Department of Mathematics and Systems Analysis

Bayesian Optimal Experimental Design in Imaging

Juha-Pekka Puska

Aalto University

DOCTORAL THESES

Aalto University publication series
DOCTORAL THESES 230/2023

Bayesian Optimal Experimental Design in Imaging

Juha-Pekka Puska

A doctoral thesis completed for the degree of Doctor of Science (Technology) to be defended, with the permission of the Aalto University School of Science, at a public examination held at the lecture hall M1 of the school on 26 January 2024 at 12:00.

Aalto University
School of Science
Department of Mathematics and Systems Analysis

Supervising professor
Professor Nuutti Hyvönen, Aalto University, Finland

Thesis advisor
Professor Nuutti Hyvönen, Aalto University, Finland

Preliminary examiners
Senior Lecturer Andrew Duncan, Imperial College London, UK
Professor Marko Vauhkonen, University of Eastern Finland, Finland

Opponent
Assistant Professor Xun Huan, University of Michigan, USA

Aalto University publication series
DOCTORAL THESES 230/2023

ISBN 978-952-64-1608-3 (printed)
ISBN 978-952-64-1609-0 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)
http://urn.fi/URN:ISBN:978-952-64-1609-0

Unigrafia Oy
Helsinki 2023

Finland

NORDIC SWAN ECOLABEL
Printed matter
4041-0619

Author
Juha-Pekka Puska

Name of the doctoral thesis
Bayesian Optimal Experimental Design in Imaging

Publisher School of Science

Unit Department of Mathematics and Systems Analysis

Series Aalto University publication series DOCTORAL THESES 230/2023

Field of research Mathematics

Manuscript submitted 13 September 2023 **Date of the defence** 26 January 2024

Permission for public defence granted (date) 8 December 2023 **Language** English

☐ **Monograph** ☒ **Article thesis** ☐ **Essay thesis**

Abstract

An inverse problem is defined as a problem that violates one of the classical criteria of a well posed problem: a solution exists, is unique, and is continuous with respect to the data in some reasonable topology. A problem that is not well posed is called *illposed*, and the development of tools to tackle illposed problems is the goal of the field of inverse problems research.

In imaging, illposedness is often an inevitable consequence of the high dimension of the unknown, compared with the measurement data. In an imaging problem, one aims to reconstruct the spatial two- or three-dimensional structure of an object of interest, leading to unknown parameters in the hundreds of thousands or beyond, while the dimension of the measurement data is determined by the number of sensors, and thus limited by physical constraints to values often at least an order of magnitude lower. Another consequence of the high dimensionality of the problem is the computational cost involved in the computations.

In imaging problems, there is also usually a cost involved in acquiring data, and thus one would naturally want to minimize the amount of data collection required. One tool for this is *optimal experimental design*, where one aims to perform the experiment in a way as to maximize the value of the data obtained. The challenge of this however, is that the search for this optimal design usually leads to a computationally challenging problem, whose size is dependent on the dimension of both the data and the unknown. Overcoming this difficulty is the main objective of this thesis. The problem can be tackled by using Gaussian approximations in the formulation of the imaging problem, which leads to practical solution formulas for the quantities of interest. In this thesis, tools are developed to enable the efficient computation of expected utilities for certain measurement designs, particularly in sequential imaging problems and for non-Gaussian prior models. Additionally, these tools are applied to medical imaging and astronomy.

Keywords Inverse problem, Bayesian modeling, optimal experimental design, computed tomography, magnetorelaxometry imaging, adaptive optics

ISBN (printed) 978-952-64-1608-3 **ISBN (pdf)** 978-952-64-1609-0

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

Location of publisher Helsinki **Location of printing** Helsinki **Year** 2023

Pages 150 **urn** http://urn.fi/URN:ISBN:978-952-64-1609-0

Tekijä
Juha-Pekka Puska

Väitöskirjan nimi
Bayesiläinen optimaalinen koesuunnittelu kuvantamisessa

Julkaisija Perustieteiden korkeakoulu

Yksikkö Matematiikan ja systeemianalyysin laitos

Sarja Aalto University publication series DOCTORAL THESES 230/2023

Tutkimusala Matematiikka

Käsikirjoituksen pvm 13.09.2023 **Väitöspäivä** 26.01.2024

Väittelyluvan myöntämispäivä 08.12.2023 **Kieli** Englanti

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

Tiivistelmä

Inversio-ongelma määritellään ongelmana, joka rikkoo yhtä hyvin asetetun ongelman klassisista kriteereistä: ongelmalla on ratkaisu, joka on yksikäsitteinen ja jatkuva datan suhteen jossain järkevässä topologiassa. Ongelman, joka ei ole hyvin asetettu, sanotaan olevan *huonosti asetettu*, ja huonosti asetettujen ongelmien ratkaisemiseen soveltuvien työkalujen kehitys on inversio-ongelmien tutkimusalan tavoite.

Kuvantamisessa tuntemattoman muuttujan korkea dimensio verrattuna mittausdatan dimensioon johtaa usein siihen, että ongelma on huonosti asetettu. Kuvantamisongelmassa yritetään rekonstruoida kohteena olevan kappaleen kaksi- tai kolmiulotteinen rakenne, minkä seurauksena tuntemattomien parametrien määrä voi olla satojatuhansia, tai jopa ylikin. Mittausdatan dimension sen sijaan määrittää sensorien lukumäärä, joten fyysisten rajoitteiden seurauksena se on yleensä vähintään kertaluokkaa matalampi. Toinen seuraus ongelman korkeasta dimensiosta on ratkaisemisen laskennallinen vaativuus.

Kuvantamisongelmassa datan keräämisestä seuraa myös kustannuksia, joten luonnollisena tavoitteena on minimoida tarvitun datan määrä. Yksi työkalu tähän on *optimaalinen koesuunnittelu*, jossa yritetään suorittaa koe niin, että saadun datan arvo olisi mahdollisimman suuri. Haasteena tässä on, että optimaalisen koeasetelman etsiminen johtaa yleensä laskennallisesti haastavaan ongelmaan, jonka koko riippuu sekä datan että tuntemattoman dimensiosta. Tämän haasteen ratkaiseminen on tämän väitöskirjan pääasiallinen tavoite. Ongelmaa voidaan lähestyä käyttämällä gaussisia approksimaatioita ongelman asettelussa, mikä johtaa käytännöllisiin ratkaisukaavoihin kohdemuuttujille. Tässä väitöskirjassa kehitetään työkaluja, jotka mahdollistavat mittausasetelmien odotusarvoisten hyötyjen tehokkaan laskennan, erityisesti sekventiaalisille kuvantamisongelmille ja ei-gaussisille priorimalleille. Lisäksi näitä työkaluja sovelletaan sekä lääketieteelliseen kuvantamiseen että tähtitieteeseen.

Avainsanat Inversio-ongelma, Bayesiläinen mallinnus, optimaalinen koesuunnittelu, tomografiakuvantaminen, tietokonetomografia, magnetorelaxometriakuvantaminen, adaptiivinen optiikka

ISBN (painettu) 978-952-64-1608-3 **ISBN (pdf)** 978-952-64-1609-0

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

Julkaisupaikka Helsinki **Painopaikka** Helsinki **Vuosi** 2023

Sivumäärä 150 **urn** http://urn.fi/URN:ISBN:978-952-64-1609-0

Preface

Tämä väitöskirja on eräänlainen päätös tai vähintäänkin välietappi matkalle, joka on vienyt minut suomalaisen koulutusjärjestelmän läpi. Erityisesti se kuvastaa aikaa, jonka olen viettänyt matematiikan laitoksella Inversio-ongelmien tutkimusryhmässä. Suurin kiitos kuuluu ohjaajalleni Nuutti Hyvöselle, joka tarjosi minulle tämän mahdollisuuden ja on ollut kärsivällisin ja ymmärtäväisin ohjaaja, jonka voin kuvitella. Kiitos kuuluu myös muille, joiden kanssa olen saanut tehdä tutkimusta tai vain jakaa arkipäiviä laitoksella.

Jo aiemmin matkan varrella Otaniemessä minulla on ollut vertaistukea vuosikurssiltani, ja heidän kanssaan olen jakanut matkan matematiikan peruskursseilta tutkijoiksi asti. Tästä ryhmästä on erityisesti mainittava Henri Salmenjoki, Kristian Moring ja Topi Paananen.

Myös vähintään yhtä korvaamaton on ollut se tuki, jonka olen saanut kavereiltani. Heidän osaltaan kyse ei ole ollut niinkään vertaistuesta - matemaatikon murheita on usein vaikea jakaa muiden kanssa. Sen sijaan he ovat tuoneet työlle vastapainoa sekä kärsivällisesti kuunnelleet ja kannustaneet. Kiitos siis teille Tomi, Aatos, Santtu, Ville, Martta, Mikko, Luukas ja monet muut.

Nuoria tutkijoita vaivaa jatkuvasti epävarmuus itsestään ja pelko riittämättömyydestään. Minulle korvaamaton apu tähän ovat olleet vanhempani, jotka ovat aina osoittaneet horjumatonta luottamusta minuun. Erityisesti isäni Pekka Puska on suurin esikuvani. Hänen esimerkkiään seuraten olen osannut keskittyä olennaiseen ja suhtautua tutkijan elämän mutkiin ja yllätyksiin riittävällä huumorilla.

Lopuksi haluaan kiittää Jeminaa, joka on ollut vierelläni koko väitöskirjatyön ajan. Olemme nyt jo luoneet unohtumattomia muistoja yhdessä, ja tulevaisuus tuo varmasti tullessaan uusia seikkailuja.

Paris, December 12, 2023,

Juha-Pekka Puska

Contents

List of Publications

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

I M Burger, A Hauptmann, T Helin, N Hyvönen and J-P Puska. Sequentially optimized projections in x-ray imaging. *Inverse Problems*, 37 075006, June 2021.

II T Helin, N Hyvönen and J-P Puska. Edge-promoting adaptive Bayesian experimental design for X-ray imaging. *SIAM Journal on Scientific Computing*, 44(3) B506-B530, May 2022.

III T Helin, N Hyvönen, J Maaninen and J-P Puska. Bayesian design of measurements for magnetorelaxometry imaging. *Inverse Problems*, 39 125020, November 2023.

IV J Nousiainen, J-P Puska, T Helin, N Hyvönen, M Kasper. The power of prediction: spatiotemporal Gaussian process modeling for predictive control in slope-based wavefront sensing. Submitted to *Journal of Astronomical Telescopes, Instruments, and Systems*, October 2023.

Author's Contribution

Publication I: "Sequentially optimized projections in x-ray imaging"

Puska implemented most of the computer codes, performed the numerical experiments, and had the main responsibility for the writing process together with Hyvönen. Hyvönen, Puska and Helin designed the algorithms and numerical experiments. All authors participated in inventing the main idea behind the article.

Publication II: "Edge-promoting adaptive Bayesian experimental design for X-ray imaging"

Puska implemented most of the computer codes, performed the numerical experiments, and had the main responsibility for the the writing process together with Hyvönen. All authors participated in designing the algorithms and the numerical experiments.

Publication III: "Bayesian design of measurements for magnetorelaxometry imaging"

The original computer codes were written by Maaninen for his master's thesis. Puska modified the codes, implemented the numerical experiments, and had the main responsibility for the the writing process together with Hyvönen. All authors participated in designing the employed algorithms and the numerical experiments.

Publication IV: "The power of prediction: spatiotemporal Gaussian process modeling for predictive control in slope-based wavefront sensing"

The main research topic was proposed by Nousiainen, Helin and Kasper, with Nousiainen having the main responsibility for the project. Puska worked with Nousiainen to develop the approach. In particular, Puska was responsible for the parts involving optimal experimental design, but he also implemented and performed other numerical experiments. All authors contributed to the writing process.

Abbreviations

AO Adaptive optics

CT Computed tomography

CM Conditional mean

DM Deformable mirror

DNN Deep neural network

DP Dynamic programming

FBP Filtered backprojection

FF-GP Frozen flow Gaussian process

KL Kullback-Leibler

LD Lagged-diffusivity

MAP Maximum a posteriori

MDP Magnetic drug targeting

MNP Magnetic nanoparticle

MRXI Magnetorelaxometry imaging

OED Optimal experimental design

TV Total variation

WAFF-GP Wind-averaged frozen flow Gaussian process

WFS Wavefront sensor

Symbols

Ω Object domain

σ Standard deviation parameter of noise

γ Standard deviation parameter of prior

π_{pr} Prior probability density

π_{post} Posterior probability density

π_{noise} Noise probability density

Γ_{pr} Prior covariance (resp. posterior, noise covariance)

x_{pr} Prior mean (resp. posterior, noise mean)

x_{MAP} Maximum a posteriori estimate

x_{CM} Conditional mean estimate

1. Introduction

A challenge of all computational scientists in their work is to be able to efficiently solve the problems considered using sophisticated numerical methods, and the field of inverse problems is no exception. An inverse problem usually involves working with high-dimensional variables, and the problems may be governed by models that are computationally expensive to evaluate. Thus the limits of computational resources may restrict the viability of many methods and prevent them from being used in practice.

Technological advance in computing machines does not seem to be able to definitively solve this problem. The increase in the speed of computer processors is slower than in the past, and it seems that physical limits will constrain even this small increase in the future.

Moreover, the well-known problems in numerics have been extensively studied by a large number of researchers and without a major breakthrough, it seems unlikely that large gains will follow from that direction. This has raised interest in another research direction: trying to improve the informativeness of the data so that one could then "do the same with less". That is, collect less data, but still get accurate estimates. The result would be lower computational demands meaning that the methods could be applied with less resources, or alternatively the fidelity could be increased, without raising the costs.

In some sense, all researchers at least implicitly try to perform their experiments so that the results would give the maximum information on the phenomenon being studied: in testing for new drugs, doctors give their patients dosages that give information on the response of the patients, atmospheric scientists position their measurement devices in places that are thought to be of special interest and will give useful results, and social scientists split subjects into test groups so that meaningful differences between the groups can be distinguished. This is the practice of designing an experiment.

Being able to get by with less data has another benefit as well, since collecting data often comes with a cost. This cost can be a literal monetary cost: paying test subjects or paying for time on expensive measurement

devices such as synchrotrons or space telescopes. It can also be other negative effects that one would wish to minimize. This commonly happens in medical applications: giving patients experimental drugs carries a risk of negative side effects and imaging patients with ionizing radiation is associated with an increased risk of cancer.

Bringing this line of thought of experiment design into more formal setting has resulted in the field of *optimal experimental design* (OED). It has been used for example to position sensors in electrical machines to determine iron loss distribution [29], monitor and control gas networks [71], optimize electrode positions in electrical impedance tomography [36], position boreholes in a groundwater contamination study [51], and find optimal measurement points to determine time of death [69].

Applying the method to an inverse problem however presents challenges. Finding an optimal design involves performing an optimization procedure in which the inverse problem is a subproblem in the complete algorithm. Since inverse problems, particularily in imaging applications can be computationally demanding, the requirement of possibly having to repeatedly solve them can make the optimization prohibitively expensive. This aspect is especially vital in problems, where the designs cannot be computed offline, that is, before the actual experiment is performed. Instead many applications are inherently sequential, and future measurement designs depend on previously measured data. Thus the design for the next measurement needs to be computed efficiently enough so that the measurements can proceed in a reasonable time. Developing efficient algorithms to compute optimal designs for various inverse problems is an area of active research that offers tools for revealing the full potential of applications that depend on solving them.

1.1 Outline of thesis

This thesis is a contribution to the line of research in optimal experimental design for inverse problems and imaging applications, and it presents algorithms to efficiently compute optimal designs for a certain class of inverse problems. The goal is that the algorithms would be quite general and could be applied to a wide range of inverse problems. Specifying the forward models for each application of course requires expert knowledge on the particular problem, but it is outside the scope of this thesis, and only problems with established models are considered. Some expert knowledge is still required though, as one needs to consider the range of admissible designs and define the design space appropriately.

First in Publication I we present the case of sequential measurements in a linear inverse problem with a Gaussian prior and likelihood model. In this setting, one considers the case where multiple measurements are

taken with the option of adjusting the measurement setup between each step. The model problem used is that of *Computed Tomography* (CT), since it has many features that make it well suited to OED. An experiment in CT involves taking multiple measurements, called *projections* from different sides of the object or patient, so that the detector is moved between each projection. CT is a widely used imaging modality and is well described by a linear model.

As each additional measurement introduces new degrees of freedom to the space of possible designs, the complexity of the optimization problem also grows. While in theory it is possible to optimize over the entire design space simultaneously, in practice computational demands restrict this approach to problems of a low dimension. Instead, we employ a greedy algorithm to optimize each step in sequence. A greedy approach cannot of course guarantee global optimality in general, but we hope to demonstrate through numerical experiments that the optimization still results in designs that are better than reference, naive, or randomized approaches.

Since the assumption of a Gaussian prior is quite limiting, and especially in imaging applications often unrealistic, in Publication II we extend the algorithm to cover the class of certain edge-promoting priors. Edge promoting priors usually describe the characteristics of the expected targets better than Gaussian priors and are widely used in practice. This modification is achieved using the *lagged diffusivity* (LD) iteration, which is a method developed for computing the mode of a posterior distribution when an edge-promoting prior is used, but it also has the attractive feature of generating a Gaussian approximation for the posterior. This again enables the efficient computation of optimal designs, as explained later.

In Publication III we employ the developed algorithm to the problem of *Magnetorelaxometry Imaging* (MRXI). This does not pose major theoretical challenges, since MRXI can also be described by a linear model and the experiment consists of sequentially measuring magnetic fields from different patterns of activations. Despite this, the results are of interest due to the different nature of the imaging modality. The way magnetic fields propagate is very different from that of x-rays, and thus it is to be expected that the optimal designs produced also differ significantly.

Publication IV investigates Bayesian modeling and optimal experimental design in the field of adaptive optics. The evolution of the atmosphere turbulence is modeled as a spatiotemporal Gaussian process and the Bayesian posterior estimate is used to produce predictions for the future state of the deformed wavefront. The behaviour of the system and the predictions are analyzed using numerical simulations and compared to less sophisticated models. In this case, the OED methodology is used to analyze the utility of past data. This is of particular interest in the application of adaptive optics, since the timescale of the dynamics of the system is in the order

of milliseconds, which places significant demands on the computational efficiency of any prediction method.

2. Inverse Problems and Imaging

The common definition of an inverse problem is that it is a problem that does not fulfill at least one of the three criteria for a well-posed problem:

- A solution exists,
- The solution is unique,
- The solution is continuous w.r.t. the data in some reasonable topology.

A problem that is not well-posed is called *illposed*. One can generally think of inverse problems as ones where the aim is to compute a source that generates the observed data [30]. An illposed problem may lack a unique solution even in the idealized case, but commonly difficulties may also arise from it being highly sensitive to measurement noise.

A linear inverse problem can often be stated as the problem of solving the operator equation

$$Tx = y, \tag{2.1}$$

where $T : \mathscr{X} \to \mathscr{Y}$ is a linear operator between two Hilbert spaces. Since the problem is illposed, we generally have $R(T) \neq \mathscr{Y}$ or $N(T) \neq \{0\}$, and thus the inverse T^{-1} does not exist. One must then look for a *generalized inverse* $T^{\dagger}$, for which $T^{\dagger}y$ is close to x in some suitable sense. Loosely speaking, this corresponds to projecting the data y onto the closure of the range of T and only looking for a solution in the orthogonal complement of the nullspace of T. Even if T^{-1} or $T^{\dagger}$ were to exist, it would not in general be continuous (or bounded), due to the problem failing the third criterion of a well-posed problem, which is typical for infinite-dimensional inverse problems.

Illposedness of the problem can be characterized by the singular values of the operator T, assuming it is compact, which is often the case. More specifically, if the operator T has singular values that are very small, any components in the data in the direction of the corresponding subspaces will be amplified in the naive solution process by a large factor.

Inverse problems often arise in situations where indirect measurements are made. The difficulty in inverse problems can be due to so-called glob-

ality, meaning that all parts of the object contribute to the measurement. Also in inverse problems one often goes against the direction of causality, such as when attempting to solve the state of a system at an earlier time. Commonly many different initial conditions may cause the same (or almost the same) measurement. In inverse problems appearing in imaging, the dimension of the data is limited by the number of sensors. Thus the dimension of the unknown will often be much higher, which makes the problem underdetermined.

Illposed problems require special solutions methods to compensate for their difficulties. *Regularization* is the property of a solution method to encourage certain *regularity* or smoothness in the solution. Since the data of an inverse problem is not generally sufficient to fully and/or stably determine the solution, a regularization method only recovers partial information about the solution, and does it as stably as possible [23]. This can also be seen as solving a related problem; one that is uniquely solvable and more robust than the original one. We can often assume a certain smoothness in the solution due to the physical properties of the unknown, for example, the spatial distribution of a density parameter. Thus regularization tends to force the solution into a lower dimensional subspace where the true unknown can be assumed to lie, and remove the illposedness of the problem.

To set the notation for this section, denote by $\Omega \subset \mathbb{R}^d$, $d = 2,3$ the domain of the object, and by $u : \Omega \to \mathbb{R}^+$ a function representing a parameter about which we intend to gather information. In the discrete case, the domain is discretized into pixels (or voxels in the 3D case) indexed as Ω_i, $i = 1,\dots,n$. The discrete version of the unknown function is the vector denoted by $U \in \mathbb{R}^n$.

2.1 Computed Tomography

Computed tomography (CT), also known as x-ray tomography, is a method that uses the ability of x-rays to penetrate through material to permit reconstruction of the interior structure of an object. It is an important tool in medicine and nondestructive testing, since it allows the object to be reconstructed at a very fine resolution [12].

In x-ray tomography, a *source* and a receiver are rotated around the object being investigated. The x-rays originating from the source pass through the object and their intensity is measured by detectors on the other side. As the rays pass through the object, they are attenuated, and the magnitude of the attenuation depends on the absorption coefficient of the object. The attenuation can be described by the Beer–Lambert law

$$I = I_0 \exp\left(- \int_L u \mathrm{d}s \right), \tag{2.2}$$

where I_0 is the initial intensity of the x-ray, u the absorption coefficient of the object/material, and L the line along which the ray travels. In computational studies one commonly takes the logarithm of the relative intensity as measurement data

$$y = \log\left(\frac{I}{I_0}\right) = -\int_L u \mathrm{d}s, \tag{2.3}$$

which makes the relation between y and u linear.

The roots of the theory of tomographic imaging go back to the work of Radon [57] on determining a function through its values along line-integrals. From this work we have the *Radon transform* that maps a function to a set of line integrals through the domain, parametrized by a polar angle θ and a distance s from the origin for the considered line $L(s,\theta)$:

$$(Rf)(s,\theta) = \int_{L(s,\theta)} f(x)\mathrm{d}x. \tag{2.4}$$

It can be shown that given complete projection data, that is, knowing the line integrals for all values of s and θ, the function f can be reconstructed. In reality, only a finite number of projections can be taken. If only relatively few projections are available, the setting is called *sparse tomography*. Also, it may be the case that only a limited subset of the projection angles is available. This situation is called *limited-angle tomography*. Both of these limitations make the problem significantly illposed. For limited-angle data, certain features of the function f cannot be stably determined from the data, whereas for finite projection angles, the function f can be perturbed in an almost arbitrary way without changing the data [66]. Both of these cases require special solution methods, since simple inversion approaches such as *filtered backprojection* (FBP) produce undesirable artifacts [27]. For a thorough introduction to the theory of CT, see also [46] and [47].

In a discrete setting, the domain is discretized into n pixels (or voxels in the 3D case), and the absorption coefficient is assumed to be constant within each pixel. Denoting by $|L \cap \Omega_i|$ the length of the intersection of the line L and the pixel Ω_i, the discrete version of the integral in (2.3) is a sum of the absorption values of each pixel, weighted by the length of the intersection of the pixel and the ray. That is, the j:th data points is given as

$$y_j = \log\left(\frac{I}{I_0}\right) \approx \sum_i u_i |L_j \cap \Omega_i|, \tag{2.5}$$

where u_i is the absorption coefficient of pixel $i \in \{1,\dots,n\}$, y_j is the data measured by sensor $j \in \{1,\dots,m\}$, and L_j denotes the line from the considered source to sensor j. For m detectors, the forward model for one projection image results in a set of m linear equations, the coefficients of which can be collected into a matrix $R \in \mathbb{R}^{m\times n}$. Furthermore, assuming k projections are taken in total, forward matrices $R_1,\dots,R_k$ can be stacked to form the

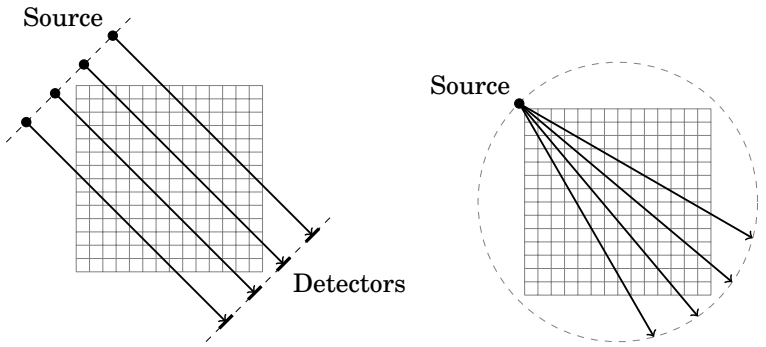

Figure 2.1. Parallel beam (left) and fan beam (right) measurement setups for CT

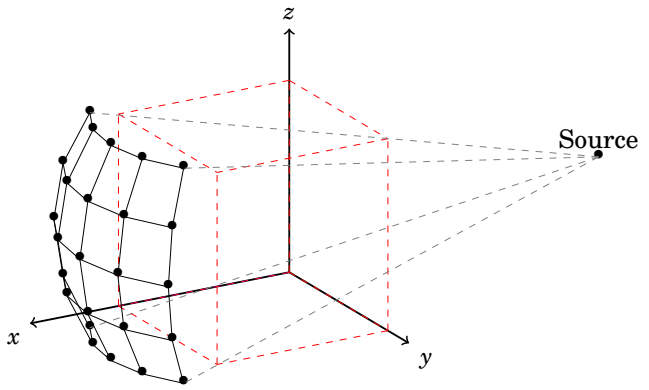

Figure 2.2. Cone beam measurement setup for 3D CT

forward matrix for the entire experiment,

$$\mathbf{R}_k = \begin{pmatrix} R_1 \\ \vdots \\ R_k \end{pmatrix} \in \mathbb{R}^{km \times n}. \tag{2.6}$$

The discretized inverse problem of CT is to recover the absorption coefficient $u \in \mathbb{R}^n$ from the measurement data $y \in \mathbb{R}^{km}$. The full set of measurement data is often called a *sinogram*, due the sine-waves that appear in the image, when the measurement vector is visualized as a 2D-image.

In the measurement setup, the x-rays can be assumed to originate from a single point source, or alternatively each detector can have a corresponding source, so that the rays proceed parallel to each other. The former setup is called *fan beam* geometry and the latter *parallel beam* geometry, both or which are illustrated in 2.1. In the numerical studies of this thesis, the 2D experiments employ a parallel beam geometry, and the 3D experiments a cone beam geometry, shown in Figure 2.2.

2.2 Magnetorelaxometry Imaging

Magnetic nanoparticles (MNPs) are particles with a core consisting of a ferro- or ferrimagnetic material, surrounded by a biocompatible coating. MNPs usually have a core diameter in the range of 1-100 nm, and due to their small size and non-toxicity, they can be safely injected into the human bloodstream. MNPs have a number of promising applications in medicine [53], [52]. Due to the magnetic properties of the core material, MNPs can be manipulated using an external magnetic field and can be used to transport therapeutic agents into desired locations in the body. This process is known as *magnetic drug targeting* (MDT). Strong alternating magnetic fields can also be used to generate heat in the MNPs. If the MNPs can be accurately transported to carcinoma sites, the heating effect can be used to damage the cancer tissue, a procedure known as magnetic hyperthermia.

Both of these approaches depend on accurate knowledge of the spatial distribution of the MNPs. One way to acquire this information is to use the *superparamagnetic* properties of the particles. Using an external magnetic field, the magnetic moments of the particles can be aligned, and thus the particles made to generate a superposition magnetic field. After the external magnetic field has been switched off, the particles do not instantly return to an unaligned state, but do so over a period of time in a process known as *magnetic relaxation*. The speed at which this happens depends on the shape and size of the particles, and is governed by two separate mechanisms: Néel relaxation [48], which depends on the magnetocrystalline anistropy of the core material, and Brownian relaxation [14], which is caused by the rotation of the particles themselves.

The relaxation curve of the magnetic field generated by the MNPs can be measured using sensors and the measurements provide information about the particle distribution. This measurement can be performed in various ways and for various purposes, but all of them fall under the concept of *magnetorelaxometry* (MRX) [70]. A diagram of an MRXI measurement setup is shown in Figure 2.3. A specific technique is using multiple activation coils to sequentially generate inhomogeneous magnetic fields, each of which is measured by a sensor array [41]. Solving the corresponding inverse problem, one can then reconstruct the spatial distribution of the nanoparticles. This is called MRX tomography [42] or *magnetorelaxometry imaging* (MRXI). Recently progress has been made on developing MRXI measurement devices suitable for human applications including a measurement setup with a target object representing a human head [8] and a human torso [64].

The inverse problem of MRXI is put into a formal mathematical setting in [24]. For the forward model, one needs to first define the magnetic field produced by the activation coils. For this, if the coil diameter ρ is assumed

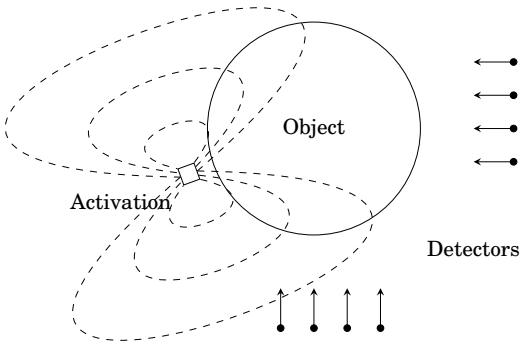

Figure 2.3. Experimental setup for MRXI. The activation coil generates a magnetic field, with the dashed lines denoting the field lines. The detectors (with their orientations being denoted by the arrows) measure the relaxation of the magnetic field of the MNPs.

to be small, the activation coil can be modeled as a dipole at a given point $a \in \mathbb{R}^3$ producing the magnetic field

$$B_a(x) = \pm \frac{\mu_0}{4\pi} \Big(\frac{3(x-a)(x-a)^\top}{|x-a|^5} - \frac{I}{|x-a|^3} \Big) \alpha + O\Big(\frac{\rho^4}{|x-a|^5} \Big), \tag{2.7}$$

where $I \in \mathbb{R}^{d\times d}$ is the identity matrix and $\alpha = \pi\rho^2 J\nu$ is the dipole moment, for which the sign depends on the direction of the current J in the loop and the chosen orientation for the unit normal of the loop ν. The latter term in (2.7) can then be dropped and also the notation simplified by including all the physical constants in α.

If the magnitude of the magnetic field generated by the activation coil is small, the magnetization $M : \Omega \to \mathbb{R}^d$ of the MNPs can be modeled to have linear dependence on B_a.

$$M(x) \propto qB_a(x)c(x), \qquad x \in \Omega, \tag{2.8}$$

where $c(x)$ is the MNP density and $q > 0$ is a constant depending on the properties of the specific MNPs used.

The magnetization of the nanoparticles combine to generate a superposition field, that can be expressed as

$$B_M(w) = \frac{\mu_0}{4\pi} \int_\Omega \Big(\frac{3(w-x)(w-x)^\top}{|w-x|^5} - \frac{I}{|w-x|^3} \Big) M(x)\,\mathrm{d}x, \qquad w \notin \overline{\Omega}. \tag{2.9}$$

Finally, a measurement sensor located at $s \in \mathbb{R}^d \setminus \Omega$ measures the magnetic field along the orientation of the sensor $\sigma \in \mathbb{R}^d$, given altogether as

$$y_{s,\sigma} = \sigma^\top B_M(s) = \sigma^\top \int_\Omega \Big(\frac{3(s-x)(s-x)^\top}{|s-x|^5} - \frac{I}{|s-x|^3} \Big) B_a(x)\,c(x)\,\mathrm{d}x. \tag{2.10}$$

Importantly, in this model the dependence of $y_{s,\sigma}$ on c is linear.

Discretizing the MNP density into pixels/voxels with constant value in each pixel, and numerically evaluating the integral gives a discrete model for the measurement after one activation. Assuming N_s sensors, this model can be written as

$$y = Kc \in \mathbb{R}^{N_s}, \tag{2.11}$$

where $c \in \mathbb{R}^{N_c}$ now represents the pixel values.

The entire experiment, consisting of k measurements with separate activations is then modeled by stacking the corresponding matrices to obtain

$$\mathbf{K} = \begin{pmatrix} K_1 \\ \vdots \\ K_k \end{pmatrix} \in \mathbb{R}^{kN_s \times N_c} \tag{2.12}$$

as the system matrix.

This model ignores a lot of the complexities encountered in a real-world setting. Mainly, the measurement devices measure the relaxation amplitude ΔB_M that is the change in the magnetic field from the moment the activation is switched off, until the point where all the MNPs have returned to an essentially random orientation. However, after the activation coil is switched off, there is a delay before the external magnetic field vanishes and the measurement can begin. This *dead time* means that the measured quantity will be slightly less than the true magnetic field generated by the particles. Also, the model does not use information about the shape of the relaxation curve, which is dependent on the properties of the MNPs used.

2.3 Adaptive optics

When imaging with ground-based telescopes, a major source of error is the distortion caused by the atmospheric turbulence. When entering the atmosphere, light from distant sources has the form of a plane wave. However, when in the atmosphere, the random mixing of air at different temperatures causes variations in the refractive index of the air, altering the optical path length of light and causing the phase of light to be distorted [59].

A way to compensate for this error is the technique of *adaptive optics* (AO) [10], where the telescope is fitted with a *deformable mirror* (DM) that can be used for real time correction of the variations in the phase of the light reaching the telescope. If the DM can be controlled to have the same shape as the distorted wavefront, but with half the amplitude, the resulting wavefront surface is transformed back into a plane wave. The phase of the light is measured with a *wavefront sensor* (WFS). Since the model for sensors used in this thesis does not measure the phase of the light itself,

but only its gradient at certain points, reconstructing the wavefront itself involves the inverse problem of inverting a certain differentiation operator.

The physical limitations of the system cause multiple sources of error: the number of sensors and actuators limits the accuracy of the compensation that can be achieved by an AO system. A finite number of WFSs cannot perfectly measure the shape of the wavefront. Moreover, a real-world deformable mirror inevitably has limited degrees of freedom and thus is unable to perform arbitrary corrections.

Another significant error contribution originates from the time delay error: in the time it takes for the wavefront to be measured, the correction to be computed and the DM to be adjusted, the turbulence profile of the atmosphere has already changed, and the correction applied by the DM is no longer optimal. *Predictive control* aims to address this, by using past measurements on the wavefront to predict its behaviour at future timesteps. Publication IV is a contribution into this line of research and uses a Gaussian process (GP) model for the evolution of the turbulence statistics [58]. While reinforcement learning methods have recently gained attention in predictive control of AO (see e.g. [49], [50], [39], [40], [55]), the GP approach enables analytical estimates on the performance of the correction, systematic incorporation of prior information, and also the use of OED methods to optimize the utilization of measurement data in the prediction.

The atmosphere turbulence is commonly assumed to consist of independent layers, each with a given turbulence strength. In this model, the layers do not mix but are only moved by the wind, with a given strength and direction for each layer. The effect on the measurement that is due to the evolution of the turbulent layers themselves is of a timescale that is assumed to be much longer than the effect of the wind, and thus the layers can be modeled as being static, apart from the lateral movement. This is called the "frozen flow" hypothesis. The work of [22] analyzes this kind of a setting in the GP context, and derives closed form estimates for the prediction error variance. However, the work neglects the spatial correlations of the turbulence, and no simulations are performed to investigate the analytical results. Publication IV of this thesis extends this line of research to spatiotemporal Gaussian models and uses the OED methodology to investigate the optimal use of data for the wavefront prediction.

3. Bayesian Methods

The so called classical methods solve an inverse problem by removing the ill-posedness by assuming a certain smoothness, regularity or property in the solution. This regularization can be made explicit e.g. in a penalty functional, or it can be an implicit feature of the numerical method used to solve the problem.

A way to make this idea more explicit, that is, to specify more clearly what kind of features we would like the unknown to have, is to cast the problem into a probabilistic setting and to define a *prior* probability distribution to encode the knowledge about the unknown. The inverse problem then becomes a problem of inferring the probability distribution of the unknown, which is treated as a random variable, instead of simply an unknown parameter. In Bayesian thinking, the randomness comes from the experimenter's lack of information about the parameter, even if it can be thought to possess a single clearly defined value in reality.

The probabilistic version of the problem is more well-posed than the original. There are many ways to think about why this is the case, but when the inverse problem is thought of as a quest for information it becomes almost obvious: even though the original problem may not be solvable, the question "What do we know about the unknown?" always has an answer, even if that answer is "Not very much". In some cases, a direct connection can be made between the Bayesian and classical methods by interpreting regularization parameters as some parameters of the prior distribution.

Solid introductions to Bayesian methods in inverse problems are [37] and [15]. The theory for Bayesian inverse problems has also been developed for the infinite-dimensional case [67], but in the following we will assume that all variables are of finite-dimension, which is also the setting of Publications I-IV.

To begin with, define the measurement, unknown parameter and noise as random variables $Y, E \in \mathbb{R}^m$ and $X \in \mathbb{R}^n$. The *forward model* $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^m$

then relates the measurement with the unknown and the noise vector

$$Y = f(X, E). \tag{3.1}$$

This relation is often simplified by assuming *additive* noise

$$Y = f(X) + E, \tag{3.2}$$

where X and E are mutually independent. In this case, the conditional density of Y given a realization $X = x$ is simply the density of the noise parameter shifted by $f(x)$, that is

$$\pi(y|x) = \pi_{\text{noise}}(y - f(x)). \tag{3.3}$$

This conditional density is called the *likelihood density*. It gives the probability of observing $Y = y$ if the unknown were to have the value $X = x$. This density encodes the model of the phenomenon being investigated.

As mentioned above, the knowledge about the unknown parameters is encoded in the prior probability density $\pi_{\text{pr}}(x)$. The core idea of Bayesian reasoning is that the prior knowledge is updated with the information gained from the measurements. Using Bayes' theorem, the prior and the likelihood can be combined to give the *posterior density*,

$$\pi_{\text{post}}(x) = \pi(x|y) = \frac{\pi(x)\pi(y|x)}{\pi(y)}, \tag{3.4}$$

which gives the "probability" of the unknown having value $X = x$ after $Y = y$ has been observed. The density in the denominator corresponds to the marginal distribution of the measurement. It is usually difficult to specify $\pi(y)$, but it is often not important in practice since it only acts as a normalizing constant.

The final step in the workflow of Bayesian reasoning is to explore the posterior density and compute from it quantities of interest. In many applications, including imaging problems, one wishes to compute a point estimate to serve as the reconstruction of the unknown. The most straightforward way is to compute the mode of the posterior, called the *Maximum a Posteriori* (MAP) estimate

$$x_{\text{MAP}} = \arg\max_x \pi_{\text{post}}(x). \tag{3.5}$$

Computing a MAP estimate usually involves solving an optimization problem, though in the linear Gaussian case, a closed form solution is available. Another possibility is the *conditional mean* (CM) of the posterior distribution, defined as

$$x_{\text{CM}} = \int_{\mathbb{R}^n} x\pi(x|y)\mathrm{d}x. \tag{3.6}$$

Finding the conditional mean involves evaluating an integral, with dimension equal to the dimension of the unknown. In imaging problems this

dimension can be up to the order of hundreds of thousands (or even more), which makes computing the CM estimate computationally demanding. Because of this, the MAP estimator is arguably most commonly used in practice.

It is common to assume the noise distribution π_{noise} to be a Gaussian with zero mean, which gives a likelihood of the form

$$\pi(y|x) = \frac{1}{\sqrt{(2\pi)^m \det(\Gamma_{\text{noise}})}} \exp\left(-\frac{1}{2}\big(y - f(x)\big)^T \Gamma_{\text{noise}}^{-1}\big(y - f(x)\big)\right), \tag{3.7}$$

where the noise covariance Γ_{noise} is often assumed to be diagonal, indicating that the components of the noise vector are mutually independent.

3.1 Choice of prior

The choice of a prior density is a nontrivial task and a large body of literature exists about formulating priors to describe the features that the object is anticipated to have. In addition, certain priors can give posterior densities for which computing the desired estimator is simpler. In the case of a linear model, that is when $f(x) = Ax$, where $A \in \mathbb{R}^{m\times n}$, an especially convenient choice is to also define the prior density as Gaussian, that is

$$\pi_{\text{pr}}(x) = \frac{1}{\sqrt{(2\pi)^n \det(\Gamma_{\text{pr}})}} \exp\left(-\frac{1}{2}\big(x - x_0\big)^T \Gamma_{\text{pr}}^{-1}\big(x - x_0\big)\right), \tag{3.8}$$

where $x_0 \in \mathbb{R}^n$ is the mean of the prior density and $\Gamma_{\text{pr}} \in \mathbb{R}^{n\times n}$ its covariance matrix. In this case, the posterior mean and covariance can be written explicitly, either as

$$\begin{aligned} \Gamma_{\text{post}} &= \left(\Gamma_{\text{pr}}^{-1} + A^T \Gamma_{\text{noise}}^{-1} A\right)^{-1}, \\ \bar{x} &= \left(\Gamma_{\text{pr}}^{-1} + A^T \Gamma_{\text{noise}}^{-1} A\right)^{-1} \left(A^T \Gamma_{\text{noise}}^{-1} y + \Gamma_{\text{pr}}^{-1} x_0\right), \end{aligned} \tag{3.9}$$

or alternatively as

$$\begin{aligned} \Gamma_{\text{post}} &= \Gamma_{\text{pr}} - \Gamma_{\text{pr}} A^T \left(A \Gamma_{\text{pr}} A^T + \Gamma_{\text{noise}}\right)^{-1} A \Gamma_{\text{pr}}, \\ \bar{x} &= x_0 + \Gamma_{\text{pr}} A^T \left(A \Gamma_{\text{pr}} A^T + \Gamma_{\text{noise}}\right)^{-1} (y - A x_0). \end{aligned} \tag{3.10}$$

The equivalence of these two forms can be shown using the Woodbury matrix identity (see e.g. [32]). An important point to note is that the computational complexity of the two forms might differ considerably if the dimension of the measurement vector is significantly smaller or larger than the dimension of the unknown. In imaging applications one usually has $m \ll n$ and thus the latter formulas are used.

The simplest choice for the prior covariance is to define it as

$$\Gamma_{\text{pr}} = \gamma^2 \mathbf{I}, \tag{3.11}$$

where $\mathbf{I} \in \mathbb{R}^{n \times n}$ is the identity matrix and γ is the standard deviation of an element in the unknown. This model assumes that all elements are mutually independent and have the same variance. A slightly more detailed model could assume that the elements are correlated, and that the correlation depends on the distance between them, as defined by some suitable metric. One possible prior covariance of this form is defined by

$$(\Gamma_{\text{pr}})_{i,j} = \gamma^2 \exp\left(-\frac{\|x_i - x_j\|^2}{2l^2}\right), \tag{3.12}$$

where $\gamma > 0$ is the pointwise standard deviation, x_i, x_j are the locations of the elements i and j in the target, and $l > 0$ is a correlation length parameter that describes the scale of the structures expected in the target.

Despite the advantages mentioned in the previous section, a Gaussian prior often does not well describe the characteristics of the unknown that one would expect to see. When imaging the human body for example, one often treats the organs as approximately constant regions, separated by sharp boundaries between them. Thus a prior that would give high probability to sharp edges and allow smooth regions between them would be desirable. A class of widely used priors that has exactly these characteristics is called *edge-promoting*, and many of them can be represented in the form

$$\Phi(x) = \int_\Omega r(\|\nabla_\xi x(\xi)\|) \mathrm{d}\xi, \tag{3.13}$$

where Ω is the domain where the unknown is located, and r is a suitable monotonically increasing function.

Choosing $r(t) = t$ gives us the well-known *total variation* (TV) prior, which was first considered in [60]. Often also a smoothened version is used, where one chooses $r(t) = \sqrt{t^2 + T^2}$ with T a suitable small parameter, to get around the non-differentiability of the absolute value function at zero. In [1], the solution is shown to be stable with respect to the parameter T, so that as $T \to 0$, the solution converges to the solution of the unperturbed functional. Another option is the *Perona-Malik* prior [54] that is obtained with the choice $r(t) = \frac{1}{2} T \log(1 + (t/T)^2)$, with $T > 0$ playing the role of a threshold for detectable edges.

As mentioned above, choosing an edge-promoting prior results in a posterior that is not Gaussian and formulas such as (3.9) and (3.10) are not available. Efficiently computing the MAP estimate in this case has been previously addressed e.g. in [7] and [31]. The main contribution of this thesis is to use these methods to enable the efficient computation of optimal experimental designs for imaging application, as explained in the next chapter.

3.2 Optimal Experimental Design

Performing a statistical experiment always involves making decisions about the experimental setup. While the setup itself does not influence the phenomenon being studied, it does impact the data that is obtained.

Designing an experiment generally involves expert judgement and established heuristics. However, one can also try to create a more precise definition for what is a good experiment. Generally the goal of an experiment is to gain information about an unknown parameter and to potentially make predictions. A good experiment is one that gives as much information about the unknown as possible and reduces the uncertainty of the prediction more than other experiments, depending of course on the ultimate goal of the considered application.

Casting this idea into a decision theoretic framework was done by [43] (see also [16], [9], [56]) in the following way: first determine the variables under the control of the researcher with respect to which the optimization will be performed, and define D to be the set of possible *designs*, for example the set of admissible sensor positions. Secondly, define a utility function $u(x, y; d)$ that gives the utility of performing an experiment with design $d \in D$, the unknown having value $x \in X$, and measuring the data $y \in Y$. Since the unknown and data are a priori of course not known, take the expected value with respect to their joint probability density conditioned on the design d,

$$U(d) = \int_Y \int_X u(x, y; d)\pi(x|d, y)\pi(y|d)\mathrm{d}x\,\mathrm{d}y. \tag{3.14}$$

This is called the *expected utility*. An optimal design d^* then is one that maximizes this expected utility:

$$d^* = \arg\max_{d \in D} U(d). \tag{3.15}$$

Finding the optimizer d^* is a highly nontrivial task and presents a number of problems. Evaluating the expected utility involves computing an integral over X and Y, both of which can be high-dimensional, which makes any quadrature based approach run into the curse of dimensionality. In any case, various approaches have been developed to efficiently find the maximizer of the expected utility. See [62] for a review of some modern algorithms.

Some research has approached the problem using sampling methods that are in a certain sense dimension independent. Examples of these are [63], [33], and [34]. However, even such approaches are usually limited to problems with dimensions much smaller than those encountered in imaging applications.

The OED problem can also be formulated in the non-Bayesian setting, in which the optimal design is usually one that minimizes the *empirical risk*

over a set of training models [25]. The paper [26] presents this approach for the nonlinear case and demonstrates the approach in numerical experiments for an electrical impedance tomography problem, where the aim is to promote sparsity in the measurements. The inverse problem itself is formulated via a penalized Tikhonov-type functional.

In some special cases it is possible to derive computationally more attractive formulas for U. One can then find approximations for more general settings that enable these formulas to be used. A popular line of research is that of utilizing Laplace approximations used e.g. in [13], [45], and [44]. The approach of utilizing the lagged diffusivity iteration studied in this thesis also follows a similar path.

In addition to efficiently evaluating the expected utility, developing a suitable method for solving for the optimum presents considerable difficulties. Often, very little can be said about the shape of the objective function. It can, and often does, have multiple local optima that make gradient based approaches perform poorly. This generally forces one to broadly cover the design space during the search process, so that the search is not stuck in a particularly nonoptimal region. At the very least, multiple initializations should be used to mitigate this problem. Various heuristics can also be used, but they are often problem dependent and do not generalize well. Thus finding good search algorithms is an important subtopic when applying OED methodologies.

The set of admissible designs is often constrained, in which case the design problem becomes a constrained optimization problem. For example, if the design parameters correspond to the positions of measurement sensors, the geometry of the considered setup may pose restrictions on the values the design parameters may take. Also the state of the unknown may be constrained in the inverse problem [61], for example, in applications where the unknown corresponds to a physical quantity, one might wish to constrain the integral of the unknown to a fixed value if its total "mass" is known in advance.

An important point to note is that the optimal designs can often be highly unintuitive. This emphasizes the usefulness of a formal approach to the problem.

Various choices for the utility function can be considered, but the two most common are the negative squared-norm

$$u(x, y; d) = -\|x - \hat{x}(y; d)\|^2, \tag{3.16}$$

where $\hat{x}$ denotes the employed posterior estimate for the unknown x after y has been measured. This utility function results in what is known as the A-optimality criterion. Another option is the *Kullback-Leibler* (KL) divergence [38]

$$u_{\mathrm{KL}}(y; d) = \int_\Omega \log\left(\frac{\pi(x|y;d)}{\pi_{\mathrm{pr}}(x)}\right) \pi(x|y;d)\mathrm{d}x \tag{3.17}$$

that measures the difference in information between the posterior and the prior. The KL divergence can be considered as a "distance" between the two densities, though it is not a true metric since it is not symmetric and does not satisfy the triangle inequality. When using the KL divergence as the utility function, the resulting criterion is known as D-optimality.

When the goal of an inverse problem is to reconstruct a physical quantity, A-optimality is perhaps the natural criterion to use, since the criteria then corresponds to the quantity that one is trying to minimize: the average (squared) reconstruction error. This is also supported by the fact that in the experiments performed in this thesis, A-optimality generally has superior performance over D-optimality. However, the KL divergence, maximizing the *expected information gain* from the measurements, can be arguably thought to have better theoretical Bayesian motivation.

In the special case of a Gaussian posterior, simpler forms for the expected utility can be derived. In the case of A-optimality, maximizing the expected utility is equivalent to minimizing the trace of the posterior covariance:

$$d^* = \arg\min_d \operatorname{tr}\left(\Gamma_{\text{post}}(d)\right) \tag{3.18}$$

Similarly, for D-optimality, one ends up minimizing the logarithm of the determinant of the posterior covariance

$$d^* = \arg\min_d \log\Big(\det\left(\Gamma_{\text{post}}(d)\right)\Big). \tag{3.19}$$

For the modalities considered in this thesis, OED for X-ray tomography has been investigated in [61]. For MRXI [18], [19] and [20] investigate optimizing the activation placements and patterns, though none of them in the Bayesian setting.

Following the formulation of Bayesian inverse problems in infinite dimensions in [67], the OED problem has also been formulated for models where the unknown parameter is infinite dimensional. Such methodology has been developed in [4], [5], [3], [6] and [2]. Formulating the problem in infinite dimensions is natural in many problems governed by PDEs, where the unknown quantity is a spatially distributed parameter field. In particular, this highlights the importance of performing the discretization in an appropriate way, since otherwise finer discretizations do not necessarily converge to the infinite-dimensional problem in the limit. In the infinite-dimensional case, defining the prior also has to be done with care, since the prior has to ensure sufficient regularity for the posterior to be defined.

This thesis largely concentrates on *sequential* OED problems, where the measurement can be thought to consist of distinct steps: after measuring some data, there is a possibility to analyze that data, before deciding on the next measurement design. If an OED problem is formulated so that future designs depend on the previously measured data, the design, or more specifically the algorithm for computing the design, is called *adaptive*.

In this thesis, the sequential OED problem is usually solved using a *greedy* (also known as *myopic*) approach, where the optimal design for each step is found considering all data measured up to that point, but disregarding the effect of any future measurements. In the Bayesian framework, this translates to using the posterior probability density as the prior for the next measurement design. This does not guarantee that the design is globally optimal over all the measurement steps, but makes the problem computationally tractable. It is possible to formulate the sequential OED problem so that one looks for a global optimum, but then the dimension of the problem grows rapidly with each additional measurement step. For the kind of models considered in this thesis, special approaches must be developed to solve for a global optimum for more than a small number of steps. An example of such is [35], where the optimization is formulated as a *dynamic programming* (DP) problem, with the value functions approximated by suitable surrogates to limit the computational burden. The paper [65] further develops this approach by using *deep neural networks* (DNNs) to represent the value and policy functions and enable the use of policy gradient techniques. A limitation of this technique is that the number of steps must be defined in advance, whereas in many practical settings the experimenter may want to terminate the experiment when he or she deems that sufficient data has been collected, e.g. by assessing the quality of the reconstruction.

3.3 Lagged diffusivity iteration

The LD iteration is a fixed point iterative scheme introduced in [68] as a way to minimize a TV-penalized least squares functional. Interpreting the regularization parameters suitably, the iteration can also be seen as a way to compute the mode of a posterior distribution when an edge-promoting prior is used. Seen in this context, the iteration generates a series of Gaussian approximations to the posterior, for which the mode will converge to the mode of the original posterior, under suitable assumptions. See [11] for an explanation of the Gaussian interpretation. The convergence of the method has been analyzed in [21] and [17], though these analyses only cover settings with limitations on the forward model. Despite this, the LD iteration has been successful in practice, and has been used for example in [31] and [28].

In computing the mode, one tries to find the maximizer for the posterior density, which is equivalent to computing the minimizer of the negative log-likelihood

$$f(U) = \frac{1}{2}\|y - RU\|^2_{\Gamma^{-1}_{\text{noise}}} + \gamma\Phi_{\text{TV}}(u), \tag{3.20}$$

where R denotes the linear forward map and the smoothened TV penalty

term is of the form

$$\Phi(u) = \int_\Omega \sqrt{(\nabla_\xi u(\xi))^2 + T^2}\mathrm{d}\xi, \tag{3.21}$$

with T being a small parameter ensuring differentiability; see (3.13).

At the maximizer, the gradient of f is zero, i.e.,

$$\nabla_U f(U) = R^T \Gamma_{\text{noise}}^{-1} R + \gamma \nabla \Phi_{\text{TV}}(u) - R^T \Gamma_{\text{noise}}^{-1} y = 0. \tag{3.22}$$

The LD iteration arises from the observation that the gradient of the penalty term can be expressed as a matrix-vector product. To achieve this, the unknown parameter is assumed to have a discretized parametric representation $u = \sum_{j=1}^n U_j \varphi_j$, where ϕ_j are suitable weakly differentiable basis functions; in what follows, capital letters denote vectors and the corresponding lower case letters the associated parametrized functions following such a parametrization.

This is the formulation used in Publications II and III. Computing the partial derivative of the penalty term gives

$$\frac{\partial \Phi_{\text{TV}}}{\partial U_i} = \int_\Omega \frac{1}{\sqrt{\|\nabla u(x)\|_2^2 + T^2}} \sum_{j=1}^n U_j \nabla \phi_j(x) \cdot \nabla \phi_i(x) \mathrm{d}x. \tag{3.23}$$

Defining a matrix $H(u) \in \mathbb{R}^{n \times n}$ as

$$\big(H(u)\big)_{i,j} = \int_\Omega \frac{1}{\sqrt{\|\nabla u(x)\|_2^2 + T^2}} \sum_{j=1}^n \nabla \phi_j(x) \cdot \nabla \phi_i(x) \mathrm{d}x, \tag{3.24}$$

the gradient in (3.22) can be given as

$$\nabla_U \Phi_{\text{TV}}(u) = H(u)U. \tag{3.25}$$

The matrix H can be interpreted as suitably discretized steady-state diffusion operator, from which the "diffusivity" part of the name originates. As for the "lagged" in "lagged diffusivity", observe first that inserting (3.25) into (3.22) results in

$$\nabla_U f(U) = \left[R^T \Gamma_{\text{noise}}^{-1} R + \gamma H(u)\right] U - R^T \Gamma_{\text{noise}}^{-1} y = 0. \tag{3.26}$$

The difficulty in solving this equation is due to the matrix H depending on u, making the equation nonlinear. To overcome this, u is replaced by its current estimate $\hat{u}$, i.e., letting it "lag" one step behind, resulting in the approximate condition

$$\nabla_U f(U) = \left[R^T \Gamma_{\text{noise}}^{-1} R + \gamma H(\hat{u})\right] U - R^T \Gamma_{\text{noise}}^{-1} y = 0. \tag{3.27}$$

To deduce the Gaussian approximations mentioned earlier, observe that the solution of (3.27) equals the posterior mean of (3.9) when the prior is defined to be Gaussian with zero mean and the covariance

$$\Gamma_{\text{pr}} = \gamma^{-1} H(\hat{u})^{-1}. \tag{3.28}$$

The LD iteration then alternates between the two steps:

- Use the current estimate $\hat{u}$ to form the matrix $H(\hat{u})$ according to (3.24).
- Solve for a new estimate $\hat{u}$ from (3.27).

Once the iteration is deemed to have converged, the final estimate $\hat{u}$ is declared as the reconstruction. From the final iterate, one also gets a covariance matrix for the approximation of the prior via (3.28) that can be used in (3.10) to compute a posterior covariance for an approximate Gaussian posterior. The posterior covariance can then be used to choose the next measurement design according to the employed OED optimality criterion as explained in the previous section.

4. Summary of Results

I: Sequentially optimized projections in x-ray imaging

In this work a general algorithm for sequential OED for finite-dimensional linear Gaussian inverse problems is developed.

For a linear inverse problem with Gaussian noise and a Gaussian prior, the posterior is also Gaussian and explicit forms for its mean and covariance can be derived. The Gaussian posterior also allows us to write the design optimization targets as functions of the posterior covariance matrix, and computationally attractive forms for the evaluation of these targets are derived. In this setting, the optimal designs do not depend on the measurement data, and they can thus be computed offline, that is, before any measurements are taken. The algorithm also allows the users to define a region of interest, such that only the unknown parameters in that spatial region are considered when determining optimal designs.

The developed algorithm is then tested on a simulated 2D X-ray tomography problem, where the aim is to reconstruct the spatial absorption coefficient of the object using measurements consisting of line integrals of the absorption through the object. The target objects may also contain obstructions through which the projections obtain no information. The design variables are chosen as the projection angle and the lateral location of the source-receiver pair in parallel beam tomography.

As an additional feature, a hyperparameter estimation step is included in the optimization, where a parameter of the prior distribution is estimated from the measurement data. In this case, the algorithm can no longer be run offline, since the (original) prior is updated after each measurement.

The numerical experiments show that optimizing the projections indeed provides better reconstructions on average, compared to random or equidistant projections. Especially in the case of obstructions, the algorithm is able to choose the projections in an intelligent way, so that imaging through the obstructions is avoided. In the considered setting the hyperparameter search also converges very quickly, and it provides a clear advantage

compared to running the algorithm with a misspecified prior model.

II: Edge-promoting adaptive Bayesian experimental design for X-ray imaging

This work extends Publication I to enable Bayesian OED also for certain non-Gaussian prior models. Of particular interest for imaging applications are edge-promoting priors such as the TV prior. Sequential OED for such priors is achieved by using the lagged diffusivity iteration to form a series of Gaussian approximations and using the resulting posterior covariance in evaluating the optimization targets presented in Publication I.

The approach is tested in numerical simulations for both two- and three-dimensional x-ray tomography. The experiments include computing average errors over a set of randomly generated targets, as well as for fixed targets such as the widely used Shepp–Logan phantom. Again, the optimization shows superior performance over reference approaches in all considered tests. The usage of an edge-promoting prior is observed to give better reconstructions over Gaussian priors in the considered imaging applications, where the targets often consist of areas with close to constant absorption values, with sharp borders between them.

It is important to note that the algorithm is not limited to the considered application, but can be applied to any finite-dimensional linear inverse problem, where the unknown is a spatial parameter field.

III: Bayesian design of measurement for magnetorelaxometry imaging

In the third publication the algorithms introduced in the first two publications are applied to another imaging methodology, namely that of MRXI. A feature that is different from Publications I and II is that in MRXI the measurement depends smoothly and explicitly on the design parameters, which facilitates the use of gradient-based optimization methods. In addition, the behavior of the measurement process in MRXI is of interest, because unlike in CT imaging, each measurement provides information about the entire object. Thus it cannot be reasoned with certainty what kinds of designs are produced by optimizing the experiment, based solely on the results in Publications I and II.

Both Gaussian and TV-prior models are investigated, and the performance of gradient-based and exhaustive search algorithms are compared. For symmetric targets, the observed results match well with intuition and the produced designs are roughly symmetrical. With other targets though, the designs clearly depend on the previous measurements if TV prior is used, and optimizing the measurement design results in smaller reconstruction errors compared to reference designs. Also, unlike in CT, the scaling of the domain is shown to have a major impact on the designs.

IV: The power of prediction: spatiotemporal Gaussian process modeling for predictive control in slope-based wavefront sensing

The fourth publication explores Bayesian modeling and Gaussian processes in adaptive optics. The evolution of the atmosphere is modeled as a series of turbulent layers that are affected by wind, and two prior models are formulated: in the first one, the wind speeds and directions are assumed to be known accurately for each layer, along with their turbulence strengths. This model is called the frozen flow GP model (FF-GP). In the second model, the wind directions are unknown and only information about the average wind strength is available. The second model is called wind-averaged frozen flow (WAFF-GP).

The performance of the two models is investigated in a predictive reconstruction task and compared to a model that only takes into account spatial correlations. The numerical tests show that both FF-GP and WAFF-GP produce significantly lower variance for the predictions compared to the non-predictive spatial model.

The spatiotemporal GP model also allows to compute reconstructions with a higher resolution than the measurement resolution for the wavefront, which helps to mitigate antialiasing errors. This effect is also demonstrated by numerical simulations.

Finally, the OED methodology is employed to investigate the utility of past data, so that the computational burden can be limited. The results show that depending on wind and noise conditions, a timeseries length of 4 to 8 steps can be sufficiently informative for the reconstructions, and utilizing longer timeseries only brings minor benefits that arguably do not justify the associated higher computational burden.

References

[1] R. Acar and C. R. Vogel. Analysis of total variation penalty methods for ill-posed problems. *Inverse Prolems*, 10:1217–1229, 1994.

[2] A. Alexanderian. Optimal experimental design for infinite-dimensional bayesian inverse problems governed by PDEs: A review. *Inverse Problems*, page 043001, 2021.

[3] A. Alexanderian, P. J. Gloor, O. Ghattas, et al. On Bayesian A- and D-optimal experimental designs in infinite dimensions. *Bayesian Analysis*, 11(3):671–695, 2016.

[4] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized l_0-sparsification. *SIAM Journal on Scientific Computing*, 36(5):A2122–A2148, 2014.

[5] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. *SIAM J. Sci. Comput.*, 38(1):A243–A272, 2016.

[6] A. Alexanderian and A. K. Saibaba. Efficient D-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems. *SIAM Journal on Scientific Computing*, 40(5):A2956–A2985, 2018.

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

[8] S. Arsalani, P. Radon, P. Schier, A. Jaufenthaler, M. Liebl, D. Baumgarten, and F. Wiekhorst. Developing magnetorelaxometry imaging for human applications. *Physics in Medicine & Biology*, 67(22):225007, 2022.

[9] A. Atkinson, A. Donev, R. Tobias, et al. *Optimum experimental designs, with SAS*, volume 34. Oxford University Press, 2007.

[10] H. W. Babcock. The possibility of compensating astronomical seeing. *Publications of the Astronomical Society of the Pacific*, 65(386):229–236, 1953.

[11] J. M. Bardsley. *Computational Uncertainty Quantification for Inverse Problems*. SIAM, Philadelphia, PA, 2018.

[12] J. K. Batenburg, F. De Carlo, L. Mancini, and J. Sijbers. Advanced x-ray tomography: experiment, modeling, and algorithms. *Measurement Science and Technology*, 29(8), 2018.

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

[14] W. F. Brown Jr. Thermal fluctuations of a single-domain particle. Physical Review. Vol. 130:5., pages 1677–1686, 1963.

[15] D. Calvetti and E. Somersalo. *An introduction to Bayesian scientific computing: ten lectures on subjective computing*, volume 2. Springer Science & Business Media, 2007.

[16] K. Chaloner and I. Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.

[17] T. F. Chan and P. Mulet. On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2):354–367, 1999.

[18] A. Coene, G. Crevecoeur, and L. Dupre. Adaptive control of excitation coil arrays for targeted magnetic nanoparticle reconstruction using magnetorelaxometry. *IEEE transactions on magnetics*, 48(11):2842–2845, 2012.

[19] A. Coene, J. Leliaert, L. Dupré, and G. Crevecoeur. Quantitative model selection for enhanced magnetic nanoparticle imaging in magnetorelaxometry. *Medical Physics*, 42(12):6853–6862, 2015.

[20] G. Crevecoeur, D. Baumgarten, U. Steinhoff, J. Haueisen, L. Trahms, and L. Dupré. Advancements in magnetic nanoparticle reconstruction using sequential activation of excitation coil arrays using magnetorelaxometry. *IEEE Transactions on Magnetics*, 48(4):1313–1316, 2012.

[21] D. C. Dobson and C. R. Vogel. Convergence of an iterative method for total variation denoising. *SIAM Journal on Numerical Analysis*, 34(5):1779–1791, 1997.

[22] N. Doelman. The minimum of the time-delay wavefront error in adaptive optics. *Monthly Notices of the Royal Astronomical Society*, 491(4):4719–4723, 2020.

[23] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375. Springer Science & Business Media, 1996.

[24] J. Föcke, D. Baumgarten, and M. Burger. The inverse problem of magnetorelaxometry imaging. *Inverse problems*, 34:115008, 2018.

[25] E. Haber, L. Horesh, and L. Tenorio. Numerical methods for experimental design of large-scale linear ill-posed inverse problems. *Inverse Problems*, 24(5):055012, 2008.

[26] E. Haber, L. Horesh, and L. Tenorio. Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems. *Inverse Problems*, 26(2):025002, 2009.

[27] K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen. Sparse tomography. *SIAM Journal on Scientific Computing*, 35(3):B644–B665, 2013.

[28] A. Hannukainen, L. Harhanen, N. Hyvönen, and H. Majander. Edge-promoting reconstruction of absorption and diffusivity in optical tomography. *Inverse Problems*, 32(1):015008, 2015.

[29] A. Hannukainen, N. Hyvönen, and L. Perkkiö. Inverse heat source problem and experimental design for determining iron loss distribution. *SIAM Journal on Scientific Computing*, 43(2):B243–B270, 2021.

[30] P. C. Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.

[31] L. Harhanen, N. Hyvönen, H. Majander, and S. Staboulis. Edge-enhancing reconstruction algorithm for three-dimensional electrical impedance tomography. *SIAM Journal on Scientific Computing*, 37:B60–B78, 2015.

[32] H. V. Henderson and S. R. Searle. On deriving the inverse of a sum of matrices. *SIAM Review*, 23(1):53–60, 1981.

[33] X. Huan. *Accelerated Bayesian experimental design for chemical kinetic models*. PhD thesis, Massachusetts Institute of Technology, 2010.

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

[35] X. Huan and Y. M. Marzouk. Sequential Bayesian optimal experimental design via approximate dynamic programming. *arXiv preprint arXiv:1604.08320*, 2016.

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

[37] J. Kaipio and E. Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.

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

[39] R. Landman, S. Y. Haffert, V. M. Radhakrishnan, and C. U. Keller. Self-optimizing adaptive optics control with reinforcement learning. In *Adaptive Optics Systems VII*, volume 11448, page 1144849. International Society for Optics and Photonics, SPIE, 2020.

[40] R. Landman, S. Y. Haffert, V. M. Radhakrishnan, and C. U. Keller. Self-optimizing adaptive optics control with reinforcement learning for high-contrast imaging. *Journal of Astronomical Telescopes, Instruments, and Systems*, 7(3):039002, 2021.

[41] M. Liebl, U. Steinhoff, F. Wiekhorst, J. Haueisen, and L. Trahms. Quantitative imaging of magnetic nanoparticles by magnetorelaxometry with multiple excitation coils. *Physics in Medicine & Biology*, 59(21):6607, 2014.

[42] M. Liebl, F. Wiekhorst, D. Eberbeck, P. Radon, D. Gutkelch, D. Baumgarten, U. Steinhoff, and L. Trahms. Magnetorelaxometry procedures for quantitative imaging and characterization of magnetic nanoparticles in biomedical applications. *Biomedical Engineering/Biomedizinische Technik*, 60(5):427–443, 2015.

[43] D. V. Lindley. *Bayesian Statistics - A Review*. SIAM, 1972.

[44] Q. Long, M. Motamed, and R. Tempone. Fast Bayesian optimal experimental design for seismic source inversion. *Computer Methods in Applied Mechanics and Engineering*, 291:123–145, 2015.

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

[46] F. Natterer. *The Mathematics of Computerized Tomography*. Vieweg+Teubner Verlag Wiesbaden, 1986.

[47] F. Natterer and F. Wübbeling. *Mathematical methods in image reconstruction*. SIAM Monographs on Mathematical Modeling and Computation. SIAM, Philadelphia, PA, 2001.

[48] L. Néel. Théorie du traînage magnétique des ferromagnétiques en grains fins avec applications aux terres cuites. *Annales de Géophys.*, 5:99–136, 1949.

[49] J. Nousiainen, C. Rajani, M. Kasper, and T. Helin. Adaptive optics control using model-based reinforcement learning. *Optics Express*, 29(10):15327–15344, 2021.

[50] J. Nousiainen, C. Rajani, M. Kasper, T. Helin, S. Haffert, C. Vérinaud, J. Males, K. Van Gorkom, L. Close, J. Long, et al. Toward on-sky adaptive optics control using reinforcement learning-model-based policy optimization for adaptive optics. *Astronomy & Astrophysics*, 664:A71, 2022.

[51] W. Nowak, F. De Barros, and Y. Rubin. Bayesian geostatistical design: Task-driven optimal site investigation when the geostatistical model is uncertain. *Water Resources Research*, 46(3), 2010.

[52] Q. Pankhurst, N. Thanh, S. Jones, and J. Dobson. Progress in applications of magnetic nanoparticles in biomedicine. *Journal of Physics D: Applied Physics*, 42(22):224001, 2009.

[53] Q. A. Pankhurst, J. Connolly, S. K. Jones, and J. Dobson. Applications of magnetic nanoparticles in biomedicine. *Journal of physics D: Applied physics*, 36(13):R167, 2003.

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

[55] B. Pou, F. Ferreira, E. Quinones, D. Gratadour, and M. Martin. Adaptive optics control with multi-agent model-free reinforcement learning. *Optics Express*, 30(2):2991–3015, Jan 2022.

[56] F. Pukelsheim. *Optimal design of experiments*. SIAM, 2006.

[57] J. Radon. Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Ber. Verh. Sächsischen Akad. Wiss. Leipzig. Math.-Phys. Kl.*, 69:262–277, 1917.

[58] C. E. Rasmussen, C. K. Williams, et al. *Gaussian processes for machine learning*. Springer, 2006.

[59] F. Roddier. *Adaptive optics in astronomy*. Cambridge University Press, 1999.

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

[61] L. Ruthotto, J. Chung, and M. Chung. Optimal experimental design for inverse problems with state constraints. *SIAM Journal on Scientific Computing*, 40(4):B1080–B1100, 2018.

[62] E. G. Ryan, C. C. Drovandi, J. M. McGree, and A. N. Pettitt. A review of modern computational algorithms for bayesian optimal design. *International Statistics Review*, 84(1):128–154, 2016.

[63] 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):585–603, 2003.

[64] P. Schier, A. Jaufenthaler, M. Liebl, S. Arsalani, F. Wiekhorst, and D. Baumgarten. Human-sized quantitative imaging of magnetic nanoparticles with nonlinear magnetorelaxometry. *Physics in Medicine & Biology*, 06 2023.

[65] W. Shen and X. Huan. Bayesian sequential optimal experimental design for nonlinear models using policy gradient reinforcement learning. *arXiv preprint arXiv:2110.15335*, 2021.

[66] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. Statistical inversion for medical x-ray tomography with few radiographs: I. general theory. *Physics in Medicine & Biology*, 48:1437–1463, may 2003.

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

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

[69] M. Weiser, Y. Freytag, B. Erdmann, M. Hubig, and G. Mall. Optimal design of experiments for estimating the time of death in forensic medicine. *Inverse Problems*, 34(12):125005, 2018.

[70] F. Wiekhorst, U. Steinhoff, D. Eberbeck, and L. Trahms. Magnetorelaxometry assisting biomedical applications of magnetic nanoparticles. *Pharmaceutical research*, 29:1189–1202, 2012.

[71] J. Yu, V. M. Zavala, and M. Anitescu. A scalable design of experiments framework for optimal sensor placement. *Journal of Process Control*, 67:44–55, 2018.

Aalto-DT 230/2023

ISBN 978-952-64-1608-3 (printed)
ISBN 978-952-64-1609-0 (pdf)
ISSN 1799-4934 (printed)
ISSN 1799-4942 (pdf)

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

BUSINESS + ECONOMY

ART + DESIGN + ARCHITECTURE

SCIENCE + TECHNOLOGY

CROSSOVER

DOCTORAL THESES