathematical model are effective factors on the response variable, but they have unknown and uncertain values in the sense that they can only be determined with the help of measurement data, which always come together with measurement errors. The goal of an inverse problem is to find the unknown parameters given some observations as the measurement data. The parameters cannot be controlled during the experiments. However, knowing the parameter values gives concrete information about the system behavior. To this end, we need to solve the inverse problem.

There are various methods for estimating the parameter values and solving the inverse problem. One can divide the statistical inference approaches into two categories: frequentist and Bayesian inference. The differences between these two approaches stem from the way the concept of probability is interpreted. In the frequentist approach, the unknown parameter is assumed to be fixed and deterministic, and probabilities are defined as long-term frequencies of occurrences of an event. The event has to occur many times. Therefore, in this approach we collect data from a sample of the population and estimate its mean as the value which agrees best with the data.

In contrast, in the Bayesian approach, the unknown parameters are assumed to be random variables. In this technique, probabilities are rooted in degrees of belief and logical support and can be used to represent uncertainties in any event, even in non-repeatable events. In Bayesian inference, we define probability distributions over possible parameter values and use the data to update the distribution, which means that beliefs are updated in response to new evidence. The updating is done by applying Bayes' theorem. In fact, the new information (e.g., experimental data) makes the probability distribution more focused around the true value of the unknown parameter. Thus confidence intervals can easily be calculated.

Prominent examples of exploiting Bayesian analysis for inverse modeling are in climate and weather forecasts [1], subsurface flow [2], life sciences [3, 4, 5], and biomedicine [6, 7, 8]. Bayesian inference tools have been so far applied successfully to many inverse problems in various applications (see for example [4, 9, 10]).

In Bayesian approach [11, 12, 13], in order to solve inverse problems, one considers the unknown parameters as random variables and calculate the posterior probability density that reflects the distribution of the parameter values based on the prior knowledge and the observations. Therefore, in this method, not a single parameter value but its probability distribution is found. This is an advantage, since probability distribution conveys information

how well the parameters can be determined. To this end, we consider the error model:

$$Y = g(\theta) + \epsilon,$$

where Y is the measurement data, g is the parameter-to-observable map and ϵ is an additive Gaussian noise with the mean 0 and covariance matrix Σ_ϵ. Moreover, it is assumed that $p(\theta)$, $p(\theta|Y)$, and $p(Y|\theta)$ are the probability density functions of the prior, posterior, and (data) sampling distributions, respectively. The density $p(Y|\theta)$ of the data provides information from the measured data to update the prior knowledge, and it is usually called the likelihood density function. The goal of Bayesian inversion is to estimate the posterior probability density function $p(\theta|Y)$, which reflects the uncertainty about the quantity of interest θ using measured data Y. In Chapter 2, we present a brief introduction to the Bayesian inversion concept and numerical algorithms including general Markov Chain Monte Carlo methods and Delayed-rejection adaptive Metropolis-Hasting (DRAM). Furthermore, we present an application of the Bayesian approach for a PDE-governed inverse problem in biology, based on our work in [4].

But, how much informative is the measurement data? When it comes to challenging fields of study in computational science and engineering, where the experiments are too expensive or rarely available to be carried out, the design of experiments plays an essential role to overcome these issues. To obtain the most informative data, the experiments must be set up under the best possible designs. The high quality and informative measurement data gives a better quantitative and qualitative understanding of the problem, which leads to the most accurate estimation of the unknown model parameters. However, the quantitative methods for optimizing data acquisition have not been much explored. In this work, we aim to obtain the designs under which the maximum amount of information can be extracted from our experiments.

In order to optimize the design of experiments, there are various optimality criteria. In statistical inversion models, the parameter(s) of interest is estimated via some proposed estimators, and their mean value as well as covariance matrix. The inverse of the covariance matrix is called the *information matrix*. Various criteria in optimal experimental design are defined based on unique real-valued summaries obtained from the information matrix, which are well-known as the *information criteria*. These criteria include

- A-optimality, which seeks to minimize the trace of the inverse of the information matrix. This criterion results in minimizing the average variance of the estimates of the parameter of interest in the model.
- D-optimality, amis to minimize the determinant of the inverse of the information matrix. This criterion leads to maximizing the Shannon information content of the parameter estimates.
- E-optimality, which maximizes the minimum eigenvalue of the information matrix.

In order to measure the information which is gained through a parameter estimation, one can use the Kullback-Leibler distance (D_{KL}) which measures the relative entropy. The expected D_{KL} is an indicator for measuring uncertainty and will be a criterion for the

optimal experimental design, the so-called *Expected Information Gain (EIG)*. The entropy of a continuous parameter θ is given by

$$H(\theta) := -\int_{\Theta} p(\theta)\log(p(\theta))d\theta,$$

where Θ is the parameter space in the inverse problem. The entropy measures the degree of disorder or lack of information carried by the pdf of θ. Low entropy indicates that few data is needed to estimate the true value of the unknown quantity which shows the less uncertainty in the estimation, whereas a high entropy means that, in order to determine the true value, a lot of data needs to be collected.

The Kullback–Leibler Divergence, D_{KL}, quantifies the discrepancy or closeness of two probability distributions, and measures the amount of information of two different pdfs. The Kullback-Leibler divergence, or simply, the KL divergence, is a non-symmetric measure of the difference between two probability distributions $p(\theta)$ and $q(\theta)$, and has been used popularly in data mining literature. The continuous version of the KL divergence between $p(\theta)$ and $q(\theta)$ is defined by

$$D_{KL}(p||q) = \int_{\Theta} p(\theta)\log\frac{p(\theta)}{q(\theta)}d\theta, \tag{1.1}$$

which is a non-negative quantity, and $D_{KL}(p||q) = 0$ if and only if $p = q$. In the case that p and q are Gaussian distributions, i.e. $p(\theta) \sim \mathcal{N}(\mu_p, \Sigma_p)$ and $q(\theta) \sim \mathcal{N}(\mu_q, \Sigma_q)$, the Kulback-Leibler distance has a closed form representation

$$D_{KL}(p||q) = \frac{1}{2}\Big(\mathrm{tr}(\Sigma_q^{-1}\Sigma_p) - n + (\mu_p - \mu_q)^T\Sigma_q^{-1}(\mu_p - \mu_q) + \log(\frac{\det\Sigma_q}{\det\Sigma_p})\Big),$$

where $n = \dim\Theta$. For a simpler case, if $\mu_p = [\mu_1, \ldots, \mu_n]^T$ and $\Sigma_p = \mathrm{diag}([\sigma_1^2, \ldots, \sigma_n^2])$ and $q(\theta) \sim \mathcal{N}(0, I)$, the KL divergence has a much simpler representation as following:

$$D_{KL}(p||q) = \frac{1}{2}\sum_{i=1}^{n}\Big(\sigma_i^2 - 1 + \mu_i^2 - \log(\sigma_i^2)\Big).$$

We consider an inverse problem with design parameters ξ and the measurement model

$$Y_\xi = g(\theta, \xi) + \epsilon,$$

where Y_ξ is the measurement data and g is the parameter-to-observable map. In this model, ϵ is an additive Gaussian noise with the mean 0 and covariance matrix Σ_ϵ. The goal is to estimate the Expected Information Gain (EIG) for a given design ξ, which is defined by

$$\mathrm{EIG}_\theta(\xi) = \mathbb{E}[D_{KL}] = \int_{\mathcal{Y}}\int_{\Theta} p(\theta|Y_\xi)\log\Big(\frac{p(\theta|Y_\xi)}{p(\theta)}\Big)\mathrm{d}\theta\, p(Y_\xi)\mathrm{d}Y_\xi,$$

where $\theta \in \Theta$ is the parameter of interest, $p(\theta)$ is the prior knowledge on θ and $p(\theta|Y_\xi)$ is the posterior distribution of the parameter of the interest. As the posterior distribution of θ is intractable, we will use Bayesian approach, and related numerical methods to estimate the EIG for given design parameters which will be explained in Chapter 2.

1.2 Electrical Impedance Tomography

Electrical impedance tomography (EIT) [14, 15, 16, 17, 18] is an imaging technology which reconstructs electrical properties of the interior of a body using surface electrode measurements. The electrical and physical properties of a human body produce great information about the body interior for the identification and characterization of inclusions, for instance, cancerous tissues. This phenomenon is exploited in EIT, where the electrical properties such as conductivity information is used to build images of the interior. This technology has attracted lots of attention since it possesses lots of practical advantages such as non-invasive, safe, portable and comparatively low cost devices. EIT technology has been applied for imaging in various medical fields; lung imaging [19, 20, 21, 22] for example for the detection of pulmonary edema, breast imaging for the detection of abnormalities [23, 24] and brain imaging for the study of epilepsy, migraine, and strokes [25, 26, 27] as a fast neuroimaging tool. Despite of all advances, this charming technology is not yet competitive with computed tomography (CT) and magnetic resonance tomography (MRI) and has not yet been adopted widely as a clinical imaging tool; EIT is an ill-posed and nonlinear inverse problem [28], which suffers from low spatial resolution [29, 30] and a poor signal-to-noise ratio (SNR) [31] compared to conventional tomographic imaging techniques such as CT and MRI. Using more electrodes increases the spatial resolution, but housing many electrodes is not always possible in real-life scenarios. Furthermore, using a smaller FEM mesh size may improve the image quality, but results in higher computational time and cost.

The domain being considered in EIT problem is usually a cross-section of the main object which contains inclusions representing a different material than the background medium and with several electrodes which are equidistantly attached to the surface of the main body. The device with this configuration will be used as the computational domain for the underlying PDE problem later. In Figure 1.1, a schematic diagram of the device with one of the measurement patterns is shown. In this pattern, a current is applied between two electrodes, and the resulting potential is measured at the rest of the electrodes.

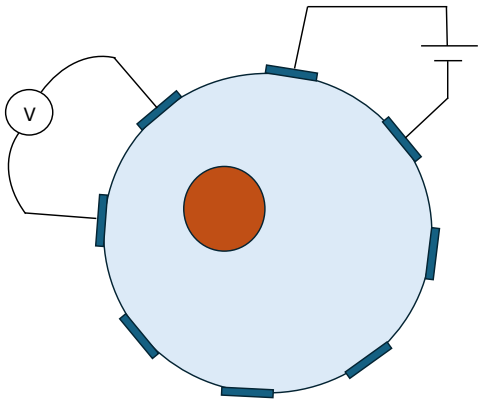

Figure 1.1: Schematic diagram of a cross-section of an EIT device.

The EIT forward problem is to find the electrostatic potential in the physical domain

and then to calculate the potential at the electrodes. Assume that $D \subset \mathbb{R}^2$, is a closed and bounded domain with a smooth boundary ∂D. Using the quasi-static approximation, the electrical field E can be represented in terms of a scalar potential u by

$$E(x) = -\nabla u(x), \tag{1.2}$$

where $x \in D$. For simplicity, we assume direct current or sufficiently low-frequency current such that the magnetic field can be neglected. The linear model of EIT problem

$$\nabla \cdot (A \nabla u) = 0, \tag{1.3}$$

where $A(x)$ is the conductivity of the tissue inside the domain. The linear model (1.3) is widely used for modeling EIT; the Electrical Impedance Tomography and Diffuse Optical Tomography Reconstruction Software EIDORS [32] (http://eidors.org) is based on this model equation.

Mathematical modeling and numerical simulations of the EIT problem including uncertainty quantification and efficiently solving the tomography inverse problem are essential for optimal and efficient design of EIT devices in imaging technology. In Chapter 3 which is based on our work in [7, 8], we develop a new PDE model for EIT and propose Bayesian inversion techniques for solving a real-world EIT inverse problem. Furthermore, the well-posedness of Bayesian inversion for the new inverse problem as well as numerical approaches for estimating the conductivity of the domain as the unknown parameter are developed in this chapter.

Beyond Bayesian approach for EIT, we explore the optimal design of experiment in real-world EIT applications. The main question is: how and where should the electrodes at the body surface be located to extract more informative measurement data? To answer this question, we will maximize the expected information gain over the EIT design space. To this end, the inversion recipe includes the presentation of optimal measurement frequency and optimal electrode placement to collect the measurement data, which lead to estimate the parameter of interest in the model, i.e., the conductivity of interior body. In Chapter 3, we also develop a double-loop Monte-Carlo (DLMC) method for the EIT problem to calculate the expected information gain as well as a stochastic optimization to obtain the optimal designs for the EIT problem.

1.3 Laplace Approximation for Electrical Impedance Tomography Inverse Problem

In order to compute the expected information gain (EIG), we apply a double-loop Monte Carlo method which is in fact a sampling-based method. It utilises the Bayes' rule and sample from the prior distribution of the parameter of interest, and the likelihood distribution to calculate the EIG. In this process, the data evidence $p(y)$ has to be also calculated through the inner loop, which results in high complexity in computations, specially when the forward problem is not easily possible to assess, for instance in PDE governed inverse problems such as EIT.

Instead of using the double-loop Monte Carlo method, which is computationally very expensive, an alternative is to exploit a Laplace approximation to the posterior to reduce

the computational complexity. This approach estimates the posterior distribution using a Gaussian approximation and accelerates the estimation of the expected information gain significantly [33, 34, 35, 36]. Laplace approximation provides an analytical expression for the posterior distribution of the unknown parameter θ by fitting a Gaussian distribution [37, 38]. The mean for the target Gaussian distribution is obtained via the Maximum a Posteriori (MAP) estimation and the covariance of the distribution is found via the Fisher information matrix, which is obtained through the Hessian inverse of the forward problem.

Assume $\theta \in \mathbb{R}^{N_p}$ is the parameter of interest, and has a multivariate Gaussian prior distribution. Further assume that the likelihood density function is defined by $p(Y_\xi|\theta) := \exp(-\frac{1}{2}\|g_\xi(\theta) - Y_\xi\|^2_{\Sigma_{\text{noise}}})$, where $Y_\xi = \{Y_\xi^i\}_{i=1}^M$ are the measurement data (which are obtained via the experiment with the design ξ), and g_ξ is the forward problem which is in most of the cases a (partial) differential equation. The noise $\epsilon = \{\epsilon_i\}_{i=1}^M$ is assumed to be a Gaussian additive noise $\epsilon \sim \mathcal{N}(0, \Sigma_{\text{noise}})$. According to the Bayes' rule, the posterior distribution of the parameter θ can be obtained by

$$p(\theta|Y_\xi) \propto P(Y_\xi|\theta)\, P(\theta).$$

which means the posterior of the parameter θ is proportional to the product of the likelihood function and the prior pdf, and it can be represented by

$$p(\theta|Y_\xi) = p(\theta|\{Y_\xi^i\}) \propto \prod_{i=1}^{M} \exp\left(-\frac{1}{2}(Y_\xi^i - g_\xi(\theta))^T \Sigma_{\text{noise}}^{-1}(Y_\xi^i - g_\xi(\theta))\right) p(\theta).$$

The Laplace approximation of the posterior distribution is a Gaussian distribution

$$p(\theta|Y_\xi) \sim \mathcal{N}(\hat{\theta}, \Sigma_{\text{post}}),$$

where $\hat{\theta}$ is the MAP estimation of the negative logarithm of the posterior pdf, and Σ_{post} is the negative inverse Hessian of the log posterior pdf evaluated at $\hat{\theta}$.

Considering the Laplace approximation of the posterior, and a fast estimation of it presented in [34, 36], the expected information gain (EIG) can be estimated in a much more faster way than the double-loop Monte Carlo method. Using the Laplace approximation, the complexity of having a nested integral will be reduced to an integral over the parameter space, which can be estimated via a standard Monte Carlo method or multilevel approaches [39, 40, 41]

In Chapter 4, we will consider a simpler version of the nonlinear EIT problem with the same domain as it is considered in Chapter 3. To solve the proposed EIT model, the first order Galerkin finite element method is used and in order to illustrate the numerical solution, a cross-section of a human right leg is assumed as the computational domain, where eight electrodes are attached to its boundary. The main domain includes three different subdomains: bone, muscle, and fat. Each subdomain has its own electrical conductivity and in the solution of the forward problem, the FEM mesh is aligned with the inclusions such that each element has a constant value for the coefficient A in the EIT model.

In Chapter 4, we will develop a Laplace approximation methodology for this PDE governed inverse problem, and will estimate the posterior distribution for the conductivity of the body interior tissues. We will also estimate the expected information gain (EIG) using the

Laplace approximation approach for this inverse problem for a given design of experiment. We expect a significant improvement in the performance in terms of computational cost in the Laplace approximation comparing to the double-loop Monte Carlo approach as the problem is reduced to a single-loop integration.

1.4 Multilevel Monte-Carlo Methods for Optimal Bayesian Experimental Design

Despite the Monte Carlo and double-loop Monte Carlo methods are very useful approaches for the estimation of expectations as well as nested expectations, they can be computationally expensive, particularly when the cost of generating individual stochastic samples is very high. This might happen when we are dealing with stochastic PDEs, or in the case of PDE-governed inverse problems. In such problems, usually there are some features to trade-off between accuracy and computational costs in the evaluation of the underlying function or equation that should be estimated for each sample in the Monte Carlo approach. In the other word, higher accuracy leads to higher computational costs. As an example, the computational cost and accuracy of the solution in a PDE with random input data depend on the finite element discretization and mesh size h. For the coarser mesh size the accuracy is low while the CPU time for solving the PDE also decreases, but for finer mesh sizes accuracy and running time will be quite high. The creative idea here is to use some levels in the Monte Carlo process to take most samples in the starting levels (with coarser mesh size) with a low accuracy and corresponding low cost, and only very few samples are taken in final levels with high accuracy and high cost. This approach and its various extensions are the so-called Multilevel Monte Carlo (MLMC) methods which greatly help to reduce the computational costs by distributing the samples among the levels.

Giles [42, 43] introduced the Multilevel Monte Carlo (MLMC) in the context of stochastic differential equations (SDEs) for option pricing, however, earlier traces are back to the work in parametric integration by Heinrich [44]. In MLMC approach, the telescoping sum is being considered as

$$\mathbb{E}[\mathcal{Z}_L] = \mathbb{E}[\mathcal{Z}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}],$$

in which each term on the right hand side can be computed using the standard Monte Carlo method with N_ℓ samples drawn in level ℓ.

Since early works, the multilevel Monte Carlo method has been successfully applied for various applications in computational science and engineering. These applications include the stochastic drift-diffusion-Poisson system for modeling nano-scale semiconductor devices [45, 46], applications to stochastic partial differential equations (SPDEs) or PDEs with random coefficients [47, 48], and computing posterior expectations in elliptic inverse problems [49]. Sinha and Chakrabarty [50] present a review of the recent developments on the Multilevel Monte Carlo (MLMC) approach with applications in option pricing and

financial risk management. They discuss how to incorporate the importance sampling in conjunction with the MLMC estimator in order to achieve variance reduction in the estimator. They also discuss efficient algorithms in order to estimate the risk measures of Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR). To this end, an adaptive sampling algorithm is constructed for efficient estimation of the nested expectations, which, in general is computationally expensive [51].

In spite of widely use of Multilevel Monte Carlo method, there are still many problems in machine learning and statistics which involve nested expectations where the conventional (multilevel) Monte Carlo (MC) estimation is not enough. Among various applications, we focus on the calculation of the expected information gain (EIG) in the Bayesian experimental design problems [52]. The general form of the EIG is the following nested integral

$$I = \int_{\mathcal{Y}} f\Big(\int_{\mathcal{X}} g(y,x)dx \Big) dy,$$

for which the standard method to estimate numerically is the double loop Monte Carlo (DLMC) estimator [33], which is defined by

$$I_{\text{DLMC}} = \frac{1}{N} \sum_{n=1}^{N} f\Big(\frac{1}{M} \sum_{m=1}^{M} g(y^{(n)}, x^{(n,m)}) \Big),$$

where $y^{(n)}$ and $x^{(n,m)}$ are obtained by i.i.d sampling in the data and parameter spaces, i.e., $\mathcal{Y}$ and $\mathcal{X}$, respectively. We will explain this approach in details in Section 2.5.

To exploit the MLMC idea to construct an efficient estimator for the EIG, first we define an increasing geometric sequence, $M_\ell = M_0\ 2^\ell$, for the number of inner samples at level ℓ, and to determine the number of outer samples in an optimal way [41]. Another option is to use the multilevel approach to estimate the single integral over the parameter space obtained through the Laplace approximation procedure [39] (see Chapter 4). In this case, one should rely on the finite element discretization mesh size h_ℓ in each level ℓ. In this approach, the coarser mesh sizes are used in early levels with more number of samples, and on the opposite side, finer FEM mesh sizes are used for the final levels in which the cost of evaluating the forward PDE problem is high due to the smaller mesh size h_ℓ.

As an extension to (multilevel) Monte Carlo methods, randomised quasi-Monte Carlo (ML)RQMC method is another sampling approach, where low-discrepancy sequences such as the Halton sequence [53] and the Sobol sequence [54] are employed to achieve variance reduction. The MLRQMC method was first introduced in [55] and developed for a PDE problem in [56]. Figure 1.2 shows visually the points sampled in 2D with different point sets.

The main difference between the (multilevel) quasi-Monte Carlo and the original one is the way the samples are chosen. Bartuska et. al. [40] developed a randomized quasi Monte Carlo method to estimate nested integrals where they exploit this approach to evaluate the expected information gain (EIG). In Chapter 5 of this thesis, we present a new approach by combining multilevel double-loop Monte Carlo and a randomised quasi Monte Carlo method, namely multilevel double-loop randomised quasi Monte Carlo (ML-DL-RQ-MC) for the estimation of the EIG, where we use Goda's [41] approach for the nested integrals.

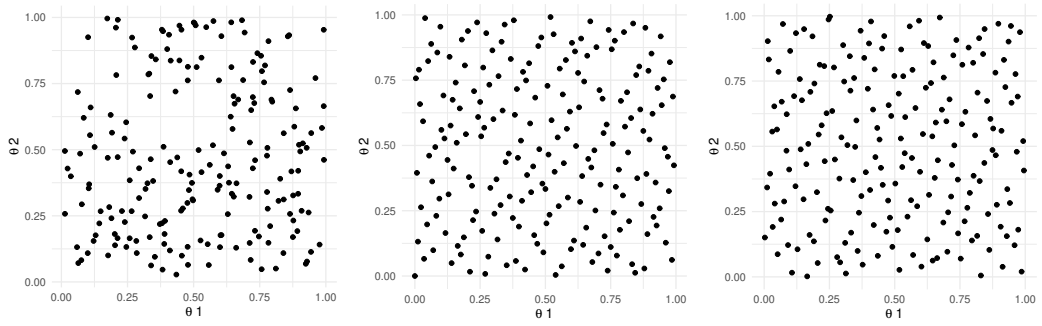

Figure 1.2: 200 points sampled in 2D with a uniform random (left), Halton (middle), Sobol (right) sequences.

1.5 Epidemiological Models

Mathematical modeling of epidemiological diseases using differential equations are of great importance in order to recognize the characteristics of the diseases and their outbreak. The coronavirus COVID-19 pandemic was a new infectious disease in recent years which emerged from China in fall 2019 and then spread around the world. The limitations in result of this outbreak did influence all people and countries in social, economical, political aspects. This pandemic spreads through (micro-) droplets and its outbreak speed was very high. The first reported case of SARS-CoV-2 was identified in Wuhan, China. The first case outside of China was reported in Thailand on 13 January 2020 [57]. Since then, this ongoing outbreak had spread all over the world [58]. In the first wave of this pandemic, till May 21st 2020, the number of 5 230 000 individuals this pandemic had infected around the world and more than 335 000 deaths was reported. Despite restrict measures and limitations by governments, this pandemic which lasted for couple of years through different waves. In the very first wave, out of more than 2 780 000 active cases around the world, 2% were critical patients, in the time of publication of our work in [59]. One of the main sources which collected the data of pandemic in each country separately was the Johns Hopkins CSSE database (https://github.com/CSSEGISandData/COVID-19). On March 10, 2023, the Johns Hopkins Coronavirus Resource Center ceased its collecting and reporting of global COVID-19 data.

The COVID-19 pandemic was first officially confirmed to have spread to Austria on 25 February 2020 and till May 21st 2020 (which is considered for the modeling at the time of publication of [59]. This period of time covers almost the first wave of the pandemic in Austria) more than 16 400 people were infected and 633 deaths and 833 active cases were reported. Various databases have reported different numbers and some of them update their daily reported data even for preceding days. For instance, the Federal Ministry of Social Affairs, Health, Care and Consumer Protection, Republic of Austria (https://info.gesundheitsministerium.at), has updated the number of deaths till May 21 to 657. However, here we have used the Johns Hopkins CSSE database. Figure 1.3 displays daily confirmed cases in Austria as well as the total cumulative count of confirmed, active,

and fatality cases in Austria. By removing deaths and recoveries from total cases, we obtain the "currently infected cases" or "active cases" (cases still awaiting an outcome).

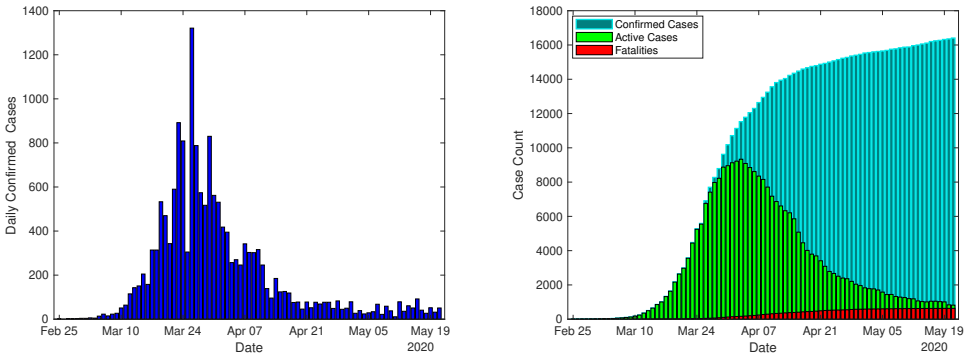

Figure 1.3: (left) Daily confirmed count of coronavirus infected cases and (right) total cumulative count of confirmed infected, active and fatality cases (till May 21st 2020) in Austria.

Infected people needed breathing assistance and a large number of them required medical treatment in an intensive care unit (ICU). Countries which were affected by COVID-19 attempted to keep the daily number of cases below the capacity of their health care system. In order to avert the disastrous inundation of hospitals, the virus had be kept from spreading fast. To this end, countries implemented protective measures such as closing schools, canceling mass gatherings, working from home (home office), self-quarantine, self-isolation, avoiding crowds, social distancing, wearing protection masks, etc.

In Chapter 6 which is based on our published work in [59], we propose Bayesian inference for the analysis of the COVID-19 data in order to estimate the crucial unknown quantities of the pandemic models. We use an adaptive MCMC method to find the probability distributions and confidence intervals of the epidemiological models parameters using the Austrian infection data. We use this analysis for the prediction of the duration of the epidemic in Austria as well as the total number of infected people and fatalities. The model validation shows a very good agreement between the computational and measurement data of infections in Austria which proves the reliability and the accuracy of the predictions. This is of great importance for making governmental decisions in implementing the measures in order to prevent the spread of the virus.

The estimated parameters and the analysis of fatalities provide useful information for decision makers and makes it possible to perform more realistic forecasts of future outbreaks. According to our Bayesian analysis for the logistic model, the growth rate and the carrying capacity are estimated respectively as 0.28 and 14 974. Moreover for the parameters of the SIR model, namely the transmission rate and recovery rate, we estimate 0.36 and 0.06, respectively. Additionally, we obtained an average infectious period of 17 days and a transmission period of 3 days for COVID-19 in Austria. We also estimate the reproduction number over time for Austria. This quantity is estimated around 3 on March 26, when the first recovery was reported. Then it decays to 1 at the beginning of April. Furthermore, we present a fatality analysis for COVID-19 in Austria, which is also of importance for

governmental protective decision making. According to our analysis, the case fatality rate (CFR) is estimated as 4% and a prediction of the number of fatalities for the coming 10 days is also presented. Additionally, the ICU bed usage in Austria indicates that around 2% of the active infected individuals are critical cases and require ICU beds. Therefore, if Austrian governmental protective measures would not have taken place and for instance if the number of active infected cases would have been around five times larger, the ICU bed capacity could have been exceeded.

1.6 Conclusions

The focus of the thesis is on the inverse UQ methods and more specifically on the Bayesian inversion and optimal experimental design for inverse problems. After giving some preliminaries in Chapter 2, we focus on developing theory and numerical methods in Bayesian inversion and optimal Bayesian experimental design to estimate posterior distribution as well as the expected information gain (Chapters 3-6). The applications include electrical impedance tomography (EIT) in medical imaging, an inverse problem in Biology and Medicine, as well as in epidemiological models. The main part of this dissertation is based of following publications of the author:

- **A. Karimi**, C. Heitzinger, Laplace Approximation for PDE Governed EIT Inverse Problems. (2024), preprint.
- **A. Karimi**, L. Taghizadeh, C. Heitzinger, Optimal Bayesian experimental design for electrical impedance tomography in medical imaging, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113489.
- L. Taghizadeh, **A. Karimi**, B. Stadlbauer, W. J. Weninger, E. Kaniusas, C. Heitzinger, Bayesian inversion for electrical-impedance tomography in medical imaging using the nonlinear Poisson–Boltzmann equation, Computer Methods in Applied Mechanics and Engineering 365 (2020) 112959.
- L. Taghizadeh, **A. Karimi**, E. Presterl, C. Heitzinger, Bayesian inversion for a biofilm model including quorum sensing, Computers in Biology and Medicine 117 (2020) 103582.
- L. Taghizadeh, **A. Karimi**, C. Heitzinger, Uncertainty quantification in epidemiological models for the COVID-19 pandemic, Computers in Biology and Medicine 125 (2020) 104011.

Bibliography

[1] P. Bauer, A. Thorpe, G. Brunet, The quiet revolution of numerical weather prediction, Nature 525 (7567) (2015) 47–55.

[2] M. A. Iglesias, K. Lin, A. M. Stuart, Well-posed Bayesian geometric inverse problems arising in subsurface flow, Inverse Problems 30 (11) (2014) 114001.

[3] J.-C. Croix, N. Durrande, M. A. Alvarez, Bayesian inversion of a diffusion model with application to biology, Journal of Mathematical Biology 83 (2) (2021) 13.

[4] L. Taghizadeh, A. Karimi, E. Presterl, C. Heitzinger, Bayesian inversion for a biofilm model including quorum sensing, Computers in Biology and Medicine 117 (2020) 103582.

[5] B. Rahnsch, L. Taghizadeh, Network-based uncertainty quantification for mathematical models in epidemiology, Journal of Theoretical Biology 577 (2024) 111671.

[6] P. Bohra, T.-a. Pham, J. Dong, M. Unser, Bayesian inversion for nonlinear imaging models using deep generative priors, IEEE Transactions on Computational Imaging 8 (2022) 1237–1249.

[7] A. Karimi, L. Taghizadeh, C. Heitzinger, Optimal Bayesian experimental design for electrical impedance tomography in medical imaging, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113489.

[8] L. Taghizadeh, A. Karimi, B. Stadlbauer, W. J. Weninger, E. Kaniusas, C. Heitzinger, Bayesian inversion for electrical-impedance tomography in medical imaging using the nonlinear Poisson–Boltzmann equation, Computer Methods in Applied Mechanics and Engineering 365 (2020) 112959.

[9] E. K. Lenzi, L. R. Evangelista, L. Taghizadeh, D. Pasterk, R. S. Zola, T. Sandev, C. Heitzinger, I. Petreska, Reliability of Poisson–Nernst–Planck anomalous models for impedance spectroscopy, The Journal of Physical Chemistry B 123 (37) (2019) 7885–7892.

[10] B. Stadlbauer, A. Cossettini, D. Pasterk, P. Scarbolo, L. Taghizadeh, C. Heitzinger, L. Selmi, Bayesian estimation of physical and geometrical parameters for nanocapacitor array biosensors, Journal of Computational Physics 397 (2019) 108874.

[11] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013.

[12] A. Tarantola, Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, Elsevier, 1987, new York.

[13] R. C. Smith, Uncertainty Quantification: Theory, Implementation, and Applications, Vol. 12, SIAM, 2013.

[14] J. G. Webster, Electrical Impedance Tomography, Taylor & Francis Group, 1990.

[15] C. W. L. Denyer, Electronics for Real-Time and Three-Dimensional Electrical Impedance Tomographs, Ph.D. thesis, Oxford Brookes University (1996).

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

[17] L. Borcea, Electrical impedance tomography, Inverse Problems 18 (6) (2002) R99.

[18] D. S. Holder, Electrical Impedance Tomography: Methods, History and Applications, CRC Press, 2004.

[19] I. Frerichs, J. Hinz, P. Herrmann, G. Weisser, G. Hahn, T. Dudykevych, M. Quintel, G. Hellige, Detection of local lung air content by electrical impedance tomography compared with electron beam CT, Journal of Applied Physiology 93 (2) (2002) 660–666.

[20] T. Meier, H. Luepschen, J. Karsten, T. Leibecke, M. Großherr, H. Gehring, S. Leonhardt, Assessment of regional lung recruitment and derecruitment during a PEEP trial based on electrical impedance tomography, Intensive Care Medicine 34 (3) (2008) 543–550.

[21] N. Harris, A. Suggett, D. Barber, B. Brown, Applications of applied potential tomography (APT) in respiratory medicine, Clinical Physics and Physiological Measurement 8 (4A) (1987) 155.

[22] A. Adler, Y. Berthiaume, R. Guardo, R. Amyot, Imaging of pulmonary edema with electrical impedance tomography, in: Proceedings of 17th International Conference of the Engineering in Medicine and Biology Society, Vol. 1, IEEE, 1995, pp. 557–558.

[23] R. J. Halter, A. Hartov, K. D. Paulsen, A broadband high-frequency electrical impedance tomography system for breast imaging, IEEE Transactions on Biomedical Engineering 55 (2) (2008) 650–659.

[24] V. Cherepenin, A. Karpov, A. Korjenevsky, V. Kornienko, A. Mazaletskaya, D. Mazourov, D. Meister, A 3D electrical impedance tomography (EIT) system for breast cancer detection, Physiological Measurement 22 (1) (2001) 9.

[25] D. Holder, A. Rao, Y. Hanquan, Imaging of physiologically evoked responses by electrical impedance tomography with cortical electrodes in the anaesthetized rabbit, Physiological Measurement 17 (4A) (1996) A179.

[26] D. Holder, Electrical impedance tomography (EIT) of brain function, Brain Topography 5 (2) (1992) 87–93.

[27] T. Tidswell, A. Gibson, R. H. Bayford, D. S. Holder, Three-dimensional electrical impedance tomography of human brain activity, NeuroImage 13 (2) (2001) 283–294.

[28] W. R. Lionheart, EIT reconstruction algorithms: pitfalls, challenges and recent developments, Physiological Measurement 25 (1) (2004) 125.

[29] A. Sinton, B. Brown, D. Barber, F. McArdle, A. Leathard, Noise and spatial resolution of a real-time electrical impedance tomograph, Clinical Physics and Physiological Measurement 13 (A) (1992) 125.

[30] J. L. Wheeler, W. Wang, M. Tang, A comparison of methods for measurement of spatial resolution in two-dimensional circular EIT images, Physiological Measurement 23 (1) (2002) 169.

[31] A. Adler, R. Guardo, A neural network image reconstruction technique for electrical impedance tomography, IEEE Transactions on Medical Imaging 13 (4) (1994) 594–600.

[32] A. Adler, W. R. Lionheart, Uses and abuses of EIDORS: an extensible software base for EIT, Physiological Measurement 27 (5) (2006) S25.

[33] K. J. Ryan, Estimating expected information gains for experimental designs with application to the random fatigue-limit model, Journal of Computational and Graphical Statistics 12 (3) (2003) 585–603.

[34] Q. Long, M. Scavino, R. Tempone, S. Wang, Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, Computer Methods in Applied Mechanics and Engineering 259 (2013) 24–39.

[35] J. Beck, B. M. Dia, L. F. Espath, Q. Long, R. Tempone, Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain, Computer Methods in Applied Mechanics and Engineering 334 (2018) 523–553.

[36] Q. Long, M. Scavino, R. Tempone, S. Wang, A Laplace method for under-determined Bayesian optimal experimental designs, Computer Methods in Applied Mechanics and Engineering 285 (2015) 849–876.

[37] R. E. Kass, L. Tierney, J. B. Kadane, Laplace's method in Bayesian analysis, Contemporary Mathematics 115 (1991) 89–99.

[38] D. J. MacKay, Information theory, inference and learning algorithms, Cambridge university press, 2003.

[39] J. Beck, B. Mansour Dia, L. Espath, R. Tempone, Multilevel double loop Monte Carlo and stochastic collocation methods with importance sampling for Bayesian optimal experimental design, International Journal for Numerical Methods in Engineering 121 (15) (2020) 3482–3503.

[40] A. Bartuska, A. G. Carlon, L. Espath, S. Krumscheid, R. Tempone, Double-loop quasi-Monte Carlo estimator for nested integration, arXiv preprint arXiv:2302.14119 (2023).

[41] T. Goda, T. Hironaka, T. Iwamoto, Multilevel Monte Carlo estimation of expected information gains, arXiv preprint arXiv:1811.07546v3 (2019).

[42] M. Giles, Improved multilevel Monte Carlo convergence using the Milstein scheme, in: Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer, 2006, pp. 343–358.

[43] M. B. Giles, Multilevel Monte Carlo path simulation, Operations research 56 (3) (2008) 607–617.

[44] S. Heinrich, Multilevel Monte Carlo methods, in: Large-Scale Scientific Computing: Third International Conference, LSSC 2001 Sozopol, Bulgaria, June 6–10, 2001 Revised Papers 3, Springer, 2001, pp. 58–67.

[45] L. Taghizadeh, A. Khodadadian, C. Heitzinger, The optimal multilevel Monte-Carlo approximation of the stochastic drift–diffusion-Poisson system, Computer Methods in Applied Mechanics and Engineering 318 (2017) 739–761.

[46] A. Khodadadian, L. Taghizadeh, C. Heitzinger, Three-dimensional optimal multilevel Monte–Carlo approximation of the stochastic drift–diffusion–Poisson system in nanoscale devices, Journal of Computational Electronics 17 (1) (2018) 76–89.

[47] A. Barth, A. Lang, Multilevel Monte Carlo method with applications to stochastic partial differential equations, International Journal of Computer Mathematics 89 (18) (2012) 2479–2498.

[48] K. A. Cliffe, M. B. Giles, R. Scheichl, A. L. Teckentrup, Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients, Computing and Visualization in Science 14 (2011) 3–15.

[49] R. Scheichl, A. M. Stuart, A. L. Teckentrup, Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems, SIAM/ASA Journal on Uncertainty Quantification 5 (1) (2017) 493–518.

[50] D. Sinha, S. P. Chakrabarty, A review of efficient multilevel Monte Carlo algorithms for derivative pricing and risk management, MethodsX 10 (2023) 102078.

[51] M. B. Giles, A.-L. Haji-Ali, Sub-sampling and other considerations for efficient risk estimation in large portfolios, arXiv preprint arXiv:1912.05484 (2019).

[52] T. Rainforth, R. Cornish, H. Yang, A. Warrington, F. Wood, On nesting Monte Carlo estimators, in: International Conference on Machine Learning, PMLR, 2018, pp. 4267–4276.

[53] J. H. Halton, Algorithm 247: Radical-inverse quasi-random point sequence, Communications of the ACM 7 (12) (1964) 701–702.

[54] I. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, USSR Computational Mathematics and Mathematical Physics 7 (4) (1967) 86–112. doi:https://doi.org/10.1016/0041-5553(67)90144-9.
URL https://www.sciencedirect.com/science/article/pii/0041555367901449

[55] M. B. Giles, B. J. Waterhouse, Multilevel quasi-Monte Carlo path simulation, Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics 8 (2009) 165–181.

[56] F. Y. Kuo, C. Schwab, I. H. Sloan, Multi-level quasi-monte carlo finite element methods for a class of elliptic pdes with random coefficients, Foundations of Computational Mathematics 15 (2015) 411–449.

[57] D. S. Hui, E. I. Azhar, T. A. Madani, F. Ntoumi, R. Kock, O. Dar, G. Ippolito, T. D. Mchugh, Z. A. Memish, C. Drosten, et al., The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—the latest 2019 novel coronavirus outbreak in Wuhan, China, International Journal of Infectious Diseases 91 (2020) 264.

[58] W. H. O. (WHO), Coronavirus Disease 2019 (COVID-19) Situation Report - 97, WHO, 2020.

[59] L. Taghizadeh, A. Karimi, C. Heitzinger, Uncertainty quantification in epidemiological models for the COVID-19 pandemic, Computers in Biology and Medicine 125 (2020) 104011.

2 Optimal Bayesian Experimental Design in Inverse Problems

In this chapter, the goal is to explain the Bayesian optimal experimental design approach for inverse problems. As this method is based on the statistical Bayesian inversion, we first give a brief overview on Bayesian inverse problems and the so-called Markov Chain Monte Carlo (MCMC) methods. Then we will switch to optimal experimental design.

2.1 Bayesian Inverse Problems

Bayesian inversion is a statistical inference approach, which is a robust and reliable technique to quantify the uncertain parameters of mathematical models including partial differential equations. In this section, first, we explain the formulation of Bayesian inverse problems in the finite-dimensional setting using Bayes theorem, and then we aim to describe the Bayesian inversion in the infinite-dimensional setting in the sense of Radon-Nikodym derivative.

Assume a given probability space (Ω, F, P), where Ω is the set of elementary events (sample space), F a σ-algebra of events, and P a probability measure. The solution of the Bayesian inverse problem is the posterior density that best reflects the distribution of the unknown parameter based on the given a prior knowledge on the parameter (prior distribution) as well as the observations (data likelihood distribution). Hence, the measurement model is defined by

$$Y = g(\theta) + \epsilon, \tag{2.1}$$

where ϵ is the measurement noise, which is a mean-zero random variable, $\theta \in \Theta$ denotes the uncertain parameters of the model, $Y \in \mathcal{Y}$ is a given random variable representing observed data or measurements, and $g\colon \Theta \to \mathcal{Y}$ is the parameter-to-observable map[1].

Bayes' Theorem in terms of probability densities can be written as

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)} \tag{2.2}$$

with

$$p(Y) := \int_{\mathbb{R}^n} p(\theta)p(Y|\theta)\mathrm{d}\theta \neq 0, \tag{2.3}$$

where the unknown parameters $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n$ and the observed data Y are realizations of the random variables of the quantity of interests and measurement data, respectively. Furthermore, $p(\theta)$, $p(\theta|Y)$, and $p(Y|\theta)$ are the probability density functions of the prior, posterior, and (data) likelihood distributions, respectively. A probability density function is density of a continuous random variable, which is used to specify the probability of the random variable falling within a particular range of values. The density $p(Y|\theta)$ of the data

provides information from the measurement data to update the prior knowledge, and it is well-known as the likelihood density function. The goal of Bayesian inversion is to estimate the posterior probability density function $p(\theta|Y)$, which reflects the uncertainty about the quantity of interest θ using measurement data Y.

Equation (2.2) gives the posterior density and summarizes our beliefs about θ after we have observed Y. Therefore, Bayes' Theorem for inverse problems can be stated as follows.

Theorem 1 (Bayes' Theorem for inverse problems [1, 2])**.** *Let $p(\theta)$ be the prior probability density function of the realizations θ of the random parameter of interest. Let Y be a realization of the measurement data. Then the posterior density of θ given the measurements Y is*

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)} = \frac{p(\theta)p(Y|\theta)}{\int_{\mathbb{R}^n} p(\theta)p(Y|\theta)\mathrm{d}\theta}. \tag{2.4}$$

Computing the integral appearing in Bayes' Theorem 1 is costly especially if the parameter space $\mathbb{R}^n$ is high-dimensional. Another problem with quadrature rules is that they require a relatively good knowledge of the support of the probability distribution, which is usually part of the information that we seek [1, 2]. In Section 2.2 we shortly discuss the numerical algorithms for Bayesian estimation, namely generating samples by Monte-Carlo methods, which do not require evaluations of the integral. An advantage of the finite-dimensional setting is the existence of the Lebesgue measure, in the sense that the prior and posterior measures possess a density with respect to the Lebesgue measure.

To describe the Bayesian approach on function spaces, we formulate Bayes' Theorem in a measure-theoretic framework, which is suitable for problems on infinite-dimensional spaces. To this end, assume that $(\Theta, \|\cdot\|_\Theta)$ (infinite-dimensional) and $(\mathcal{Y}, \|\cdot\|_\mathcal{Y})$ (possibly infinite-dimensional) are separable Banach spaces and $\theta \in \Theta$ is a random variable distributed according to prior measure μ_0 on Θ, in which our prior beliefs about the unknown parameter θ are described. We assume the distribution of the measurement error ϵ (data likelihood) is defined by

$$\pi(Y|\theta) := \pi(Y - g(\theta)). \tag{2.5}$$

Then, the posterior probability measure μ^Y for $\theta \in \Theta$ given $Y \in \mathcal{Y}$, is defined by

$$\pi(\theta|Y) = \frac{\pi_0(\theta)\pi(Y - G(\theta))}{\int_\Theta \pi_0(\theta)\pi(Y - g(\theta))\mathrm{d}\theta} \tag{2.6}$$

using Bayes' formula, where π_0 and π are the prior and posterior density functions and correspond to the probability measures μ_0 and μ^Y, respectively. Thus, we have

$$\pi(\theta|Y) \propto \pi_0(\theta)\pi(Y - g(\theta)) \tag{2.7}$$

with a constant of proportionality depending only on Y.

As in infinite-dimensional spaces there is no density with respect to the Lebesgue measure, Bayes' rule should be interpreted as providing the Radon-Nikodym derivative between

the posterior measure $\mu^Y(\mathrm{d}\theta) = P(\mathrm{d}\theta|Y)$ (with density $\pi(\theta|Y)$) and the prior measure $\mu_0(\mathrm{d}\theta) = P(\mathrm{d}\theta)$ (with density π_0), yielding

$$\frac{\mathrm{d}\mu^Y}{\mathrm{d}\mu_0}(\theta) \propto \pi(Y - g(\theta)). \tag{2.8}$$

Without loss of generality, since the density π is nonnegative, we can view the right-hand side as the exponential of the negative of $\Phi(\theta, Y)$, where $\Phi\colon \Theta \times \mathcal{Y} \to \mathbb{R}$ is a potential function [3]. Hence, equation (2.8) can be rewritten as

$$\frac{\mathrm{d}\mu^Y}{\mathrm{d}\mu_0}(\theta) \propto \exp\left(-\Phi(\theta, Y)\right), \tag{2.9}$$

which gives a more general definition for the posterior measure in the infinite-dimensional setting using the Radon-Nikodym derivative, and generalizes the Bayes rule for the function spaces. According to this equation, the prior measure should be absolutely continuous with respect to the posterior measure, i.e., $\mu_0 \ll \mu^Y$.

More precisely, the posterior measure μ^Y can be formulated as

$$\frac{\mathrm{d}\mu^Y}{\mathrm{d}\mu_0}(\theta) = \frac{1}{C(Y)} \exp\left(-\Phi(\theta, Y)\right), \tag{2.10}$$

where $C(Y)$ is a normalization constant and chosen such that μ^Y is a probability measure, i.e.,

$$C(Y) := \int_\Theta \exp\left(-\Phi(\theta, Y)\right) \mathrm{d}\mu_0(\theta). \tag{2.11}$$

Furthermore, we assume that $\mu_0(\Theta) = 1$ holds for the infinite-dimensional separable Banach space Θ. This was a brief explanation on the finite- and infinite-dimensional setting of Bayesian inverse problems. For more details, we refer to [1, 3].

It is worth to mention that usually the posterior distribution is intractable and has no closed form. However, it is well known that Gaussian prior and likelihood measures, and a linear parameter-to-observable map g lead to a Gaussian posterior measure $\mathcal{N}(m_{\text{post}}, \Sigma_{\text{post}})$. More precisely, we assume a Gaussian prior $\mu_0 = \mathcal{N}(m_0, \Sigma_0)$, and that the data is given by

$$Y = g(\theta) + \epsilon = A\theta + \epsilon, \qquad \epsilon \sim (0, \Sigma_\epsilon), \tag{2.12}$$

where A is the matrix corresponding to the linear parameter-to-observable map g. Then the posterior measure has a Gaussian distribution with the following covariance and mean:

$$\Sigma_{\text{post}} := \left(A^T \Sigma_\epsilon^{-1} A + \Sigma_0^{-1}\right)^{-1} \tag{2.13}$$

and

$$m_{\text{post}} := \Sigma_{\text{post}}(A^T \Sigma_\epsilon^{-1} Y + \Sigma_0^{-1} m_0). \tag{2.14}$$

2.1.1 Well-posedness of Bayesian Inversion

In this subsection, we give a brief overview for the well-posedness of the Bayesian inversion according to [3]. To this end, we first define the Hellinger distance between probability measures on Θ, and then we define the well-posedness of Bayesian inverse problems.

Definition 1 (Hellinger distance)**.** *Suppose μ_1, μ_2 and ν denote probability measures on Θ such that $\mu_1, \mu_2 \ll \nu$. Then, the Hellinger distance between μ_1 and μ_2 is defined by*

$$\mathrm{d}_{\mathrm{Hell}}(\mu_1, \mu_2) := \Big(\frac{1}{2}\int_\Theta \Big(\sqrt{\frac{\mathrm{d}\mu_1}{\mathrm{d}\nu}} - \sqrt{\frac{\mathrm{d}\mu_2}{\mathrm{d}\nu}}\Big)^2 \mathrm{d}\nu\Big)^{\frac{1}{2}}. \tag{2.15}$$

Definition 2 (Wellposedness)**.** *The Bayesian inverse problem given by* (2.10)*, for any prior measure μ_0 on Banach space Θ and any likelihood potential function Φ, is well-posed if the following two properties are satisfied:*

1. *(Well-definedness) There exists a unique posterior probability measure μ^Y given by* (2.10)*, where $\mu^Y \ll \mu_0$.*

2. *(Stability) The posterior measure μ^Y is Lipschitz continuous with respect to the data Y in the Hellinger metric, i.e., there exists a positive constant $C = C(r)$ such that*

$$\mathrm{d}_{\mathrm{Hell}}(\mu^{Y_1}, \mu^{Y_2}) \leq C\|Y_1 - Y_2\|_{\mathcal{Y}}. \tag{2.16}$$

   for all $Y_1, Y_1 \in \mathcal{Y}$ with $\max(\|Y_1\|_{\mathcal{Y}}, \|Y_2\|_{\mathcal{Y}}) < r$, and for any $r > 0$.

The following assumptions ([3, Assumption 2.6]) on the log-likelihood function Φ guarantee that the corresponding Bayesian inverse problem in the infinite-dimensional setting is well-posed.

Assumptions 1. *Assume that Θ and $\mathcal{Y}$ are Banach spaces, and potential function $\Phi\colon \Theta \times \mathcal{Y} \to \mathbb{R}$ satisfies the following properties:*

1. *For an $\alpha_1 > 0$ and every $r > 0$, there is a constant $M(\alpha_1, r) \in \mathbb{R}$ such that for all $\theta \in \Theta$ and $y \in \mathcal{Y}$ with $\|Y\|_{\mathcal{Y}} < r$*

$$\Phi(\theta; Y) \geq M - \alpha_1\|\theta\|_\Theta.$$

2. *For every $r > 0$, there exists a constant $K = K(r) > 0$ such that for all $\theta \in \Theta$ and $Y \in \mathcal{Y}$ with $\max\{\|\theta\|_\Theta, \|Y\|_{\mathcal{Y}}\} < r$*

$$\Phi(\theta; Y) \leq K.$$

3. *For every $r > 0$, there exists a constant $L = L(r) > 0$ such that for all $\theta_1, \theta_2 \in \Theta$ and $Y \in \mathcal{Y}$ with $\max\{\|\theta_1\|_\Theta, \|\theta_2\|_\Theta, \|Y\|_{\mathcal{Y}}\} < r$*

$$\big|\Phi(\theta_1; Y) - \Phi(\theta_2; Y)\big| \leq L\|\theta_1 - \theta_2\|_\Theta.$$

4. *For an $\alpha_2 > 0$ and for any $r > 0$, there exists a positive constant $C = C(\alpha_2, r)$ such that for all $\theta \in \Theta$ and $Y_1, Y_2 \in \mathcal{Y}$ with* $\max(\|Y_1\|_{\mathcal{Y}}, \|Y_2\|_{\mathcal{Y}}) < r$,

$$\left|\Phi(\theta; Y_1) - \Phi(\theta; Y_2)\right| \leq \exp(\alpha_2 \|\theta\|_{\Theta} + C)\|Y_1 - Y_2\|_{\mathcal{Y}}. \tag{2.17}$$

The following two theorems state the well-defindness and stability of the Bayesian posterior measure if Assumptions 1 hold true.

Theorem 2 (Well-definedness, [3, Thm. 4.1])**.** *Suppose that Θ and $\mathcal{Y}$ are Banach spaces and the likelihood potential function Φ satisfies Assumptions 1 (1), (2) and (3). Further assume that μ_0 is a Gaussian measure with $\mu_0(\Theta) = 1$. Then, the posterior measure μ^Y in* (2.10) *is a well-defined probability measure on Θ.*

Theorem 3 (Stability, [3, Thm. 4.2])**.** *Suppose that Θ and $\mathcal{Y}$ are Banach spaces and the likelihood potential function Φ satisfies Assumptions 1 (1), (2) and (4). Further assume that μ_0 is a Gaussian measure with $\mu_0(\Theta) = 1$, and that $\mu^Y \ll \mu_0$ with Radon-Nikodym derivative given by* (2.10)*. Then, the posterior measure μ^Y is Lipschitz continuous in the data Y with respect to the Hellinger distance.*

If the observation space is finite-dimensional, the well-posedness Assumptions 1 on the potential function are reduced to the following two properties [3]:

Assumptions 2. *Assume that Θ is a Banach space with $\mu_0(\Theta) = 1$, and the parameter-to-observable map $g\colon \Theta \to \mathbb{R}^m$ satisfies the following two conditions:*

1. *For every $\epsilon > 0$, there exists a constant $M = M(\epsilon) \in \mathbb{R}$ such that for all $\theta \in \Theta$*

$$\|g(\theta)\|_{\Sigma_\epsilon} \leq \exp(M + \epsilon\|\theta\|_{\Theta}).$$

2. *For every $r > 0$, there exists a constant $L = L(r) > 0$ such that for all $\theta_1, \theta_2 \in \Theta$ with* $\max\{\|\theta_1\|_{\Theta}, \|\theta_2\|_{\Theta}\} < r$,

$$\left\|g(\theta_1) - g(\theta_2)\right\|_{\Sigma_\epsilon} \leq L\|\theta_1 - \theta_2\|_{\Theta}.$$

Then, the likelihood potential satisfies the well-posedness Assumptions 1.

2.2 Markov-Chain Monte-Carlo Methods

Markov-chain Monte-Carlo (MCMC) methods are a class of Monte-Carlo methods with the general idea of constructing Markov chains whose stationary distribution is the posterior density of the unknown parameter, given a prior density and the measurement data [1]. In the following subsections, first we introduce the standard random-walk Metropolis-Hastings (MH) and then an adaptive MH algorithm.

2.2.1 Standard Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm is an MCMC algorithm to draw samples from a desired distribution by building a Markov-chain of accepted values (out of proposed values) for the unknown parameter as a posteriori distribution. In this algorithm, the first state of the chain θ_0 is given and the new state θ_k, $k = 1, 2, \ldots, N$ of the chain is constructed based on the previous state θ_{k-1}. To this end, a new value θ^* is proposed using the proposal density function $J(\theta^*|\theta_{k-1}) = \mathcal{N}(\theta_{k-1}, \sigma_n^2)$, where σ_n is the proposal covariance. Admissibility of this proposed value is tested by means of calculating the acceptance ratio $\alpha(\theta^*|\theta_{k-1})$, which is defined by

$$\alpha(\theta^*|\theta_{k-1}) = \min\left(1, \frac{p(\theta^*|Y)}{p(\theta_{k-1}|Y)} \cdot \frac{J(\theta_{k-1}|\theta^*)}{J(\theta^*|\theta_{k-1})}\right), \tag{2.18}$$

where $p(\theta|Y)$ and J are respectively posterior and the proposal distributions. Applying Bayes' Theorem of inverse problems, we calculate α as

$$\alpha(\theta^*|\theta_{k-1}) = \min\left(1, \frac{p(Y|\theta^*)p(\theta^*)}{p(Y|\theta_{k-1})p(\theta_{k-1})} \cdot \frac{J(\theta_{k-1}|\theta^*)}{J(\theta^*|\theta_{k-1})}\right), \tag{2.19}$$

where $J(\theta_{k-1}|\theta^*) = J(\theta^*|\theta_{k-1})$ for symmetric proposal functions and $p(\theta)$ is a given prior distribution. Furthermore, $p(Y|\theta)$ is the likelihood distribution which is defined by

$$p(Y|\theta) = N(Y, \sigma_{\mathrm{L}}^2) = \frac{1}{(2\pi\sigma_{\mathrm{L}}^2)^{n/2}} e^{-S_\theta/2\sigma_{\mathrm{L}}^2}, \tag{2.20}$$

where σ_{L} is the likelihood covariance, $S_\theta := \sum_{i=1}^{\mathrm{L}} (Y_i - g(\theta))^2$ is the sum of squares error and $g(\theta)$ denotes the parameter-dependent model response. If the proposed value is admissible, it is accepted as θ_k, otherwise the old value is kept. The mechanism of acceptance and the evolution of the chain are clearly described in Algorithm 1, which is an adaptive MCMC algorithm and will be explained in the next subsection. More details can be found for example in [4, 5, 6, 7] and we refer the reader especially to [1, Chapter 8].

Although the convergence speed is determined by the choice of a good proposal distribution, at least tens or hundreds of thousands of samples are necessary to converge to the target distribution. Choosing the optimal proposal scaling is a crucial issue and affects the MCMC results; if the covariance of the proposal distribution is too small, the generated Markov chain moves too slowly, and if it is too large, the proposals are rejected. Hence, optimal proposal values should be found to avoid both extremes, which leads to adaptive MCMC methods [8, 9, 10]. In the following section, we will consider an adaptive algorithm that helps sample from potentially complicated distributions.

2.2.2 Delayed-Rejection Adaptive-Metropolis algorithm

Searching for a good proposal value can be done manually through trial and error, but this becomes intractable in high dimensions. Therefore, adaptive algorithms that find optimal proposal scales automatically are advantageous. The delayed-rejection adaptive-Metropolis (DRAM) algorithm is an efficient adaptive MCMC algorithm [9]. It is based on the combination of two powerful ideas to modify the Markov-chain Monte-Carlo method,

namely adaptive Metropolis (AM) [11, 12] and delayed-rejection (DR) [13, 14], which are used as global and local adaptive algorithms, respectively. AM finds an optimal proposal scale and updates the proposal covariance matrix, while DR updates the proposal value when θ^* is rejected.

The basic idea of the DR algorithm is that, if the proposal θ^* is rejected, delayed rejection (DR) provides an alternative candidate θ^{**} as a second-stage move rather than just retaining the previous value θ_{k-1}. This process is called delayed rejection, which can be done for one or many stages. Furthermore, the acceptance probability of the new candidate(s) is calculated. Therefore, in the DR process, the previous state of the chain is updated using the optimal parameter scale or proposal covariance matrix that has been calculated via the AM algorithm.

The AM algorithm is a global adaptive strategy, where a recursive relation is used to update the proposal covariance matrix. In this algorithm, we take the Gaussian proposal centered at the current state of the chain θ_k and update the chain covariance matrix at the k-th step using

$$V_k = s_n \mathrm{Cov}(\theta_0, \theta_1, \ldots, \theta_{k-1}) + \varepsilon I_n, \tag{2.21}$$

where s_n is a design parameter and depends only on the dimension n of the parameter space. This parameter is specified as $s_n := 2.38^2/n$ as the common choice for Gaussian targets and proposals [15], as it optimizes the mixing properties of the Metropolis-Hastings search in the case of Gaussians. Furthermore, I_n denotes the n-dimensional identity matrix, and $\varepsilon \geq 0$ is a very small constant to ensure that V_k is not singular, and in most cases it can be set to zero [9].

The adaptive Metropolis algorithm employs the recursive relation

$$V_{k+1} := \frac{k-1}{k} V_k + \frac{s_n}{k} \left(k\overline{\theta}_{k-1}\overline{\theta}_{k-1}^\top - (k+1)\overline{\theta}_k\overline{\theta}_k^\top + \theta_k\theta_k^\top \right)$$

to update the proposal covariance matrix, where the sample mean $\overline{\theta}_k$ is calculated recursively by

$$\overline{\theta}_k = \theta_k + \frac{k}{k+1}(\overline{\theta}_{k-1} - \theta_k).$$

A second-stage candidate θ^{**} is chosen using the proposal function

$$J_2(\theta^{**}|\theta_{k-1}, \theta^*) := N(\theta_{k-1}, \gamma_2^2 V_k), \tag{2.22}$$

where V_k is the covariance matrix produced by the adaptive algorithm (AM) as the covariance of the first-stage and $\gamma_2 < 1$ is a constant. The probability of accepting the second-stage candidate, having started at θ_{k-1} and rejected θ^*, is

$$\alpha_2(\theta^{**}|\theta_{k-1}, \theta^*) := \min\left(1, \frac{p(\theta^{**}|Y)J(\theta^*|\theta^{**})(1-\alpha(\theta^*|\theta^{**}))}{p(\theta_{k-1}|Y)J(\theta^*|\theta_{k-1})(1-\alpha(\theta^*|\theta_{k-1}))}\right), \tag{2.23}$$

where α is the acceptance probability (2.24) in the non-adaptive approach. The acceptance probability is computed so that reversibility of the posterior Markov chain is preserved (for more details see for example [1, Section 8.6]). The DRAM technique is summarized in Algorithm 1.

Algorithm 1 The DRAM algorithm

Initialization:
Choose the first state of the chain θ_0 such that $p(\theta_0) > 0$.
Choose the number N_{samples} of samples or iterations.
Choose the parameter ε.
Choose the initial proposal covariance matrix V_0 (diagonal or symmetric).
Choose the factor γ (often $\gamma := 1/5$) for the second-stage proposal distribution.
for $k = 1 : N_{\text{samples}}$ **do**

1. (Adaptivity:) The covariance matrix V_k in the k-th step is updated by (2.21).
2. A first-stage proposal θ^* is generated from $J(\theta^*|\theta_{k-1}) := N(\theta_{k-1}, V_k)$.
3. The new value θ^* is accepted with probability

$$\alpha(\theta^*|\theta_{k-1}) = \min\left(1, \frac{p(\theta^*|Y)}{p(\theta_{k-1}|Y)} \cdot \frac{J(\theta_{k-1}|\theta^*)}{J(\theta^*|\theta_{k-1})}\right). \tag{2.24}$$

4. If the new state is accepted, we set $\theta_k = \theta^*$. Otherwise:
   a) (Delayed rejection:) A second-stage proposal θ^{**} is generated from proposal density
   $$J_2(\theta^{**}|\theta_{k-1}, \theta^*) := N(\theta_{k-1}, \gamma_2^2 V_k), \tag{2.25}$$
   where V_k is the adapted covariance matrix.
   b) The new value θ^{**} is accepted with probability
   $$\alpha_2(\theta^{**}|\theta_{k-1}, \theta^*) := \min\left(1, \frac{p(\theta^{**}|y)J(\theta^*|\theta^{**})(1 - \alpha(\theta^*|\theta^{**}))}{p(\theta_{k-1}|Y)J(\theta^*|\theta_{k-1})(1 - \alpha(\theta^*|\theta_{k-1}))}\right). \tag{2.26}$$
   c) If the new state is accepted, we set $\theta_k := \theta^{**}$, otherwise $\theta_k := \theta_{k-1}$.

end for

2.2.3 An Application in Biology

In order to understand the growth of biofilms in biology, we have developed a mathematical model based on parabolic partial differential equations (PDE) that describes the time-dependent evolution of the size of the biofilm [16]. The model contains parameters such as growth rate and cooperation that cannot be determined directly from experimental data. We therefore solve the corresponding inverse problem to estimate these important parameter values. Here, we use Markov-chain Monte-Carlo techniques and specifically the DRAM algorithm for a simplified version of the original model.

This section is a part of author's work in [16].

The original model (1-3) in [16] can be rewritten in polar coordinates as

$$\frac{\partial u}{\partial t} = A\frac{\partial^2 u}{\partial r^2} + \frac{A}{r}\frac{\partial u}{\partial r} + \alpha u(1-u/\beta) - \gamma(t)(1-v)u, \tag{2.27a}$$

$$\frac{\partial v}{\partial t} = \rho \max\big(0, \arctan\big(\mu(q(u)-\nu)\big)\big)v - \kappa v^2, \tag{2.27b}$$

where

$$q(u)(t,r) := \big(u(t,\cdot) * rG(\cdot)\big)(r) \tag{2.28}$$

and

$$G(r) := \frac{1}{2\pi\sigma^2}\exp\left(-\frac{r^2}{2\sigma^2}\right), \tag{2.29}$$

with the initial conditions

$$u(t=0,r) = \exp(-r^2/10), \tag{2.30a}$$

$$v(t=0,r) = \exp(-r^2/10), \tag{2.30b}$$

which are symmetric, and with zero Neumann boundary condition for u. Here, A, α, β, ρ, κ, ν and μ denote positive constants which describe the growth of the biofilm. For more detail on the model, we refer to our work in [16]. We have implemented the PDE model (2.27)–(2.30) of growth and degradation of biofilms including quorum sensing by means of the method of lines (MOL) and present the results in the the rest of this section. MOL is a technique for solving partial differential equations by discretizing in all but one dimension and then integrating the semi-discrete problem as a system of ODEs. Here, the discretization in space is done utilizing the finite-difference method, and the resulting ODE system in time is solved using a multistep solver based on numerical differentiation formulas (NDFs). Here, we present simulation results for the model (2.27)–(2.30). We assume a circular dish with diameter 35 mm centered at the origin as the growth space and the computational domain for the biofilms, which are initially located at the center of the dish. Biofilm growth is monitored for about six hours. The goal is to calculate the area covered by the biofilm every hour. In these simulations, the initial biofilm was located at the center of the dish. We find the largest circle in the computational domain where the concentration is above a certain threshold and calculate the relative coverage using the area of this circle.

Fig. 2.1 shows concentrations of biofilms for three different growth rates $\alpha = 1.01$, 1.03 and 1.05. In fact, if two populations of bacteria start to grow with the same diffusion constant but different growth rates, the one with higher growth rate creates a much more concentrated and thicker biofilm. Furthermore, the numerical results show larger coverage area for a bigger growth rate.

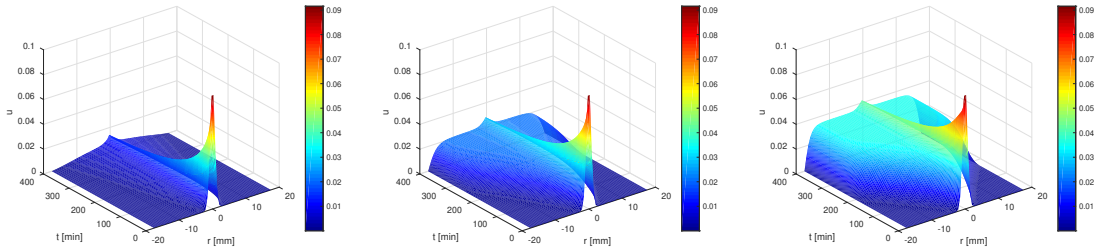

Figure 2.1: Biofilm concentration for $\alpha = 1.01$ (left), $\alpha = 1.03$ (middle), and $\alpha = 1.05$ (right).

The approach here to solve the biofilm inverse problem is Bayesian PDE inversion [1, 17, 18]. In this method, we consider the unknown parameters as random variables and calculate the posterior probability density that reflects the distribution of the parameter values based on the observations. Therefore, in this method, not a single parameter value but its probability distribution is found. This is an advantage, since probability distribution conveys information how well the parameters can be determined.

Fig. 2.2 illustrates the 9D Bayesian estimation results and it displays the marginal histograms of the resulted posterior distribution of nine parameters of the biofilm model together with their mean values. The results give us confidence intervals for each of the unknown quantities as well.

The correlation between some of the model parameter pairs is illustrated in Figure 2.3 by showing two dimensional histograms of the posterior distribution of the pairs calculated by the DRAM algorithm. For the rest of the pairs, similar histograms have been obtained.

In order to verify the response of the presented biofilm model, we compare the simulated coverage using the optimal parameter values with the experimental data. Figure 2.4 illustrates this assessment and shows a very good agreement between the simulations and the measurements, which proves the efficiency of the presented mathematical model as well as the robustness of the applied inverse methods for parameter estimation.

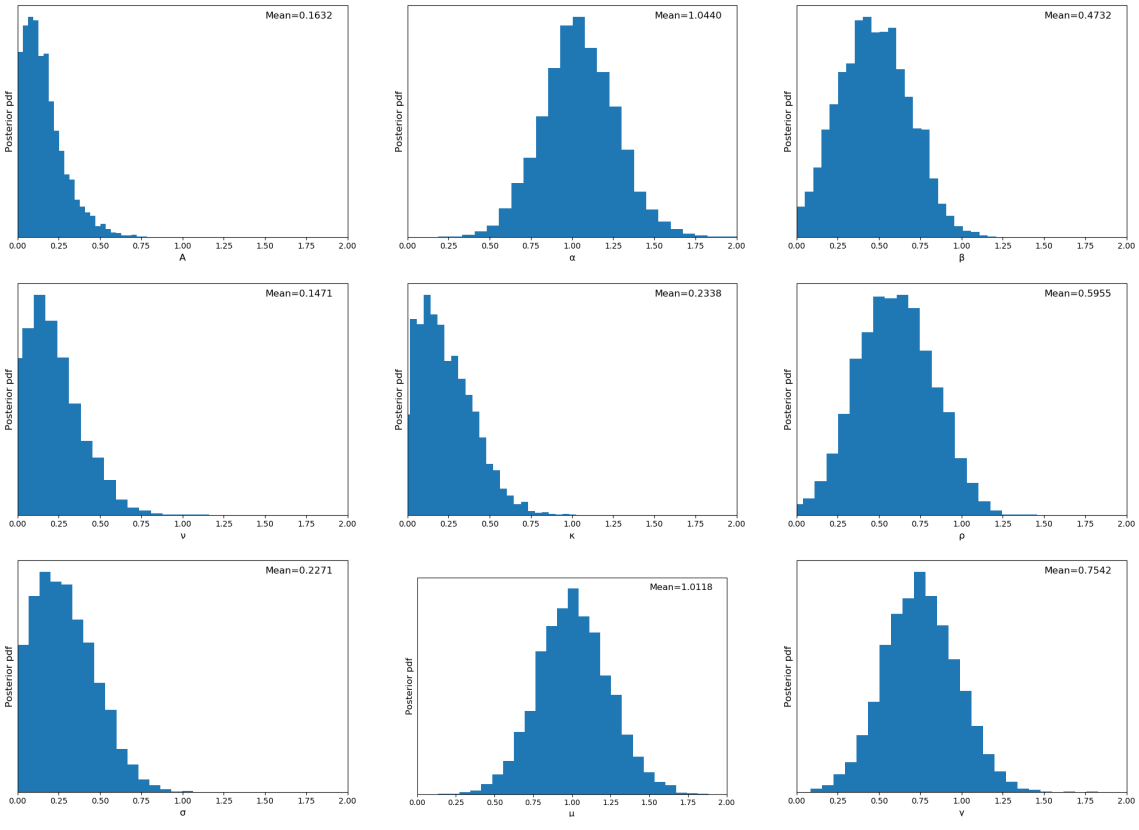

Figure 2.2: Marginal histograms of posterior distributions of the parameters in the biofilm model using the DRAM algorithm with 30 000 samples.

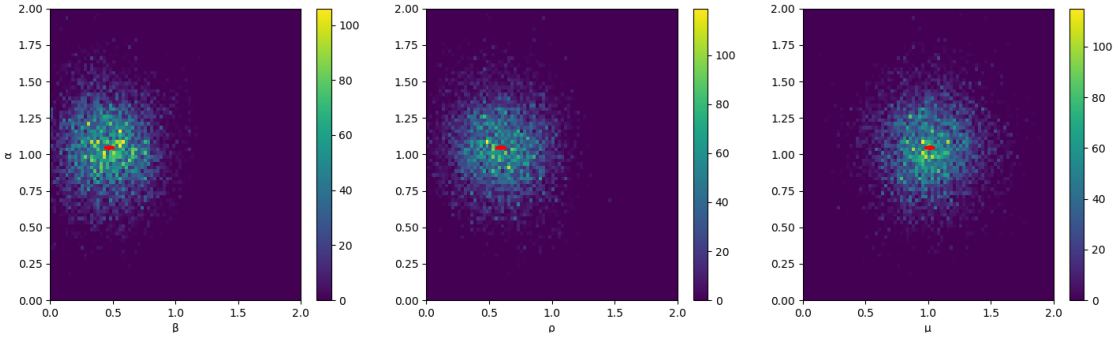

Figure 2.3: Two dimensional histograms of posterior distributions of the parameter pairs in the biofilm model displaying the correlation between some of the pairs.

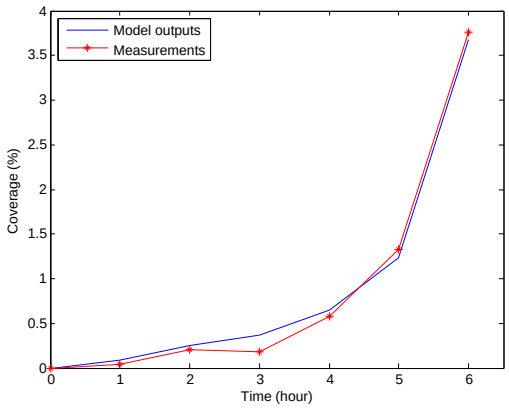

Figure 2.4: The model assessment: experimental data versus simulated coverage.

2.3 Optimal Experimental Design

The goal of the optimal experimental design (OED) [1, 19, 20, 21] is to reduce the uncertainty in the Bayesian inversion to get the most accurate estimation of the unknown parameter. This is especially of great importance when the experiments to collect measurement data are expensive, delicate, or time-consuming to perform. In such cases, OED plays an essential role in accurate parameter estimation and uncertainty quantification. OED has various criteria, while they all try to measure uncertainty, they differ for linear and nonlinear Bayesian inverse problems.

For instance, the A-optimal criterion evaluates the average variance of the estimated parameters, which should be minimized to find the optimal design parameter. This example gives the idea behind the OED, which is a two-stage approach; first, quantification of the uncertainty measure as the design criterion, and second, minimization of the uncertainty measure. Furthermore, in an OED problem, a Bayesian inverse problem is involved, which can be challenging to deal with if it is a nonlinear Bayesian inverse problem for functions (see for instance [22, 23]), due to the nonlinearity as well as the so-called curse of dimensionality. In particular, in the case of PDE-based OED problems, the forward model is expensive to solve, and therefore the OED problem can be prohibitively costly.

We consider the following model:

$$Y_\xi^i = g(\theta, \xi) + \epsilon_i, \tag{2.31}$$

where Y_ξ^i is the ith $(i = 1, 2, \ldots, M)$ s-dimensional response vector, $g : \mathbb{R}^d \times \mathbb{R}^r \to \mathbb{R}^s$ is a nonlinear function in terms of parameter θ and design parameter ξ. Here, ϵ_i is a Gaussian noise with the mean 0 and covariance matrix Σ_ϵ. The noises $\epsilon = \{\epsilon_i\}_{i=1}^M$ are assumed to be independent and identically distributed (i.i.d.). Suppose that we are able to generate M synthetic data points $Y_\xi = \{Y_\xi^i\}_{i=1}^M$ using the same design parameter and repetitive experiments. We should note that there data points are M i.i.d. samples of the random variable Y_ξ (measurement data), and in the model-based problems we can generate them given true θ_*, specific value for ξ and $\{\epsilon_i\}_{i=1}^M$.

There are various types of OED criteria for Bayesian inverse problems, including A-optimality and D-optimality, as well as the information-based optimality criterion, namely the expected information gain.

2.4 Optimal Experimental Design Criterion: Expected Information Gain

This section is mainly based on [24, 25] and aims to clearly define one of the most important criteria for the quality of measurement data in Bayesian Experimental Design, so-called Excepted Information Gain (EIG).

Assume that θ is a random variable that represents the uncertain quantity of interest. We can reduce the uncertainty of θ with the help of collected measurement data Y_ξ through carrying out some experiments under an experimental setup ξ. A very known way to compute the uncertainty of θ is to measure its information entropy [26]. The aim of Bayesian experimental designs is to find an optimal experimental setup ξ_* which maximizes the expected information gain, that is, the expected amount of the information entropy reduction about θ. In the case that ξ is a continuous variable, the straightforward way would be to evaluate the derivative of the expected information gain with respect to ξ, in order to search for a maximizer ξ_*. Otherwise, an accurate evaluation of the expected information gain for given ξ plays the essential role in constructing optimal Bayesian experimental designs.

Definition 3. *Let θ be a random variable and $p(\theta)$ denotes the probability density function of θ. The information entropy of the random variable θ is given by*

$$-\mathbb{E}_\theta[\log p(\theta)]. \tag{2.32}$$

Note that we usually consider $p(\theta)$ as the prior probability density function of θ in Bayesian inversion and optimal experimental design context. Therefore, $-\mathbb{E}_\theta[\log p(\theta)]$ is indeed unconditional (prior) entropy of the θ. Respectively, after collecting data Y_ξ, the conditional (posterior) information entropy of θ is defined the same way as $-\mathbb{E}_{\theta|Y_\xi}[\log p(\theta|Y_\xi)]$, where $p(\theta|Y_\xi)$ denotes the posterior probability density function of θ given Y_ξ. The expectation in the conditional (posterior) information is taken with respect to posterior distribution of θ, i.e. $p(\theta|Y_\xi)$, which is conditional on the measurement data Y_ξ. Thus the expected conditional information entropy of θ by collecting measurement data Y_ξ is

$$\mathbb{E}_{Y_\xi}\left[-\mathbb{E}_{\theta|Y_\xi}[\log p(\theta|Y_\xi)]\right]. \tag{2.33}$$

In 1956, Lindley [26] suggested for the first time the use of Shannon expected information gain [27], which is based on Bayes' Theorem and the Kullback-Leibler (KL) divergence [28, 29] from the posterior to the prior probability density functions of the unknown parameters. The expected information gain $\mathrm{EIG}_\theta(\xi)$ is the average amount of the reduction

of the information entropy about θ by collecting measurement data Y_ξ.

$$\begin{aligned}\mathrm{EIG}_\theta(\xi) &:= -\mathbb{E}_\theta[\log p(\theta)] - \mathbb{E}_{Y_\xi}\left[-\mathbb{E}_{\theta|Y_\xi}[\log p(\theta|Y_\xi)]\right] \\ &= \mathbb{E}_{Y_\xi}\left[-\mathbb{E}_{\theta|Y_\xi}[\log p(\theta)] + \mathbb{E}_{\theta|Y_\xi}[\log p(\theta|Y_\xi)]\right] \\ &= \mathbb{E}_{Y_\xi}\mathbb{E}_{\theta|Y_\xi}\left[\log \frac{p(\theta|Y_\xi)}{p(\theta)}\right]. \end{aligned} \tag{2.34}$$

The inner expectation is the Kullback-Leibler divergence between prior $p(\theta)$ and $p(\theta|Y_\xi)$, which is denoted by D_{KL}. The criterion $\mathrm{EIG}_\theta(\xi)$ is to estimate the efficiency of the proposed experiment setup ξ. The larger value of $\mathrm{EIG}_\theta(\xi)$ means that the gathered data Y_ξ is more informative about θ.

Based on measure theoretic approach, if Θ and $\mathcal{Y}$ are respectively the support of $p(\theta)$ and $p(Y_\xi)$, the Kullback-Leibler divergence is as follows,

$$D_{KL} = \int_\Theta p(\theta|Y_\xi) \log\Big(\frac{p(\theta|Y_\xi)}{p(\theta)}\Big)\mathrm{d}\theta, \tag{2.35}$$

and the expected information gain is calculated by

$$\mathrm{EIG}_\theta(\xi) = \mathbb{E}[D_{KL}] = \int_{\mathcal{Y}}\int_\Theta p(\theta|Y_\xi) \log\Big(\frac{p(\theta|Y_\xi)}{p(\theta)}\Big)\mathrm{d}\theta\, p(Y_\xi)\mathrm{d}Y_\xi. \tag{2.36}$$

Remark 1. *As we fix the design ξ in the rest of this chapter, we omit the subscript ξ and simply write g, Y, EIG in the formulas. In the case that there is no confusion of different designs ξ, we stick to the simpler model*

$$Y = g(\theta) + \epsilon. \tag{2.37}$$

Since $Y - g(\theta)$ follows the probability distribution of ϵ, it is easy to compute $p(Y|\theta)$ for given θ and Y.

Remark 2 (Closed-form expression for the likelihood entropy)**.** *Let $\epsilon_i \sim \mathcal{N}(0, \Sigma_\epsilon)$, $1 \le i \le M$, where Σ_ϵ is a diagonal matrix in $\mathbb{R}^{s\times s}$ with entries $\sigma^2_{\epsilon\{j,j\}}$, $1 \le j \le s$. The likelihood entropy is only depends on ϵ and can be computed in closed form*

$$-\frac{M}{2}\sum_{j=1}^{s}\Big(\log\Big(2\pi\sigma^2_{\epsilon\{j,j\}}\Big) + 1\Big). \tag{2.38}$$

For more details, we refer to [30, 31].

2.5 Numerical Methods for Quantification of the Expected Information Gain

In this section, we present numerical methodology to compute EIG for a given design ξ. The known approach is the Double-Loop Monte Carlo Method (DLMC) which is basically for calculating nested integrals. This approach is also called Nested Monte Carlo (NMC). To calculate EIG, the key point is to sample from the posterior distribution. As sampling directly from the posterior distribution of the θ is usually hard, the expected information gain in Equation (2.36) has no closed form and must be approximated numerically. Using Bayes' rule

$$p(\theta|Y) = \frac{p(\theta)p(Y|\theta)}{p(Y)} = \frac{p(\theta)p(Y|\theta)}{\mathbb{E}_\theta[p(Y|\theta)]}$$

we rewrite (2.34) as [32]

$$\begin{aligned}
\text{EIG} &= \mathbb{E}_Y \mathbb{E}_{\theta|Y}\left[\log \frac{p(Y|\theta)}{\mathbb{E}_\theta[p(Y|\theta)]}\right] \\
&= \mathbb{E}_Y \mathbb{E}_{\theta|Y}\left[\log p(Y|\theta)\right] - \mathbb{E}_Y \mathbb{E}_{\theta|Y}\left[\log \mathbb{E}_\theta[p(Y|\theta)]\right] \\
&= \mathbb{E}_Y \mathbb{E}_{\theta|Y}\left[\log p(Y|\theta)\right] - \mathbb{E}_Y\left[\log \mathbb{E}_\theta[p(Y|\theta)]\right] \\
&= \mathbb{E}_\theta \mathbb{E}_{Y|\theta}\left[\log p(Y|\theta)\right] - \mathbb{E}_Y\left[\log \mathbb{E}_\theta[p(Y|\theta)]\right]. \qquad (2.39)
\end{aligned}$$

Correspondingly, based on the measure theory approach, we can rewrite (2.36):

$$\begin{aligned}
\mathbb{E}[D_{KL}] &= \int_{\mathcal{Y}} \int_\Theta \log\left(\frac{p(\theta|Y)}{p(\theta)}\right) p(\theta|Y) d\theta p(Y) dY \\
&= \int_{\mathcal{Y}} \int_\Theta \log\left(\frac{p(Y|\theta)}{p(Y)}\right) p(Y|\theta) d\theta p(\theta) dY \\
&= \int_\Theta \int_{\mathcal{Y}} \Big(\log(p(Y|\theta)) - \log(p(Y))\Big) p(Y|\theta) dY \, p(\theta) d\theta. \qquad (2.40)
\end{aligned}$$

Monte Carlo sampling can then be used to estimate the integral in Equation (2.40). Assume that $\theta^{(k)}$ are samples drawn from the prior $p(\theta)$ and $Y^{(k)}$ are drawn from the conditional distribution $p(Y|\theta = \theta^{(k)})$ (i.e. likelihood). Therefore

$$\mathbb{E}[D_{KL}] \approx \frac{1}{n_{\text{out}}} \sum_{k=1}^{n_{\text{out}}} \Big(\log(p(Y^{(k)}|\theta^{(k)})) - \log(p(Y^{(k)}))\Big). \qquad (2.41)$$

Here, n_{out} is the number of samples in the outer Monte Carlo estimate. The evidence evaluated at $Y^{(k)}$ can be again approximated using another importance sampling estimate:

$$p(Y^{(k)}) = \int_\Theta p(Y^{(k)}|\theta^{(k)}) p(\theta) d\theta \approx \frac{1}{n_{\text{in}}} \sum_{l=1}^{n_{\text{in}}} p(Y^{(k)}|\theta^{(k,l)}) \qquad (2.42)$$

where $\theta^{(\cdot,l)}$ are drawn from the prior $p(\theta)$ and n_{in} is the number of samples in the "inner" Monte Carlo sum. The combination of Equations (2.41) and (2.42) yields an estimator for EIG:

$$\text{EIG} = \mathbb{E}[D_{KL}] = I \approx I_{\text{DLMC}} = \frac{1}{n_{\text{out}}} \sum_{k=1}^{n_{\text{out}}} \log\Big(\frac{p(Y^{(k)}|\theta^{(k)})}{\frac{1}{n_{\text{in}}}\sum_{l=1}^{n_{\text{in}}} p(Y^{(k)}|\theta^{(k,l)})}\Big). \tag{2.43}$$

Algorithm 2 shows the procedure of calculating EIG using the double loop Monte Carlo method [33]. This method is also called in the literature as Nested Monte Carlo (NMC) method [24].

Algorithm 2 Double Loop Monte Carlo to compute Expected Information Gain (EIG)

Require: Design parameter ξ, Prior pdf $p(\theta)$ and likelihood $p(Y|\theta)$;
Ensure: Expected Information Gain with respect to parameter θ and design ξ

1: Generate a sample $(\theta_1, \ldots, \theta_N)$ of size N from the $p(\theta)$

Outer Loop:

2: Choose an index set of size $n_{\text{out}} \leq N$ randomly from the range 1 to N to get samples for the outer loop $(\theta_1, \ldots, \theta_{n_{\text{out}}})$
3: Obtain the noisy measurement data, $\{Y^{(k)}\}_{k=1}^{n_{\text{out}}}$, by solving forward problem and calculate the likelihood probability $p(Y^{(k)}|\theta^{(k)})$, (using probability distribution of ϵ)

Inner Loop:

4: Choose an index set of size $n_{\text{in}} \leq N$ randomly from the range 1 to N to get samples for the inner loop $(\theta_{1,l}, \ldots, \theta_{n_{\text{out}},l})$ for $l = 1, 2, \ldots, n_{\text{in}}$

EIG:

5: Compute the Expected Information Gain, $\mathbb{E}[D_{KL}]$ using (2.43)

2.6 Error and Variance Analysis of the Double-Loop Monte Carlo Method

The outer loop in the DLMC estimator (2.43) is in fact a Monte Carlo estimator. Generally, the standard MC estimator is unbiased and its variance decreases by increasing the number of samples. However, here in the DLMC estimator, the outer MC estimator has a bias related to the error of the inner integral estimator. In this section, we are going to present an upper bound for the bias of DLMC estimator, i.e. $\text{bias}_{\text{DLMC}} := |I - \mathbb{E}[I_{\text{DLMC}}]|$.

In the case that we are dealing with a PDE governed inverse problem with h as the discretization size in PDE solution (as the forward map), we should also consider a numerical bias for the estimator due to the mesh discretization. In this case, the the total error is

$$|I - I_{\text{DLMC},h}| \leq |I - \mathbb{E}[I_{\text{DLMC}}]| + |\mathbb{E}[I_{\text{DLMC}} - I_{\text{DLMC},h}]|. \tag{2.44}$$

in which the numerical bias follows as

$$|\mathbb{E}[I_{\mathrm{DLMC}} - I_{\mathrm{DLMC},h}]| = C_1 h^{\eta} + o(h^{\eta}), \tag{2.45}$$

where $C_1 > 0$ is a constant. Assuming that each evaluation of the forward map g_h, has the computational work of $h^{-\gamma}$, the average computational work of the double-loop Monte Carlo estimator is

$$W_{\mathrm{DLMC}} \propto n_{\mathrm{in}} n_{\mathrm{out}} h^{-\gamma}.$$

In order to have a shorter notation, We define $\hat{p}_{n_{\mathrm{in}}}(Y) := \frac{1}{n_{\mathrm{in}}} \sum_{l=1}^{n_{\mathrm{in}}} p(Y|\theta^l)$. In the rest of this section, we present estimations for the bias and variance of the DLMC estimator through following theorems. For more details on the bias and variance of DLMC estimator, we refer to [25, 34, 35].

Theorem 4. *The bias of the DLMC estimator I_{DLMC} can be estimated by:*

$$|I - \mathbb{E}[I_{\mathrm{DLMC}}]| = \frac{C_2}{n_{\mathrm{in}}} + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big), \tag{2.46}$$

where

$$C_2 = \frac{1}{2}\mathbb{E}\Big[\mathrm{Var}\Big[\frac{p(Y|\theta)}{p(Y)}|Y\Big]\Big].$$

Proof. We will show that DLMC bias is proportional to the number of samples used for the inner loop in the estimator. To this end, we have

$$\begin{aligned}
|I - \mathbb{E}[I_{\mathrm{DLMC}}]| &= \Big|\mathbb{E}\Big[\frac{1}{n_{\mathrm{out}}}\sum_{k=1}^{n_{\mathrm{out}}} \log\Big(\frac{p(Y^k|\theta^k)}{p(Y^k)}\Big)\Big] - \mathbb{E}\Big[\frac{1}{n_{\mathrm{out}}}\sum_{k=1}^{n_{\mathrm{out}}} \log\Big(\frac{p(Y^k|\theta^k)}{\hat{p}_{n_{\mathrm{in}}}(Y^k)}\Big)\Big]\Big| \\
&= \Big|\mathbb{E}\Big[\log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)\Big] - \mathbb{E}\Big[\log\big(p(Y)\big)\Big]\Big|
\end{aligned}$$

Applying the second-order Taylor expansion of $\log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)$ around $\mathbb{E}[\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)]$ yields

$$\begin{aligned}
\mathbb{E}\Big[\log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)\Big] &= \mathbb{E}\Big[\log\big(\mathbb{E}\big[\hat{p}_{n_{\mathrm{in}}}(Y)\big]\big)\Big] \\
&\quad - \frac{1}{2\big(\mathbb{E}\big[\hat{p}_{n_{\mathrm{in}}}(Y)\big]\big)^2}\mathbb{E}\Big[\big(\hat{p}_{n_{\mathrm{in}}}(Y) - \mathbb{E}\big[\hat{p}_{n_{\mathrm{in}}}(Y)\big]\big)^2\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big) \\
&= \log(p(Y)) - \frac{1}{2p^2(Y)}\mathbb{E}\Big[\big(\hat{p}_{n_{\mathrm{in}}}(Y) - p(Y)\big)^2\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big).
\end{aligned}$$

Therefore, we conclude

$$\begin{aligned}
|I - \mathbb{E}[I_{\mathrm{DLMC}}]| &= \frac{1}{2}\mathbb{E}\Big[\frac{1}{p^2(Y)}\mathbb{E}\Big[\big(\hat{p}_{n_{\mathrm{in}}}(Y) - p(Y)\big)^2\Big|Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big) \\
&= \frac{1}{2}\mathbb{E}\Big[\frac{1}{p^2(Y)}\mathrm{Var}\Big[\hat{p}_{n_{\mathrm{in}}}(Y)\Big|Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big) \\
&= \frac{1}{2}\mathbb{E}\Big[\frac{1}{p^2(Y)}\frac{1}{{n_{\mathrm{in}}}^2}\mathrm{Var}\Big[\sum_{l=1}^{n_{\mathrm{in}}} p(Y|\theta^l)\Big|Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big) \\
&= \frac{1}{2}\mathbb{E}\Big[\frac{1}{p^2(Y)}\frac{1}{{n_{\mathrm{in}}}^2}\sum_{l=1}^{n_{\mathrm{in}}}\mathrm{Var}\Big[p(Y|\theta^l)\Big|Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big) \\
&= \frac{1}{2}\mathbb{E}\Big[\frac{1}{p^2(Y)}\frac{n_{\mathrm{in}}}{{n_{\mathrm{in}}}^2}\mathrm{Var}\Big[p(Y|\theta)\Big|Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big) \\
&= \frac{1}{2n_{\mathrm{in}}}\mathbb{E}\left[\frac{\mathrm{Var}\Big[p(Y|\theta)\Big|Y\Big]}{p^2(Y)}\right] + \mathcal{O}\Big(\frac{1}{{n_{\mathrm{in}}}^2}\Big),
\end{aligned}$$

which completes the proof.

□

Theorem 5. *The variance of the DLMC estimator I_{DLMC} is estimated by:*

$$\mathrm{Var}[I_{\mathrm{DLMC}}] = \frac{C_3}{n_{\mathrm{out}}} + \frac{C_4}{n_{\mathrm{out}}n_{\mathrm{in}}} + \mathcal{O}\Big(\frac{1}{n_{\mathrm{out}}{n_{\mathrm{in}}}^2}\Big), \tag{2.47}$$

where

$$C_3 = \mathrm{Var}\Big[\log\Big(\frac{p(Y|\theta)}{p(Y)}\Big)\Big],$$

$$C_4 = \Big(1 + \mathbb{E}\Big[\log\Big(\frac{p(Y|\theta)}{p(Y)}\Big)\Big]\Big)\,\mathbb{E}\Big[\mathrm{Var}\Big[\frac{p(Y|\theta)}{p(Y)}|Y\Big]\Big] - \mathbb{E}\Big[\log\Big(\frac{p(Y|\theta)}{p(Y)}\Big)\mathrm{Var}\Big[\frac{p(Y|\theta)}{p(Y)}|Y\Big]\Big].$$

Proof. Using the law of total variance we have

$$\begin{aligned}
\mathrm{Var}[I_{\mathrm{DLMC}}] &= \mathrm{Var}\Big[\frac{1}{n_{\mathrm{out}}}\sum_{k=1}^{n_{\mathrm{out}}}\log\Big(\frac{p(Y^k|\theta^k)}{\hat{p}_{n_{\mathrm{in}}}(Y^k)}\Big)\Big] \\
&= \frac{1}{n_{\mathrm{out}}}\mathrm{Var}\Big[\mathbb{E}\Big[\log(p(Y|\theta)) - \log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)\Big|\theta, Y\Big]\Big] \\
&+ \frac{1}{n_{\mathrm{out}}}\mathbb{E}\Big[\mathrm{Var}\Big[\log(p(Y|\theta)) - \log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)\Big|\theta, Y\Big]\Big] \\
&= \frac{1}{n_{\mathrm{out}}}\mathrm{Var}\Big[\mathbb{E}\Big[\log(p(Y|\theta)) - \log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)\Big|\theta, Y\Big]\Big] + \frac{1}{n_{\mathrm{out}}}\mathbb{E}\Big[\mathrm{Var}\Big[\log\big(\hat{p}_{n_{\mathrm{in}}}(Y)\big)\Big|Y\Big]\Big].
\end{aligned} \tag{2.48}$$

Applying the second-order Taylor expansion of $\log\left(\hat{p}_{n_{\text{in}}}(Y)\right)$ around $\mathbb{E}[\left(\hat{p}_{n_{\text{in}}}(Y)\right)]$ for the first term of (2.48) yields

$$
\begin{aligned}
&\frac{1}{n_{\text{out}}}\text{Var}\Big[\mathbb{E}\Big[\log(p(Y|\theta)) - \log\left(\hat{p}_{n_{\text{in}}}(Y)\right)\Big|\theta, Y\Big]\Big] \\
=&\frac{1}{n_{\text{out}}}\text{Var}\Big[\mathbb{E}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right) - \frac{1}{2\,p^2(Y)}\left(\frac{1}{n_{\text{in}}}\sum_{l=1}^{n_{\text{in}}} p(Y|\theta^l) - p(Y)\right)^2\Big|\theta, Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{n_{\text{out}}{n_{\text{in}}}^2}\Big) \\
=&\frac{1}{n_{\text{out}}}\text{Var}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right) - \frac{1}{2\,n_{\text{in}}}\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big] + \mathcal{O}\Big(\frac{1}{n_{\text{out}}{n_{\text{in}}}^2}\Big) \\
=&\frac{1}{n_{\text{out}}}\left(\text{Var}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\Big] + \frac{1}{4\,{n_{\text{in}}}^2}\text{Var}\Big[\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big]\right. \\
&\left.-\frac{1}{n_{\text{in}}}\text{Cov}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right), \text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big]\right) + \mathcal{O}\Big(\frac{1}{n_{\text{out}}{n_{\text{in}}}^2}\Big) \\
=&\frac{1}{n_{\text{out}}}\left(\text{Var}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\Big] + \frac{1}{4\,{n_{\text{in}}}^2}\text{Var}\Big[\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big]\right. \\
&-\frac{1}{n_{\text{in}}}\left(\mathbb{E}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big]\right. \\
&\left.\left.-\ \mathbb{E}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\Big]\mathbb{E}\Big[\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big]\right)\right) + \mathcal{O}\Big(\frac{1}{n_{\text{out}}{n_{\text{in}}}^2}\Big) \\
=&\frac{1}{n_{\text{out}}}\text{Var}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\Big] \\
&-\frac{1}{n_{\text{out}}n_{\text{in}}}\left(\mathbb{E}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big] - \mathbb{E}\Big[\log\left(\frac{p(Y|\theta)}{p(Y)}\right)\Big]\mathbb{E}\Big[\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big]\right) \\
&+\mathcal{O}\Big(\frac{1}{n_{\text{out}}{n_{\text{in}}}^2}\Big).
\end{aligned}
$$

For the second term of (2.48) we have:

$$
\begin{aligned}
&\frac{1}{n_{\text{out}}}\mathbb{E}\Big[\text{Var}\Big[\log\left(\hat{p}_{n_{\text{in}}}(Y)\right)\Big|Y\Big]\Big] \\
=&\frac{1}{n_{\text{out}}}\mathbb{E}\Big[\text{Var}\Big[\log(p(Y)) + \frac{1}{p(Y)}\left(\frac{1}{n_{\text{in}}}\sum_{l=1}^{n_{\text{in}}} p(Y|\theta^l) - p(Y)\right)\Big|Y\Big]\Big] \\
=&\frac{1}{n_{\text{out}}}\mathbb{E}\Big[\text{Var}\Big[\frac{1}{p(Y)}\left(\frac{1}{n_{\text{in}}}\sum_{l=1}^{n_{\text{in}}} p(Y|\theta^l) - p(Y)\right)\Big|Y\Big]\Big] \\
=&\frac{1}{n_{\text{out}}}\mathbb{E}\Big[\text{Var}\Big[\frac{1}{n_{\text{in}}}\frac{\sum_{l=1}^{n_{\text{in}}} p(Y|\theta^l)}{p(Y)}\Big|Y\Big]\Big] \\
=&\frac{1}{n_{\text{out}}{n_{\text{in}}}^2}\mathbb{E}\Big[\text{Var}\Big[\frac{\sum_{l=1}^{n_{\text{in}}} p(Y|\theta^l)}{p(Y)}\Big|Y\Big]\Big] \\
=&\frac{1}{n_{\text{out}}n_{\text{in}}}\mathbb{E}\Big[\text{Var}\Big[\frac{p(Y|\theta)}{p(Y)}\Big|Y\Big]\Big],
\end{aligned}
$$

which completes the proof.

□

Bibliography

[1] R. C. Smith, Uncertainty Quantification: Theory, Implementation, and Applications, Vol. 12, SIAM, 2013.

[2] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer Science & Business Media, 2006.

[3] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451–559.

[4] W. R. Gilks, S. Richardson, D. Spiegelhalter, Markov Chain Monte Carlo in Practice, Chapman & Hall, 1996.

[5] C. P. Robert, G. Cassella, Monte Carlo Statistical Methods, Springer Verlag, 1999.

[6] M. Sambridge, K. Mosegaard, Monte Carlo methods in geophysical inverse problems, Reviews of Geophysics 40 (3) (2002) 1–29.

[7] K. Mosegaard, M. Sambridge, Monte Carlo analysis of inverse problems, Inverse Problems 18 (3) (2002) R29.

[8] J. S. Rosenthal, et al., Optimal proposal distributions and adaptive MCMC, Handbook of Markov Chain Monte Carlo 4 (2011).

[9] H. Haario, M. Laine, A. Mira, E. Saksman, DRAM: efficient adaptive MCMC, Statistics and Computing 16 (4) (2006) 339–354.

[10] S. L. Cotter, G. O. Roberts, A. M. Stuart, D. White, MCMC methods for functions: modifying old algorithms to make them faster, Statistical Science (2013) 424–446.

[11] H. Haario, E. Saksman, J. Tamminen, Adaptive proposal distribution for random walk Metropolis algorithm, Computational Statistics 14 (3) (1999) 375–396.

[12] H. Haario, E. Saksman, J. Tamminen, et al., An adaptive Metropolis algorithm, Bernoulli 7 (2) (2001) 223–242.

[13] L. Tierney, A. Mira, Some adaptive Monte Carlo methods for Bayesian inference, Statistics in Medicine 18 (1718) (1999) 2507–2515.

[14] P. J. Green, A. Mira, Delayed rejection in reversible jump Metropolis–Hastings, Biometrika 88 (4) (2001) 1035–1053.

[15] A. Gelman, G. O. Roberts, W. R. Gilks, et al., Efficient Metropolis jumping rules, Bayesian Statistics 5 (599-608) (1996) 42.

[16] L. Taghizadeh, A. Karimi, E. Presterl, C. Heitzinger, Bayesian inversion for a biofilm model including quorum sensing, Computers in Biology and Medicine 117 (2020) 103582.

[17] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013.

[18] A. Tarantola, Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation, Elsevier, 1987, new York.

[19] A. Atkinson, A. Donev, R. Tobias, Optimum experimental designs, with SAS, Vol. 34, OUP Oxford, 2007.

[20] E. G. Ryan, C. C. Drovandi, J. M. McGree, A. N. Pettitt, A review of modern computational algorithms for Bayesian optimal design, International Statistical Review 84 (1) (2016) 128–154.

[21] A. Alexanderian, Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review, Inverse Problems 37 (4) (2021) 043001.

[22] A. Alexanderian, P. J. Gloor, O. Ghattas, On Bayesian A-and D-optimal experimental designs in infinite dimensions (2016).

[23] A. Alexanderian, N. Petra, G. Stadler, O. Ghattas, A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems, SIAM Journal on Scientific Computing 38 (1) (2016) A243–A272.

[24] T. Goda, T. Hironaka, T. Iwamoto, Multilevel Monte Carlo estimation of expected information gains, arXiv preprint arXiv:1811.07546v3 (2019).

[25] Q. Long, M. Scavino, R. Tempone, S. Wang, Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, Computer Methods in Applied Mechanics and Engineering 259 (2013) 24–39.

[26] D. V. Lindley, et al., On a measure of the information provided by an experiment, The Annals of Mathematical Statistics 27 (4) (1956) 986–1005.

[27] C. E. Shannon, A mathematical theory of communication, Bell System Technical Journal 27 (3) (1948) 379–423.

[28] S. Kullback, Information theory and statistics, Courier Corporation, 1997.

[29] S. Kullback, R. A. Leibler, On information and sufficiency, The Annals of Mathematical Statistics 22 (1) (1951) 79–86.

[30] P. Tsilifis, R. G. Ghanem, P. Hajali, Efficient Bayesian experimentation using an expected information gain lower bound, SIAM/ASA Journal on Uncertainty Quantification 5 (1) (2017) 30–62.

[31] A. Bartuska, A. G. Carlon, L. Espath, S. Krumscheid, R. Tempone, Double-loop quasi-Monte Carlo estimator for nested integration, arXiv preprint arXiv:2302.14119 (2023).

[32] X. Huan, Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, Journal of Computational Physics 232 (1) (2013) 288–317.

[33] C. Ben Issaid, Bayesian Optimal Experimental Design Using Multilevel Monte Carlo, Masters Thesis: King Abdullah University of Science and Technology, 2015.

[34] J. Beck, B. M. Dia, L. F. Espath, Q. Long, R. Tempone, Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain, Computer Methods in Applied Mechanics and Engineering 334 (2018) 523–553.

[35] K. J. Ryan, Estimating expected information gains for experimental designs with application to the random fatigue-limit model, Journal of Computational and Graphical Statistics 12 (3) (2003) 585–603.

3 Optimal Experimental Design for Electrical Impedance Tomography

In this chapter, following the introductory in Section 1.2, we develop optimal Bayesian inversion for electrical impedance tomography (EIT) technology for robust estimation of electrical properties of body tissues and to put this promising imaging technology more into practice. We optimize Bayesian experimental design by maximizing the expected information gain in the Bayesian inversion process in order to design optimal experiments and obtain the most informative data about the unknown parameters. We present optimal experimental designs including optimal frequency and optimal electrode configuration, all of which result in the most accurate estimation of the unknown quantities to date and high-resolution EIT medical images, which are crucial for diagnostic purposes. Numerical results show the efficiency of the proposed optimal Bayesian inversion method for the EIT inverse problem.

This chapter is based on the author's work in [1, 2].

3.1 The Linear Model

The EIT forward problem is to find the electrostatic potential in the physical domain and then to calculate the electrical current flowing through the electrodes. Ignoring magnetic effects and assuming no internal current source in EIT problems, the (complex-valued) linear model is the Poisson equation

$$\nabla \cdot (A \nabla u) = 0, \tag{3.1}$$

where

$$A(x, \omega) := \sigma(x, \omega) + i\omega\epsilon(x, \omega) \tag{3.2}$$

is the admittivity, and σ and ϵ are the electric conductivity and permittivity, respectively. Also, ω is the frequency of the electrical current. Since we restrict the present discussion to static fields, i.e., $\omega \to 0+$, the admittivity is real and coincides with the static conductivity, i.e., $A = \sigma(x)$, which we use here. Moreover, the impedivity $\rho = 1/A$ is just the resistivity of the body [3]. The linear model (3.1) is widely used for modeling EIT; the Electrical Impedance Tomography and Diffuse Optical Tomography Reconstruction Software EIDORS [4] is based on this model equation.

To find the boundary conditions, we assume that there are L contact electrodes e_ℓ, which are attached to the surface of the body, i.e., $e_\ell \subset \partial D$, $1 \leq \ell \leq L$, such that $\overline{e}_\ell \cap \overline{e}_k = \emptyset$ for $\ell \neq k$. We assume that the electrodes conduct perfectly, and thus the tangential electrical

field vanishes along the electrodes. Then, possible boundary conditions on the electrodes are the Dirichlet boundary conditions

$$u(x) = U_\ell, \quad x \in e_\ell, \quad 1 \leq \ell \leq L. \tag{3.3}$$

We also assume that no current flows in and out of the body between the electrodes, which leads to the zero Neumann boundary conditions

$$\frac{\partial u(x)}{\partial \mathbf{n}} = 0, \quad x \in \partial D \backslash \bigcup_{\ell=1}^{L} e_\ell. \tag{3.4}$$

3.2 The Nonlinear Model

To extend the model, we consider free charges of the background medium in the equation. Assuming that the energies of all ions in the electric field are distributed according to a Boltzmann distribution, we define the charge density of free charges in the system by

$$f_{\text{free}} := \eta(\exp(-\beta u) - \exp(\beta u)) = -2\eta \sinh(\beta u),$$

where η is the ion accessibility function. The constant β is defined as $\beta := 1/U_T$, where U_T is the thermal voltage at room temperature and it is defined by $U_T := k_B T/q$ in terms of the Boltzmann constant k_B and the temperature T, and $q > 0$ the elementary charge. Here, we assume that the charge of single positive and negative charge carriers are the same. In the definition of charge density of free charges, the exponential terms stem from the Boltzmann distributions for two species of ions, which leads to sinh in the model.

Adding charge density of free charges f_{free} to the fixed charges f_{fixed} (for simplicity we denote it by f in the equation), we arrive at the nonlinear Poisson-Boltzmann equation

$$-\nabla \cdot (A(x)\nabla u(x)) = f(x) - 2\eta(x)\sinh(\beta u(x))$$

as the extended model for the EIT problem. Therefore, the new forward problem describing EIT is to find the potential u in the main object D, given the conductivity A, the voltage pattern $U = (U_1, \ldots, U_L)^T$, the ion accessibility function η, and the concentration f of fixed charges, that solves the (real-valued) nonlinear elliptic PDE

$$-\nabla \cdot (A(x)\nabla u(x)) + 2\eta(x)\sinh(\beta u(x)) = f(x) \qquad \forall x \in D, \tag{3.5a}$$

$$u(x) = U_\ell \qquad \forall x \in e_\ell, \tag{3.5b}$$

$$\frac{\partial u(x)}{\partial \mathbf{n}} = 0 \qquad \forall x \in \partial D \backslash \bigcup_{\ell=1}^{L} e_\ell. \tag{3.5c}$$

As mentioned before, in every measurement pattern a potential is applied to the electrodes and the resulting electrical current on the rest of the electrodes is measured. The electrical current flowing through the electrodes in the EIT problem [5] is calculated by

$$I_\ell = \int_{e_\ell} A \frac{\partial u(x)}{\partial \mathbf{n}} ds, \qquad 1 \leq \ell \leq L. \tag{3.6}$$

The nonlinear Poisson-Boltzmann equation (3.5) has a unique solution and a pointwise estimate for the solution of the equation has been presented in [6]. The required assumptions for the existence of a unique solution are listed below.

Assumptions 3. *The conductivity* $A\colon D \to \mathbb{R}^{2\times 2}$ *and the voltage simulation pattern* $\{U_\ell\}_{\ell=1}^{L} \in \mathbb{R}^L$ *satisfy the following assumptions:*

1. *The coefficient* $A\colon D \to \mathbb{R}^{2\times 2}$ *is a piecewise constant-valued matrix, which is uniformly elliptic and satisfies*

$$A \in L^\infty(D;\mathbb{R}^{2\times 2}), \quad \operatorname{ess\,inf}_{x\in D} A(x) = A^- > 0 \tag{3.7}$$

   and contains the conductivity of the inclusion and the background medium, as the two materials are different in their physical properties.

2. *The voltages applied to the electrodes are chosen such that* $\sum_{\ell=1}^{L} U_\ell = 0$.

To solve the proposed EIT model, the first order Galerkin finite element method is used and in order to illustrate the numerical solution, a cross-section of a human right leg is assumed as the computational domain, where eight electrodes are attached to its boundary. Figure 3.1 (right) shows the domain with a finite element discretization. In this figure (left), three different subdomains are displayed: the first subdomain is a circular bone (in dark blue) surrounded by the second subdomain, the muscle (in brown), which itself consists of many partitions, and the rest is fat (in blue). Each subdomain has its own electrical conductivity and in the solution of the forward problem the FEM mesh is aligned with the inclusions such that each element has a constant value for the coefficient A.

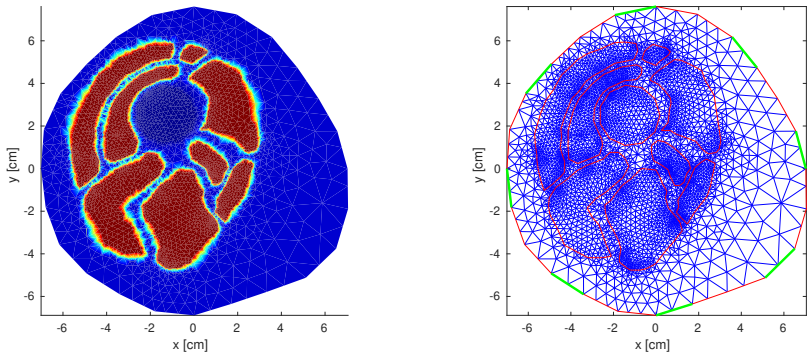

Figure 3.1: EIT domain: cross-section of a right leg illustrating a schematic of the three subdomains (left) and the discretization by a FEM mesh (right) used for solving the forward problem.

Assuming that voltages of ± 10 V are applied to the injection electrode pair under the neighboring/adjacent method, Figure 3.2 displays the obtained electrical potential as the solution of the model equation on the simulation domain for eight injection patterns.

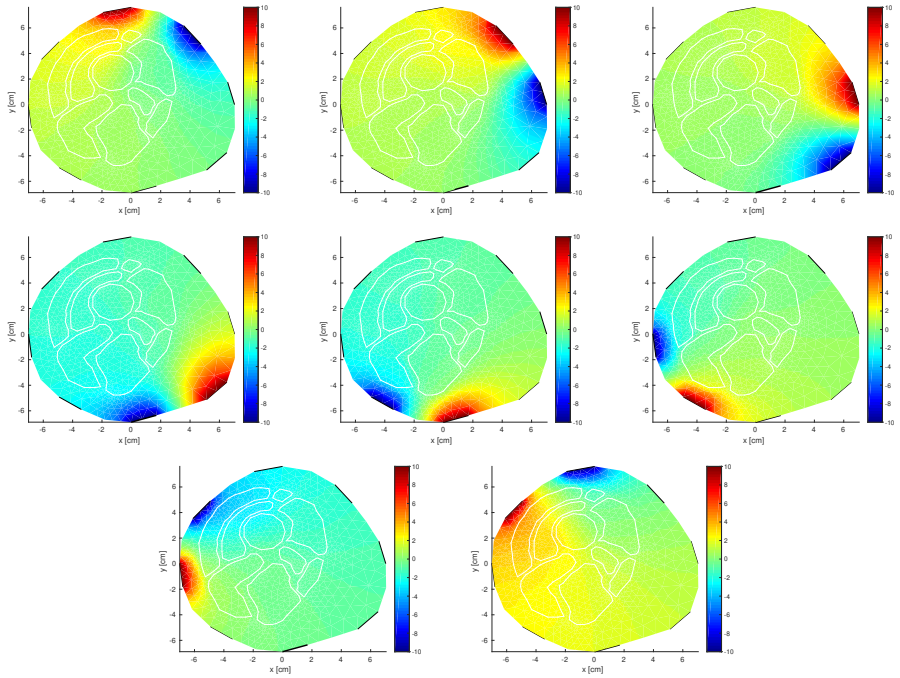

Figure 3.2: Solution of the nonlinear forward problem (in Volt) using the neighboring injection method illustrated in 8 patterns in the frequency of 50 kHz for a cross-section of a right leg.

3.3 Electrical Impedance Tomography Inverse Problem

The Electrical Impedance Tomography (EIT) inverse problem is to reconstruct the electrical and physical properties of the body interior, given the electrode measurements on its surface. To solve this inverse problem, we propose Bayesian inversion techniques, in which the solution of the inverse problem is the posterior density that best reflects the distribution of the parameters based on the observations [7].

In [2], we solved the EIT inverse problem corresponding to the nonlinear model (3.5) on the cross-section of a right leg using real-world surface electrode measurements in order to identify the electrical conductivity of muscle tissue inside the leg. To this end, we developed a Bayesian inversion theory for the proposed EIT model and proposed an adaptive Markov-chain Monte-Carlo (MCMC) method in order to solve the EIT inverse problem.

One of the main challenges is the sensitivity of the conductivity on the boundary measurements, which results from the severely ill-posed and nonlinear EIT inverse problem. In this chapter, we aim to deal with this challenge by optimizing the EIT experimental designs. More precisely, the goal is to determine the optimal frequency and electrode configuration in EIT based on the statistical interpretation of the reconstruction problem. The chapter is dedicated to the Bayesian inversion and optimal Bayesian experimental design for EIT by means of maximizing the expected information gain.

The first goal here is to show that the posterior measure μ^y of the form (2.10) is well-defined and that the problem is well-posed with respect to its dependence on the data. To this end, the function $\Phi\colon X \times Y \to \mathbb{R}$ should have essential properties, namely lower and upper bounds and the Lipschitz property in θ and y. As this function is defined in terms of the function $g\colon X \to \mathbb{R}^m$, it is sufficient to prove the following properties of the function g corresponding to the inverse problem of interest. This implies that $\Phi\colon X \times \mathbb{R}^m \to \mathbb{R}$ satisfies Assumption 2.6 in [8] with $(Y, \|\cdot\|_Y) = (\mathbb{R}^m, |\cdot|)$.

Assumptions 4. *The function $g\colon X \to \mathbb{R}^m$ has the following properties.*

1. *For every $\varepsilon > 0$, there exists an $M(\varepsilon) \in \mathbb{R}$ such that the inequality*

$$|g(\theta)| \leq \exp(\varepsilon \|\theta\|_X^2 + M(\varepsilon)) \tag{3.8}$$

   holds for all $\theta \in X$.

2. *For every $r > 0$, there exists a $K(r) > 0$ such that the inequality*

$$|g(\theta_1) - g(\theta_2)| \leq K(r)\|\theta_1 - \theta_2\|_X \tag{3.9}$$

   holds for all $\theta_1, \theta_2 \in X$ with $\max(\|\theta_1\|_X, \|\theta_2\|_X) < r$.

To prove the well-definedness of the posterior measure and well-posedness of the EIT Bayesian inversion, we need to verify if the bounds and Lipschitz properties in Assumptions 4

hold true when g is given by the solution of the (real-valued) nonlinear elliptic PDE

$$-\nabla \cdot (A(x)\nabla u(x)) + 2\eta(x)\sinh(\beta u(x)) = 0 \qquad \forall x \in D, \tag{3.10a}$$

$$u(x) = U \qquad \forall x \in \partial D_D, \tag{3.10b}$$

$$\frac{\partial u(x)}{\partial \mathbf{n}} = 0 \qquad \forall x \in \partial D_N, \tag{3.10c}$$

where ∂D_D and ∂D_N denote the Dirichlet and Neumann boundaries. In (3.10), the solution u can be the real or the imaginary part of the solution of the complex-valued model, i.e., either $\Re(u)$ or $\Im(u)$. Here, for the sake of simplicity, we denote the solution of the real-valued model (3.10) by u as well. We present the results in Subsection 3.4 as a proposition.

The nonlinear Poisson-Boltzmann equation (3.10) has a unique solution. Moreover a pointwise estimate for the solution of the equation has been presented in [6], which will be used later.

3.4 Well-Posedness of Bayesian Inversion

Here, the main results including the well-definedness of the posterior measure and well-posedness of the EIT Bayesian inversion for the new model (3.10) are presented. To this end, the validity of Assumptions 4 should be verified in the sense that a parameter dependent bound for the solution of the nonlinear model equation must be found and the Lipschitz property of the solution must hold true as well. We have collected our theoretical findings for the new model in Proposition 1. For the linear model a similar bound and the Lipschitz property has been already proved in [9].

Proposition 1. *Suppose the real-valued nonlinear elliptic equation* (3.10) *holds in the bounded domain $D \subset \mathbb{R}^n$, $n \in \{2, 3\}$, with a smooth boundary ∂D and $A := e^\theta =: \eta$, where $\theta \in L^\infty(D)$. Then the estimate*

$$\|u\|_{H^1} \leq He^{2\|\theta\|_{L^\infty}} \tag{3.11}$$

holds for all $\theta \in L^\infty(D)$, and the estimate

$$\|u_1 - u_2\|_{H^1} \leq Se^{4\max(\|\theta_1\|_{L^\infty(D)}, \|\theta_2\|_{L^\infty(D)})}\|\theta_1 - \theta_2\|_{L^\infty(D)} \tag{3.12}$$

holds for all $\theta_1, \theta_2 \in L^\infty(D)$, where $H = H(\|\nabla \overline{U}\|_{L^2(D)})$ and $S = S(\|\nabla \overline{U}\|_{L^2(D)})$ are functions and $\overline{U} \in L^2(D)$ is the Dirichlet lift of U.

Proof. We substitute $v := u - \overline{U}$ in (3.10), where $\overline{U}$ is the Dirichlet lift of g. It is defined by

$$\overline{U} := \begin{cases} U & \text{on } \partial D, \\ \text{arbitrary} & \text{in } D \end{cases} \tag{3.13}$$

such that $\overline{U} \in L^2(D)$.

In order to find estimates (3.11) and (3.12), we take the inner product with any $v \in H_0^1(D)$, which leads to

$$I := \left| \int A\nabla v \cdot \nabla v \right| = \left| -\int A\nabla \overline{U} \cdot \nabla v - \int 2\eta \sinh(\beta(\overline{U} + v))v \right|,$$

where $A = e^{\theta}$. Using $\eta = e^{\theta}$ and $\sinh(\beta(\overline{U}+v)) = (e^{\beta(\overline{U}+v)} - e^{-\beta(\overline{U}+v)})/2$ as well as the triangle inequality, we find

$$e^{-\|\theta\|_{L^\infty}}\|\nabla v\|_{L^2}^2 \leq I \leq \left|\int e^{\theta}\nabla\overline{U}\cdot\nabla v\right| + \left|\int e^{\theta}e^{\beta(\overline{U}+v)}v\right| + \left|\int e^{\theta}e^{-\beta(\overline{U}+v)}v\right|.$$

Using the Cauchy-Schwarz and Poincaré inequalities and $\kappa \leq u = \overline{U} + v \leq \lambda$, which is a pointwise estimate [6] for the solution of the Poisson-Boltzmann equation, we can write

$$e^{-\|\theta\|_{L^\infty}}\|\nabla v\|_{L^2}^2 \leq I \leq e^{\|\theta\|_{L^\infty}}(\|\nabla\overline{U}\|_{L^2} + C_p e^{|\beta\lambda|} + C_p e^{|\beta\kappa|}\|\nabla v\|_{L^2}),$$

where C_p is a Poincaré constant and κ and λ are constants. Therefore, we calculate

$$\begin{aligned}\|\nabla u\|_{L^2} &\leq \|\nabla v\|_{L^2} + \|\nabla\overline{U}\|_{L^2} \\ &\leq e^{2\|\theta\|_{L^\infty}}(\|\nabla\overline{U}\|_{L^2} + C_p e^{|\beta\lambda|} + C_p e^{|\beta\kappa|}) + \|\nabla\overline{U}\|_{L^2} \\ &= (1+e^{2\|\theta\|_{L^\infty}})\|\nabla\overline{U}\|_{L^2} + C_p e^{2\|\theta\|_{L^\infty}}(e^{|\beta\lambda|} + e^{|\beta\kappa|}) \\ &\leq 2C_p e^{2\|\theta\|_{L^\infty}}(\|\nabla\overline{U}\|_{L^2} + e^{|\beta\lambda|} + e^{|\beta\kappa|}) \\ &= H_0 e^{2\|\theta\|_{L^\infty}},\end{aligned} \tag{3.14}$$

where $H_0 := 2C_p(\|\nabla\overline{U}\|_{L^2} + e^{|\beta\lambda|} + e^{|\beta\kappa|})$. Using the definition of H^1-norm and inequality (3.14), we can write

$$\frac{1}{(1+C_p^2)^{1/2}}\|u\|_{H^1} \leq \|\nabla u\|_{L^2} \leq H_0 e^{2\|\theta\|_{L^\infty}},$$

where C_p is a Poincaré constant. This leads to

$$\|u\|_{H^1} \leq H e^{2\|\theta\|_{L^\infty}}, \tag{3.15}$$

where $H = H_0(1+C_p^2)^{1/2}$.

To prove inequality (3.12), we assume that u_1 and u_2 satisfy (3.10) with coefficients $A_1 = e^{\theta_1} = \eta_1$ and $A_2 = e^{\theta_2} = \eta_2$. Hence, subtracting the term $\nabla\cdot(e^{\theta_1}\nabla u_2)$, the difference $u_2 - u_1$ satisfies the equation

$$\nabla\cdot(e^{\theta_1}\nabla(u_2-u_1)) = \nabla\cdot((e^{\theta_1}-e^{\theta_2})\nabla u_2) + e^{\theta_2}(e^{\beta u_2} - e^{-\beta u_2}) - e^{\theta_1}(e^{\beta u_1} - e^{-\beta u_1}).$$

Taking the inner product of this equation with $u_2 - u_1$ leads to

$$\begin{aligned}I :&= \left|\int e^{\theta_1}\nabla(u_2-u_1)\cdot\nabla(u_2-u_1)\right| \\ &= \left|\int (e^{\theta_1}-e^{\theta_2})\nabla u_2\cdot\nabla(u_2-u_1) + (e^{\theta_2}(e^{\beta u_2}-e^{-\beta u_2}) - e^{\theta_1}(e^{\beta u_1}-e^{-\beta u_1}))(u_2-u_1)\right|,\end{aligned}$$

where the inequalities

$$\begin{aligned}e^{-\|\theta_1\|_{L^\infty}}\|\nabla(u_2-u_1)\|_{L^2}^2 \leq I \leq &\left|\int (e^{\theta_1}-e^{\theta_2})\nabla u_2\cdot\nabla(u_2-u_1)\right| \\ &+ \left|\int (e^{\theta_2}-e^{\theta_1})(e^{\beta\max(\lambda_1,\lambda_2)} - e^{-\beta\max(\lambda_1,\lambda_2)})(u_2-u_1)\right|\end{aligned}$$

hold because of the triangle inequality and the pointwise estimates $\kappa_i \leq u_i \leq \lambda_i$, $i \in \{1,2\}$, for the solution of the nonlinear Poisson-Boltzmann equation [6], where κ_i and λ_i are constants.

Now we use the Cauchy-Schwarz and Poincaré inequalities to find

$$e^{-\|\theta_1\|_{L^\infty}}\|\nabla(u_2-u_1)\|_{L^2} \leq I \leq \|e^{\theta_1}-e^{\theta_2}\|_{L^\infty}(\|\nabla u_2\|_{L^2} + C_p|e^{\beta\max(\lambda_1,\lambda_2)} - e^{-\beta\max(\lambda_1,\lambda_2)}|),$$

where C_p is a Poincaré constant. Since $e^{\theta(x)}$ $(x \in D)$ is continuously differentiable, it is Lipschitz continuous by the Weierstrass Theorem. Thus we have

$$\|e^{\theta_2} - e^{\theta_1}\|_{L^\infty} \leq \|\theta_1 - \theta_2\|_{L^\infty} e^{\max(\|\theta_1\|_{L^\infty},\|\theta_2\|_{L^\infty})}. \tag{3.16}$$

Using inequalities (3.14) and (3.16), we obtain

$$\begin{aligned} e^{-\|\theta_1\|_{L^\infty}}\|\nabla(u_2-u_1)\|_{L^2} &\leq I \\ &\leq \|\theta_1-\theta_2\|_{L^\infty} e^{\max(\|\theta_1\|_\infty,\|\theta_2\|_\infty)}\Big[e^{2\|\theta_2\|_\infty}\big(\|\nabla\overline{g}\|_{L^2} \\ &\quad + C_p e^{|\beta\lambda_2|} + C_p e^{|\beta\kappa_2|}\big) + C_p|e^{\beta\max(\lambda_1,\lambda_2)} - e^{-\beta\max(\lambda_1,\lambda_2)}|\Big], \end{aligned}$$

which leads to

$$\begin{aligned} \|\nabla(u_2-u_1)\|_{L^2} &\leq \|\theta_1-\theta_2\|_{L^\infty} e^{4\max(\|\theta_1\|_\infty,\|\theta_2\|_\infty)}\Big[\|\nabla\overline{g}\|_{L^2} \\ &\quad + C_p e^{|\beta\lambda_2|} + C_p e^{|\beta\kappa_2|} + C_p|e^{\beta\max(\lambda_1,\lambda_2)} - e^{-\beta\max(\lambda_1,\lambda_2)}|\Big] \\ &= S_0\|\theta_1-\theta_2\|_{L^\infty} e^{4\max(\|\theta_1\|_\infty,\|\theta_2\|_\infty)}, \end{aligned}$$

where $S_0 := \|\nabla\overline{g}\|_{L^2} + C_p e^{|\beta\lambda_2|} + C_p e^{|\beta\kappa_2|} + C_p|e^{\beta\max(\lambda_1,\lambda_2)} - e^{-\beta\max(\lambda_1,\lambda_2)}|$. Now, we can write

$$\begin{aligned} \|u_1-u_2\|_{H^1} &\leq (1+C_p^2)^{1/2}\|\nabla(u_1-u_2)\|_{L^2} \\ &\leq (1+C_p^2)^{1/2}\Big[\|\nabla\overline{g}\|_{L^2} + C_p e^{|\beta\lambda_2|} + C_p e^{|\beta\kappa_2|} \\ &\quad + C_p|e^{\beta\max(\lambda_1,\lambda_2)} - e^{-\beta\max(\lambda_1,\lambda_2)}|\Big]\, e^{4\max(\|\theta_1\|_\infty,\|\theta_2\|_\infty)}\|\theta_1-\theta_2\|_{L^\infty} \\ &= Se^{4\max(\|\theta_1\|_\infty,\|\theta_2\|_\infty)}\|\theta_1-\theta_2\|_{L^\infty}, \end{aligned}$$

where $S := S_0(1+C_p^2)^{1/2}$, which completes the proof for the nonlinear equation (3.10). □

Remark 3. *The quantity $\|\nabla\overline{U}\|_{L^2(D)}$ is non-zero even if the Dirichlet datum U is a constant. The new variable $v := u - \overline{U}$ is defined such that $\overline{U} = U$ on ∂D and $\overline{U}$ is arbitrary in D. Therefore in the (realistic) case of non-constant U, $\|\nabla\overline{U}\|_{L^2(D)}$ is non-zero in D. Hence, even if $f = 0$, the quantities F, H, and S are non-zero.*

We summarize the above results for nonlinear elliptic inverse problems in the following theorem:

Theorem 6. *Assume $g(\theta) = u$ is the observation operator representing the solution u of the real-valued nonlinear equation* (3.10) *in the bounded domain $D \subset \mathbb{R}^n$, $n \in \{2, 3\}$, with a smooth boundary ∂D and $A = \exp(\theta) = \eta$ with $\theta \in L^\infty(D)$. Then the estimates*

$$|g(\theta)| \leq H \exp\left(2\|\theta\|_{L^\infty(D)}\right) \tag{3.17}$$

and

$$|g(\theta_1) - g(\theta_2)| \leq S \exp\left(4 \max\{\|\theta_1\|_{L^\infty(D)}, \|\theta_2\|_{L^\infty(D)}\}\right) \|\theta_1 - \theta_2\|_{L^\infty(D)} \tag{3.18}$$

hold, where $H = H(\|\nabla \overline{U}\|_{L^2})$ and $S = S(\|\nabla \overline{U}\|_{L^2})$, which have been defined in Proposition 1.

The estimates (3.17) and (3.18) for real and imaginary parts of the solution yield estimates for the complex-valued equation (3.10) as well. Therefore, Assumptions 4 are satisfied for the general EIT model equations (3.10), where U is constant at each contact at the surface of the device. This will lead us to the one of the main results of this chapter, i.e., well-definedness and well-posedness of the Bayesian inversion problem for linear and nonlinear elliptic problems including the EIT inverse problem.

The posterior probability measure μ^y defined by (2.10) is well-defined if we show that the measure is normalizable. For well-posedness, the continuity of the posterior measure in the Hellinger metric with respect to the data must be shown. We state the two main theorems resulting from Proposition 1. The reader is referred to [8] for the proofs.

Theorem 7 (Well-definedness of the posterior measure [8, Theorem 4.1])**.** *Let g satisfy Assumptions 4 and assume that the prior measure μ_0 is a Gaussian measure satisfying $\mu_0(X) = 1$. Then the posterior measure μ^y given by* (2.10) *is a well-defined probability measure.*

The following theorem states the well-posedness for inverse problems by showing Lipschitz continuity of the posterior measure in the Hellinger metric with respect to changes in the data.

Theorem 8 (Well-posedness of the Bayesian inverse problem [8, Theorem 4.2])**.** *Let g satisfy Assumptions 4. Assume also that the prior measure μ_0 is a Gaussian measure satisfying $\mu_0(X) = 1$ and that the measure is absolutely continuous, $\mu^y \ll \mu_0$, with its Radon-Nikodym derivative given by* (2.10) *for each $y \in Y$.*

Then the posterior measure μ^y is Lipschitz continuous in the data y with respect to the Hellinger distance, i.e., if μ^y and $\mu^{y'}$ are two measures corresponding to data y and y', then there exists $\alpha = \alpha(r) > 0$ such that the inequality

$$d_{\text{Hell}}(\mu^y, \mu^{y'}) \leq \alpha \|y - y'\|_Y$$

holds for all y and y' with $\max\{\|y\|_Y, \|y'\|_Y\} < r$.

3.5 Numerical Results

In the context of Bayesian inversion theory [3, 7, 10, 11], there are two approaches to optimization of Bayesian experimental design: optimizing posterior covariance functionals with respect to the design/hyper parameters [12], and optimizing the Bayesian expected information gain (EIG) [13, 14, 15, 16].

The first approach is based on the notion of distinguishability of the measurements. In the distinguishability approach, the optimality criterion is defined by maximizing the difference between measurements corresponding to two different predetermined parameter distributions and minimizing the volume of the posterior density. The second approach to optimize Bayesian inversion is to maximize the information about model uncertainty. In this approach, the expected information gain is taken as the design criterion into account and the optimal experiments are found from a continuously parametrized design space by inferring the model parameters from noisy data/observations in a Bayesian setting.

As the expected information gain has no closed form, it should be approximated numerically. The numerical evaluation of the design criterion is computationally expensive since it involves the calculation of nested integrals on one hand, and searching the whole design space to find the maxima on the other hand. This evaluation is even harder and in some cases infeasible when, for instance, a higher-dimensional parameter or design space or an expensive forward solver is involved. Here, we propose a double-loop Monte-Carlo approximation, introduced in Section 2.5, for the evaluation of the expected information gain of the EIT model parameters in a finite-dimensional case.

We consider the following model:

$$Y_\xi = g_\xi(\theta) + \epsilon, \tag{3.19}$$

where g_ξ is our forward PDE model for EIT which is in terms of unknown conductivity of muscle θ, Y_ξ is a 64-dimensional response vector which are voltages measured in electrodes. Here, ϵ is a Gaussian noise with the mean 0 and covariance matrix Σ_ϵ. The Kullback-Leibler divergence is as follows,

$$D_{KL} = \int_\Theta p(\theta|Y_\xi) \log\Big(\frac{p(\theta|Y_\xi)}{p(\theta)}\Big) \mathrm{d}\theta,$$

and the expected information gain is calculated by

$$\mathrm{EIG}_\theta(\xi) = \mathbb{E}[D_{KL}] = \int_\mathcal{Y} \int_\Theta p(\theta|Y_\xi) \log\Big(\frac{p(\theta|Y_\xi)}{p(\theta)}\Big) \mathrm{d}\theta \, p(Y_\xi) \mathrm{d}Y_\xi.$$

The Double-Loop Monte Carlo method which was introduced in Section 2.5 is used to calculate the expected information gain $\mathrm{EIG}_\theta(\xi)$ for a given design ξ.

The goal of optimal Bayesian experimental design (OBED) is to maximize the expected information gain over the design space $\mathcal{D}$ to find the optimal experimental design parameter(s)

$$\xi^* = \arg\max_{\xi \in \mathcal{D}} \mathrm{EIG}_\theta(\xi)$$

by providing the most informative observations.

3.5.1 Quantification of the Design Criterion

The goal here is to find optimal experimental designs including optimal frequency, and optimal electrode placement in EIT model. In this section, we implement the DLMC method (see Algorithm 2 in Section 2.5) to compute the expected information gain for the example of a cross-section of a right leg as a real-world application of EIT. The goal is to extract the electrical conductivity of muscle inside the right leg using the optimal Bayesian inversion mechanism using the nonlinear Poisson-Boltzmann equation as the EIT forward model.

In Figure 3.3 (top), the calculated information gain corresponding to the 8 injection patterns used in Figure 3.2 are illustrated. According to this result, the information that we can obtain from different EIT design patterns is not the same. Figure 3.3 (bottom) displays the information gain for different numbers of samples used in Bayesian inversion process. According to this figure, there is no significant difference between the information gain from 10k samples and more than 10k in our EIT problem. That makes it possible to use the lower value 10k as the number of samples in our calculations. This prior knowledge regarding an optimal number of samples will lead us to efficient and low-cost calculations without losing (much) information.

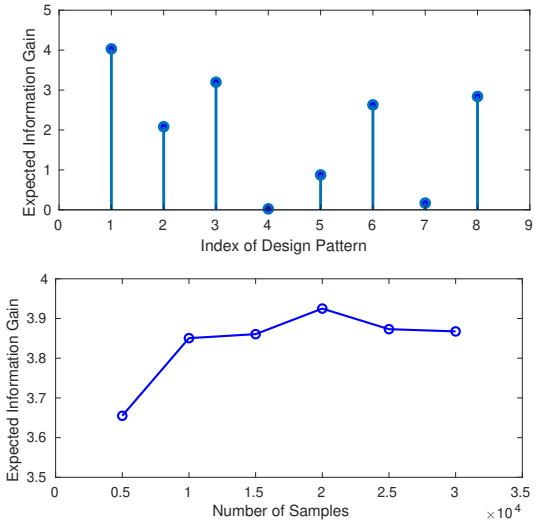

Figure 3.3: The expected information gain for different design patterns used in Figure 3.2 in order to extract the conductivity of the muscle as a parameter of interest (top). The expected information gain vs. different numbers of samples (bottom).

3.5.2 Stochastic Optimization

As mentioned, an optimal experimental design aims to maximize the worth of information of an experiment for statistical inference. This optimal process is of importance, especially for experiments that are long or expensive to perform. Hence, in order to solve the optimization problem of Bayesian experimental design (3.5), iterative search strategies and in particular stochastic optimization methods [17, 18, 19, 20] instead of deterministic ones for high-

dimensional problems are proposed, since the objective function is noisy. Here, we develop the simultaneous perturbation stochastic approximation (SPSA) algorithm [21, 22] for the stochastic optimization of the noisy EIG, which is approximated by the double-loop Monte-Carlo method.

SPSA is a gradient-based optimization algorithm in the stochastic setting, which is used when only measurements of the objective function are available but no gradient information. The SPSA algorithm can be written in the general recursive stochastic approximation (SA) form

$$\xi_{k+1} = \xi_k - a_k \mathbf{g}_k(\xi_k), \tag{3.20}$$

where k is the iteration number and $\mathbf{g}_k(\xi_k)$ is the simultaneous perturbation estimate of the gradient, which is approximated by finite differences using only two random perturbations, i.e.,

$$\mathbf{g}_k(\xi_k) = \frac{y(\xi_k + c_k\mathbf{\Delta}_k) - y(\xi_k - c_k\mathbf{\Delta}_k)}{2c_k}[\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, \ldots, \Delta_{kp}^{-1}]^T, \tag{3.21}$$

where $\mathbf{\Delta}_k = [\Delta_{k1}, \Delta_{k2}, \ldots, \Delta_{kp}]^T$ is the mean-zero p-dimensional random perturbation vector with i.i.d. entries drawing from a user-specified distribution satisfying certain conditions [17, Sections 7.3 and 7.4], and y is the objective function. Furthermore a_k is a non-negative scalar gain coefficient and c_k is a positive scalar, which are defined by

$$a_k = \frac{a}{(C+k+1)^\alpha} \quad \text{and} \quad c_k = \frac{c}{(k+1)^\gamma},$$

where a, C, α, c and γ are parameters of the algorithm and their values can be found in [21]. The advantage of this stochastic approximation method is that in order to approximate the gradient only two function evaluations per step are required, while in a full finite difference scheme $2p$ evaluations are needed.

Algorithm 3 The optimal Bayesian inversion algorithm

Choose $\xi_1 \in \mathcal{D}$ as the initial vector of the EIT design parameters.
for $k = 1 : N_{\text{iter}}$ **do**
1. Calculate the expected information gain for the design parameter ξ_k using the DLMC approximation (Algorithm 2).
2. Terminate the algorithm if the number of iterations has reached N_{iter}, express ξ_k as the optimal design.
3. Update ξ_k to a new value ξ_{k+1} using the recursive stochastic approximation form (see Section 3.5.2):
$$\xi_{k+1} = \xi_k - a_k \mathbf{g}_k(\xi_k).$$
4. Replace ξ_k by ξ_{k+1}.
end for

3.5.3 Optimal Frequency for Electrical Impedance Tomography Measurements

In this section, numerical results of optimal Bayesian inversion for the estimation of optimal frequency for the EIT inverse problem, namely the extraction of the conductivity of a

human leg muscle, are presented. Variation in the applied frequency as an EIT experimental design results in different Bayesian analysis. In Figure 3.4, the posterior distribution of the generated Markov-chains for four different frequencies are illustrated. According to these results, the chains obtained for different frequencies have different characteristics and contain different amounts of information. These results motivated us to calculate the

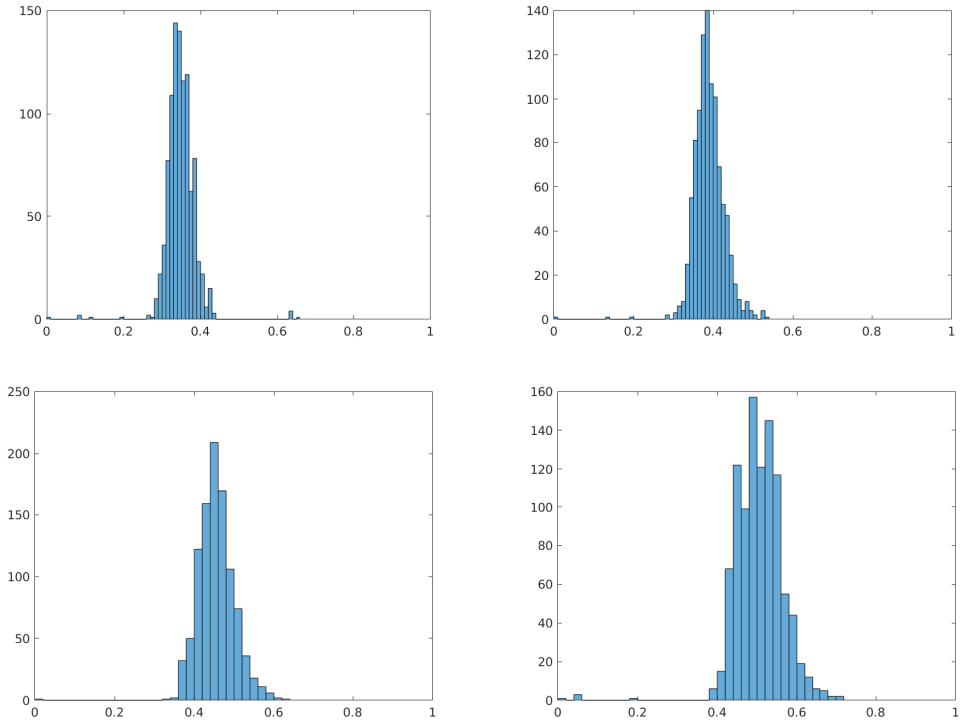

Figure 3.4: Posterior distributions of the muscle conductivity for frequencies $10, 200, 500$ and 900 kHz.

expected information gain of each frequency in the range from 1 kHz to 1 MHz, which was very costly and time-consuming. The results are displayed in Figure 3.5.

To overcome the problem of cost and complexity of calculations in order to reach a real-time experimental design in the EIT problem on one hand, and according to the noisy objective function (see Figure 3.5) on the other hand, we employed an stochastic optimization method, namely the simultaneous perturbation stochastic approximation (SPSA), which was explained in Section 3.5.2. This method is applied to the problem of maximization of the expected information gain in order to obtain optimal and more informative experimental designs. According to the numerical results, the optimal frequency for the described EIT problem is estimated as 162 kHz. Table 3.1 includes the results for different frequencies.

In the optimal frequency, the conductivity of the muscle as the mean of posterior chain is estimated more accurately and the distribution of the posterior has the highest expected information gain compared to other non-optimal frequencies. Figure 3.6 shows the posterior

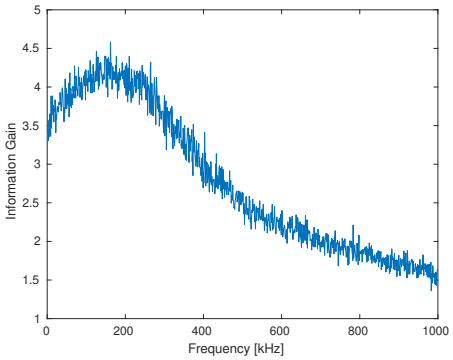

Figure 3.5: The calculated expected information gains for different frequencies.

Frequency kHz	Conductivity (True value)	Conductivity (Estimated)	Error of Estimation	Expected Information Gain
10	0.3410	0.3500	0.0090	3.7819
162 (optimal)	0.3754	0.3811	0.0057	4.4557
200	0.3840	0.3905	0.0065	4.1200
500	0.4460	0.4535	0.0075	2.4660
900	0.4950	0.5069	0.0119	1.7007

Table 3.1: Estimated mean values of muscle conductivity for various frequencies using Bayesian inference and corresponding calculated values of expected information gain.

distribution and trajectory of the parameter of interest at the optimal frequency 162 kHz in our EIT problem.

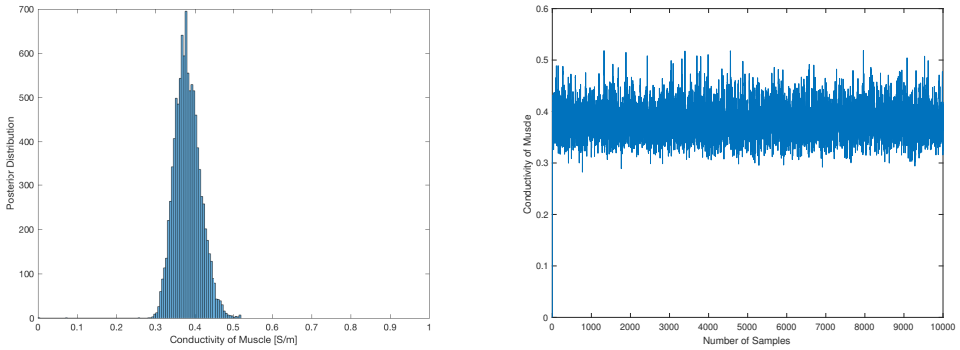

Figure 3.6: The posterior distribution (left) and trace plot (right) of muscle conductivity for the obtained optimal frequency 162 kHz.

3.5.4 Optimal Electrode Configuration for Electrical Impedance Tomography Measurements

Due to the nonlinearity and ill-posedness of the EIT inverse problem, any small variation in the electrode configuration leads to a big change in the gained information in Bayesian extraction process. In this section, we aim to find the optimal electrode configuration including the optimal placement of injection and measurement electrodes as well as the optimal measurement strategies using the optimal Bayesian inversion methodology that we proposed in Section 3.5.2 and Algorithm 3. In Figure 3.7, various electrodes placement are displayed. The calculated information gain for the extraction of the muscle conductivity and mean of the estimated posterior chain for the unknown quantity in each case are indicated in the figures as well. The numerical experiments confirm the sensitivity of the EIT inverse problem on the electrode configuration and measurement strategy.

Figure 3.8 depicts the mean of the posterior chains generated in the optimization process versus the corresponding expected information gains. This figure illustrates the convergence of the means to the true value of the muscle conductivity, i.e., 0.3754, while reaching the maximum information gain using the optimal electrode configurations chosen by the optimization algorithm.

Clustering the Electrode Configurations

Since the objective function is noisy, it is hard to decide exactly which configuration is eventually the optimal one. According to various injection patterns, we need to use two and even more (near) optimal electrode configurations for the process of the measurement and image reconstruction. Therefore, we split the obtained configurations into the optimal and non-optimal ones and to this end we apply data clustering methods. The K-means algorithm is one of the most preferred clustering tools in data mining for many real world

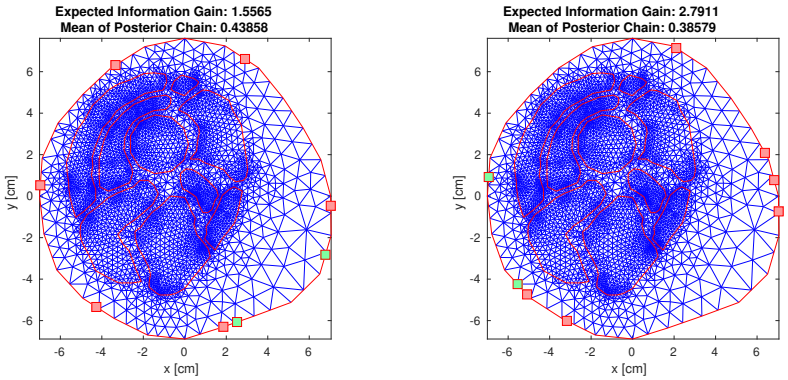

Figure 3.7: Various electrode positioning together with the corresponding expected information gains and means of the estimated posterior chains are illustrated. Green electrodes are the injection pairs and red ones are measurement electrodes.

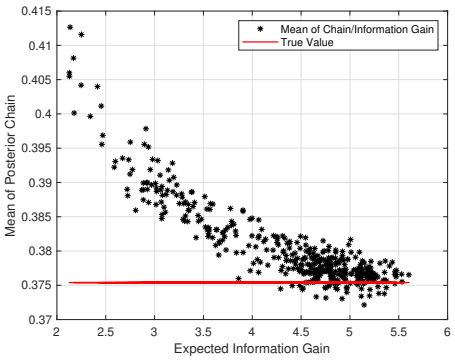

Figure 3.8: Mean of the posterior chains in the optimization process versus the corresponding expected information gains.

applications [23, 24]. Although this clustering method is simple to understand and deploy, it suffers from some limitations; due to the use of the Euclidean distance, K-means assumes that the data space is isotropic in the sense that distances are unchanged by translations and rotations and therefore each cluster consists of the data points lying within a sphere around the centroid. One of the limitations of K-means is that this method implicitly assumes all clusters have the same radius while sometimes they are very clearly identifiable. Additionally, K-means implicitly assumes each cluster has the same volume in data space in the sense that each cluster contains the same number of data points, while there might be significant differences in clusters density. Another limitation is that small changes in the data result in small changes to the position of the cluster centroid, which is assumed as average of the data points in a cluster. Therefore, outliers can significantly impair the results of K-means. Moreover, in the method of K-means it is assumed that the number K of clusters is given, while this is rarely the case in practice. These limitations lead to inaccurate conclusions about the structure in the data.

To overcome these limitations, we apply maximum a-posteriori Dirichlet process mixtures (MAP-DP) algorithm [25]. This method is statistically rigorous and it is based on nonparametric Bayesian Dirichlet process mixture modeling, which is a probabilistic model and involves fitting a probability density model to the data, while K-means is essentially geometric. This method for instance estimates the number of necessary clusters for given data points using Bayesian nonparametric (BNP) models [26] as well as it separates outliers from the data efficiently and thus the centroid is estimated accurately. Figure 3.9 demonstrates the optimal and non-optimal electrode configuration clusters using the two methods, i.e., K-means and MAP-DP algorithms. This figure clearly illustrates the efficiency of the MAP-DP compared to the K-means by more accurately identifying the optimal configurations.

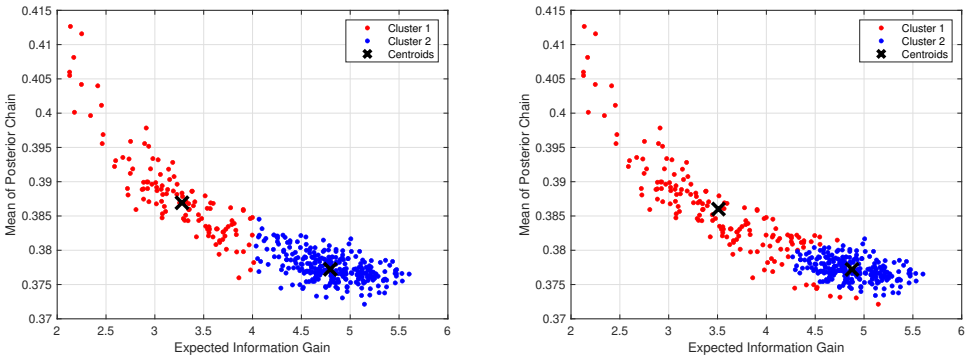

Figure 3.9: Optimal (blue) and non-optimal (red) electrode configuration clusters using K-means clustering method (left) and MAP-DP method (right).

In Figure 3.10, the standard uniform electrode configuration versus two optimal design patterns with high expected information gains are displayed. The optimal patterns are proposed by the applied optimization method on Bayesian results of the EIT inverse problem and they belong to the optimal cluster. Due to these results, injection from the top-left

of the computational domain with measurement electrodes which are close together leads to a larger information gain and more accurate estimation of the muscle conductivity, whose true value is 0.3754, while uniformly distributed electrodes result in a moderate information gain and an inaccurate estimation of the unknown quantity compared to the optimal configurations. All experiments are performed in the optimal frequency of 162 kHz, which was calculated in subsection 3.5.3.

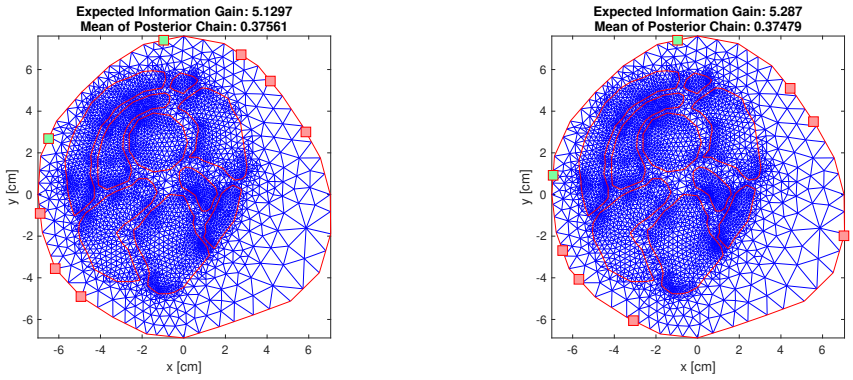

Figure 3.10: Standard uniform (top) electrode configuration versus optimal (bottom) configurations with the highest expected information gains and estimations of the muscle conductivity closest to its true value.

The results of various electrode configurations shown in Figure 3.7 and 3.10 including the corresponding expected information gain and the estimated value for the muscle conductivity using Bayesian inversion are summarized in Table 3.2. The comparison shows the efficiency of the presented optimal Bayesian inversion for the EIT inverse problem.

Electrode configuration	Expected information gain	Estimated muscle conductivity
non-optimal I (Figure 3.7-left)	1.5565	0.4386
non-optimal II (Figure 3.7-right)	2.7911	0.3858
uniform (Figure 3.10-top)	3.8077	0.3831
optimal I (Figure 3.10-bottom left)	5.1297	0.3756
optimal II (Figure 3.10-bottom right)	5.2870	0.3748

Table 3.2: Estimated expected information gains and mean of Markov chains for the muscle conductivity (true value: 0.3754) for various electrode configurations.

3.6 Conclusions

We developed an optimal algorithm for Bayesian inversion for the EIT problem, which has applications in medical imaging. The new inversion methodology presents an optimal EIT experimental design recipe including the optimal frequency and electrode configuration by maximizing the expected information gain in the inversion process, which results in accurate

estimation of EIT model unknowns and high-resolution medical images. To this end, we developed a double-loop Monte-Carlo method and simultaneous perturbation stochastic approximation for the EIT inverse problem. The comparison of numerical results for various electrode configurations indicates the efficiency of the proposed optimal Bayesian inversion methodology for the EIT problem. It is valuable to mention that the optimal sensor placement would be affected by the object to be imaged and updates to the configuration would be needed as the object is imaged. In clinical practice, electrode configurations should be implemented by more advanced EIT devices.

Bibliography

[1] A. Karimi, L. Taghizadeh, C. Heitzinger, Optimal Bayesian experimental design for electrical impedance tomography in medical imaging, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113489.

[2] L. Taghizadeh, A. Karimi, B. Stadlbauer, W. J. Weninger, E. Kaniusas, C. Heitzinger, Bayesian inversion for electrical-impedance tomography in medical imaging using the nonlinear Poisson–Boltzmann equation, Computer Methods in Applied Mechanics and Engineering 365 (2020) 112959.

[3] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer Science & Business Media, 2006.

[4] A. Adler, W. R. Lionheart, Uses and abuses of EIDORS: an extensible software base for EIT, Physiological Measurement 27 (5) (2006) S25.

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

[6] S. Baumgartner, C. Heitzinger, Existence and local uniqueness for 3D self-consistent multiscale models for field-effect sensors, Commun. Math. Sci. 10 (2) (2012) 693–716.

[7] R. C. Smith, Uncertainty Quantification: Theory, Implementation, and Applications, Vol. 12, SIAM, 2013.

[8] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451–559.

[9] M. Dashti, A. M. Stuart, Uncertainty quantification and weak approximation of an elliptic inverse problem, SIAM Journal on Numerical Analysis 49 (6) (2011) 2524–2542.

[10] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013.

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

[12] J. Kaipio, A. Seppänen, E. Somersalo, H. Haario, Posterior covariance related optimal current patterns in electrical impedance tomography, Inverse Problems 20 (3) (2004) 919.

[13] P. Tsilifis, R. G. Ghanem, P. Hajali, Efficient Bayesian experimentation using an expected information gain lower bound, SIAM/ASA Journal on Uncertainty Quantification 5 (1) (2017) 30–62.

[14] Q. Long, M. Scavino, R. Tempone, S. Wang, Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, Computer Methods in Applied Mechanics and Engineering 259 (2013) 24–39.

[15] D. Smyl, D. Liu, Optimizing electrode positions in 2d electrical impedance tomography using deep learning, IEEE Transactions on Instrumentation and Measurement (2020).

[16] N. Hyvonen, A. Seppänen, S. Staboulis, Optimizing electrode positions in electrical impedance tomography, SIAM Journal on Applied Mathematics 74 (6) (2014) 1831–1851.

[17] J. C. Spall, Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, Vol. 65, John Wiley & Sons, 2005.

[18] J. C. Spall, et al., Multivariate stochastic approximation using a simultaneous perturbation gradient approximation, IEEE Transactions on Automatic Control 37 (3) (1992) 332–341.

[19] X. Huan, Y. Marzouk, Gradient-based stochastic optimization methods in Bayesian experimental design, International Journal for Uncertainty Quantification 4 (6) (2014).

[20] X. Huan, Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, Journal of Computational Physics 232 (1) (2013) 288–317.

[21] J. C. Spall, An overview of the simultaneous perturbation method for efficient optimization, Johns Hopkins APL Technical Digest 19 (4) (1998) 482–492.

[22] J. C. Spall, Implementation of the simultaneous perturbation algorithm for stochastic optimization, IEEE Transactions on Aerospace and Electronic Systems 34 (3) (1998) 817–823.

[23] P. Berkhin, A survey of clustering data mining techniques, In Grouping Multidimensional Data: Recent Advances in Clustering (2006) 25–71.

[24] S. Lloyd, Least squares quantization in PCM, IEEE Transactions on Information Theory 28 (2) (1982) 129–137.

[25] Y. P. Raykov, A. Boukouvalas, F. Baig, M. A. Little, What to do when k-means clustering fails: a simple yet principled alternative algorithm, PloS ONE 11 (9) (2016) 1–28.

[26] N. L. Hjort, C. Holmes, P. Müller, S. G. Walker, Bayesian Nonparametrics, Vol. 28, Cambridge University Press, 2010.

4 Laplace Approximation for the Electrical Impedance Tomography Inverse Problem

Electrical impedance tomography (EIT) is the problem of determining the electrical conductivity distribution of an unknown medium using surface electrode measurements attached to the boundary of the object. The EIT inverse problem for image reconstruction is a nonlinear and severely ill-posed problem [1, 2, 3]. Bayesian inversion [4, 5, 6, 7] is an approach to tackle ill-posed and/or ill-conditioned inverse problems. This approach provides a solution that quantifies uncertainty by assigning a probability to each possible value of the unknown parameter and also incorporates prior information and beliefs about the parameter [8].

In this chapter, we specifically focus on the Laplace approximation approach for the EIT inverse problem for Bayesian estimation of the posterior distribution as well as to estimate the expected information gain (EIG) for a given design of experiment. The goal of optimal experimental design is to find the best design under which the most informative measurement data can be obtained. There are two approaches for the optimization of the Bayesian experimental design: optimizing posterior covariance functionals with respect to the design/hyper parameters [9], and optimizing the Bayesian expected information gain (EIG) [10, 11, 12, 13].

4.1 Electrical Impedance Tomography Model and the Forward Solution

Assume that our EIT problem domain D is surrounded by some electrodes which are attached to its surface ∂D. Here, we consider a linearized version of the Complete Electrode Model (CEM) of nonlinear elliptic PDE (3.10) for EIT [14, 15], which is defined by

$$-\nabla \cdot (A(x)\nabla u(x)) + 2\eta(x)u(x) = 0 \qquad \text{in} \quad D, \tag{4.1a}$$

$$\int_{e_\ell} A\frac{\partial u}{\partial \mathbf{n}} ds = I_\ell, \qquad \text{on} \quad e_\ell, \quad \ell = 1, 2, \dots L \tag{4.1b}$$

$$u + c_\ell A\frac{\partial u}{\partial \mathbf{n}} = U_\ell \qquad \text{on} \quad e_\ell, \quad \ell = 1, 2, \dots L \tag{4.1c}$$

$$\frac{\partial u}{\partial \mathbf{n}} = 0 \qquad \text{on} \quad \partial D \backslash \cup_{\ell=1}^{L} e_\ell. \tag{4.1d}$$

This model is a combination of the continuum model and the shunt model we used in Chapter 3. The Robin boundary condition $u + c_\ell A\frac{\partial u}{\partial \mathbf{n}} = U_\ell$ considers the formation of a highly resistive, thin layer at the interface between the surface ∂D and the electrode's surface e_ℓ. In this model, c_ℓ is involved to account this resistance of flow into the body D in the measured potential U_{D_ℓ}.

Let V_h be a finite-dimensional subspace of $H^1(D)$ with continuous piecewise linear basis functions $\{\varphi_j\}_{j=1}^N$ corresponding to the nodal points $\{x_j\}_{j=1}^N$ such that $\varphi_j(x_i) = \delta_{ij}$ for $i, j \in \{1, 2, \ldots, N\}$. To present finite element solution of the PDE, we obtain weak formulation of (4.1a) as

$$\int_D A\nabla u \cdot \nabla v \, dx - \int_{\partial D} A\frac{\partial u}{\partial \mathbf{n}} v \, ds + \int_D 2\eta uv \, dx = 0, \tag{4.2}$$

where $v \in H^1(D)$ is a test function. Note that

$$\int_{\partial D} A\frac{\partial u}{\partial \mathbf{n}} v \, ds = \int_{\cup_{\ell=1}^L e_\ell} A\frac{\partial u}{\partial \mathbf{n}} v \, ds + \int_{\partial D \setminus \cup_{\ell=1}^L e_\ell} A\frac{\partial u}{\partial \mathbf{n}} v \, ds = \sum_{\ell=1}^L \int_{e_\ell} A\frac{\partial u}{\partial \mathbf{n}} v \, ds,$$

because $\int A\frac{\partial u}{\partial \mathbf{n}} v \, ds = 0$ on Neumann boundary. Therefore, weak formulation (4.2) turns to

$$\int_D A\nabla u \cdot \nabla v \, dx - \sum_{\ell=1}^L \int_{e_\ell} A\frac{\partial u}{\partial \mathbf{n}} v \, ds + \int_D 2\eta uv \, dx = 0. \tag{4.3}$$

Now for Robin boundary condition

$$\begin{aligned}
& u + c_\ell A\frac{\partial u}{\partial \mathbf{n}} = U_\ell, \\
\Rightarrow \quad & A\frac{\partial u}{\partial \mathbf{n}} = \frac{1}{c_\ell}(U_\ell - u), \\
\Rightarrow \quad & \int_{e_\ell} (U_\ell - u) V_\ell \, ds = \int_{e_\ell} c_\ell A\frac{\partial u}{\partial \mathbf{n}} V_\ell \, ds, \\
& \qquad\qquad = c_\ell V_\ell \int_{e_\ell} A\frac{\partial u}{\partial \mathbf{n}} \, ds \\
& \qquad\qquad = c_\ell V_\ell I_\ell, \\
\Rightarrow \quad & \sum_{\ell=1}^L \frac{1}{c_\ell} \int_{e_\ell} (U_\ell - u) V_\ell \, ds = \sum_{\ell=1}^L V_\ell I_\ell.
\end{aligned} \tag{4.4}$$

where $V_\ell = v|_{e_\ell}$. Therefore,

$$\sum_{\ell=1}^L \frac{1}{c_\ell} \int_{e_\ell} (U_\ell - u) V_\ell \, ds - \sum_{\ell=1}^L V_\ell I_\ell = 0. \tag{4.5}$$

Substituting (4.4) in (4.3) yields

$$\int_D A\nabla u \cdot \nabla v \, dx + \int_D 2\eta uv \, dx - \sum_{\ell=1}^L \frac{1}{c_\ell} \int_{e_\ell} (U_\ell - u) v \, ds = 0,$$

which is indeed

$$\int_D A\nabla u \cdot \nabla v \, dx + \int_D 2\eta uv \, dx + \sum_{\ell=1}^L \frac{1}{c_\ell} \int_{e_\ell} uv \, ds = \sum_{\ell=1}^L \frac{1}{c_\ell} \int_{e_\ell} U_\ell v \, ds, \tag{4.6}$$

We define $U_D|_{e_\ell} = U_\ell$, $\ell = 1, 2, \ldots, L$ and $U_D = 0$ on $D \backslash \cup_{\ell=1}^{L} e_\ell$. We will continue with the simplified version of (4.6) as following

$$\int_D A\nabla u \cdot \nabla v \, dx + \int_D 2\eta uv \, dx + \int_{\partial D_D} uv \, ds = \int_{\partial D_D} U_D v \, ds, \tag{4.7}$$

where $v \in H^1(D)$ is a test function, and $\partial D_D = \cup_{\ell=1}^{L} e_\ell$. Now we define the finite element discretized approximation of the functions A, η, U_D and u by

$$\tilde{A}(x) := \sum_{i=1}^{N} A(x_i)\varphi(x) \approx A(x), \qquad \tilde{\eta}(x) := \sum_{i=1}^{N} \eta(x_i)\varphi(x) \approx \eta(x),$$
$$\tilde{U}_D(x) := \sum_{i=1}^{N} U_D(x_i)\varphi(x) \approx U_D(x), \qquad \tilde{u}(x) := \sum_{i=1}^{N} u(x_i)\varphi(x) \approx u(x).$$

Let $\tilde{A} = [\tilde{A}_1, \tilde{A}_2, \ldots, \tilde{A}_N]^T$ to be a vector for which $\tilde{A}_i = \tilde{A}(x_i), , i = 1, \ldots, N$. Similarly, for $\tilde{\eta}, \tilde{U}_D$ and $\tilde{u}$. We write nodal representation of (4.7) as following

$$\sum_{j=1}^{N}\sum_{i=1}^{N} \Big(\tilde{A}_i \int_D \varphi_i \nabla\varphi_j \cdot \nabla v \, dx\Big)\tilde{u}_j + \sum_{j=1}^{N}\sum_{i=1}^{N} \Big(\tilde{\eta}_i \int_D 2\varphi_i\varphi_j v \, dx\Big)\tilde{u}_j$$
$$+ \sum_{j=1}^{N} \Big(\int_{\partial D_D} \varphi_j v \, ds\Big)\tilde{u}_j = \sum_{i=1}^{N} \Big(\int_{\partial D_D} \varphi_i v \, ds\Big)\tilde{U}_{D_i}.$$

Now assume that the discretized test function in one of basis functions, i.e. $v \in \{\varphi_i\}_{i=1}^{N}$. In order to derive Jacobian and Hessian of the EIT forward problem, we assemble the global mass matrix $M \in \mathbb{R}^{N\times N}$ and the stiffness matrix $S \in \mathbb{R}^{N\times N}$ by local matrices M^i and $S^i \in \mathbb{R}^{N\times N}$. Then we let

$$S^i_{kj} = \int_D \varphi_i \nabla\varphi_j \cdot \nabla\varphi_k \, dx, \qquad M^i_{kj} = \int_D 2\,\varphi_i\varphi_j\varphi_k \, dx,$$
$$T_{kj} = \int_{\partial D_D} \varphi_j\varphi_k \, ds, \qquad f^i_k = \int_{\partial D_D} \varphi_i\varphi_k \, ds.$$

Therefore, we assemble

$$K = \sum_{i=1}^{N} \tilde{A}_i S^i + \sum_{i=1}^{N} \tilde{\eta}_i M^i + T \qquad \text{and} \qquad b = \sum_{i=1}^{N} \tilde{U}_{D_i} f^i.$$

Thus the finite element approximation the solution, i.e. $\tilde{u}$ can be obtained by solving the linear equation

$$K\tilde{u} = b. \tag{4.8}$$

We note that K is a function of the parameter of interest θ and the design parameter ξ. In the next subsections we will present the calculation of Jacobian and Hessian of the forward EIT problem with respect to θ, which are vital for the optimization purpose as well as in Laplace approximation of the posterior.

4.2 Electrical Impedance Tomography Inverse Problem

In order to investigate EIT inverse problem, consider the observation model

$$Y_\xi = g_\xi(\theta) + \epsilon,$$

where $Y_\xi = \{Y_\xi^i\}_{i=1}^M$ are i.i.d. samples of the measurement variable Y_ξ, i.e. measurement data, which are electrode voltage measurement. The map $g_\xi(\theta)$ presents the forward problem, which in our case is a PDE, which mapping the parameter of the interest $\theta = \log A(x)$ (conductivity of interior tissues) to the electrode measurements obtained from electrodes attached to the surface of the underlying part of the body. The design parameter is the electrode configurations before starting to take a measurement data. The noise $\epsilon = \{\epsilon_i\}_{i=1}^M$ is assumed to be an additive Gaussian noise $\epsilon \sim \mathcal{N}(0, \Sigma_{\text{noise}})$, where $\Sigma_{\text{noise}} \in \mathbb{R}^{L\times L}$.

Based on the Bayes' rule, the posterior distribution of the parameter θ can be obtained as

$$p(\theta|Y_\xi) \propto P(Y_\xi|\theta)\, P(\theta).$$

In other words, the posterior distribution of the parameter θ is proportional to the product of the likelihood function and the prior pdf, which can also be represented as

$$p(\theta|Y_\xi) = p(\theta|\{Y_\xi^i\}) \propto \prod_{i=1}^{M} \exp\left(-\frac{1}{2} r_i(\theta)^T \Sigma_{\text{noise}}^{-1} r_i(\theta)\right) p(\theta), \tag{4.9}$$

where r_i is the residual for the ith measurement, i.e. $r_i(\theta) = Y_\xi^i - g(\theta)$.

In this chapter, the inversion parameter $\theta = [\theta_1, \theta_2, \ldots, \theta_{N_p}]^T$ is assumed to have a multivariate Gaussian prior distribution, i.e. $p(\theta) \sim \mathcal{N}(\mu_{\text{prior}}, \Sigma_{\text{prior}})$. Therefore, $p(\theta) := \exp(-\frac{1}{2}\|\theta - \mu_{\text{prior}}\|^2_{\Sigma_{\text{prior}}})$ is the prior density function. The likelihood density function is also defined by $p(Y_\xi|\theta) := \exp(-\frac{1}{2}\|g_\xi(\theta) - Y_\xi\|^2_{\Sigma_{\text{noise}}})$. Therefore, the posterior density function can be written as

$$p(\theta|Y_\xi) \propto \exp\Big(-\frac{1}{2}\Big[\|g_\xi(\theta) - Y_\xi\|^2_{\Sigma_{\text{noise}}} + \|\theta - \mu_{\text{prior}}\|^2_{\Sigma_{\text{prior}}}\Big]\Big).$$

Furthermore, we define

$$F(\theta) := -\log\left(p(\theta|Y_\xi)\right). \tag{4.10}$$

Thus, using the definition of the posterior pdf and residual function we obtain

$$F(\theta) = -\log\left(\prod_{i=1}^{M} \exp\left(-\frac{1}{2} r_i(\theta)^T \Sigma_{\text{noise}}^{-1} r_i(\theta)\right) p(\theta)\right) = \frac{1}{2}\sum_{i=1}^{M} r_i(\theta)^T \Sigma_{\text{noise}}^{-1} r_i(\theta) - h(\theta) + C, \tag{4.11}$$

where $h(\theta) = \log(p(\theta))$ and C is a constant.

Theorem 9. *Assume that $Y_\xi = \{Y_\xi^i\}_{i=1}^M$ is the set of M measurement data (response) generated using the model (2.31) with a nonlinear forward function g is a nonlinear function in terms of parameter θ, and $\{\epsilon_i\}_{i=1}^M$ independent and identically distributed (i.i.d.) Gaussian noises with the mean 0 and covariance matrix Σ_{noise}. Then, the Laplace approximation of the posterior leads to a Gaussian pdf*

$$\tilde{p}(\theta|Y_\xi) = \frac{1}{\sqrt{2\pi}^d |\Sigma_{\text{post}}|^{\frac{1}{2}}} \exp\left(-\frac{(\theta-\hat{\theta})^T \Sigma_{\text{post}}^{-1}(\theta-\hat{\theta})}{2}\right), \tag{4.12}$$

where

$$\hat{\theta} = \arg\min_\theta \left(\frac{1}{2}\sum_{i=1}^M (Y_\xi^i - g(\theta))^T \Sigma_{\text{noise}}^{-1}(Y_\xi^i - g(\theta)) - h(\theta)\right), \tag{4.13}$$

and

$$\Sigma_{\text{post}} = \left(\nabla\nabla F(\hat{\theta})\right)^{-1} = \left(H(\hat{\theta})^T \Sigma_{\text{noise}}^{-1} E_s + M J(\hat{\theta})^T \Sigma_{\text{noise}}^{-1} J(\hat{\theta}) - \nabla\nabla h(\hat{\theta})\right)^{-1}, \tag{4.14}$$

where J and H are Jacobian and Hessian of the parameter-to-observational map g, respectively. Also, $E_s = E_s(\hat{\theta}) = \sum_{i=1}^M r_i(\hat{\theta})$.

Proof. The maximum a posteriori point is a point in the parameter space in which $F(\theta)$ has the minimum value, i.e.

$$\begin{aligned}\hat{\theta} = \arg\min_\theta F(\theta) &= \arg\min_\theta \left(\frac{1}{2}\sum_{i=1}^M r_i(\theta)^T \Sigma_{\text{noise}}^{-1} r_i(\theta) - h(\theta)\right) \\ &= \arg\min_\theta \left(\frac{1}{2}\sum_{i=1}^M (Y_\xi^i - g(\theta))^T \Sigma_{\text{noise}}^{-1}(Y_\xi^i - g(\theta)) - h(\theta)\right).\end{aligned} \tag{4.15}$$

$F(\theta)$ is in fact the negative logarithm of the posterior pdf of θ. Now, let $\tilde{F}(\theta)$ be the second-order Taylor expansion of $F(\theta)$ around a given $\hat{\theta}$ (MAP estimation). Then

$$\tilde{F}(\theta) \simeq F(\hat{\theta}) + \nabla F(\hat{\theta})(\theta-\hat{\theta}) + \frac{1}{2}(\theta-\hat{\theta})^T \nabla\nabla F(\hat{\theta})(\theta-\hat{\theta}). \tag{4.16}$$

where $F(\hat{\theta})$ is a constant and it does not influence the shape of the approximated posterior. Furthermore, $\hat{\theta}$ is the maximum a posteriori point of the θ, so $\nabla F(\hat{\theta}) = 0$.

For the functional $F(\theta) = \frac{1}{2}\sum_{i=1}^M r_i(\theta)^T \Sigma_{\text{noise}}^{-1} r_i(\theta) - h(\theta)$ we compute the Jacobian of F as

$$\begin{aligned}J_F(\theta) := \nabla F(\theta) &= \sum_{i=1}^M J(\theta)^T \Sigma_{\text{noise}}^{-1} r_i(\theta) - \nabla h(\theta) \\ &= J(\theta)^T \Sigma_{\text{noise}}^{-1} \sum_{i=1}^M r_i(\theta) - \nabla h(\theta) \\ &= J(\theta)^T \Sigma_{\text{noise}}^{-1} E_s(\theta) - \nabla h(\theta),\end{aligned} \tag{4.17}$$

and the Hessian of F as

$$H_F(\theta) := \nabla\nabla F(\theta) = H(\theta)^T \Sigma_{\text{noise}}^{-1} \sum_{i=1}^{M} r_i(\theta) + J(\theta)^T \Sigma_{\text{noise}}^{-1} \sum_{i=1}^{M} J(\theta) - \nabla\nabla h(\theta)$$
$$= H(\theta)^T \Sigma_{\text{noise}}^{-1} E_s(\theta) + M J(\theta)^T \Sigma_{\text{noise}}^{-1} J(\theta) - \nabla\nabla h(\theta), \tag{4.18}$$

where J and H are Jacobian and Hessian of the parameter-to-observational (forward) map g, respectively. Also, $E_s = E_s(\theta) = \sum_{i=1}^{M} r_i(\theta)$. This completes the proof. □

In the following Theorem, Long et al. [11] presented an approximation for $\hat{\theta}$ based on the true value for parameter θ_*:

Theorem 10 ([11])**.** *Considering optimization problem* (4.15) *and using the first order approximation the gradient of the objective function, one can obtain an approximation for* $\hat{\theta}$*:*

$$\hat{\theta} = \theta_* - \left(M J^T \Sigma_{\text{noise}}^{-1} J + H^T \Sigma_{\text{noise}}^{-1} E_s - \nabla\nabla h(\theta_*) \right)^{-1} \left(J^T \Sigma_{\text{noise}}^{-1} E_s \right) + \mathcal{O}_p(\frac{1}{M}) \tag{4.19}$$

where θ_* *is the true value for the unknown parameter* θ*! Here,* $J = J(\theta_*)$*,* $H = H(\theta_*)$ *and* $E_s = E_s(\theta_*)$*.*

To follow this approach, one needs to know the true value for the unknown parameter in advance, which is not the case in many inverse problems. If we use the true value θ_* as the initial point, as in [11], we will have a biased approach which gives a seemingly accurate estimation of the MAP point $\hat{\theta}$ using one-step approximation (4.19). However, the implementations show that the approximation is not accurate if we use an arbitrary initial value θ_{init} to estimate the MAP, even for a simple nonlinear problem with an one-dimensional parameter space. We borrow the following example from [11, 16] to clarify the issue.

Example 1 (One-dimensional Parameter Space)**.** *Let* y *is the scalar response of the following nonlinear model with one-dimensional parameter space:*

$$Y_\xi = g(\theta, \xi) + \epsilon = \theta^3 \xi^2 + \theta \exp(-|0.2 - \xi|) + \epsilon, \tag{4.20}$$

where $\epsilon \sim \mathcal{N}(0, 10^{-2})$ *is Gaussian noise, the prior is chosen to be the uniform distribution* $\mathcal{U}(0, 1)$*. Here,* θ *is the parameter of interest. Let* $\xi = 0.2$ *be the design parameter and* $\theta_t = 1$ *be the true value for the parameter* θ*. Let* $\{Y_\xi^i\}_{i=1}^M$ *for* $M = 1, 5, 10$ *be the response of the model with* $\theta = \theta_t = 1$ *with the Gaussian noise* ϵ*. For a moment, assume that we do not have the true value for* θ *and we would like to estimate the true value using* MAP *estimation, given the response* $\{Y_\xi^i\}_{i=1}^M$*.*

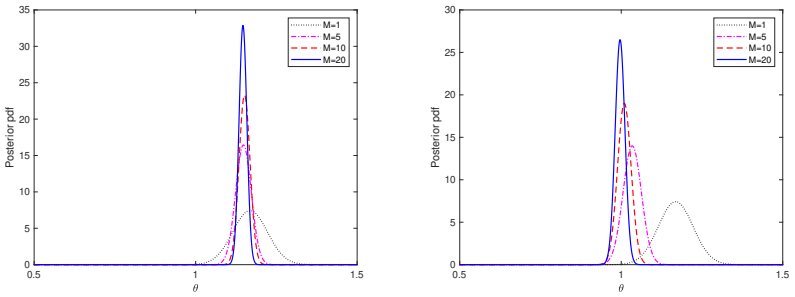

Figure 4.1: Posterior pdf: One-step (left) vs. Iterative MAP estimation (right). $\theta_t = 1$ is the true value for the parameter of interest. The initial value for θ in both cases one-step and iterative is $\theta_{\text{init}} = 1.4$ and design parameter $\xi = 0.2$.

We start with different initial values and try to compute the MAP point for Example 1 using one-step estimation (4.19). Figure 4.2 shows the posterior distributions with one-step (left) versus iterative (right) estimations of the MAP, starting with a same arbitrary initial value θ_{init} for theta, while the true value of θ_* is 1. The iterative model performs better.

4.3 Jacobian and Hessian calculations of the Electrical Impedance Tomography Forward Mapping

Here we aim to derive the Jacobian (gradient) and Hessian of the parameter-to-observable map g in our EIT model (4.1). g is an Elliptic PDE which maps the conductivity field to the voltages at the electrodes that we measure them as observations:

$$g : A(x) \xmapsto{\text{PDE}} u(x) \xmapsto{\text{SaE}} \text{voltage at electrodes},$$

where SaE (Solution at Electrodes) is a projection that calculates voltage at the electrodes. The partial derivative of the solution with respect to the parameter of interest is obtained by the partial derivative of the finite element solution (4.8) of the EIT model, i.e. $\tilde{u} = K^{-1}b$. Using the rule for inverse matrix derivatives, we obtain

$$\frac{\partial \tilde{u}}{\partial \tilde{A}_i} = \frac{\partial K(\tilde{A}_i)^{-1}b}{\partial \tilde{A}_i} = -K^{-1}\frac{\partial K}{\partial \tilde{A}_i}K^{-1}b = -K^{-1}\frac{\partial K}{\partial \tilde{A}_i}\tilde{u}. \tag{4.21}$$

To calculate the Jacobian of g with respect to the conductivity, we use SaE operator, i.e. $\frac{\partial g}{\partial \tilde{A}_i} = \text{SaE}\ \frac{\partial \tilde{u}}{\partial \tilde{A}_i}$. Here, SaE is a matrix with as many rows as electrodes and as many columns as nodal points. We recall the assembly of $K = \sum_{i=1}^{N} \tilde{A}_i S^i + \tilde{\eta}_i M^i + T$ and assume $\tilde{A} = \tilde{\eta}$ which guarantees the well-posedness of the Bayesian inversion of the EIT problem. We obtain

$$\frac{\partial K}{\partial \tilde{A}_i} = S^i + M^i. \tag{4.22}$$

It is worth to mention that the conductivity field has the lognormal probability distribution and $\theta = \log \tilde{A}(x)$. Therefore, the Jacobian of the g with respect to θ (with Gaussian posterior distribution) is

$$J(\theta) := \frac{\partial \tilde{u}}{\partial \theta} = \frac{\partial \tilde{u}}{\partial \tilde{A}} \frac{\partial \tilde{A}}{\partial \theta}, \tag{4.23}$$

where $\frac{\partial \tilde{A}}{\partial \theta}$ is a matrix whose (k, n)-th element is $\delta_{kn} \exp(\theta_k)$.

As mentioned in (4.18), in order to compute the Hessian of the functional $-\log(p(\theta|Y_\xi))$ we need to calculate Hessian of the parameter-to-observable map, i.e. H. To this end, we continue with (4.21) and will compute

$$H_{s,t} := \begin{pmatrix} \frac{\partial^2 \tilde{u}}{\partial \tilde{A}_s^2} & \frac{\partial^2 \tilde{u}}{\partial \tilde{A}_s \partial \tilde{A}_t} \\ \\ \frac{\partial^2 \tilde{u}}{\partial \tilde{A}_t \partial \tilde{A}_s} & \frac{\partial^2 \tilde{u}}{\partial \tilde{A}_t^2} \end{pmatrix} \quad s, t = 1, 2, \ldots, N_p, \tag{4.24}$$

where the second order partial derivative of the solution voltage with respect to conductivity $\tilde{A}_s$ and $\tilde{A}_t$ for $s, t = 1, 2, \ldots, N_p$ are obtained by

$$\frac{\partial^2 \tilde{u}}{\partial \tilde{A}_s \partial \tilde{A}_t} = K^{-1} \frac{\partial K}{\partial \tilde{A}_s} K^{-1} \frac{\partial K}{\partial \tilde{A}_t} \tilde{u} + K^{-1} \frac{\partial K}{\partial \tilde{A}_t} K^{-1} \frac{\partial K}{\partial \tilde{A}_s} \tilde{u} - \begin{cases} K^{-1} \frac{\partial^2 K}{\partial \tilde{A}_s^2} \tilde{u} & s = t, \\ 0 & \text{else.} \end{cases} \tag{4.25}$$

In our EIT model and its finite element solution $\frac{\partial^2 K}{\partial \tilde{A}_s^2} \tilde{u} = 0$ due to (4.22). To work with our parameter of interest θ with Gaussian distribution in the Laplace approximation approach, we calculate the eternities of Hessian with respect to θ as follows

$$H_{s,t}(\theta) = H_{s,t} \exp(\theta_s) \exp(\theta_t) + \delta_{st} J_t(\theta), \tag{4.26}$$

where $J_t(\theta)$ is the column number t of $J(\theta)$ calculated in (4.23).

Now with the calculated Jacobian and Hessian for the forward EIT problem as the parameter to the observable map, we are ready to continue the Laplace approximation approach to carry out an optimization for the MAP estimation as well as the computation of the covariance of the posterior distribution (see Theorem 9).

4.4 Laplace Approximation for the Expected Information Gain

In optimal experimental design problems, we aim to find the best design parameter that gives us the most information gain. This optimal design setup in real-world experiments leads us to the most informative measurement data that helps us to estimate the parameter of interest of the problem in a more accurate way. To this end, we need to compute the

expected information gain over all possible $(\theta_t, \{Y_\xi^i(\theta_t)\})$ where $\theta_t \in \Theta$ (parameter space) and $\{Y_\xi^i(\theta_t)\}$ obtained by the model (2.31). Note, the pair $(\theta_t, \{Y_\xi^i(\theta_t)\})$ is generated as synthetic data which is not necessarily the true value for the unknown parameter θ given a fixed measurement data $\{Y_\xi^i\}$ obtained by a real-world experiment. Based on this fact, all approximations and discussions presented in the Section 4.2 are valid when we use θ_t as initial value (as data is generated synthetically, the true value θ_t is known) and based on one-step estimation (4.19), $\hat{\theta} = \theta_t + \mathcal{O}_p(\frac{1}{\sqrt{M}})$ as $M \to \infty$.

As the approximated posterior pdf is normal, a closed-form approximation for D_{KL} is available with the help of the logarithm of the ratio posterior to prior which is involved in the definition of Kullback-Leibler divergence.

Theorem 11 ([11])**.** *The information gain (Kullback-Leibler divergence) for the model parameter given the measurement data is obtained as a closed form*

$$D_{KL} = -\frac{1}{2}\log((2\pi)^d|\Sigma_{\text{post}}|) - \frac{d}{2} - h(\hat{\theta}) - \frac{\text{tr}(\Sigma_{\text{post}}\nabla\nabla h(\hat{\theta}))}{2}, \tag{4.27}$$

where $\text{tr}(\cdot)$ *is the trace operator.*

Using the closed form of the Kullback-Leibler divergence presented in (4.27) and applying the one-step MAP estimation (4.19), the Expected Information Gain (EIG) can be computed by

$$\text{EIG}_\theta(\xi) = \mathbb{E}[D_{KL}] = \int_\Theta \left[-\frac{1}{2}\log((2\pi)^d|\Sigma_{\text{post}}|) - \frac{d}{2} - h(\theta_t)\right] p(\theta_t)d\theta_t. \tag{4.28}$$

In order to compute the expected information gain, using the expression (4.28), we use numerical techniques such as the Monte Carlo methods or the sparse grid quadrature.

4.5 Numerical Results

Here, we present some numerical results for the computation of expected information gain. In Example 1 with the one-dimensional parameter space and design parameter $\xi \in [0, 1]$, we calculate the Expected Information Gain (EIG) via integration (4.28) over the parameter space $\theta \in [0.5, 1.5]$ (See Figure 4.2).

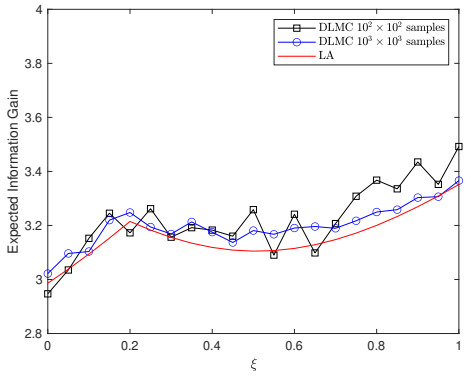

Figure 4.2: Expected Information Gain (EIG) for Example 1 for different designs ξ calculated with double-loop Monte Carlo method as well as Laplace Approximation method.

In electrical impedance tomography (EIT) method of imaging, the surface electrodes are attached to the skin around the body part being examined. Electrical conductivity inside the body varies considerably among various biological tissues and can be inferred using surface electrode measurements. To solve our proposed EIT model, we use the first order Galerkin finite element method on the same domain in Chapter 3, i.e. a cross-section of a human right leg as our computational domain. On the boundary of the domain there are eight electrodes attached with a fix distance from each other (See Figure 1.1). Here in this chapter, our focus is on the computation of the EIG for a given design, not optimizing designs to achieve the best design of for experiment. Therefore, we fix our design on the standard configuration of the electrodes which are attached equidistantly.

Figure 4.3 (right) shows the domain with a finite element discretization. In this figure (left), three different subdomains are displayed: the first subdomain is a circular bone (in dark blue) surrounded by the second subdomain, the muscle (in brown), which itself consists of many partitions, and the rest is fat (in blue). Each subdomain has its own electrical conductivity and in the solution of the forward problem the FEM mesh is aligned with the inclusions such that each element has a constant value for the coefficient A.

Injecting current to an electrode pair under the neighboring/adjacent method, and solving the corresponding model using finite element method yields the solution $u(x)$ (electrical potential) at the all nodal points x in the domain and other electrodes. We repeat the process for all possible neighboring options and combine all solutions of different PDEs in one single vector that we call it measurement data. This measurement data might be noisy. We use additive random noise with Gaussian distribution to simulate this noise. Therefore, given the conductivity of the tissues of the body as the parameter A in the model, we obtain a 40 dimensional measurement data with some noise, which is our forward model. Figure 4.4 (Top Left) shows the noisy measurements versus different values for muscle conductivity.

For the corresponding inverse problem, we narrow our focus on a low dimensional problem and consider only two parameters of interest, namely, conductivity of muscle and the conductivity of fat. This reducing here is necessary as dealing with high dimensional EIT problem and calculating EIG is not an easy task from computational point of view, specially

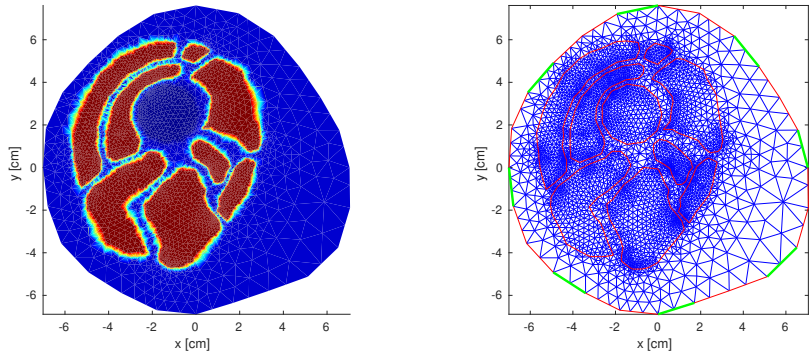

Figure 4.3: EIT domain: cross-section of a right leg illustrating a schematic of the three subdomains (left) and the discretization by a FEM mesh (right) used for solving the forward problem.

in our case that in each evaluating our forward mapping we should solve 8 PDEs of the model (4.1) to obtain the measurements. With this reduce, our inverse problem is similar to inclusion detection in EIT [17]. The main goal is here to estimate the expected information gain EIG using Laplace approximation. We implemented the methodology presented in the previous section, to estimate posterior distribution for the parameters of interest as well as calculating the EIG using (4.28).

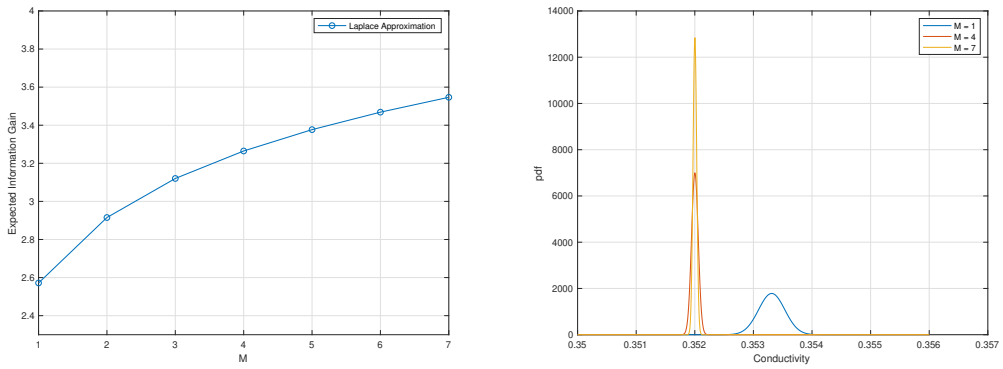

Figure 4.4: Left: Expected Information Gain (EIG) for EIT problem computed by Laplace approximation vs. number of measurements M. Right: Posterior distribution of muscle conductivity estimated with Laplace approximation using the different number of measurements M.

The conductivity of biological tissues like bone, muscle, and fat can vary depending on factors such as water content, density, and composition. At 50 kHz frequency, the conductivities of these tissues can be approximately in the range of 0.01 to 0.1 S/m for bone conductivity, 0.2 to 0.6 S/m for muscle and approximately in the range of 0.03 to 0.1 S/m for fat. These values are approximated and can vary based on factors such as tissue

hydration, temperature, and the specific characteristics of the individual being measured. We note that different sources may provide slightly different ranges for tissue conductivities [18, 19, 20]. Here in this thesis, we generate and use synthetic data for the simulations in which the true value for the muscle and fat conductivities are assumed to be 0.352 and 0.04, respectively.

The experiment can be done for several times to obtain noisy measurement data $Y_\xi = \{Y_\xi^i\}_{i=1}^M$ which are in fact i.i.d. samples of the response random variable which has Gaussian distribution with the mean $g(\theta)$ and the covariance Σ_{noise}. With a higher number of measurements (i.e. more repeat in the experiment) we can obtain better results with less uncertainty for the inverse problem while the computation cost is also higher. Figure 4.4 (Right) shows the posterior distribution for the muscle conductivity for different numbers of measurement repeats M. It is obvious that with more repeat of experiments we obtain better results with more accurate MAP estimation and less uncertainty which are represented by the variance of the unknown parameter.

For estimating EIG, a Monte Carlo approximation is carried out to compute the integration presented in Equation (4.28). Figure 4.4 (Left) shows the estimated EIG for different numbers of measurement repeats. Increasing the measurement repeats, the expected information gain is also increases which is shown in this Figure.

For the EIT problem with two unknown parameters, i.e. muscle and fat conductivities, we implemented Laplace approximation method. Figure 4.5 (Left) shows the posterior joint distribution of unknown parameters of the EIT inverse problem (true value for fat conductivity is 0.04 and for muscle is 0.352). We calculated the EIG by both double-loop Monte Carlo (Section 2.5) and Monte Carlo Laplace approximation (carrying out a Monte Carlo approximation for (4.28)). Figure 4.5 (Right) shows the expected information gain estimated with these approaches for different number of samples. Note that the number of samples here in the x-axis is the number of samples which are used for both inner and outer loops in the double-loop Monte Carlo method, and are the same as the number of samples used for the Monte Carlo approximation in order to calculate the integration in Equation (4.28).

Computational complexity of double-loop Monte Carlo and Laplace approximation methods have been investigated in recent years [11, 21]. Laplace approximation method provides a closed-form approximation of the inner integral resulting in reducing the estimation problem to a single-loop integration, and therefore a significant reduction in computational costs. Figure 4.6 shows a comparison of computational cost between Monte Carlo Laplace approximation and the double-loop Monte Carlo methods.

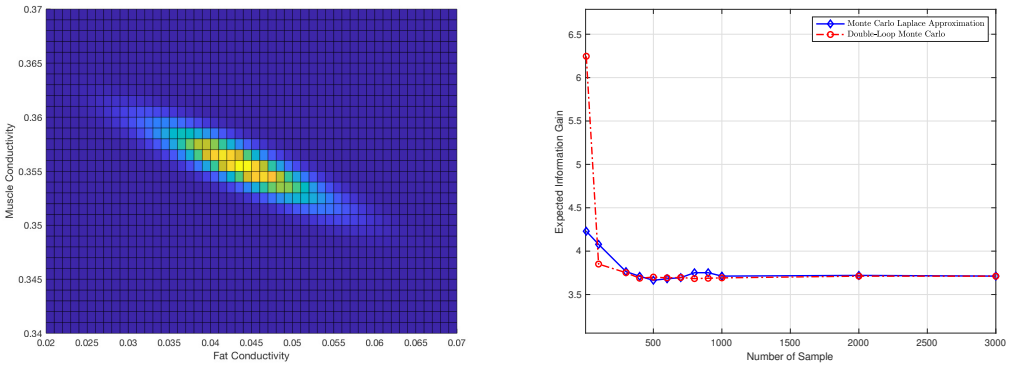

Figure 4.5: Left: Posterior distribution of unknown parameters of the EIT inverse problem (true value for fat conductivity is 0.04 and for muscle is 0.352). Right: Expected Information Gain (EIG) for the EIT problem with two unknown parameters computed by Monte Carlo Laplace Approximation and Double-Loop Monte Carlo methods for different numbers of samples.

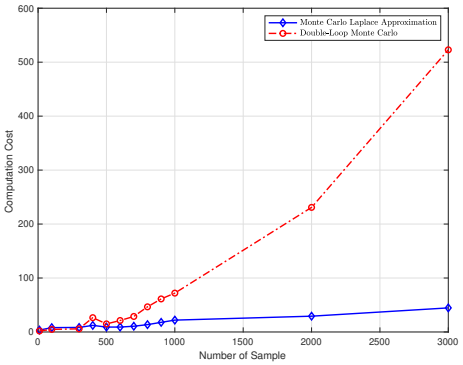

Figure 4.6: Computation cost for Monte Carlo Laplace Approximation and Double-Loop Monte Carlo methods in estimating EIG for the EIT problem.

4.6 Conclusions

In this chapter, we developed a Laplace approximation methodology for a PDE governed EIT inverse problem. To this end, we considered a linearised version of the nonlinear EIT model presented in [14, 15], and estimated the posterior distribution for the conductivity, as well as the expected information gain (EIG) using Laplace approximation approach. Comparing CPU times shows the superiority of the Laplace approximation versus the double-loop Monte Carlo approach.

Bibliography

[1] F. Colibazzi, D. Lazzaro, S. Morigi, A. Samoré, Learning nonlinear electrical impedance tomography, Journal of Scientific Computing 90 (1) (2022) 58.

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

[3] S. Mansouri, Y. Alharbi, F. Haddad, S. Chabcoub, A. Alshrouf, A. A. Abd-Elghany, Electrical impedance tomography–recent applications and developments, Journal of Electrical Bioimpedance 12 (1) (2021) 50–62.

[4] R. C. Smith, Uncertainty Quantification: Theory, Implementation, and Applications, Vol. 12, SIAM, 2013.

[5] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, D. B. Rubin, Bayesian Data Analysis, Chapman and Hall/CRC, 2013.

[6] J. Kaipio, E. Somersalo, Statistical and Computational Inverse Problems, Springer Science & Business Media, 2006.

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

[8] F. G. Waqar, S. Patel, C. M. Simon, A tutorial on the Bayesian statistical approach to inverse problems, APL Machine Learning 1 (4) (2023).

[9] J. Kaipio, A. Seppänen, E. Somersalo, H. Haario, Posterior covariance related optimal current patterns in electrical impedance tomography, Inverse Problems 20 (3) (2004) 919.

[10] P. Tsilifis, R. G. Ghanem, P. Hajali, Efficient Bayesian experimentation using an expected information gain lower bound, SIAM/ASA Journal on Uncertainty Quantification 5 (1) (2017) 30–62.

[11] Q. Long, M. Scavino, R. Tempone, S. Wang, Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, Computer Methods in Applied Mechanics and Engineering 259 (2013) 24–39.

[12] D. Smyl, D. Liu, Optimizing electrode positions in 2d electrical impedance tomography using deep learning, IEEE Transactions on Instrumentation and Measurement (2020).

[13] N. Hyvonen, A. Seppänen, S. Staboulis, Optimizing electrode positions in electrical impedance tomography, SIAM Journal on Applied Mathematics 74 (6) (2014) 1831–1851.

[14] L. Taghizadeh, A. Karimi, B. Stadlbauer, W. J. Weninger, E. Kaniusas, C. Heitzinger, Bayesian inversion for electrical-impedance tomography in medical imaging using the nonlinear Poisson–Boltzmann equation, Computer Methods in Applied Mechanics and Engineering 365 (2020) 112959.

[15] A. Karimi, L. Taghizadeh, C. Heitzinger, Optimal Bayesian experimental design for electrical impedance tomography in medical imaging, Computer Methods in Applied Mechanics and Engineering 373 (2021) 113489.

[16] X. Huan, Y. M. Marzouk, Simulation-based optimal Bayesian experimental design for nonlinear systems, Journal of Computational Physics 232 (1) (2013) 288–317.

[17] S. Ren, M. Soleimani, Y. Xu, F. Dong, Inclusion boundary reconstruction and sensitivity analysis in electrical impedance tomography, Inverse Problems in Science and Engineering 26 (7) (2018) 1037–1061.

[18] H. P. Schwan, Electrical properties of tissue and cell suspensions, in: Advances in biological and medical physics, Vol. 5, Elsevier, 1957, pp. 147–209.

[19] W. T. Joines, Y. Zhang, C. Li, R. L. Jirtle, The measured electrical properties of normal and malignant human tissues from 50 to 900 mhz, Medical physics 21 (4) (1994) 547–550.

[20] D. Miklavčič, N. Pavšelj, F. X. Hart, Electric properties of tissues, Wiley encyclopedia of biomedical engineering (2006).

[21] A. Bartuska, L. Espath, R. Tempone, Laplace-based strategies for bayesian optimal experimental design with nuisance uncertainty, arXiv preprint arXiv:2310.10783 (2023).

5 Multilevel Monte-Carlo Methods for Optimal Bayesian Experimental Design

In this chapter, following the introduction presented in Section 1.4, we focus on the multilevel methods for the numerical estimation of the expected information gain (EIG) which is in fact a nested integral. Before we start the multilevel methods for double loop integrals, we first go through the basic theory of the multilevel Monte Carlo (MLMC) method.

5.1 Multilevel Monte Carlo Method

Multilevel Monte Carlo method was introduced by Giles in 2008 [1, 2] and since then has been widely used in different areas of computational science and engineering [3, 4, 5, 6]. In this section, we give an overview on the multilevel Monte Carlo method based on [2, 5, 6, 7] to which we will refer in the rest of the thesis.

Multilevel Monte Carlo method is an approach to compute expected values that arise in stochastic simulations and uncertainty quantification. This method is as Monte Carlo method based on random sampling, but samples are taken from different levels. In order to use multilevel method, one needs to define different levels of accuracy in which the solution is tending to the true solution when the number of levels increases. However, the cost of computation increases while the accuracy improves.

Let $\mathcal{Z}$ be a random variable (for example, parameter of interest in inverse problems) which can not be sampled exactly, and let $\mathcal{Z}_0, \mathcal{Z}_1, \dots$ be a sequence of random variables which approximate $\mathcal{Z}$ with increasing accuracy but also with increasing cost. As an example, consider the solution of a stochastic differential equation which one can approximate it using finite element approximation $\mathcal{Z}_{h_\ell}$, where h_ℓ is the finite element discretization size defined as $h_\ell = h_{\ell-1}/2$ (we call $\ell = 0, 1, 2, \dots$ to be different levels of accuracy). In this case, $\mathcal{Z}_{h_\ell} \to \mathcal{Z}$ while $\ell \to \infty$.

The goal here is to estimate $\mathbb{E}[\mathcal{Z}]$ efficiently. Assume that $\mathcal{Z}_L$ is an approximation of $\mathcal{Z}$. The standard Monte Carlo estimator is defined by

$$\mathbb{E}_{\mathrm{MC}}[\mathcal{Z}_L] = \frac{1}{N} \sum_{i=1}^{N} \mathcal{Z}_L^{(i)}, \tag{5.1}$$

where $\mathcal{Z}_L^{(i)}$ are random samples of $\mathcal{Z}_L$. The Mean Square Error (MSE) of the Monte Carlo

estimator is given by:

$$\begin{aligned}\mathrm{MSE}_{\mathrm{MC}} &= \mathbb{E}\Big[\big(\mathbb{E}_{\mathrm{MC}}[\mathcal{Z}_L] - \mathbb{E}[\mathcal{Z}]\big)^2\Big] \\ &= \Big(\mathbb{E}[\mathcal{Z}_L - \mathcal{Z}]\Big)^2 + \frac{\mathrm{Var}[\mathcal{Z}_L - \mathcal{Z}]}{N},\end{aligned} \tag{5.2}$$

where Var is the variance and $\mathbb{E}[\mathcal{Z}_L - \mathcal{Z}]$ is the bias.

Now consider the telescoping sum

$$\mathbb{E}[\mathcal{Z}_L] = \mathbb{E}[\mathcal{Z}_0] + \sum_{\ell=1}^{L} \mathbb{E}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]. \tag{5.3}$$

In the multilevel Monte Carlo method, one can compute each term on the right hand side of the telescopic sum (5.3) using standard Monte Carlo method with N_ℓ samples drawn in level ℓ. In this case,

$$\hat{\mathcal{Z}}_L := \mathbb{E}_{\mathrm{MLMC}}[\mathcal{Z}_L] = \sum_{\ell=0}^{L} \frac{1}{N_\ell} \sum_{i=1}^{N_\ell} \big(\mathcal{Z}_\ell^{(i)} - \mathcal{Z}_{\ell-1}^{(i)}\big), \tag{5.4}$$

where $\mathcal{Z}_\ell^{(i)}$ are samples of $\mathcal{Z}$ drawn in the level of accuracy ℓ. If we define

$$\Delta\mathcal{Z}_\ell := \begin{cases} \mathcal{Z}_\ell - \mathcal{Z}_{\ell-1} & \ell > 0, \\ \mathcal{Z}_\ell & \ell = 0, \end{cases} \tag{5.5}$$

then the multilevel Monte Carlo estimator can be written as

$$\mathbb{E}_{\mathrm{MLMC}}[\mathcal{Z}_L] = \sum_{\ell=0}^{L} \mathbb{E}[\Delta\mathcal{Z}_\ell]. \tag{5.6}$$

We use $\mathcal{Z}_{-1} = 0$ when needed. The MSE of the MLMC estimator is given by

$$\begin{aligned}\mathrm{MSE}_{\mathrm{MLMC}} &= \mathbb{E}\Big[\big(\mathbb{E}_{\mathrm{MLMC}}[\mathcal{Z}_L] - \mathbb{E}[\mathcal{Z}]\big)^2\Big] \\ &= \Big(\mathbb{E}[\mathcal{Z}_L - \mathcal{Z}]\Big)^2 + \sum_{\ell=0}^{L} \frac{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]}{N_\ell}.\end{aligned} \tag{5.7}$$

For simplicity, we define $V_\ell = \mathrm{Var}[\mathcal{Z}_L - \mathcal{Z}]$ and note that

- $\mathbb{E}[\hat{\mathcal{Z}}_L] = \mathbb{E}[\mathcal{Z}_L]$ and
- $\mathrm{Var}[\hat{\mathcal{Z}}_L] = \sum_{\ell=0}^{L} \frac{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]}{N_\ell} = \sum_{\ell=0}^{L} \frac{V_\ell}{N_\ell}$.

Theorem 12 (Computational Cost of MLMC [2])**.** *Let $\mathcal{Z}_\ell$ be an approximation of random variable $\mathcal{Z}$ in level ℓ with the computational cost C_ℓ and variance* Var_ℓ*. Assume that there are positive constants $\alpha, \beta, \gamma, c_1, c_2, c_3$ such that $\alpha \geq \min(\beta, \gamma)/2$ and*

- $|\mathbb{E}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]| \leq c_1 2^{-\alpha\ell}$,
- $\mathrm{Var}_\ell \leq c_2 2^{-\beta\ell}$,
- $C_\ell \leq c_3 2^{-\gamma\ell}$,

then there is a positive constant c_4 *such that for any* $\varepsilon < exp(-1)$ *there are* L *and* N_L *for which* MLMC *estimator has a mean square error less than* ε^2, *i.e.* $\mathbb{E}[\mathcal{Z}_L - \mathcal{Z}] < \varepsilon^2$. *The computational complexity of the estimator has the the bound*

$$\mathbb{E}[C] \leq \begin{cases} c_4\varepsilon^{-2} & \beta > \gamma, \\ c_4\varepsilon^{-2}(\log\varepsilon)^2 & \beta = \gamma, \\ c_4\varepsilon^{-2-(\gamma-\beta)/\alpha} & \beta < \gamma. \end{cases} \tag{5.8}$$

We note that the samples drawn in consequent levels $\ell-1$ and ℓ can be somehow correlated. Therefore $\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]$ would decrease when the level ℓ increases. This leads to a smaller number of necessary samples in higher levels. That means the most samples are drawn in smaller levels, and since the computational cost of $\mathcal{Z}_\ell$ is cheaper in smaller levels, then the overall computational cost can be reduced compared to the standard Monte Carlo method.

Remark 4. *In order to ensure that* $\mathrm{MSE}_{\mathrm{MLMC}}$ *is less than* ε^2, *it is sufficient to show that both* $(\mathbb{E}[\hat{\mathcal{Z}}_L - \mathcal{Z}])^2$ *and* $\mathrm{Var}[\hat{\mathcal{Z}}_L]$ *are less than* $\varepsilon^2/2$. *To this end, the number of levels is chosen so to have* $(\mathbb{E}[\hat{\mathcal{Z}}_L - \mathcal{Z}])^2 \leq \varepsilon^2/2$ *while number of samples in each level,* N_ℓ, *is chosen to ensure* $\mathrm{Var}[\hat{\mathcal{Z}}_L] \leq \varepsilon^2/2$.

In order to find the optimal number of samples N_ℓ in each level, one should solve an optimization problem that minimizes the total cost of the estimator for a fixed variance. The optimization problem is

$$\{N_\ell^*\}_{\ell=0}^L = \arg\min_{N_\ell} \sum_{\ell=0}^{L} N_\ell C_\ell \tag{5.9}$$

s.t

$$\sum_{\ell=0}^{L} \frac{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]}{N_\ell} = \frac{\varepsilon^2}{2}.$$

Using Lagrange multipliers method, the problem is reformulated as the minimization problem

$$\min \sum_{\ell=0}^{L} N_\ell C_\ell + \mu^2 \left(\frac{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]}{N_\ell} - \frac{\varepsilon^2}{2} \right), \tag{5.10}$$

which gives the optimal number of samples in each level:

$$N_\ell^* = \mu \sqrt{\frac{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}]}{C_\ell}}.$$

Therefore, to achieve a total variance ε^2, we need to have

$$\mu = \frac{1}{\varepsilon^2} \sum_{\ell=0}^{L} \sqrt{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}] C_\ell}.$$

Thus, the total cost is given by

$$C = \frac{1}{\varepsilon^2} \left(\sum_{\ell=0}^{L} \sqrt{\mathrm{Var}[\mathcal{Z}_\ell - \mathcal{Z}_{\ell-1}] C_\ell} \right)^2.$$

5.2 Multilevel Monte Carlo Method for Nested Integrals

In this section, we extend the idea of multilevel Monte Carlo method to estimate the nested integrals of the form

$$I = \int_{\mathcal{Y}} f\Big(\int_{\mathcal{X}} g(y, x) dx \Big) dy, \tag{5.11}$$

where the square-integrable function f is nonlinear and twice differentiable with respect to y. As an specific application, we focus on the case that the function f is logarithmic in which the nested integral is to estimate the expected information gain (EIG) (see Equation (2.40)).

In Section 2.5, we presented a double loop Monte Carlo estimator for calculating the EIG based on the Bayes' rule, which is defined by

$$\mathrm{EIG}_{\mathrm{DLMC}} = \frac{1}{N} \sum_{n=1}^{N} \Big[\log \Big(p(Y^{(n)} | \theta^{(n,0)}) \Big) - \log \Big(\frac{1}{M} \sum_{m=1}^{M} p(Y^{(n)} | \theta^{(n,m)}) \Big) \Big]. \tag{5.12}$$

Here we try to briefly introduce the multilevel Monte Carlo estimator for the expected information gain, but for more details, we refer to [5, 6]. The random variable that is used in the EIG formula is

$$\mathcal{Z} := \log p(Y|\theta) - \log \mathbb{E}[p(Y|\theta)], \tag{5.13}$$

where Y is distributed conditionally on the random variable θ. The second term of (5.13) includes $\mathbb{E}_\theta[p(Y|\theta)]$ which is an expected value with respect to the parameter $\theta \sim p(\theta)$, and

it has no closed form to compute. To deal with this, we define a sequence of approximations $\mathcal{Z}_0, \mathcal{Z}_1, \ldots$ of $\mathcal{Z}$ with increasing accuracy but also increasing computational cost,

$$\mathcal{Z}_\ell := \log p(Y|\theta) - \log\Big(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell} p(Y|\theta^{(m)})\Big), \tag{5.14}$$

where $M_0 < M_1 < \ldots$ is an increasing sequence. We also define

$$\mathcal{S}^{M_\ell}(Y) := \frac{1}{M_\ell}\sum_{m=1}^{M_\ell} p(Y|\theta^{(m)}),$$

and we assume that $\{M_\ell\}_{\ell=0}^{\infty}$ is a geometric sequence for which $M_\ell = M_0 2^\ell$. Based on the definition of $\Delta\mathcal{Z}_\ell$ we have

$$\Delta\mathcal{Z}_\ell := \begin{cases} \log \mathcal{S}^{M_{\ell-1}}(Y) - \log \mathcal{S}^{M_\ell}(Y) & \ell > 0, \\ \log p(Y|\theta) - \log \mathcal{S}^{M_0}(Y) & \ell = 0. \end{cases} \tag{5.15}$$

Goda [6] suggested that the first $M_{\ell-1}$ random samples of θ used in the computation of $\log \mathcal{S}^{M_\ell}(Y)$ is also used in the $\log \mathcal{S}^{M_{\ell-1}}(Y)$. There are also techniques for better coupling of $\mathcal{Z}_\ell$ and $\mathcal{Z}_{\ell-1}$ to avoid more computational costs [6, 7, 8]. Figure 5.1 illustrates the behavior of the variances of $\mathcal{Z}_\ell$ and $\Delta\mathcal{Z}_\ell$ which shows a decay in variance of $\Delta\mathcal{Z}_\ell$, while that of $\mathcal{Z}_\ell$ is constant.

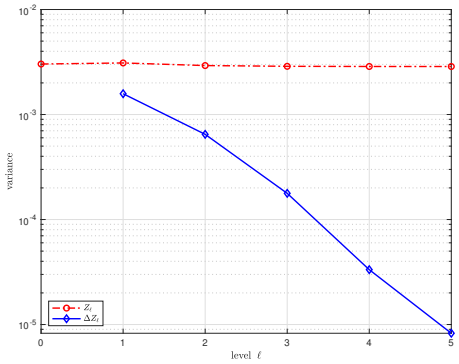

Figure 5.1: Behavior of the empirical variances of both $\mathcal{Z}_\ell$ and $\Delta\mathcal{Z}_\ell$. Here we plot the variance as a function of level. While the variance of $\mathcal{Z}_\ell$ is almost constant, the variance of $\Delta\mathcal{Z}_\ell$ decays as the level increases.

For our EIT problem explained in Chapter 4, we developed a multilevel Monte Carlo methodology to calculate the expected information gain for the given design of experiment. As it is shown in the Figure 5.1, the variance of $\Delta\mathcal{Z}_\ell$ decays as the level increases, while the variance of $\mathcal{Z}_\ell$ is almost constant.

In the rest of this chapter, we turn our focus on developing a new approach for calculating EIG. This new method is called Multilevel Double Loop Randomized Quasi Monte Carlo (ML-DL-RQ-MC), where we use low-discrepancy point sets, namely randomized quasi points, for sampling.

5.3 Multilevel Randomized Quasi-Monte Carlo for Calculating of Expected Information Gain

In this section, we aim to develop a multilevel randomized quasi Monte-Carlo method for the calculation of the expected information gain in Bayesian inverse problems. In fact, the goal is the calculation of the expected value of the random variable

$$\mathcal{Z} := \log p(Y|\theta) - \log \mathbb{E}[p(Y|\theta)],$$

which is presented in Equation (5.13). Before that, we give a brief overview of the Randomized Quasi Monte Carlo nested integrals.

Quasi-Monte Carlo method (QMC) is an extension to Monte Carlo method in which low-discrepancy sequences (also called quasi-random sequences or sub-random sequences) are used for sampling purposes. In this approach, quasi-Monte Carlo uses a low-discrepancy sequence such as the Halton sequence [9], the Sobol sequence [10], or other options and tries to reduce the variance.

In the calculation of nested integration where a nonlinear function is involved, sometimes there is no closed-form solution. As explained before, such nested integrations are common in engineering problems, such as optimal experimental design. Using the Monte Carlo method to approximate both integrals leads to a double-loop Monte Carlo estimator (see Section 2.5), which is often prohibitively expensive and estimation of the outer integral has bias relative to the variance of the inner integrand. In the case that the the inner integrand is not exact and is calculated approximately, then an additional bias will be involved the estimation of the outer integral. To overcome this issue, one can use variance reduction methods, e.g. importance sampling, to reduce the computational cost and to make computations more affordable. Another solution is to use deterministic low-discrepancy sequences instead of random samples. This approach leads us to quasi-Monte Carlo techniques [11]. Beyond this, randomizing the low-discrepancy sequences simplifies the error analysis of the proposed double-loop quasi-Monte Carlo estimator.

In [12], Bartuska et. al. developed a randomized quasi Monte Carlo method to estimate nested integrals namely Double Loop Quasi-Monte Carlo (DLQMC) in which randomized low-discrepancy Sobol sequences are used for sampling. The authors present the asymptotic error bounds and obtain the optimal number of samples for both integral approximations. In the rest of this section, we focus on the Double Loop Randomized Quasi-Monte Carlo (DLRQMC) method to calculate the nested integrals.

Let's consider again the nested integral (5.11) where the square-integrable function f is nonlinear and twice differentiable with respect to y. Here, $\mathcal{Y}$ and $\mathcal{X}$ are in fact hypercubes in finite dimensional spaces. The standard method to approximate this nested integral is via the Double Loop Monte Carlo (DLMC) estimator [13], is defined by

$$I_{\text{DLMC}} = \frac{1}{N}\sum_{n=1}^{N} f\Big(\frac{1}{M}\sum_{m=1}^{M} g(y^{(n)}, x^{(n,m)})\Big), \tag{5.16}$$

where $y^{(n)}$ and $x^{(n,m)}$ are sampled i.i.d. from the uniform distributions on hypercubes $[a,b]^{d_1}$ and $[a,b]^{d_2}$, respectively. We have already used the Double Loop Monte Carlo (DLMC) estimator for estimating expected information gain. See Section 2.5 for more details.

Generally in the case of single integrals, the standard MC estimator is unbiased, and its variance decreases by increasing the number of samples. But here, the outer MC estimator has a bias related to the variance of the inner integral estimator [12]. Therefore, we need many inner and outer samples to keep the bias and variance of this estimator under control. This would limit significantly practical problems, Specially for computationally demanding problems, such as PDE-governed inverse problems, in which a finite element method (FEM) approximation g_h with discretization parameter h is used to evaluate forward mapping g.

Definition 4 (DLQMC estimator [12])**.** *The DLQMC estimator of the nested integral* (5.11) *is defined as follows:*

$$I_{\text{DLQMC}} = \frac{1}{N}\sum_{n=1}^{N} f\Big(\frac{1}{M}\sum_{m=1}^{M} g(y^{(n)}, x^{(n,m)})\Big), \tag{5.17}$$

where the square-integrable function $f : \mathbb{R} \to \mathbb{R}$ is nonlinear, and $g : [0,1]^{d_1} \times [0,1]^{d_2} \to \mathbb{R}$ is square-integrable. The sample points have the following shape:

$$\begin{aligned} y^{(n)} &= \{\xi_{d_1}^{(n)}, \rho_{d_1}\}, \qquad 1 \le n \le N \\ x^{(n,m)} &= \{\xi_{d_2}^{(m)}, \rho_{d_2}^{(n)}\}, \qquad 1 \le n \le N, 1 \le m \le M, \end{aligned} \tag{5.18}$$

where $\xi_{d_1} \in [0,1]^{d_1}$ and $\xi_{d_2} \in [0,1]^{d_2}$ are chosen from a low-discrepancy sequence.

Note that $y^{(n)}$ and $x^{(n,m)}$ consist of a deterministic component $\xi_{d_i} \in [0,1]^{d_i}, i = 1,2$ and a random component $\rho_{d_i}, i = 1,2$ [14]. An option is to choose Sobol sequence as low-discrepancy sequence which uses a base of two to form successively finer uniform partitions of the unit interval. The difference between the estimators (5.16) and (5.17) lies only in the points used to evaluate the function to be integrated.

In the context of Bayesian inverse problems, Monte-Carlo methods are used to approximate posterior distribution of the quantity of interest, which is usually intractable. The mean squared error (MSE) of the approximation by the Monte-Carlo method is of the rate $O(N^{-1})$

in terms of N number of samples, and an accuracy of ϵ needs $O(\epsilon^{-2})$ samples. One approach to mitigate this issue is to use random points with better quality than the MC points in the sense of a better coverage of the random space and a reduced MSE. To this end, quasi Monte-Carlo (QMC) points with the (MSE) error rate of $O(N^{-2}\log N^{2d-2})$ have been investigated (see for instance [15, 16, 17]). The disadvantage of QMC methods is that their MSE depends on the dimension d of the random variable. That means for high dimensions, one need a big number of samples to reach the same accuracy as the MC, which leads to high computational costs. To overcome this issue, randomized quasi Monte-Carlo points are proposed [18], which use randomly shifted low-discrepancy points and the corresponding optimal rate is independent of the dimension of the random variable. For instance there exist shifted rank-1 lattice rules, which are constructed by the component-by-component (CBC) algorithm, and has the convergence rate of $O(N^{-2+\delta})$ for any $\delta > 0$ [19].

The randomized quasi-Monte-Carlo methods are constructed using a uniformly distributed d-dimensional random shift $\rho \sim U[0,1)^d$, on the QMC low discrepancy sequences (for instance Halton and Sobol sequences). In particular, a randomized rank-1 lattice rule [20] is defined by

$$x_j^{(i)} := \frac{j}{N}\lambda + \rho^{(i)} \mod 1, \quad j \in \{1,\dots,N\}, \quad i \in \{1,\dots,M\}, \tag{5.19}$$

where N is the number of random points, M the number of random shifts, and $\lambda \in \mathbb{R}^d$ is a d-dimensional generating vector. The idea of developing Double Loop Quasi Monte Carlo (DLQMC) is to use N low-discrepancy points for outer loop and M points with importance sampling for the inner loop of the double-loop Monte Carlo estimator. For the randomized version, one can use one randomization of the outer samples in the estimator and utilize low-discrepancy points with randomizations.

We can also improve the computational efficiency of the sampling method by using the multilevel idea proposed by Heinrich [21] for the integration and Giles [2] for the stochastic differential equations.

Here in this thesis, we develop a Multilevel Randomized Quasi-Monte Carlo (MLRQMC) method for evaluating nested integrals. For the low-discrepancy sequences, unlike [12], instead of Sobol points, we exploit the generalized Fibonacci grid [22, 23] to be our low-discrepancy point sets and will transform them to optimal deterministic Gaussian sampling sets to be used for our importance sampling purposes. The advantage of the generalized Fibonacci grid is that the low-discrepancy will be preserved when we transform the uniform quasi points to the Gaussian ones. That is important because we usually deal with Gaussian priors in uncertainty quantification problems, more specifically in the estimation of the EIG, where we use Laplace-approximation-based importance sampling for the inner loop to avoid numerical underflow. Figure 5.2 illustrates low-discrepancy Gaussian points generated by the generalized Fibonacci grid due to given the target mean and covariance. The MLRQMC approach is a sampling method, where we use the multilevel idea for the randomized quasi Monte-Carlo (RQMC) sampling method. In this method, randomly shifted quasi Monte-Carlo points are used instead of Monte-Carlo points as in multilevel Monte-Carlo approach.

To define our Multilevel Double-loop Quasi-Monte Carlo (MLDLQMC), we consider the sequence of approximations $\mathcal{Z}_0, \mathcal{Z}_1, \dots$ of $\mathcal{Z}$ with increasing accuracy but also increasing computational cost. We apply the same notation introduced in MLMC estimator for EIG as

$$\mathcal{Z}_\ell := \log p(Y|\theta) - \log\Big(\frac{1}{M_\ell}\sum_{m=1}^{M_\ell} p(Y|\theta^{(m)})\Big) = \log p(Y|\theta) - \log \mathcal{S}^{M_\ell}(Y), \tag{5.20}$$

where $M_0 < M_1 < \dots$ is an increasing sequence. In this case, the MLDLQMC is

$$I_{\text{MLDLQMC}} := \\ \frac{1}{M_0}\sum_{n=1}^{N_0}\left[\log(Y^n|\theta^n) - \log \mathcal{S}^{M_0}(Y^n)\right] + \sum_{\ell=1}^{L}\frac{1}{N_\ell}\sum_{n=1}^{N_\ell}\left[\log \mathcal{S}^{M_{\ell-1}}(Y^n) - \log \mathcal{S}^{M_\ell}(Y^n)\right], \tag{5.21}$$

where $N_0 > N_1 > \dots > N_L$ are the number of low-discrepancy samples chosen for each level. The samples for each inner loop are quasi-random Gaussian [24] points generated using a Gaussian distribution, e.g. the posterior distribution. Laplace approximation-based posterior distribution explained in Theorem 9 (See Section 4.2) is used for this purpose. For a faster estimation of MAP, we use one-step approximation presented in Theorem 4.19 (See also [25])

$$\hat{\theta}^n = \theta^n - \Big(J(\theta^n)^T\Sigma_{\text{noise}}^{-1}J(\theta^n) + H(\theta^n)^T\Sigma_{\text{noise}}^{-1}E_s(\theta^n) - \nabla\nabla\log(p(\theta^n))\Big)^{-1}\left(J(\theta^n)^T\Sigma_{\text{noise}}^{-1}E_s(\theta^n)\right)$$

where θ^n are the low-discrepancy quasi points generated for the outer loop. Here, J and H Jacobian and Hessian of the $-g$ (forward problem), i.e., $J(\theta) = -\nabla g(\theta)$, and $H(\theta) = -\nabla\nabla g(\theta)$. Also, the residual E_s is defined by $E_s(\theta^n) = Y - g(\theta^n)$. For the covariance to generate quasi-random Gaussian points related to each low-discrepancy point to be used for inner loops, we use

$$\hat{\Sigma}^n = \Big(J(\hat{\theta}^n)^T\Sigma_{\text{noise}}^{-1}J(\hat{\theta}^n) - \nabla\nabla\log(p(\hat{\theta}^n))\Big)^{-1}.$$

As the multilevel estimator is a variance reduction approach, one could investigate the optimal size for the number of samples at each level by minimizing the total variance of the estimator. A similar optimization problem is done in [6].

5.4 Generalized Fibonacci Grid and Quasi-Random Gaussian Points

Quasi-random sequences, such as Sobol or Halton, usually deal with the uniform distribution in $[0,1]^d$ in which the goal is to approximate the uniform distribution by finite sets of points.

But for many applications in uncertainty quantification and optimal experimental design purposes, we need to work with Gaussian distribution and sampling. In order to obtain low-discrepancy points which correspond to the Gaussian distribution, there are various options.

Having uniform quasi points in hand, one can transfer to Gaussian points using inverse transform sampling to generate pseudo-random samples from any probability distribution given its cumulative distribution function. Inverse transformation sampling takes $u \in [0, 1]$, and then returns the smallest $x \in \mathbb{R}$ such that $\mathcal{F}(x) \geq u$ where $\mathcal{F}$ is the target cumulative distribution function [26]. The Box–Muller transform was developed as a more computationally efficient alternative to the inverse transform sampling method [27, 28]. Gauss–Hermite quadrature points are alternatives to use ideally for scalar integrals of polynomial-like functions multiplied with a univariate Gaussian density function [29]. The required points increases exponentially with the number of dimensions.

As a recent work, Frisch and Hanebeck [22, 23] use the generalized Fibonacci grid as the low-discrepancy point set to present an optimal deterministic Gaussian sampling. For univariate Gaussian samples, the inverse Gaussian distribution is applied to transform uniform samples into Gaussian ones. For multivariate Gaussians, this scalar transformation is applied along the directions of the eigenvectors of the covariance matrix, respectively. They utilities the Fibonacci grid, as it can be anisotropically re-scaled along the main axes while preserving the uniformity of points. Instead of colliding, Fibonacci grid points automatically get new neighbors, depending on the amount of re-scaling. The steps for a discrepancy preserving mapping that transforms the uniform density on $[0, 1]$ to arbitrary Gaussian densities, is described in [22]. Here, we use these low-discrepancy Gaussian points for our quasi-Monte Carlo sampling in the new multilevel approach for the estimation of expected information gain.

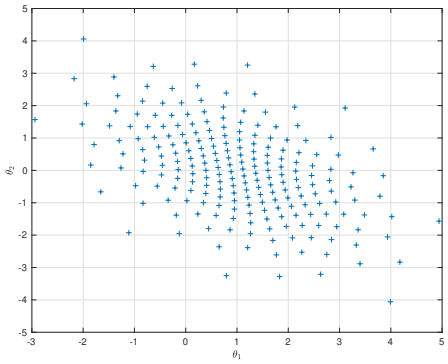

Figure 5.2: Low-discrepancy Gaussian points generated by Fibonacci grid points. Here, a sample set of 200 quasi-Gaussian points is generated with respect to the prior distribution.

5.5 Numerical Results

In this section, we implement our methodology of MLDLQMC for a linear inverse problem for which the analytical value of the EIG is available. Consider following linear model

$$Y = A\theta + \epsilon, \tag{5.22}$$

in which $\theta \in \mathbb{R}^d$ is the vector of parameters of the interest, $A \in \mathbb{R}^{w\times d}$ forward mapping, and $Y \in \mathbb{R}^w$ the measurement data. The noise ϵ is Gaussian with the mean 0 and covariance matrix Σ_ϵ. We also assume that we are given a prior knowledge on θ by a multivariate Gaussian distribution $\mathcal{N}(\mu_{\text{prior}}, \Sigma_{\text{prior}})$. Under the assumption of Gaussian prior and Gaussian noise, the posterior of θ is also Gaussian, and the analytical solution for the maximum a posteriori (MAP) estimation (See Theorem 9 and Equation (4.13)) as well as the covariance matrix of the posterior are available. In this case, the mean of the posterior distribution of θ is given by

$$\hat{\theta} = \Sigma_{\text{post}}\big(A^T\Sigma_\epsilon^{-1}Y + \Sigma_{\text{prior}}^{-1}\mu_{\text{prior}}\big),$$

and the covariance of the posterior is

$$\Sigma_{\text{post}} = (H + \Sigma_{\text{prior}}^{-1})^{-1},$$

where $H = A^T\Sigma_\epsilon^{-1}A$ is the Hessian of the negative logarithm of the posterior pdf [30]. Therefore, the expected information gain (EIG) for this linear inverse problem admits the closed form

$$\text{EIG} = \frac{1}{2}\log\det\Big(N_{\text{ex}}\Sigma_\epsilon^{-1}A\Sigma_{\text{prior}}A^T + I\Big), \tag{5.23}$$

where N_{ex} is the number of experiments repeats [6]. However, [30, 31] present a similar but slightly different closed form for EIG using prior-preconditioned Hessian that includes both data and prior information. We recall the benchmark linear problem presented in [6] and implement numerically our multilevel double-loop quasi-Monte Carlo method on this linear inverse problem. To this end, we assume $\theta = [\theta_1, \theta_2]^T$ to be the parameter of interest from two-dimensional real-valued parameter space, and define

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 4 \end{bmatrix}, \qquad \Sigma_\epsilon = \begin{bmatrix} 0.1 & -0.05 & 0 \\ -0.05 & 0.1 & -0.05 \\ 0 & -0.05 & 0.1 \end{bmatrix}.$$

The measurement data Y is calculated using the model 5.22. We also assume we have a prior knowledge given by a Gaussian distribution with the mean and covariance

$$\mu_{\text{prior}} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \text{and} \quad \Sigma_{\text{prior}} = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix},$$

respectively. For this setting, Figure 5.3 shows the true expected information gain (EIG) calculated for different numbers of experiment repeats.

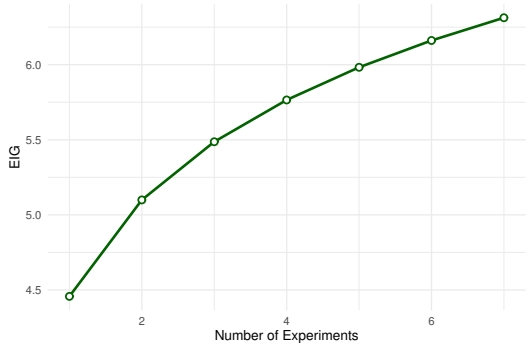

Figure 5.3: True expected information gain (EIG) vs. Number of Experiments (N_{ex}) for the linear inverse problem with the given setting.

For the double-loop Monte Carlo method, we follow the Algorithm 2. The random samples for the outer loop are drawn using the prior distribution which is a multivariable normal distribution. Furthermore, we exploit the fact that the mean and covariance of the posterior distribution are available as a closed form. We use the posterior distribution $p(\theta|Y)$ as the importance measure to draw important samples for the inner loops. To develop double-loop quasi-Monte Carlo method, we apply the low-discrepancy Gaussian points generated with generalized Fibonacci grid considering the mean and covariance of the posterior.

Figure 5.4 shows the total cost as the number of samples versus the accuracy of the presented estimators. As seen in this figure, the double-loop quasi-Monte Carlo and multilevel double-loop quasi-Monte Carlo have better performance against the standard double-loop Monte Carlo regarding the number of samples in total they need to reach the specific accuracy level.

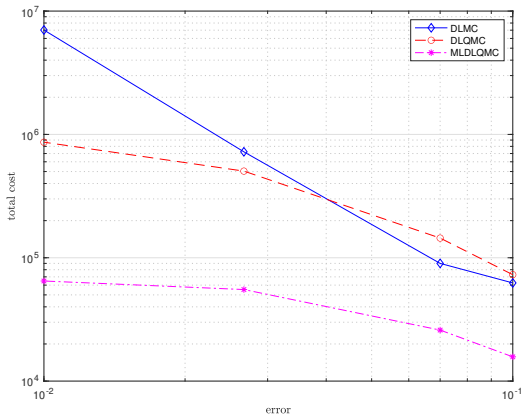

Figure 5.4: Error vs. total cost (total number of samples needed to reach the specific accuracy level).

Bibliography

[1] M. Giles, Improved multilevel Monte Carlo convergence using the Milstein scheme, in: Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer, 2006, pp. 343–358.

[2] M. B. Giles, Multilevel Monte Carlo path simulation, Operations research 56 (3) (2008) 607–617.

[3] L. Taghizadeh, A. Khodadadian, C. Heitzinger, The optimal multilevel Monte-Carlo approximation of the stochastic drift–diffusion-Poisson system, Computer Methods in Applied Mechanics and Engineering 318 (2017) 739–761.

[4] A. Khodadadian, L. Taghizadeh, C. Heitzinger, Three-dimensional optimal multilevel Monte–Carlo approximation of the stochastic drift–diffusion–Poisson system in nanoscale devices, Journal of Computational Electronics 17 (1) (2018) 76–89.

[5] J. Beck, B. Mansour Dia, L. Espath, R. Tempone, Multilevel double loop Monte Carlo and stochastic collocation methods with importance sampling for Bayesian optimal experimental design, International Journal for Numerical Methods in Engineering 121 (15) (2020) 3482–3503.

[6] T. Goda, T. Hironaka, T. Iwamoto, Multilevel Monte Carlo estimation of expected information gains, arXiv preprint arXiv:1811.07546v3 (2019).

[7] M. B. Giles, Multilevel Monte Carlo Methods, Acta numerica 24 (2015) 259–328.

[8] M. B. Giles, T. Goda, Decision-making under uncertainty: using MLMC for efficient estimation of EVPPI, Statistics and computing 29 (2019) 739–751.

[9] J. H. Halton, Algorithm 247: Radical-inverse quasi-random point sequence, Communications of the ACM 7 (12) (1964) 701–702.

[10] I. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, USSR Computational Mathematics and Mathematical Physics 7 (4) (1967) 86–112. doi:https://doi.org/10.1016/0041-5553(67)90144-9.
URL https://www.sciencedirect.com/science/article/pii/0041555367901449

[11] R. E. Caflisch, Monte Carlo and quasi-Monte Carlo methods, Acta numerica 7 (1998) 1–49.

[12] A. Bartuska, A. G. Carlon, L. Espath, S. Krumscheid, R. Tempone, Double-loop quasi-Monte Carlo estimator for nested integration, arXiv preprint arXiv:2302.14119 (2023).

[13] K. J. Ryan, Estimating expected information gains for experimental designs with application to the random fatigue-limit model, Journal of Computational and Graphical Statistics 12 (3) (2003) 585–603.

[14] F. J. Hickernell, Lattice rules: how well do they measure up?, in: Random and quasi-random point sets, Springer, 1998, pp. 109–166.

[15] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, 1992.

[16] P. L'Ecuyer, C. Lemieux, Recent advances in randomized quasi-Monte Carlo methods, Modeling Uncertainty: An Examination of Stochastic Theory, Methods, and Applications (2002) 419–474.

[17] G. Leobacher, F. Pillichshammer, Introduction to Quasi-Monte Carlo Integration and Applications, Springer, 2014.

[18] J. Dick, F. Y. Kuo, I. H. Sloan, High-dimensional integration: the quasi-Monte Carlo way, Acta Numerica 22 (2013) 133–288.

[19] F. Kuo, Component-by-component constructions achieve the optimal rate of convergence for multivariate integration in weighted Korobov and Sobolev spaces, Journal of Complexity 19 (3) (2003) 301–320.

[20] R. Cools, F. Y. Kuo, D. Nuyens, Constructing embedded lattice rules for multivariate integration, SIAM Journal on Scientific Computing 28 (6) (2006) 2162–2188.

[21] S. Heinrich, Multilevel Monte Carlo methods, in: Large-Scale Scientific Computing: Third International Conference, LSSC 2001 Sozopol, Bulgaria, June 6–10, 2001 Revised Papers 3, Springer, 2001, pp. 58–67.

[22] D. Frisch, U. D. Hanebeck, The generalized Fibonacci grid as low-discrepancy point set for optimal deterministic Gaussian sampling, Journal of Advances in Information Fusion 18 (1) (2023) 16–34.

[23] D. Frisch, U. D. Hanebeck, Deterministic gaussian sampling with generalized Fibonacci grids, in: 2021 IEEE 24th International Conference on Information Fusion (FUSION), IEEE, 2021, pp. 1–8.

[24] O. Teytaud, S. Gelly, Dcma: yet another derandomization in covariance-matrix-adaptation, in: Proceedings of the 9th annual conference on Genetic and evolutionary computation, 2007, pp. 955–963.

[25] Q. Long, M. Scavino, R. Tempone, S. Wang, Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations, Computer Methods in Applied Mechanics and Engineering 259 (2013) 24–39.

[26] P. L'Ecuyer, Non-uniform random variate generations. (2011).

[27] P. E. Kloeden, E. Platen, A survey of numerical methods for stochastic differential equations, Stochastic Hydrology and Hydraulics 3 (1989) 155–178.

[28] G. E. Box, M. E. Muller, A note on the generation of random normal deviates, The annals of mathematical statistics 29 (2) (1958) 610–611.

[29] N. Steen, G. Byrne, E. Gelbard, Gaussian quadratures for the integrals, Mathematics of Computation 23 (107) (1969) 661–671.

[30] K. Wu, P. Chen, O. Ghattas, An efficient method for goal-oriented linear Bayesian optimal experimental design: Application to optimal sensor placemen, arXiv preprint arXiv:2102.06627 (2021).

[31] A. Alexanderian, P. J. Gloor, O. Ghattas, On Bayesian A- and D-Optimal Experimental Designs in Infinite Dimensions, Bayesian Analysis 11 (3) (2016) 671 – 695.

6 Uncertainty Quantification in Epidemiological Models

Mathematical modeling of epidemiological diseases using differential equations are of great importance in order to recognize the characteristics of the diseases and their outbreak. The procedure of modeling consists of two essential components: the first component is to solve the mathematical model numerically, the so-called forward modeling. The second component is to identify the unknown parameter values in the model, which is known as inverse modeling and leads to identifying the epidemiological model more precisely.

The main goal of this chapter is to develop the forward and inverse modeling of the coronavirus (COVID-19) pandemic using novel computational methodologies in order to accurately estimate and predict the pandemic. This leads to governmental decisions support in implementing effective protective measures and prevention of new outbreaks. To this end, we use the logistic equation and the SIR (susceptible-infected-removed) system of ordinary differential equations to model the spread of the COVID-19 pandemic.

For the inverse modeling, we propose Bayesian inversion techniques, which are robust and reliable approaches, in order to estimate the unknown parameters of the epidemiological models. We deploy an adaptive Markov-chain Monte-Carlo (MCMC) algorithm for the estimation of a posteriori probability distribution and confidence intervals for the unknown model parameters as well as for the reproduction number. We perform our analyses on the publicly available data for Austria to estimate the main epidemiological model parameters and to study the effectiveness of the protective measures by the Austrian government.

This chapter is based on the author's work in [1].

6.1 Mathematical Models for COVID-19

Predictive mathematical models are essential for the quantitative understanding of epidemics and for supporting decision makers in order to implement the most effective and protective measures. Many mathematical models for the spread of infectious diseases [2, 3, 4, 5] and in particular for the novel COVID-19 [6, 7, 8, 9, 10] have been presented and analyzed. Here we start with the logistic equation as a preliminary model for epidemics and continue with the SIR model and its extensions [11, 12, 13, 14, 15, 16].

6.1.1 The Logistic Model

The logistic equation is a nonlinear ordinary differential equation, which is used for modeling population growth. This ODE is also well-known as logistic growth model and is given by

$$y'(t) = \alpha y(t)\big(1 - \frac{y(t)}{\beta}\big), \qquad y(0) = y_0, \tag{6.1}$$

where $y_0 \neq 0$ is the initial population size (initial number of confirmed cases), y denotes population size (total accumulated confirmed cases) and t time. Furthermore, α and β are respectively the growth rate (infection rate) and the carrying capacity (maximum number of confirmed cases), which are positive constants.

The solution to the logistic model equation is

$$y(t) = \frac{\beta y_0}{y_0 + (\beta - y_0)e^{-\alpha t}},$$

which can be rewritten as

$$y(t) = \frac{\beta}{1 + Ae^{-\alpha t}},$$

where

$$A = \frac{\beta - y_0}{y_0}.$$

The inflection point represents the time that maximal rate of confirmed cases (growth rate) occurs. The inflection point of the logistic function is calculated as

$$t^* = \frac{\ln(A)}{\alpha},$$

where the estimated number of infected people is $\beta/2$. However, there are generalizations of the naïve logistic function, and we will present numerical results based on them in Section 6.3.2.

6.1.2 The Susceptible-Infected-Removed Model

The susceptible-infected-removed (SIR) model is an epidemiological model that computes the number of people infected with a contagious disease in a closed population over time. The Kermack-McKendrick model [17, 18, 19] is one of the SIR models, which is defined by the system

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \tag{6.2a}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \tag{6.2b}$$

$$\frac{dR}{dt} = \gamma I \tag{6.2c}$$

of ordinary differential equations, where β and γ are the infection and recovery rates, respectively. The model consists of three components: S for the number of susceptible, I

for the number of infectious, and R for the number of recovered or deceased (or immune) individuals. Furthermore, N denotes the constancy of population, i.e.

$$S(t) + I(t) + R(t) = N. \tag{6.3}$$

Moreover, the dynamics of the infectious class depends on the basic reproduction number, which is defined as

$$R_0 := \frac{\beta}{\gamma}.$$

If the reproduction number is high, the probability of a pandemic is high, too. This number is also used to estimate the herd immune threshold (HIT). If the reproduction number multiplied by the percentage of susceptible individuals is equal to 1, it shows an equilibrium state and thus the number of infectious people is constant.

Additionally, the recovery period is defined by

$$t_1 = \frac{1}{\gamma}$$

and describes the average days to recover from infection. The transmission period in the sense of the average days to transmit the infection to a person is defined by

$$t_2 = \frac{1}{\beta}.$$

However in a population with vital dynamics, new births can provide more susceptible individuals to the population, which sustain an epidemic or allow new introductions to spread in the population. Taking the vital dynamics into account, the SIR model is extended to

$$\frac{dS}{dt} = \mu N - \frac{\beta IS}{N} - \nu S, \tag{6.4a}$$

$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I - \nu I, \tag{6.4b}$$

$$\frac{dR}{dt} = \gamma I - \nu R, \tag{6.4c}$$

where μ and ν denote the birth and death rates, respectively. To maintain a constant population, we assume $\mu = \nu$ is the natural mortality rate.

6.2 Bayesian Inversion for the Model Parameters

We propose Bayesian inversion methods, in which probabilities are used as a general concept to represent the uncertainty in the model parameters in order to solve the backward/inverse problem of COVID-19, i.e., the problem of accurate estimation of the epidemiological model parameters as well as the reproduction ratio. Bayesian inference in the context of the statistical inversion theory is based on Bayes' Theorem and represents the uncertainty probabilistically by defining a probability distribution over the possible values of the

parameters and uses sample data to update this distribution. Bayesian analysis, in contrast to traditional inverse methods, is a robust inversion technique for determining parameters, yields the (a posteriori) probability distribution, and has the advantage of updating the prior knowledge about the unknown quantity using the measurement/observation data, giving confidence intervals for the unknowns instead of providing a single estimate. We have already successfully applied Bayesian inversion techniques to various PDE models in engineering and medicine in order to identify parameters (see for instance [20, 21, 22, 23]).

6.3 Numerical Results

In this section, we present simulation results of Bayesian inversion and the adaptive MCMC method (see Algorithm 1) for the two epidemic models, namely the logistic and the SIR models, using the data of the COVID-19 outbreak in Austria. The results include model parameter estimation, model validation and outbreak forecasting.

6.3.1 Parameter Estimation

According to Bayesian analysis, the unknown parameters of the logistic and SIR models using the data of COVID-19 outbreak in Austria were found and summarized in Table 6.1 and Table 6.3, respectively. These tables show the confidence intervals for the models parameters as well as the mean of the obtained Markov chains in the Bayesian inference.

Parameter	Description	Estimated mean	Confidence interval 95%
α	Growth rate	0.28	[0.23, 0.33]
β	Carrying capacity	14 974	[12 703, 17 244]

Table 6.1: Estimated confidence intervals and mean of Markov chains for the parameters of the logistic model using Bayesian inversion method for Austria.

Inflection point t^*	Infected cases at inflection point	
Estimated	Estimated	Actual
March 27	7 486	7 697

Table 6.2: Estimated and actual number of the infected cases in Austria at the inflection point of the logistic model.

Furthermore, Tables 6.2 and 6.4 include temporal quantities such as inflection time of the outbreak estimated using the Bayesian inference for the logistic and SIR models. According to our analysis, March 27 is estimated as the inflection point of the outbreak in Austria, when the maximal rate of confirmed cases (growth rate) occurs. The estimated total number of confirmed cases till the inflection point agrees with measured data (see Table 6.2).

Figure 6.2 illustrates marginal histograms of posterior distribution for the three quantities of interest in the SIR model, namely β, γ and R_0 using Bayesian analysis.

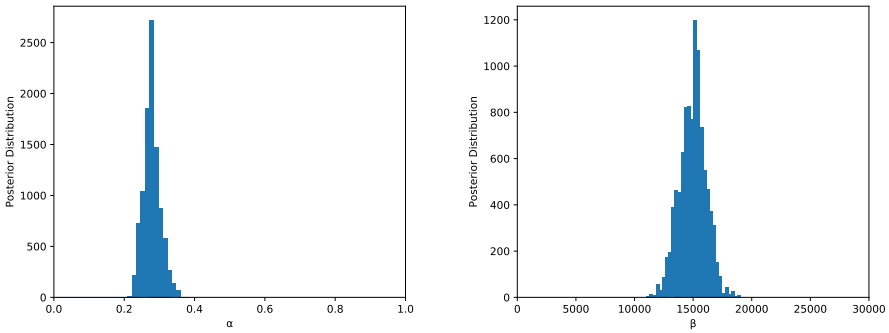

Figure 6.1: The marginal histograms of posterior distribution for the two quantities of interest in the logistic model, namely α and β.

Parameter	Description	Estimated mean	Confidence interval 95%
β	Transmission rate	0.36	[0.32, 0.39]
γ	Recovery rate	0.06	[0.03, 0.09]

Table 6.3: Estimated confidence intervals and means of Markov chains for the parameters of the SIR model using Bayesian inversion method for Austria.

Quantity	Description	Average Estimation (days)
t_1	Infectious period	16.7
t_2	Transmission period	2.8

Table 6.4: Estimated temporal quantities for Austria using the SIR model.

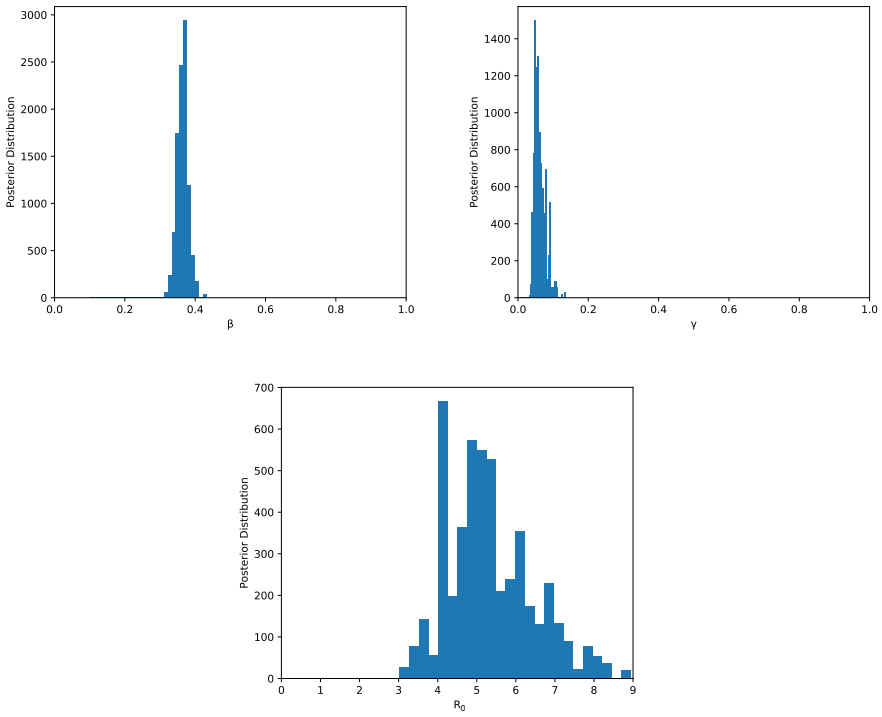

Figure 6.2: The marginal histograms of posterior distribution for the three quantities of interest in the SIR model, namely β, γ and R_0.

6.3.2 Model Validation

Here, we aim to validate the logistic and SIR models for forecasting the COVID-19 outbreak by comparing the Bayesian simulation results and the actual data. Figure 6.3 illustrates the actual number of infected individuals in Austria till now, as well as the estimated number of infected people according to the Bayesian inversion for the logistic equation.

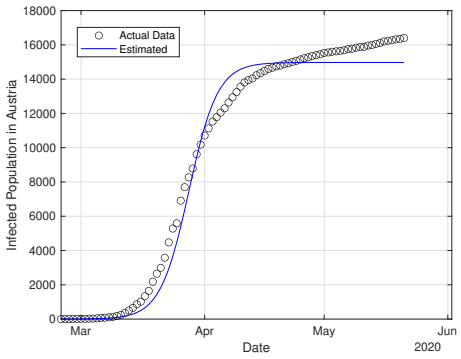

Figure 6.3: Estimated total cumulative count of coronavirus confirmed infected cases in Austria using Bayesian analysis for the logistic model versus actual or measured infected population.

Figure 6.4 displays a similar estimation using the Bayesian inference for the SIR model, which shows a very good agreement between the measurements and the simulation.

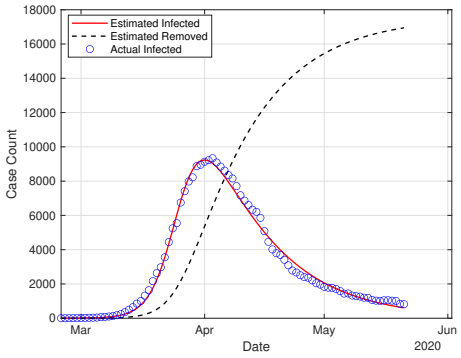

Figure 6.4: Estimated total count of coronavirus infected and recovered cases using Bayesian analysis for the SIR model as well as actual confirmed active cases in Austria.

Figure 6.5 illustrates the number of infected people estimated using the Gompertz function and a generalized logistic function. The Gompertz function is a sigmoid function which describes population growth and it is defined by

$$G(t) := ae^{-be^{-ct}}, \tag{6.5}$$

where a is an asymptote, and b and c are the negative growth rates. The second derivative of $G(t)$ gives the tipping point $t = -\log(-b))/c$ in which the growth of daily cases will start

to decrease. In contrast to the logistic function, the Gompertz function does not have a symmetrical first derivative at the inflection point.

As another more complex generalization of logistic method, we apply the model

$$P(t) := \frac{M}{(1 + e^{-at+b})^{\alpha}}, \tag{6.6}$$

where a is growth rate and M is the carrying capacity. Furthermore, b is multiplication of the growth rate and unknown inflection time. Moreover, $\alpha = 1/\nu$, where $\nu > 0$ affects near which asymptote maximum growth occurs.

The comparison shows a better agreement between the measured data and both of the Gompertz estimation and the generalized logistic function than the naive logistic function (see Figures 6.3 and 6.5).

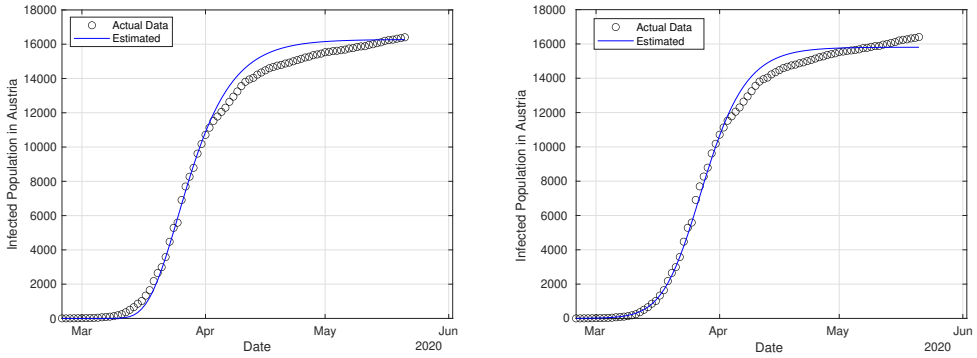

Figure 6.5: Estimated total cumulative count of infected cases using the Gompertz function (left) and a generalized logistic function (right) versus the observed infected population in Austria.

Quantifying the uncertain parameters such as the reproduction number leads to calculate the average number of days to recover from the infection and gives useful information about properly and accurately implementing protective measures in order to prevent the spread of the virus. Furthermore, the parameter identification in the epidemic model makes it possible to predict the length of the pandemic, the number of infected individuals and the fatality rate.

In Figure 6.6, the actual and estimated infection rates are depicted. This rate is defined by

$$\text{infection rate} := \frac{\Delta I_n}{I_{n-1}}, \qquad n \in \{2, 3, \ldots\}, \tag{6.7}$$

where $\Delta I_n = I_n - I_{n-1}$, I_n and I_{n-1} are infected population of consequent times (e.g. in days or weeks), which are obtained using the estimated infection from the SIR model. Figure 6.7 shows the estimated and actual reproduction number $R(t)$ at day t for Austria, i.e. the average number of people someone infected at time t would infect over their entire infectious lifespan. In order to calculate $R(t)$, we apply the formula

$$R(t) = \frac{I_{\text{inc}}(t)}{\sum_{\tau=1}^{t_1^*} \omega(t) I_{\text{inc}}(t-\tau)}, \tag{6.8}$$

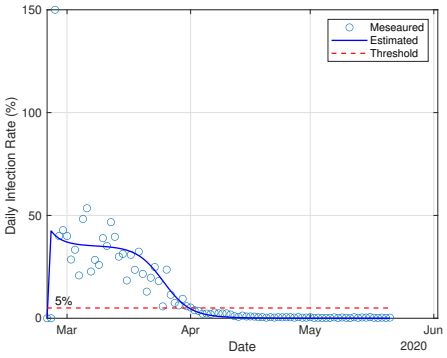

Figure 6.6: Infection rate in Austria.

where $I_{\text{inc}}(t)$ is the number of incident cases at time t and $t_1^* = 17$ is the average estimated infectious period in days (see Table 6.4). The normal distributed function ω specifies the so-called infectivity profile during the infectious period [24, 25]. The first recovery in Austria was reported on March 26, where the reproduction number is estimated around 3. This estimated quantity decays to 1 in the beginning of April, and since then remains below 1 (till the time of writing this paper on May 21, 2020). The measured $R(t)$ is below 1 in April, then oscillates around 1 for the rest of time frame in this study. In this figure, the threshold of $R = 1$ is also displayed in the sense that there is no immediate public health emergency any more when the reproduction number is below this threshold.

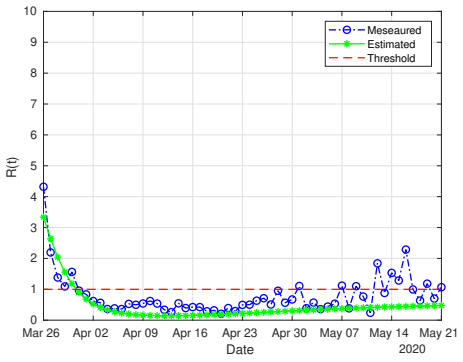

Figure 6.7: Reproduction number in Austria.

6.3.3 Fatality Analysis

Social-distancing and other protective measures started in Austria around March 16, 2020 in order to slow down the outbreak and consequently to prevent an increase in fatalities by keeping the cases that require hospitalization below the capacity of the healthcare system. In Figure 6.8, the daily fatalities in Austria as well as the relative change in fatalities are illustrated.

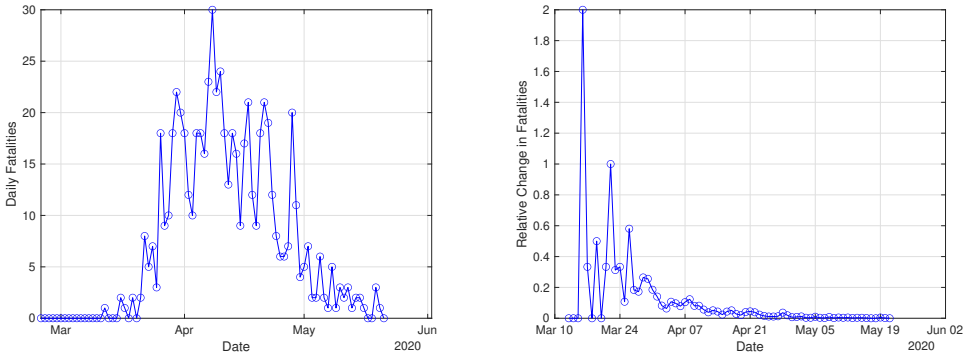

Figure 6.8: Fatalities in Austria including the daily fatalities (left) and the relative change in fatalities (right).

Applying the fatality ratio as well as confirmed infected cases, we present a fatality analysis which is of importance for governmental protective decision making. In epidemiology, a case fatality rate (CFR) is the proportion of deaths from a certain disease compared to the total number of people diagnosed with the disease for a certain period of time. Figure 6.9 depicts the case fatality rate (CFR), which is defined by

$$\text{CFR} := \frac{\text{fatalities}}{\text{fatalities} + \text{recoveries}}, \tag{6.9}$$

and we call it the true CFR here, in contrast to the naive CFR defined by

$$\text{naive CFR} := \frac{\text{fatalities}}{\text{infections}}. \tag{6.10}$$

The straightforward calculations using the recorded data in Austria show that both CFR and naive CFR converge to the same value of $\text{CFR}^* = 0.04$ (Figure 6.9).

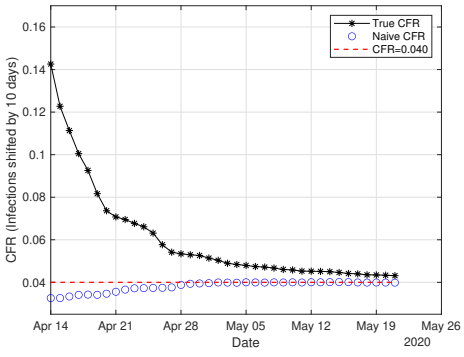

Figure 6.9: Case fatality rate (CFR) in Austria.

We can roughly predict the fatalities using the confirmed infections, a shift, and the CFR. The shift is approximately equal to the time between infection and death/recovery (currently

average estimated to be 17 days (see Table 6.4)) minus the incubation time (currently estimated to be 5–6 days [26]) minus 1 day for testing and reporting (see Figure 6.10). The average time between infection and death is reported approximately 17.8 days in [27]. The estimator of fatalities in Austria is defined by

$$\text{fatality cases} := \text{CFR}^* \times \text{confirmed infected cases (shifted by 10 days)}.$$

Figure 6.10 shows a good agreement between the estimated fatalities and true values in Austria.

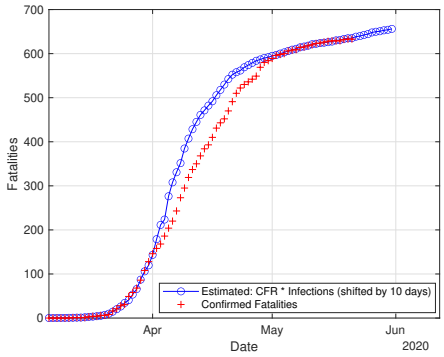

Figure 6.10: Prediction of number of fatalities in Austria.

6.3.4 The Impact of Governmental Protective Measures

Here, in order to study the effect of the protective measures implemented by the Austrian government, we compare the infection rate and the infected population in different time intervals with and without implementing the measures. Although public health measures were in place from March 16 to control the spread of COVID-19, Austrians started to practice social distancing in advance. Table 6.5 shows the weekly infection rate in Austria and how it decays in subsequent weeks. The comparison between the estimated infection rates in subsequent weeks before and after implementing the protective measures highlights the importance and effectiveness of the measures such as social-distancing and lock-down in controlling and slowing down the spread of COVID-19.

The main goal of protective measures and lock-down is to "flatten the curve", i.e., to decrease the infection rate so that the healthcare system is kept from becoming overwhelmed with too many critical cases at the same time. In countries where the counterfactual scenario i.e., no public health interventions is applied, for instance in Sweden, the ICU demand is estimated to be almost 20 times higher than the intensive care capacity in the country and a much larger number of deaths is predicted [28]. However according to Institute for Health Metrics and Evaluation (IHME) database (http://www.healthdata.org/), France, Italy, and Spain, where lock-downs were enacted, also experienced problems due to the limited numbers of ICU beds. In early April, these countries experienced an overload in their health care systems.

	without measures	with measures			
Time interval	March before 18th	March 18–24	March 25–31	April 1–7	April after 7th
Infection rate	7.30	3.97	1.93	1.2	≈ 1
Infected population	1 332	5 283	10 180	12 639	14 000–17 000

Table 6.5: Infection rate (see Equation (6.7)) and total cumulative infected population at the end of different time intervals with and without implementing the protective measures in Austria.

The fatality forecast in Section 6.3.3 is valid as long as protective measures are in place, otherwise the number of fatalities will increase due to a large number of infected people and the limit in the capacity of intensive care unit (ICU) beds as the number of intensive cases increases dramatically. According to the website of the Federal Ministry of Social Affairs, Health, Care and Consumer Protection of Austria, around 700 ICU beds have been available on average for COVID-19 patients since the beginning of the pandemic. Figure 6.11 illustrates the average number of available ICU beds for COVID-19 patients and the reported number of occupied ICU beds in Austria. In early April around 26% and in mid-May around 5% of all available intensive care beds for COVID-19 patients were occupied by these patients. This amount of ICU beds usage shows that approximately 2% of the active infected individuals were critical cases and required ICU beds. If Austrian governmental protective measures would not have taken place and the reproduction number would have remained as before the measures, active infected cases would have increased dramatically. In the case the number of active infected cases would have been around five times larger, the ICU bed capacity could have been exceeded according to the rate of around 2% critical cases in Austria.

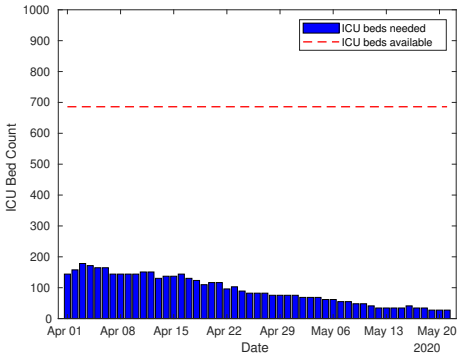

Figure 6.11: The average number of available ICU beds for COVID-19 patients vs. the reported number of occupied ICU beds in Austria.

6.4 Conclusions

In this work, we developed an adaptive Bayesian inversion for epidemiological models, namely the logistic and the SIR models, in order to solve the inverse problem of estimating unknown quantities for the novel coronavirus COVID19. Quantifying the uncertainties in these models is essential since it leads to describe the characteristics of the epidemics on one hand and accurately forecasting the pandemic on the other hand. The proposed inversion recipe is robust and yields probability distributions and confidence intervals for the unknown parameters of the epidemic models including the growth rate of the outbreak and transmission and recovery rates as well as the reproduction number, whose quantification is crucial for decision makers.

We applied our methodology to the publicly available data for Austria to estimate the main epidemiological model parameters and to present a fatality analysis, all of which are of great importance for the government and decision-makers to adopt the most efficient and effective protective measures in order to prevent human and economic damages. We also validated the presented models by comparing the simulated and measured data, whose results show a very good agreement.

Based on Bayesian analysis for the logistic model, the means of the growth rate α and the carrying capacity β are estimated respectively as 0.28 and 14 974. Furthermore for these parameters, 95% confidence intervals of $[0.23, 0.33]$ and $[12\,703, 17\,244]$ are obtained. Moreover for the parameters of the SIR model, namely the transmission rate β and recovery rate γ the means of 0.36 and 0.06 as well as the 95% confidence intervals of $[0.32, 0.39]$ and $[0.03, 0.09]$ are inferred. Additionally, we obtained the infectious period of 17 days and transmission period of 3 days for COVID-19 in Austria. The first recovery in Austria was reported on March 26, where the reproduction number is estimated around 3. This estimated quantity decays to 1 in the beginning of April, and since then remains below 1 (till the time of writing this paper on May 21, 2020). The measured $R(t)$ is below 1 in April, then oscillates around 1 for the rest of time frame in this study.

Analyzing data of infected, recovered and death cases, we obtained that the case fatality rate (CFR) has converged to the value 4%. This estimation makes it possible to forecast the fatalities in the coming 10 days. According to our analysis, the total number of death in Austria is estimated as 633 in May 21, which perfectly matches the measured data, according to the Johns Hopkins CSSE database.

Furthermore, we estimated the infection rate for consequent weeks starting from before implementing the protective measures, which shows a significant decay after the measures are in place. Moreover, the ICU bed usage shows that approximately 2% of the active infected individuals were critical cases and needed ICU beds. If Austrian governmental protective measures would not have taken place and the reproduction number would remain as before the measures, active infected cases would increase dramatically. In the case that the number of active infected cases was around five times larger, the ICU bed capacity could have been exceeded according to the rate of around 2% critical cases in Austria. These results indicate the impact of the measures such as social distancing and lock-down in controlling the spread of COVID-19.

Bibliography

[1] L. Taghizadeh, A. Karimi, C. Heitzinger, Uncertainty quantification in epidemiological models for the COVID-19 pandemic, Computers in Biology and Medicine 125 (2020) 104011.

[2] R. M. Anderson, B. Anderson, R. M. May, Infectious diseases of humans: dynamics and control, Oxford University Press, 1992.

[3] O. Diekmann, J. A. P. Heesterbeek, Mathematical epidemiology of infectious diseases: model building, analysis and interpretation, Vol. 5, John Wiley & Sons, 2000.

[4] H. W. Hethcote, The mathematics of infectious diseases, SIAM Review 42 (4) (2000) 599–653.

[5] F. Brauer, C. Castillo-Chavez, C. Castillo-Chavez, Mathematical models in population biology and epidemiology, Vol. 2, Springer, 2012.

[6] T. Zhou, Q. Liu, Z. Yang, J. Liao, K. Yang, W. Bai, X. Lu, W. Zhang, Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV, Journal of Evidence-Based Medicine (2020).

[7] L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong, Epidemic analysis of COVID-19 in China by dynamical modeling, arXiv preprint arXiv:2002.06563 (2020).

[8] Y. Chen, J. Cheng, Y. Jiang, K. Liu, A time delay dynamical model for outbreak of 2019-nCoV and the parameter identification, Journal of Inverse and Ill-posed Problems 28 (2) (2020) 243–250.

[9] I. Nesteruk, Statistics based predictions of coronavirus 2019-nCoV spreading in mainland China, MedRxiv (2020).

[10] V. Verma, R. K. Vishwakarma, A. Verma, D. C. Nath, H. T. Khan, Time-to-death approach in revealing chronicity and severity of COVID-19 across the world, PLoS ONE 15 (5) (2020) e0233074.

[11] G. Chowell, N. W. Hengartner, C. Castillo-Chavez, P. W. Fenimore, J. M. Hyman, The basic reproductive number of Ebola and the effects of public health measures: the cases of Congo and Uganda, Journal of Theoretical Biology 229 (1) (2004) 119–126.

[12] H. Wang, Z. Wang, Y. Dong, R. Chang, C. Xu, X. Yu, S. Zhang, L. Tsamlag, M. Shang, J. Huang, et al., Phase-adjusted estimation of the number of coronavirus disease 2019 cases in Wuhan, China, Cell Discovery 6 (1) (2020) 1–8.

[13] W. Jumpen, B. Wiwatanapataphee, Y. Wu, I. Tang, A SEIQR model for pandemic influenza and its parameter identification, International Journal of Pure and Applied Mathematics 52 (2) (2009) 247–265.

[14] X. Bai, Optimization of prognostication model about the spread of Ebola based on SIR model, in: 2016 6th International Conference on Machinery, Materials, Environment, Biotechnology and Computer, Atlantis Press, 2016.

[15] C. Anastassopoulou, L. Russo, A. Tsakris, C. Siettos, Data-based analysis, modelling and forecasting of the COVID-19 outbreak, PloS ONE 15 (3) (2020) e0230405.

[16] G. Giordano, F. Blanchini, R. Bruno, P. Colaneri, A. Di Filippo, A. Di Matteo, M. Colaneri, Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy, Nature Medicine (2020) 1–6.

[17] W. O. Kermack, A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character 115 (772) (1927) 700–721.

[18] W. O. Kermack, A. G. McKendrick, Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity, Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character 138 (834) (1932) 55–83.

[19] W. O. Kermack, A. G. McKendrick, Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 141 (843) (1933) 94–122.

[20] B. Stadlbauer, A. Cossettini, D. Pasterk, P. Scarbolo, L. Taghizadeh, C. Heitzinger, L. Selmi, Bayesian estimation of physical and geometrical parameters for nanocapacitor array biosensors, Journal of Computational Physics 397 (2019) 108874.

[21] L. Taghizadeh, A. Karimi, B. Stadlbauer, W. J. Weninger, E. Kaniusas, C. Heitzinger, Bayesian inversion for electrical-impedance tomography in medical imaging using the nonlinear Poisson–Boltzmann equation, Computer Methods in Applied Mechanics and Engineering 365 (2020) 112959.

[22] L. Taghizadeh, A. Karimi, E. Presterl, C. Heitzinger, Bayesian inversion for a biofilm model including quorum sensing, Computers in Biology and Medicine (2019) 103582.

[23] E. K. Lenzi, L. R. Evangelista, L. Taghizadeh, D. Pasterk, R. S. Zola, T. Sandev, C. Heitzinger, I. Petreska, Reliability of Poisson–Nernst–Planck anomalous models for impedance spectroscopy, The Journal of Physical Chemistry B 123 (37) (2019) 7885–7892.

[24] C. Fraser, Estimating individual and household reproduction numbers in an emerging epidemic, PloS One 2 (8) (2007) e758.

[25] T. Hotz, M. Glock, S. Heyder, S. Semper, A. Böhle, A. Krämer, Monitoring the spread of COVID-19 by estimating reproduction numbers over time, ArXiv preprint arXiv:2004.08557 (2020).

[26] W. H. O. (WHO), Coronavirus Disease 2019 (COVID-19) Situation Report - 73, WHO, 2020.

[27] R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. G. Walker, H. Fu, et al., Estimates of the severity of coronavirus disease 2019: a model-based analysis, The Lancet Infectious Diseases (2020).

[28] H. Sjödin, A. F. Johansson, Å. Brännström, Z. Farooq, H. K. Kriit, A. Wilder-Smith, C. Åström, J. Thunberg, J. Rocklöv, Covid-19 health care demand and mortality in Sweden in response to non-pharmaceutical (NPIs) mitigation and suppression scenarios, MedRxiv https://doi. org/10.1101/2020.03 20 (2020).