

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Computational Engineering and Technical Physics
Technomathematics

Juha Tamminen

BAYESIAN OPTIMAL DESIGN IN X-RAY TOMOGRAPHY

Master's Thesis

Examiners: Professor Tapio Helin
 Professor Lassi Roininen

Supervisors: Professor Tapio Helin

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Computational Engineering and Technical Physics
Technomathematics

Juha Tamminen

BAYESIAN OPTIMAL DESIGN IN X-RAY TOMOGRAPHY

Master's Thesis

2020

43 pages, 12 figures.

Examiners: Professor Tapio Helin
 Professor Lassi Roininen

Keywords: X-ray tomography, Bayesian optimal design, A-optimality, D-optimality

This thesis focuses on Bayesian optimal design and its applications to X-ray tomography. X-ray tomography is a well-known setting where X-rays are projected at an object. X-ray intensity data is recorded on the other side of the object by sensors. Data sets from different angles form a linear inverse problem. Solving the problem allows mapping of the composition of an object. Applying Bayesian optimal design and sequential algorithm to this problem improves accuracy of reconstruction when the composition of the object is unknown. This thesis covers two design models: A- and D-optimal designs. Objects of the thesis are to present the designs and test them in a controlled environment against a random method of projection angle selection. Both A- and D-optimal designs perform better compared to the random method, but the A-optimal design was superior to all methods. D-optimal design provides worse results during the early iterations, but eventually D-optimality is able to catch up to A-optimality and after that both methods are equal. D-optimal design places projection angles next to previous ones whereas A-optimality places the angles more evenly, which is the cause between the differences in accuracy.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Engineering Science
Laskennallinen tekniikka ja teknillinen fysiikka
Teknillinen matematiikka

Juha Tamminen

BAYESIN OPTIMAALINEN SUUNNITTELU RÖNTGENTOMOGRAFIASSA

Diplomityö

2020

43 sivua, 12 kuvaa.

Tarkastajat: Professori Tapio Helin
 Professori Lassi Roininen

Avainsanat: Röntgentomografia, Bayesin optimaalinen suunnittelu, A-optimaalisuus, D-optimaalisuus

Diplomityö keskittyy Bayesin optimaaliseen suunnitteluun ja sen sovelluksiin röntgentomografiassa. Röntgentomografia on hyvin tunnettu tieteen haara, jossa röntgensäteet heijastetaan esineelle ja intensiteetti arvot tallennetaan anturien avulla kohteen toisella puolella. Eri kulmista otetut arvot muodostavat lineaarisen käänteisen ongelman. Ongelman ratkaiseminen mahdollistaa kohteen koostumuksen kartoittamisen. Bayesin optimaalisen suunnittelun soveltaminen tähän ongelmaan parantaa jälleenrakennuksen tarkkuutta, kun kohteen koostumus on tuntematon. Diplomityö kattaa kaksi suunnittelumallia: A- ja D-optimaalinen suunnittelu. Diplomityön tavoitteena on esitellä suunnittelumallit, testata ne hallitussa ympäristössä ja verrata niitä satunnaiseen menetelmään. Molemmat mallit toimivat paremmin verrattuna satunnaiseen menetelmään, mutta A-optimaalisuus oli paras menetelmä kaikista. D-optimaalisuus tuottaa huonommat tulokset varhaisissa iteraatioissa, mutta lopulta D-optimaalisuus kykenee saavuttamaan A-optimaalisuuden tason. Eroavaisuudet tuloksissa johtuvat kulmien valintatavasta.

PREFACE

First I want to thank my university for the enjoyable five years of study. It has been truly enjoyable and I have learned so much about mathematics and programming. I like applying mathematics to real life problems and the thesis subject proposed by my supervisor Tapio Helin was an excellent one. I also want to thank Tapio Helin for his assistance and guidance.

Lappeenranta, July 31, 2020

Juha Tamminen

CONTENTS

1	INTRODUCTION	9
1.1	Background	9
1.2	Objectives and delimitations	10
1.3	Structure of the thesis	10
2	HISTORY	11
2.1	X-Ray tomography	11
2.2	Bayesian optimal design	12
3	THEORY	14
3.1	Bayesian inverse problems	14
3.2	X-ray tomography	15
3.2.1	Radon transform	15
3.2.2	Discretized model	17
3.3	Experimental design	18
3.4	Bayesian optimal design	18
3.4.1	A-optimality	19
3.4.2	D-optimality	21
3.4.3	Posterior	23
4	SEQUENTIAL OPTIMIZATION	26
5	EXPERIMENTS	28
5.1	Programming environment	28
5.2	Data	28
5.3	Evaluation criteria	29
5.4	Description of experiments	29
5.4.1	Experiment 1	32
5.4.2	Experiment 2	32
5.4.3	Experiment 3	33
5.5	Results	34
5.5.1	Results 1	34
5.5.2	Results 2	36
5.5.3	Results 3	38
6	DISCUSSION	41
6.1	Current study	41
6.2	Future work	41

	6
7 CONCLUSION	42
REFERENCES	43

LIST OF SYMBOLS AND ABBREVIATIONS

A	Data matrix for pixel vector
$A^T A$	A positive semi-definite matrix
A_{mk}	Definition for $ L_m \cap P_{ij} $
B	Auxiliary matrix
BC	Bayes cost
P_{OD}	Bayesian optimal design
β	Proportion of real sensors
C	Cholesky decomposition
c	Defines a constant $(-\frac{1}{2} \log(\det \Gamma_{pr}))$
CM	Conditional mean
CT	Computed tomography
d	A constant
d'	A constant $(\text{tr}(A\Gamma_{pr}A^T))$
dS	Lebesgue surface measure
ϵ	Noise model
f	Model function
G	X-ray projection matrix
Γ	Covariance matrix
h	Differential entropy
I	Intensity
I_0	Initial intensity
KL	Kullback-Leibler divergence
k	Number of iterations
L	Hyperplane
l	Correlation length
μ	Mass absorption coefficient
MS	Mean square
m	Number of projection angles
N	Number of pixels per edge
ν	Total number of virtual sensors
n	Dimension
R	Matrix with model parameter values
ROI	Region of interest
Ω	Object domain
ω	Unit vector
Φ	Optimality criteria

Φ_A	A-optimal design criteria
Φ_D	D-optimal design criteria
p	Design variable
P_{ij}	Pixel [i,j]
$\pi_{post}(x)$	Posterior density
$\pi_{pr}(x)$	Prior density
$\pi(x, y)$	Joint probability density of x and y
\mathbb{R}	Set of real numbers
STD	Standard deviation
σ	Control parameter for noise
σ_{noise}	Standard deviation of noise
\sum	Sum
u	Mean value of μ
x	Pixel vector
y	Absorption data
2-D	Two-dimensional

1 INTRODUCTION

1.1 Background

The thesis is about X-ray tomography and relatively new approach of using Bayesian optimal design in order to improve the results of X-ray tomography. X-ray tomography is mainly used in clinical and industrial fields, where the objective is to analyse some part of the human body/industrial product by using radiation absorption to our advantage. The thesis focuses on medical imaging.

X-ray tomography is a method to analyse material properties of a given object. X-rays are projected at a specified object and the radiation travels through the object while intensity of the rays falls due to absorption. The absorption depends on the density of the object and the distance travelled. Sensors to record the change in intensity are placed on the other side of the object. By taking multiple measurements along a set axis gives us a single data set from one angle. Scanning the object from multiple angles will give us more data. Knowing the initial intensity, the intensity after absorption and the angles of measurements allows us to form a linear inverse problem. Pixel grid is used to define our accuracy or in this case the resolution of the image. Denser grid will lead to better image quality, but it also requires more data to be solvable. Solving the problem allows us to find the mass absorption coefficient for every pixel in the object domain. The coefficient changes based on material and can be used to predict the composition of the object, which allows us to see if there is anything unusual on the inside e.g. a tumour inside a human or a weakness in an industrial component (air pocket).

If we knew the composition of the object we could find the optimal set of angles to solve the inverse. Since, we do not know the composition before measurements, the best thing to do is to optimize the next angle of measurement. Why would we want to optimize the process in the first place? We could just take random angles and eventually we could solve the inverse. Besides computational time we mainly want to minimize the amount of radiation when dealing with human subjects. Maximizing image resolution quality and minimizing radiation dosage is essential in clinical industry. This is where Bayesian optimal design will help us. Bayesian optimal design aims to optimize the process of data gathering as new samples are being taken and actively find the next optimal angle of measurement based on current data. The thesis covers two optimal design models called A- and D-optimal designs.

1.2 Objectives and delimitations

The objectives of the thesis are to cover the mathematical theory related to standard X-ray tomography and how this can be improved by applying Bayesian optimal design methods, and studying the effects of A- and D-optimal designs on image quality. The image quality will be analysed based on accuracy and variance of results. To make the experiments and theory easier we make a few assumptions. We assume that only absorption occurs when X-rays travel through the object, so scattering or ricocheting from the predicted path does not occur. We also assume mass absorption coefficient to be proportional to the density of the material.

Research questions:

1. What are A- and D-optimal designs?
2. How do they differ from each others?
3. What are their strengths and weaknesses?

1.3 Structure of the thesis

The thesis covers the essential information about X-ray tomography, the application of Bayesian optimal design for it and some experiments regarding the use of A- and D-optimal designs. Chapter 2 summarizes history of X-ray tomography and how we have arrived at A- and D-optimal designs. Chapter 3 contains the theory regarding Bayesian inverse, X-ray tomography and Bayesian optimal design. In X-ray tomography Section we will cover radon transform and the discretized model. The Bayesian optimal design Section covers information regarding A- and D-optimal designs that can be used to enhance image quality. After that we go over sequential optimization algorithm in chapter 4. Following this Chapter we have Chapter 5, which covers the experiments with A- and D-optimal designs. It will cover the programming environment, used data, descriptions and results of experiments. Chapter 6 covers the discussion of the results and possible future work. Finally, Chapter 7 summarizes the key points of the thesis. References are presented at the end.

2 HISTORY

Section 2.1 is based on references [1, 2] which cover history of computed tomography. The history review is mostly based on Simone Carmignato et al. work on "Industrial X-ray computed tomography" (2018) which covers early history of computed tomography and technological advancements with different radiation sources and scanners.

Section 2.2 uses references [3–5], but mainly Anirban DasGuptas article of Bayesian optimal designs. The section highlights key points in the history of Bayesian optimal design.

2.1 X-Ray tomography

The beginning of X-ray tomography can be traced to 1914, when Polish radiologist Karol Mayer presented the first idea of tomography [2]. In 1917 the mathematical theory of Radon transform was published by Johann Radon, which is the foundation of current X-ray tomography [1]. The first patent related to X-ray computed tomography (CT) was made by Gabriel Frank in 1940. A major advancement at the time, but the lack of computational techniques and power greatly hindered the method. The progress halted for a couple of decades, but during the 1960s more progress was made in the field of tomography. In 1961 William H. Oldendorf published experiments with Gamma radiation [1]. The goal was to simulate a human head by using nails that surrounded another nail and see if it was possible to gain information about the internal nail while it was surrounded by similar structure (effectively making direct examination impossible by conventional radiography). The next step in tomographic field was taken by David Kuhl and Roy Edwards in 1963 when they introduced transverse tomography [1]. In their experiments they effectively performed manual backprojection operation, which they later managed to digitalize. At the time they did not have a way for reconstruction, but it is possible now and it is the basis of modern emission CT technology.

The first CT scanner was detailed in 1963 by Cormack and during the same time period Godfrey Hounsfield would formulate his own ideas regarding X-ray scanning. In 1967 the first prototype of a clinical CT scanner was build by Hounsfield [1]. It used low intensity gamma rays. Acquiring a data set took nine days and 2.5 h of computing. Later it was modified to use high intensity X-ray source, thus improving the scan time and accuracy. Commercial version was released in 1971 [1]. A Nobel Prize in Physiology

and Medicine was given for both Cormack and Hounsfield for the invention of CT technology [1]. Soon after this companies specializing in radiology would invade the field and cause rapid acceleration of CT technology. Second, third and fourth generation CT scanners became available with increased accuracy along new methods, such as magnetic resonance, which does not use ionizing radiation.

Nowadays CT scanners are mainly divided to clinical and industrial ones. Industry can use higher intensity radiation sources and longer scan times for increased accuracy when performing scans on various materials. Clinical scanners specialize in human subjects, which requires controlling the amount of radiation received during scanning. New algorithms and methods to reconstruct and analyse data are being developed constantly to make these scanners more efficient and accurate.

2.2 Bayesian optimal design

The field of optimal designs can be tracked back to 1918 when Kristine Smith presented optimal designs to be used for polynomial models [3, 4]. Some claim that the first paper on optimal designs was published in 1815 by Joseph Diaz Gergonne [5]. Gergonne's paper is related to optimal designs of polynomial regression where as Smith's paper is on optimal designs of polynomial models. It is debatable whether Gergonne's solution can be counted as optimal design the way we know it, but during the 1900s optimal designs saw a lot of development so Smith's paper can be seen as more important to the field of optimal designs.

During the 1900s experimental design saw steady development and multiple optimality criterias were developed for optimal designs. One of the key contributors was Abraham Wald. Wald's paper (1943) on efficient design had great influence to his peers and it affected the field for three decades [4]. One of the most popular theories during the time was Kiefer-Wolfowitz theory (1959) which introduced approximate design. The theory tied probability theory to polynomial regression and other mathematical theories. It was so impactful that it is used even today [4]. Following the introduction of approximate designs the field of optimal design was split into multiple branches, such as integer or qualitative design.

The Bayesian decision theory started to form during the 70s and in 1984 Kathryn Chaloner developed a Bayesian optimal design approach for linear regression [4]. Following Chaloner, other researchers started to work with Bayesian optimal designs. Pilz provided a solid

footing by studying linear regression and many improvements were made by other scientist including El-Krunz, DasGupta and Studden [4]. At this point linear models had a lot of information available, but non-linear models remained problematic as linear designs performed poorly even when there was only small amount of non-linearity [4]. Designs are constantly evolving and more experimenting is done with non-linear models. Today Bayesian optimal design can be applied to X-ray tomography and other fields.

3 THEORY

In this section we will cover theory regarding Bayesian inverse problems and X-ray tomography including Radon transform by Kaipio and Somersalo [6]. Experimental design uses Chaloner et al. [7]. After that we will present Bayesian optimal design theory which includes optimality criterias of A- and D-optimal designs and posterior formation following multiple sources, but mainly according to Burger et al. [8] and Alen Alexanderian et al. [9, 10]. The posterior for A- and D-optimal designs follows Burger et al. [8] exclusively.

3.1 Bayesian inverse problems

In Bayesian setting all parameters are treated as continuous random variables, thus making it possible to express them with probability densities. In the classical inverse problem, we have measurements from two quantities. In our case that is prior information from initial intensity of the X-ray I_0 and the measured intensity I after passing through object domain Ω . These measurements also contain noise. If there is no knowledge of the type of noise it is most often assumed to be white noise and the model can be expressed in the following form

$$y = f(x, \epsilon) = Rx + \epsilon, \quad \epsilon \sim \mathcal{N}(0, \sigma_{noise}^2 I), \quad (1)$$

where f is the model function containing the model parameters x , and the measurement noise and unknown parameters ϵ . Since we are using random variables and their realizations, we must be able to differentiate between these two. We will choose commonly used notation. For random variables we will use capital letters and lower case for realizations i.e.

$$Y = f(X, E). \quad (2)$$

Joint probability density of X and Y is denoted by $\pi(x, y)$ and the marginal distribution of the unknown X satisfies

$$\int_{\mathbb{R}^m} \pi(x, y) dy = \pi_{pr}(x). \quad (3)$$

According to Bayesian theory, we should have some information about random variable x before making measurements. This information can be presented as a probability density $\pi_{pr}(x)$ known as prior density. With this the posterior density can be expressed as

$$\pi_{post}(x) = \pi(x|y_{obs}) = \frac{\pi(y_{obs}|x)\pi_{pr}(x)}{\pi(y_{obs})}, \quad (4)$$

where $\pi(x|y_{obs})$ is the likelihood function and $\pi(y_{obs})$ is the marginal density that normalizes our posterior density. Now we shall simply denote $y = y_{obs}$, since we will use observed values (our data points). The marginal density can then be written as

$$\pi(y) = \int_{\mathbb{R}^n} \pi(x, y) dx = \int_{\mathbb{R}^n} \pi(y|x) \pi_{pr}(x) dx. \quad (5)$$

The Bayes formula is comprised of three things. Finding a probability distribution that best describes prior information and thus prior density π_{pr} , finding the likelihood function $\pi(y|x)$ that is used to express the interaction between the measurement outcome with a given $X = x$ and finally developing methods to evaluate posterior probability density.

Gaussian probability distributions are normally distributed random variables which follow Gaussian curves. Mean and variance are the parameters used to define a specific Gaussian curve. Since our posterior (4) is Gaussian we can represent the covariance matrix and the mean in the following form [9]

$$\Gamma_{post} = (\Gamma_{pr}^{-1} + R^T \Gamma_{noise}^{-1} R)^{-1}, \quad \hat{x} = \Gamma_{post} (\Gamma_{pr}^{-1} \bar{x} + R^T \Gamma_{noise}^{-1} y). \quad (6)$$

Using the Woodbury matrix identity we can rewrite these in an alternative form

$$\begin{aligned} \Gamma_{post} &= \Gamma_{pr} - \Gamma_{pr} R^T (R \Gamma_{pr}^{-1} R^T + \Gamma_{noise})^{-1} R \Gamma_{pr}, \\ \hat{x} &= \bar{x} + \Gamma_{pr} R^T (R \Gamma_{pr}^{-1} R^T + \Gamma_{noise})^{-1} (y - R \bar{x}), \end{aligned} \quad (7)$$

which can be computationally easier depending on the ratio of matrices [8]. If the Γ_{noise} is significantly smaller than the Γ_{pr} then the Woodbury conversion is going to speed up our computations. However, if we are interested in the inverse of covariance Γ_{post}^{-1} then the first form (6) is better.

3.2 X-ray tomography

3.2.1 Radon transform

The classical X-ray tomography problem is centered around Radon transform. We have a bounded domain $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, representing an object ($n = 3$) or a cross-sectional slice ($n = 2$). The object is scanned with a point-like X-ray source along an axis. The most commonly used model assumes that the X-ray passes through the object without ricocheting or scattering (e.g. only absorption occurs) and it is detected on the other

side of the object by an X-ray film or a digital sensor. If we also assume that the mass absorption coefficient is proportional to the density of the material, then the attenuation of the intensity I along a line segment dl can be written as

$$dI = -\mu I dl, \quad (8)$$

where $\mu = \mu(p) \geq 0, p \in \Omega$ is the mass absorption of the material [6]. If I_0 is the original intensity of the X-ray source then received intensity on the other side of the object can be calculated as such

$$I = I_0 e^{-\int_l \mu(p) dl} \quad (9)$$

and by taking a logarithm we can calculate the logarithmic ratio of intensities by mass coefficient integral

$$\log \left(\frac{I}{I_0} \right) = \int_{I_0}^I \frac{dI}{I} = - \int_l \mu(p) dl(p). \quad (10)$$

The problem can then be stated as: If we know the integrals (10) can we find out μ ? Task is to estimate values of μ from a set of integrals along straight lines which pass through an object domain Ω . Depending on how many lines of integration data are available the nature of the problem can be either complete or incomplete. In an ideal case we have a complete data set where all possible lines are available and we are able to solve μ uniquely. If we do not have the complete data set then we have to make some approximations. In practical applications we never have nearly enough data for a complete problem, which makes the problem always incomplete and thus approximations are essential in finding a decent solution.

Let μ be a piecewise continuous function in a domain $\Omega \subset \mathbb{R}^n$ and L a hyperplane with a dimension of $n - 1$. If $\omega \in \mathbb{S}^{n-1}$ is a unit vector normal to the plane, we can represent L as

$$L = L(\omega, s) = \{p \in \mathbb{R}^n | \langle \omega, p \rangle = s\}, \quad (11)$$

The Radon transform of the μ at (ω, s) is the integral along the hyperplane $L(\omega, s)$

$$R\mu(\omega, s) = \int_{L(\omega, s)} \mu(p) dS(p), \quad (12)$$

where dS is the Lebesgue surface measure on the plane $L(\omega, s)$. The Radon transform also defines μ uniquely

$$R_\omega \mu : \mathbb{R} \rightarrow \mathbb{R}, \quad s \rightarrow R\mu(\omega, s) \quad (13)$$

where $\omega \in \mathbb{S}^{n-1}$ is fixed.

3.2.2 Discretized model

Analytic inversion methods rely on complete projection data. However, in all real life applications complete projection data is not available for example due to radiation intake limits on live tissue or limited time to take measurements on a moving conveyor belt. Discretization of the problem is required by computers to make calculations possible for a given grid. The discretized observation model is then solved approximately.

In a two dimensional case this means dividing the image area into pixels and approximating the density of a piecewise constant function. Define the image area as a unit rectangle $[0, 1] \times [0, 1]$ and divide it into $N \times N$ pixels

$$\mu(p) = \mu_{ij}, \text{ when } p \in P_{ij}, 1 \leq i, j \leq N \quad (14)$$

where P_{ij} denotes the pixel area

$$P_{ij} = \left(\frac{j-1}{N}, \frac{j}{N} \right) \times \left(\frac{i-1}{N}, \frac{i}{N} \right), \quad (15)$$

i.e. the image is given in matrix form by μ_{ij} . To get the data in vector form, we stack the matrix values in a vector x with a length of N^2 so that the rows are stacked

$$x((i-1)N + j) = \mu_{ij}, \quad 1 \leq i, j \leq N. \quad (16)$$

If we use L_1, \dots, L_M to denote X-ray lines passing through our object then for the m th line, L_m , we have

$$y_m = \int_{L_m} \mu(p) dp = \sum_{i,j=1}^N |L_m \cap P_{ij}| \mu_{ij} = \sum_{k=1}^{N^2} A_{mk} x_k, \quad (17)$$

where $|L_m \cap P_{ij}|$ is the length of the intersection between the pixel P_{ij} and the line L_m , and we defined

$$A_{mk} = |L_m \cap P_{ij}|, \quad k = (i-1)N + j. \quad (18)$$

The model (17) can then be written as

$$Ax = y, \quad (19)$$

giving us the discretized model for the X-ray imaging as a linear matrix problem. The matrices are typically of large scale, but also very sparse. Depending on the problem (e.g. restricted angle/number of lines) the inverse can become extremely undetermined.

3.3 Experimental design

Experimental design or design of experiments is about making decisions of every aspect of an experiment beforehand [7]. It is important to keep in mind available resources and design the best set of experiments based on the known limitations. It is also important to design the experiments around one's objective. What is our goal or information we are after? What are the key features that depend on the problem? Depending on the problem we will also have different types of limitations.

Before data collection we have to make some decisions. Decisions can include parameter values, length of experiments, experimental units, sample sizes and much more [7]. What is the range of our experiments or where do we restrict ourselves. Sample size can affect computation time or analysis of data. How do we know how much data is required for accurate results? Accuracy can be adjusted for computer programs, but for real data it needs to be specified before data collection. This includes taking into account accuracy and error of chosen sensor. These are all typical decisions we have to deal with to get meaningful data.

How does this relate to Bayesian optimal design. Experimental design has always considered prior information when deciding on what type of experiments are used. Bayesian approach takes a different approach and assumes prior information to behave as a probability function, thus incorporating some degree of uncertainty in the model. Bayesian optimal designs are a class of experimental designs and the purpose of the Bayesian optimal design is to make optimal decisions under uncertainty. The optimization is done by choosing an optimization criteria that are optimal based on some statistical criteria. In a case where we have multiple experimental sets, rather than using the same model for every set we can use the previous tests to adjust our model for the next experimental sets. This can be used in regular experimental design, but optimal designs can produce more accurate results since they are optimal.

3.4 Bayesian optimal design

Bayesian optimal design is based around choosing an optimization criteria. Multiple models exist, but in this paper we focus on A- and D-optimality. A-optimality ("average") minimizes the trace of the information matrix. D-optimality (determinant) maximizes the determinant of the information matrix.

In Bayesian optimal design our domain Ω consists of three types of regions/pixels:

- Region of interest (ROI): the area we are interested in
- Obstruction: a region that interferes with our analysis of ROI
- Background: the rest of the object.

This is a typical case in medical applications where we are only interested in a small part of body (i.e. an organ, a tumour or a bone) that is surrounded by tissue. Tissue in this case would be considered background. Metal objects, such as screws and nails would be considered as obstructions which block X-rays. Screws and nails are commonly used when fixing broken bones and we can assume that their positions are well-known. In industrial setting we could be looking for defective products.

Optimal design models can be used to solve problems with respect to different optimality criterias and thus reveal different information. Choosing the right optimality criteria for the problem is important. Bayesian optimal design is centered around optimizing the so-called Bayes cost (BC). The optimal design is defined as

$$P_{OD} = \arg \min_p \text{BC}_\Phi(p), \quad (20)$$

where Bayes cost is given by

$$\text{BC}_\Phi(p) = \int \int \Phi(x, y; p) \pi_{post}(x|y; p) dx \pi(y; p) dy, \quad (21)$$

for some loss function $\Phi : \mathbb{R} \rightarrow \mathbb{R}$. Bayes cost is defined with this general equation using optimality criteria, posterior and prior information. Changing the optimality criteria allows us to change the target of optimization and this will lead to different results. We will differentiate between different optimality criterias by defining $\Phi_A(p)$ and $\Phi_D(p)$ as A-optimal and D-optimal criteria respectively.

3.4.1 A-optimality

In the following theorem we prove that A-optimal design can be calculated by taking trace of the posterior covariance matrix.

Theorem 3.1. *Let $\Phi(x, y; p) = \Phi_A(p) = \|A(x - x_{CM}(y))\|^2$. It follows that $\text{BC}_{\Phi_A}(p) = \text{tr}(A\Gamma_{post}(p)A^T)$.*

Proof. A-optimal design is defined with mean square estimator instead of conditional mean, but in this case they are the same. This can be shown by writing [6]

$$\begin{aligned}\mathbb{E}\{\|X - \hat{x}\|^2|y\} &= \mathbb{E}\{\|X\|^2|y\} - \|\mathbb{E}\{X|y\}\|^2 + \|\mathbb{E}\{X|y\} - \hat{x}\|^2 \\ &\geq \mathbb{E}\{\|X\|^2|y\} - \|\mathbb{E}\{X|y\}\|^2.\end{aligned}\quad (22)$$

This equality holds only if $\hat{x}(y) = \mathbb{E}\{X - E\{X|y\}\} = x_{CM}$. Using this the estimated error becomes zero

$$\mathbb{E}\{X - x_{CM}\} = \mathbb{E}\{X - \mathbb{E}\{X|y\}\} = 0. \quad (23)$$

This in turn means that our conditional mean minimizes the trace of the covariance. A-optimality criteria was defined in equation (3.1) and using this the Bayes cost is

$$\text{BC}_{\Phi_A}(p) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \|A(x - x_{CM}(y))\|^2 \pi_{post}(x|y; p) dx \pi(y; p) dy. \quad (24)$$

The matrix norm can be presented through trace and the conditional mean was shown to be equal to the mean value of the posterior. Substituting these into the equation (24) gives us

$$\text{BC}_{\Phi_A}(p) = \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \text{tr}(A(x - \hat{x}(y; p))(x - \hat{x}(y; p))^T A^T) \pi(x|y; p) dx \pi(y; p) dy, \quad (25)$$

where $A^T A$ is a positive semi-definite matrix. Since the trace is linear and scalars do not affect it, we can move it outside integrals

$$\begin{aligned}\text{BC}_{\Phi_A}(p) &= \text{tr}(A \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} (x - \hat{x}(y; p))(x - \hat{x}(y; p))^T \pi(x|y; p) dx \pi(y; p) dy A^T) \\ &= \text{tr}(A \int_{\mathbb{R}^m} \Gamma_{post}(p) \pi(y; p) dy A^T) \\ &= \text{tr}(A \Gamma_{post}(p) A^T).\end{aligned}\quad (26)$$

A-optimality is then defined as

$$P_{OD} = \arg \min_p \text{tr}(A \Gamma_{post}(p) A^T) \quad (27)$$

□

3.4.2 D-optimality

Kullback-Leibler divergence (KL) is used in deriving D-optimal design and it is defined as

$$D_{KL}(P\|Q) = \int p(x) \log \left(\frac{p(x)}{q(x)} \right) dx, \quad (28)$$

where P and Q are continuous random variables [10]. However, Kullback-Leibler divergence can also be presented in explicit form if we are dealing with multivariate normal distributions, such as Gaussian, in n-dimensional case the expression would be

$$D_{KL}(\mu_{post,n} \|\mu_{pr,n}) = \frac{1}{2} \left[\text{tr}(\Gamma_{pr}^{-1} \Gamma_{post}) + (u_{pr} - u_{post})^T \Gamma_{pr}^{-1} (u_{pr} - u_{post}) - n + \log \left(\frac{\det \Gamma_{post}}{\det \Gamma_{pr}} \right) \right], \quad (29)$$

where μ_{post} and μ_{pr} are Gaussian distributions with means u_{post} and u_{pr} , and covariance matrices Γ_{post} and Γ_{pr} respectively [10].

In the following theorem we show that D-optimal design can be calculated by taking logarithmic determinant from the posterior covariance matrix.

Theorem 3.2. *Let $\Phi(x, y; p) = \Phi_D(p) = \log \left(\frac{\pi_{post}(x|y;p)}{\pi_{pr}(x)} \right)$. It follows that $BC_{\Phi}(p) = \frac{1}{2} \log \{ \det \Gamma_{post}(p) \} + c$, where c is a constant.*

Proof. We apply Gaussian distribution to equation (28) [7, 9]

$$\begin{aligned} D_{KL}(X|y \parallel X; p) &= \int_{\mathbb{R}^n} \log \left(\frac{\pi(x|y; p)}{\pi_{pr}(x)} \right) \pi(x|y; p) dx \\ &= \int_{\mathbb{R}^n} (\log(\pi(x|y; p)) - \log(\pi_{pr}(x))) \pi(x|y; p) dx. \end{aligned} \quad (30)$$

D-optimal design aims to maximize the expected value of $D_{KL}(X|y \parallel X; p)$ over all measurements. Minimization criteria is then defined as

$$BC_{\Phi_D}(p) = \mathbb{E}^y(-D_{KL}(X|y \parallel X; p)) = - \int_{\mathbb{R}^m} D_{KL}(X|y \parallel X; p) \pi(y; p) dy. \quad (31)$$

Differential entropy of $X|y$ is given as

$$h(X|y) = - \int_{\mathbb{R}^n} \log(\pi(x|y; p)) \pi(x|y; p) dx. \quad (32)$$

This can also be represented in explicit form due to $X|y$ being Gaussian

$$h(X|y) = \frac{n}{2} + \frac{n}{2} \log(2\pi) + \frac{1}{2} \log(\det(\Gamma_{post}(p))), \quad (33)$$

where n is the number of dimensions [7]. The expected value of the second term in equation (30) can also be given in explicit form [7]

$$\begin{aligned} \mathbb{E}^y \left(\int_{\mathbb{R}^n} \log(\pi_{pr}(x)) \pi(x|y; p) dx \right) &= \int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \log(\pi_{pr}(x)) \pi(x, y; p) dx dy \\ &= \int_{\mathbb{R}^n} \int_{\mathbb{R}^m} \pi(y|x; p) dy \log(\pi_{pr}(x)) \pi_{pr}(x) dx \\ &= \int_{\mathbb{R}^n} \log(\pi_{pr}(x)) \pi_{pr}(x) dx = -h(X) \\ &= -\frac{n}{2} - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det \Gamma_{pr}) \end{aligned} \quad (34)$$

where the final form is $-h(X)$ using the Gaussian random variable X . It is also note worthy that it does not depend on the design variable p unlike the first term. Since $h(X|y)$ does not depend on y we can give D-optimal design in the following form

$$\begin{aligned} \text{BC}_{\Phi_D}(p) &= \mathbb{E}^y(h(X|y)) - h(X) \\ &= \int_{\mathbb{R}^m} h(X|y) dy - h(X) \\ &= \int_{\mathbb{R}^m} \left(\frac{n}{2} + \frac{n}{2} \log(2\pi) + \frac{1}{2} \log(\det(\Gamma_{post}(p))) \right) dy \\ &\quad - \frac{n}{2} - \frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\det \Gamma_{pr}) \\ &= \frac{1}{2} (\log(\det \Gamma_{post}(p)) - \log(\det \Gamma_{pr})) \\ &= \frac{1}{2} (\log(\det \Gamma_{post}(p))) + c, \end{aligned} \quad (35)$$

where $c = -\frac{1}{2} \log(\det \Gamma_{pr})$ is a constant. The optimal design only depends on the design variable p . The prior is not affected by the p and is thus a constant. D-optimal design is then

$$P_{OD} = \arg \min_p \left\{ \frac{1}{2} \log(\det \Gamma_{post}(p)) + c \right\}. \quad (36)$$

□

3.4.3 Posterior

The posterior is needed to evaluate both A- and D-optimal designs, but there are some problems in (3.2) and (3.1). The main problem with both A- and D-optimality target functions is that they tend to have many local minima. We need a suitable method to evaluate optimization targets. To do this we can use a preassigned ROI to write the posterior covariance matrix in the following form

$$\Gamma_{post} = \begin{bmatrix} \Gamma_{ROI} & \Gamma_{mix} \\ \Gamma_{mix}^T & \Gamma_{rest} \end{bmatrix} = \begin{bmatrix} (\Gamma_{post})_{11} & (\Gamma_{post})_{12} \\ (\Gamma_{post})_{21} & (\Gamma_{post})_{22} \end{bmatrix} \quad (37)$$

where Γ_{ROI} and Γ_{rest} stand for the covariance matrices for the pixels belonging to ROI and the rest of the image [8]. Using this we can write the minimization problem in the following form

$$\text{BC}_{\Phi_D}(p) := \log(\det(\Gamma_{ROI}(p))). \quad (38)$$

Since we are only interested in the information gained from ROI, we can get rid of uninteresting pixels by deleting the rows and columns corresponding to these pixels. We can further change the posterior matrix by using Schur complement of Γ_{ROI} [6, 9, 11]

$$\begin{aligned} \det(\Gamma_{post}) &= \det \begin{bmatrix} \Gamma_{ROI} & \Gamma_{mix} \\ 0 & \Gamma_{rest} - \Gamma_{mix}^T \Gamma_{ROI}^{-1} \Gamma_{mix} \end{bmatrix} \\ &= \det(\Gamma_{ROI}) \det(\Gamma_{rest} - \Gamma_{mix}^T \Gamma_{ROI}^{-1} \Gamma_{mix}), \end{aligned} \quad (39)$$

where $\Gamma_{rest} - \Gamma_{mix}^T \Gamma_{ROI}^{-1} \Gamma_{mix}$ is the Schur complement which means that the inverse will be

$$\Gamma_{post}^{-1} =: \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} = \begin{bmatrix} * & * \\ * & (\Gamma_{rest} - \Gamma_{mix}^T \Gamma_{ROI}^{-1} \Gamma_{mix})^{-1} \end{bmatrix}. \quad (40)$$

Putting everything together and we have

$$\det(\Gamma_{ROI}) = \frac{\det(\Gamma)}{\det(\Gamma_{rest} - \Gamma_{mix}^T \Gamma_{ROI}^{-1} \Gamma_{mix})} = \frac{\det(\Sigma_{22})}{\det(\Sigma)}, \quad (41)$$

where both $\det(\Sigma)$ and $\det(\Sigma_{22})$ depend on the design variable p and to evaluate D-optimality calculating the logarithms is needed.

$$\Sigma(p) = \Gamma_{pr}^{-1} + R(p)^T \Gamma_{noise}^{-1} R(p), \quad (42)$$

$$\Sigma_{22}(p) = (\Gamma_{pr}^{-1})_{22} + R_2(p)^T \Gamma_{noise}^{-1} R_2(p), \quad (43)$$

where $R(p) = [R_1(p), R_2(p)] \in \mathbb{R}^{m \times n}$ is a columnwise partitioning. $R_1(p)$ corresponds to pixels in ROI and $R_2(p)$ for the rest of the pixels. The determinants for (41) can then be defined as

$$\det(\Sigma(p)) = \det(R(p)\Gamma_{pr}R(p)^T + \Gamma_{noise}) \det(\Gamma_{noise}^{-1}) \det(\Gamma_{pr}^{-1}), \quad (44)$$

$$\det(\Sigma_{22}(p)) = \det(R_2(p)((\Gamma_{pr}^{-1})_{22})^{-1}R_2(p)^T + \Gamma_{noise}) \det(\Gamma_{noise}^{-1}) \det((\Gamma_{pr}^{-1})_{22}). \quad (45)$$

The design variable p is only found in the first determinants, which contains small ($m \times m$) matrices. We now have to use the Cholesky decomposition for the p independent matrix $(\Gamma_{pr}^{-1})_{22}$. The Cholesky decompositions are

$$C(p)C(p)^T = R(p)\Gamma_{pr}R(p)^T + \Gamma_{noise}, \quad (46)$$

$$\tilde{C}(p)\tilde{C}(p)^T = R_2(p)((\Gamma_{pr}^{-1})_{22})^{-1}R_2(p)^T + \Gamma_{noise}. \quad (47)$$

Finally, we get

$$\log(\det(\Gamma_{ROI}(p))) = 2 \sum_{j=1}^m (\log(\tilde{d}_{jj}(p)) - \log(d_{jj}(p))) + d, \quad (48)$$

where $\tilde{d}_{jj}(p)$ and $d_{jj}(p)$ are the diagonal elements of $\tilde{C}(p)$ and $C(p)$, respectively, and $d \in \mathbb{R}$ is a constant value. With a specific ROI and posterior we can improve accuracy of our computations and better evaluate our local minima.

The posterior for A-optimality can be formed by following almost the same steps as in D-optimal case. In A-optimal case we have to minimize

$$\text{BC}_{\Phi_A}(p) := \text{tr}(A\Gamma_{post}(p)A^T) \quad (49)$$

for a given $A \in \mathbb{R}^{l \times n}$. The Cholesky decomposition is used again

$$C(p)C(p)^T = R(p)\Gamma_{pr}R(p)^T + \Gamma_{noise}. \quad (50)$$

An auxiliary matrix

$$B(p) = C(p)^{-1}R(p)\Gamma_{pr}A^T \quad (51)$$

is used to formulate the final answer. By using an auxiliary matrix and the definition in (7), the A-optimal Bayes cost is

$$\text{BC}_{\Phi_A}(p) = \text{tr}(A\Gamma_{pr}A^T - B(p)^T B(p)) = \text{tr}(A\Gamma_{pr}A^T) - \text{tr}(B(p)^T B(p)). \quad (52)$$

In other words

$$\mathbf{BC}_{\Phi_A}(p) = d' - \sum_{j=1}^m \sum_{k=1}^l (B_{jk}(p))^2, \quad (53)$$

where $d' = \text{tr}(A\Gamma_{pr}A^T)$ is a constant. Evaluating the constant d' is not necessary for the minimization, but calculating it allows us to compare minimal values of different iterations when running the sequential optimization algorithm (presented in the next chapter).

4 SEQUENTIAL OPTIMIZATION

This section covers the basic sequential algorithm used in the study. The algorithm is used by Burger et al. [8] in a different set of tests to observe differences between the A- and D-optimal designs.

Sequential optimization algorithm optimizes the parallel beam projections using prior information. We initialize the algorithm by defining image discretization ($N \times N$ pixels), number of sensors n and projection count per sensor, covariance matrices Γ_{pr} and Γ_{noise} , iteration count and shape of the ROI. Defining how the design parameter $p \in \mathbb{R}^2$ parametrizes projection beams is also required. Sequential optimization means that the calculated posterior probability density of the k th iteration is then used as the prior for the next $(k + 1)$ th iteration. During the optimization iterations either A- or D-optimal target function is evaluated. The prior covariance can be formed from the posterior according to either equation (6) or (7).

Algorithm 1

Choose the covariances Γ_0 and Γ_{pr} for the initial Gaussian prior $\mathcal{N}(x_0, \Gamma_0)$ and the noise model $\mathcal{N}(0, \Gamma_{noise})$. Select the ROI, the grid for p and the number of iterations K . Initialize $\Gamma_{pr} = \Gamma_0$ and $P = []$.

for $k=1, \dots, K$ **do**

for all p on the optimization grid **do**

 Evaluate $\Phi_D(p)$ or $\Phi_A(p)$

end for

 Find the minimizer p_k of the considered optimization target.

 Append $P = [P, p_k]$.

 Form the posterior covariance $\Gamma_{post}(p_k)$.

 Set $\Gamma_{pr} = \Gamma_{post}(p_k)$.

end for

return P

The matrix $P \in \mathbb{R}^{2 \times K}$ defines the optimized projection angles. These are locally optimal projections and they are not globally optimal. Optimizing the projections sequentially one by one is computationally more efficient and easier than simultaneously trying to find projections which are jointly optimal.

To improve the algorithm we can use adaptive ROI where ROI can be altered between iterations based on previous results. This would not increase computation time, but would require an expert to give feedback between iterations and change the ROI accordingly. This would add to increase total computation time, but otherwise the algorithm would stay the same. Adaptive ROI would have the benefit of making the ROI more accurate as iterations proceed and allow to change focus or target area as new information is introduced. This could be useful in practical applications where an operator is always present and the number of selected projection angles is low.

5 EXPERIMENTS

5.1 Programming environment

The experiments used ready made MATLAB codes used in [8]. The programming environment consisted of Microsoft Windows 10 (64 bit) operating system and MATLAB R2019b version.

MATLAB has been developed by The MathWorks -company. It is a computer program meant to be used for numeric computations and mathematical coding. It offers variety of tools to handle functions and matrices, depicting data figures, algorithm and user interface implementation, and interaction with other programming languages. It is also possible to install a variety of different types of toolboxes, but it was not necessary in this study.

5.2 Data

No real world data was used in the study. The datasets consisted of images (2-D matrices) with $N \times N$ pixels in unit interval. The images represented the prior distribution of mass absorption coefficient (approximately thickness of material) per pixel. This was the ground truth for the experiments. The images were also divided into three separate regions: ROI, obstruction and background.

The initial prior covariance for different areas is of the following form

$$(\Gamma_0)_{i,j} = \sigma^2 \exp\left(-\frac{|x_i - x_j|^2}{2l^2}\right), \quad (54)$$

where l is the correlation length and σ is the control parameter for noise. During the tests the background and ROI covariances are assumed to have a common prior. The obstruction is assumed to have a different prior which dampens the radiation more than the ROI or the background. This is done by increasing the expected absorption compared to other areas. The expected absorption ratio for obstruction is set to 500 and 1 for the ROI and background. Standard deviation for noise is then defined as

$$\sigma_{noise} = \log(1 + \sigma \exp(G\alpha)), \quad (55)$$

where α is a vector with expected absorption coefficients for every pixel and G is a sparse

X-ray projection matrix. G holds information of all the different combinations of line integrals and possible angles.

5.3 Evaluation criteria

Evaluation of results was conducted by comparing accuracy of methods using mean reconstruction error, standard deviation and visual inspection. The reconstruction error is evaluated by taking Euclidean 2-norm between posterior and a sample inside ROI

$$\text{Mean error} = \frac{1}{M} \sum_{m=1}^M \|\text{Sample}_m(\text{ROI}) - \text{Post}_m(\text{ROI})\|_2, \quad (56)$$

where M is the number of samples compared to reconstruction of posterior. Samples also affect the generation of reconstruction so we are not comparing the same reconstruction to multiple samples. The samples and reconstructions are compared within ROI pixels and accuracy outside ROI is irrelevant. Mean reconstruction error is then evaluated by doing this to a number of samples and averaging over all of them. Standard deviations were compared with visual inspection, but these were found to be almost identical between A- and D-optimal designs. The only difference being A-optimality having a head start. To make observing projection angle selection easier, histograms from the first 16 projection angles were made. The first 16 angles were selected based on the results of experiment two as they accounted for most of the decreased error. The results were compared with each others and the main evaluation criteria was the mean reconstruction error during iterations.

5.4 Description of experiments

Our experiments will be related to analysis of differences between A- and D-optimal designs, their accuracy and the variables affecting the results. We test how they find optimal angles in different settings and test their accuracy against a random method. The following parameters and their default values:

- $N = 50$: number of pixels per edge (resolution)
- $\nu = 20$: total number of virtual sensors
- $\beta = 1$: proportion of real sensors

- $m = 100$: number of available projection angles
- $\sigma = 0.005$: noise level
- $k = 40$: number of optimized projection angles
- R: region case (shape of ROI, obstruction & background)

Number of pixels determines our grid and image size. Proportion of real sensors β is the ratio between the amount of sensors and sensor placements. Virtual sensors times proportion of real sensors gives us the amount of real sensors e.g. if β is less than 1 then we will have less sensors, but the same amount of sensor placement options. In that case we can optimize sensor placement when there are more positions than sensors. This was not necessary in this study. Number of projection angles m gives us the amount of possible angles spread around 180° . The angles are evenly spread and are considered as possible projection angles, which the algorithm optimizes. The 0° angle is going to be a projection from bottom to top, 90° is going to be from left to right and 180° is going to be from top to bottom. Noise level σ describes the ratio between noise level and the intensity of the source. The number of optimized projection angles k gives us the number of projection angles that are going to be selected by the algorithm from the available projection angles defined by m . One simple accuracy test depicted in figure 1 was conducted to gauge the appropriate number of angles.

Mean error test is conducted separately using previously calculated angles and settings with a randomized prior. To ensure accuracy M samples are taken from the prior. All cases used $M = 10000$ samples. Errors are calculated for every sample and then the mean value is taken. Mean errors of each method are plotted and used to determine accuracies of different methods. 20 iterations or less seems to be a good cut-off point based on the results in figure 1, but depending on the difficulty of the test more or less iterations are needed for a good reconstruction. For this reason we are using 40 iterations for our tests.

The accuracy of A- and D-optimal methods were tested against random method. The random method involves randomizing the measurement angle for each iteration, where as A- and D-optimal designs find the optimal angle for the next measurement according to sequential optimization algorithm. By comparing the results with the ground truth after every iteration, we can calculate the accuracy of each method. The accuracy depends on the number of iterations and the accuracy increases with every iteration, but we are interested in accuracy of A- and D-optimal designs compared against each others e.g. How much does the accuracy increase with every iteration and can the other method catch up?

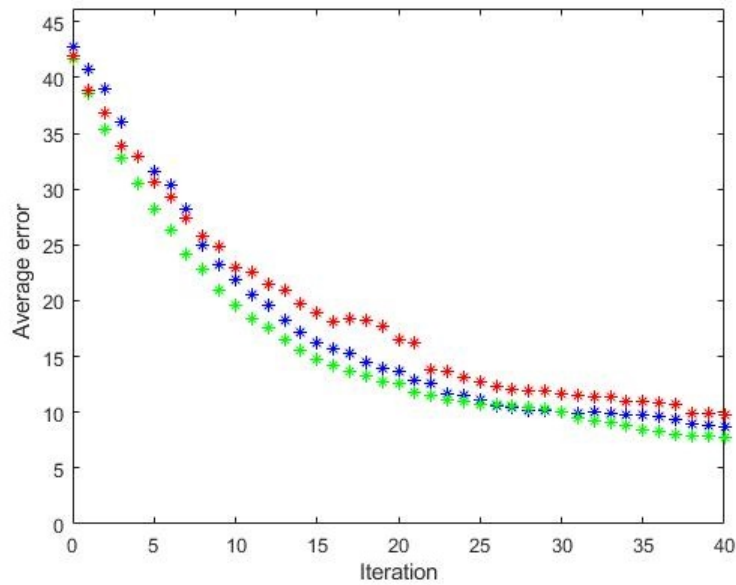


Figure 1. Mean reconstruction error in ROI with 40 iterations. Green is A-optimality, blue is D-optimality and red is random method. The y-axis represents the mean reconstruction error calculated with (56) using 10000 samples. The x-axis represents iterations or the number of optimized projection angles. The first angle is the starting angle, which is selected randomly and the following angles are optimized using the sequential algorithm. The random method selects projection angles randomly and is expected to perform worse when compared against A- and D-optimal designs.

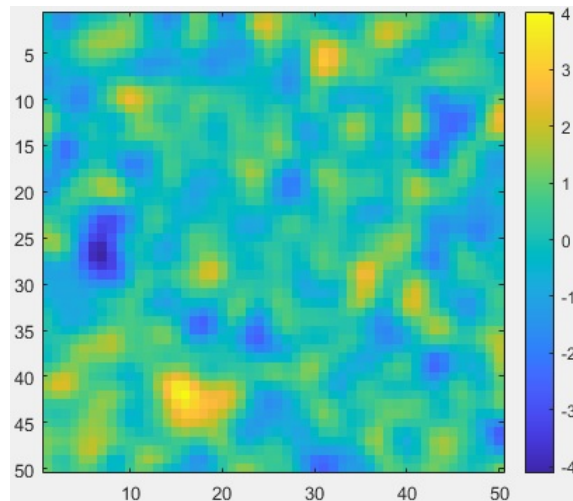


Figure 2. Example draw from prior distribution when only ROI and background are present. The prior is formed according to (54).

5.4.1 Experiment 1

The first experiment is a sanity check with full ROI (no obstructions or background). Full ROI should have the same effect as full background, but due to algorithm optimizing ROI it does not work when ROI is not present. However, both ROI and background have a common prior so there will not be an issue. With this setting the angles are expected to be evenly spaced.

For this test we are using 50×50 grid which gives us 2500 pixels. 20 virtual sensors and 100 line integrals per sensor are used to scan the grid. The parallel beam source-receiver pair covers 98% of the width of the grid which results in almost full scans from any angle. Restricting this would make the scan area smaller, but would put more line integrals on the selected area. We are also limiting the number of available projection angles to 100. These are evenly spaced around 180° . The noise level σ is set to 0.005 and the number of optimized angles k is set to 40. Essentially meaning that from 100 available angles we are going to find 40 optimized angles. The point of the experiment is to see that the algorithm is not biased and is behaving as expected.

5.4.2 Experiment 2

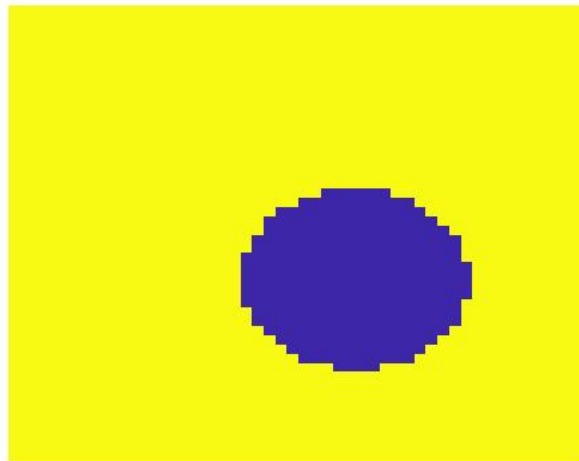


Figure 3. ROI (purple) and background (yellow) placement for experiment two. The ROI is a circular disc placed off center to avoid even angles. Even though the ROI is placed off center we might still get results with even angles due to wide parallel beam.

The second experiment is going to be with a circular ROI placed off center and the rest is background as in figure 3. The ROI is circular, but not symmetric due to discretization. The placement is off center so the selected projection angles should not be evenly spaced, but due to almost full scans it is very likely that the angles will be evenly spaced. The experiment aims to highlight differences between A- and D-optimal designs in an easy case without obstructions with a small ROI. The parameters are going to be the same as in the first experiment. The point of the test is to observe differences between methods and accuracy when a small ROI is introduced. The random method is expected to perform worse compared to sanity check.

5.4.3 Experiment 3

The experiment three is going to introduce an obstruction near the ROI. The experimental setup is otherwise the same as in experiment two, but now the obstruction is blocking the left side of the ROI. This is going to restrict projection angles as selecting angles from the left/right is going to provide less information or no information at all about the ROI compared to top/bottom angles. The expected absorption for the obstruction is set to 500, which is significantly higher than the ROI or the background which are set to 1. Other parameters are going to be the same as in experiment two. The point of the experiment is to test how A- and D-optimal designs react to obstruction and how they optimize their angles. Both methods are expected to avoid placing projection angles near the obstruction. The obstruction covers roughly 40° of unusable angles (full shadow) and roughly 120° of half usable angles (80° half shadow around 40° full shadow).

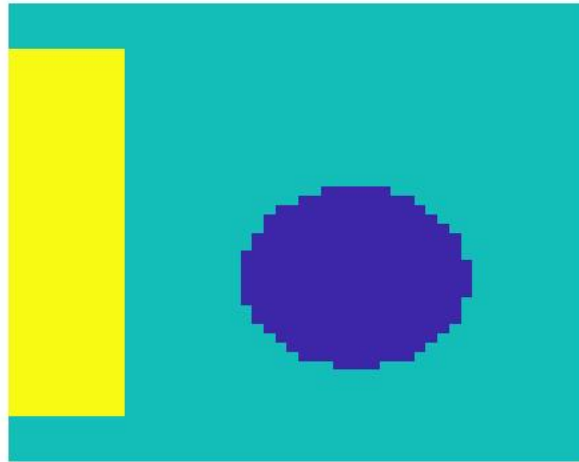


Figure 4. ROI (purple), obstruction (yellow) and background (blue) placement for experiment three. The ROI is similar to experiment two, but now we have placed an obstruction near it. A- and D-optimal designs are expected to avoid obstruction when optimizing projection angles.

5.5 Results

5.5.1 Results 1

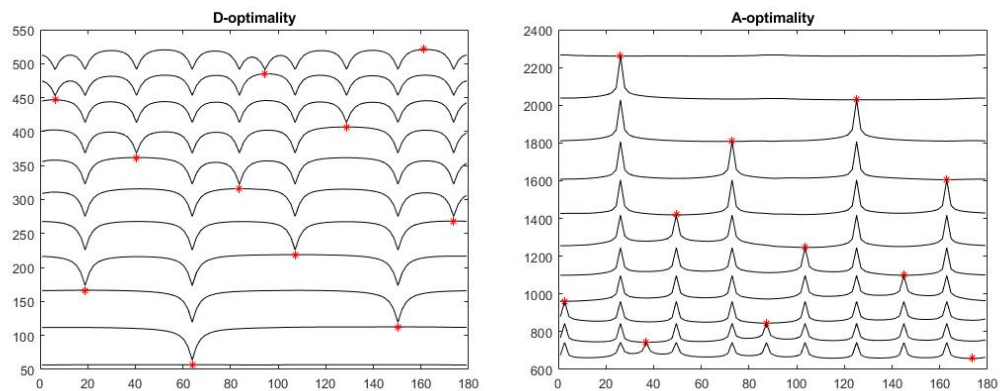


Figure 5. Experiment 1. D-optimal (left) and A-optimal (right) targets as a function of the projection angle. The first 11 target functions and projection angles are shown (one starting angle and ten optimized angles). Stars denote the selected angles and the iterations proceed from bottom to top for D-optimality and from top to bottom for A-optimality.

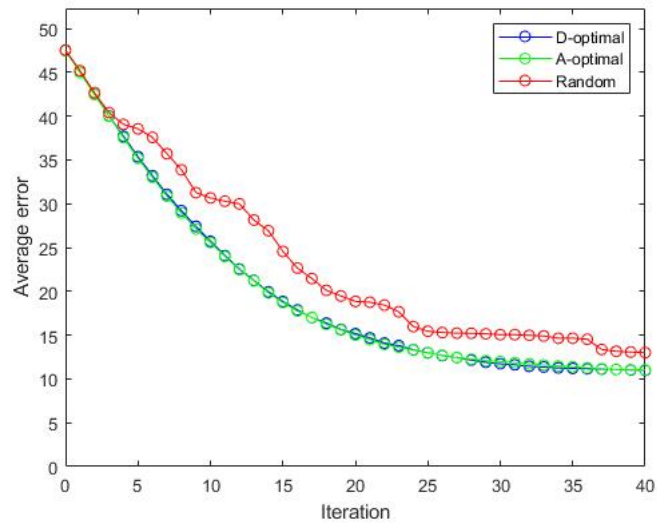


Figure 6. Experiment 1. Mean reconstruction error in ROI by iterations. Reconstruction error is estimated by comparing projection angles with a sample of 10000 randomized prior distributions. Error analysis is conducted for all the samples and the mean value of all errors is plotted.

Figure 5 depicts the projection angle selection. The iterations proceed from bottom to top for D-optimality and from top to bottom for A-optimality. Stars depict the chosen angles and the lines depict A- and D-optimal target functions. D-optimality maximises and A-optimality minimises the respective target functions. The target functions give values for every possible angle and selecting a new projection angle is done by maximizing/minimizing the current target function. We can see in figure 5 that when a new projection angle is selected the angles close to previous projection angles are less likely to be selected as a new projection angle.

Experiment one angles are evenly spaced as expected in figure 5. In both A- and D-optimal cases the projection angles take the midpoint between previously chosen angles. Even angles also make the results very similar. A- and D-optimal designs are practically the same and the random method has slightly worse results due to angles being random and uneven. This leads to A- and D-optimal designs to be superior compared to the random method in the sanity test.

5.5.2 Results 2

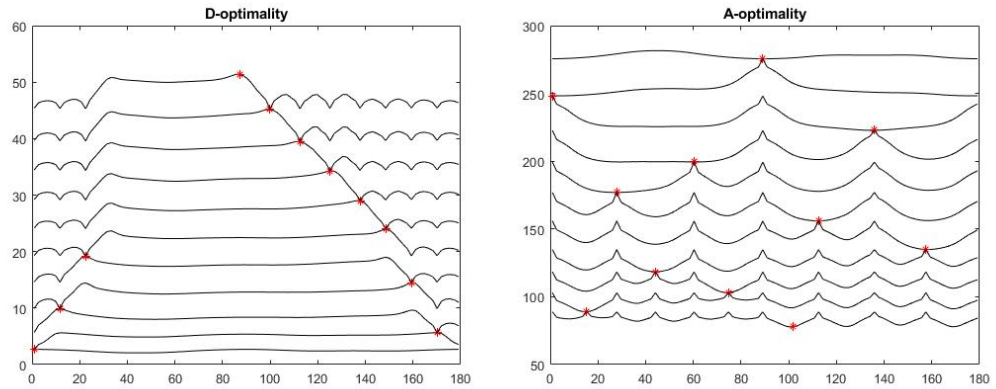


Figure 7. Experiment 2. D-optimal (left) and A-optimal (right) targets as a function of the projection angle. The first 11 target functions and projection angles are shown (one starting angle and ten optimized angles). The Stars denote the selected angles and the iterations proceed from bottom to top for D-optimality and from top to bottom for A-optimality.

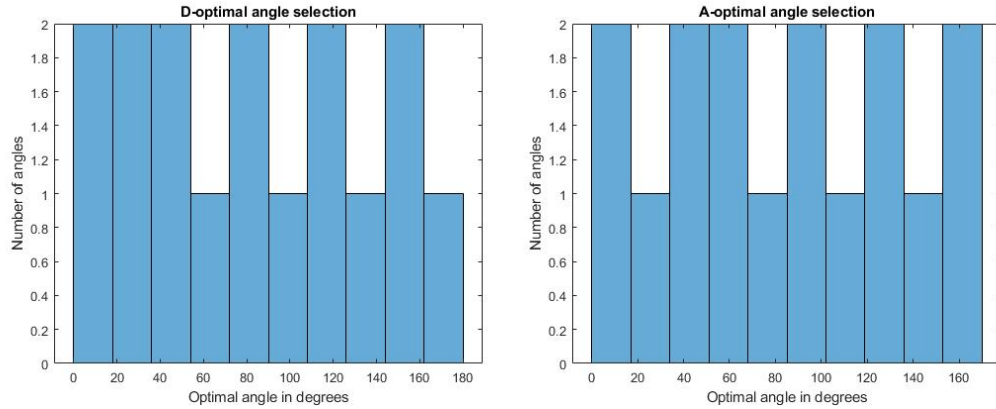


Figure 8. Experiment 2. The first 16 projection angles (one starting angle and 15 optimized angles) presented as histograms with ten bins. The first 16 angles were selected to be the point of interest since they account for the highest decrease in errors and both A- and D-optimal designs are equal at that point.

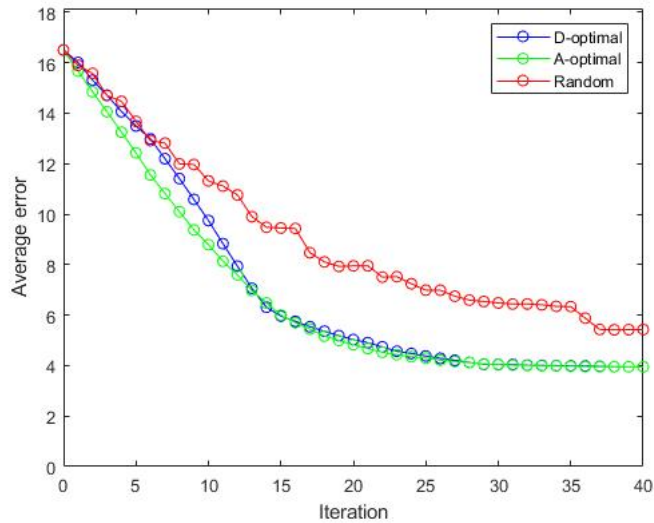


Figure 9. Experiment 2. Mean reconstruction error in ROI by iterations. Reconstruction error is estimated with a sample of 10000 randomized prior distributions. Error analysis is conducted for all the samples and the mean value of all errors is plotted.

Experiment two used circular ROI placed off center which changes the results. A- and D-optimal designs are no longer equal. A-optimal design is now the best method with the lowest mean error, but D-optimality catches up at iteration 14 and after that the results are similar. Similar results were observed in [8] with different experimental settings where during the first ten iterations A-optimality seems a lot better compared to D-optimality. However, in this test we are using higher iteration count to observe differences past ten iterations. Both methods show diminishing returns past 15 projection angles and is likely the reason D-optimality is able to catch up. Sequential optimization design does not seem to work that well with D-optimality when compared to A-optimality, but as long as iteration count is not too low D-optimality is able to produce just as good results as A-optimality. The random method is behaving randomly and at the start it is able to keep up with A- and D-optimal designs, but when the iterations proceed further the chance of placing projection angles close to each others becomes higher. This leads to random method becoming the least accurate of all methods.

Interestingly the overall mean error has decreased dramatically from the sanity test, most likely due to having to optimize less pixels due to smaller ROI, but the accuracy is also most likely higher in the ROI and worse outside it when compared to sanity check. In sanity check the errors are dispersed throughout the entire image, but in this test the smaller ROI concentrates algorithms efforts towards smaller area. This makes the reconstruction better in the target area, but outside it becomes worse. This shows how important it is

to have a proper ROI. Having a small ROI is also a risk, since we might miss something important outside it, but it makes optimization easier and provides more accurate results inside the targeted area.

Due to circular ROI projection angles are still quite evenly spaced in figure 8, but compared to sanity test it is not perfect. We are of course not aiming for symmetrical angles, but optimal ones. For A-optimality the symmetry breaks at the fourth projection angle which is no longer evenly spaced compared to sanity test. D-optimality sees a drastic change in figure 7 as it places the most preferable values next to previously chosen angles and between these angles we have slightly less preferable angles. This leads to "staircase" like structure in angle selection. D-optimality maximizes determinant and is not trying to reduce the overall error, which leads to falling behind at the start compared to A-optimality. However, D-optimality is able to catch up to A-optimality in later iterations which can be seen in both histograms 8 and mean reconstruction error in figure 9. Both methods are able to demonstrate the optimal angle selection process for a given ROI and the differences between these methods make it clear that A-optimal design is better compared to D-optimality.

5.5.3 Results 3

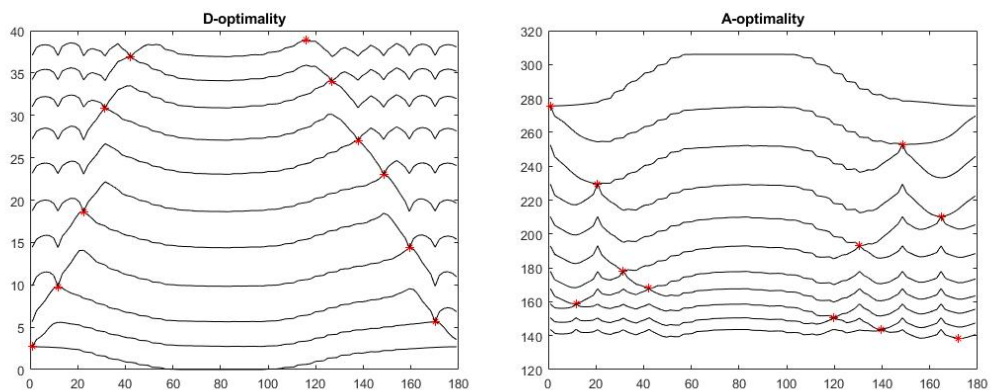


Figure 10. Experiment 3. D-optimal (left) and A-optimal (right) targets as a function of the projection angle. The first 11 target functions and projection angles are shown (one starting angle and ten optimized angles). Stars denote the selected angles and the iterations proceed from bottom to top for D-optimality and from top to bottom for A-optimality.

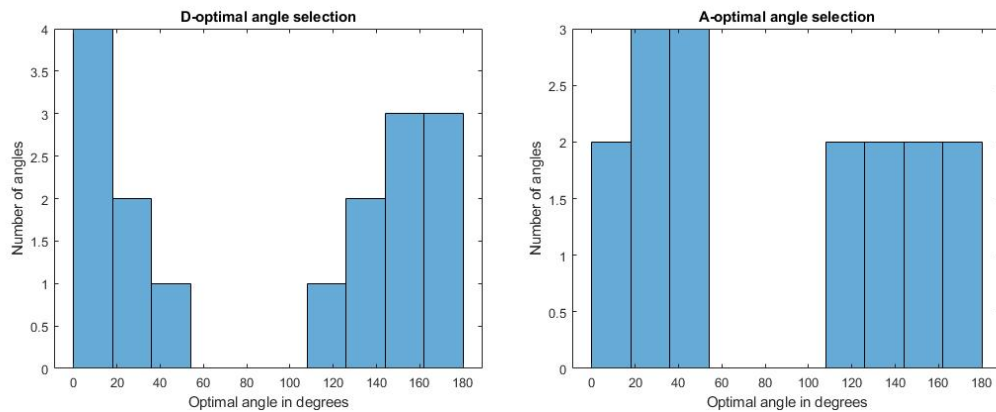


Figure 11. Experiment 3. The first 16 projection angles (one starting angle and 15 optimized angles) presented as histograms with ten bins. The 16 angles are used to get comparable results to experiment two even if the methods are equal at an earlier state.

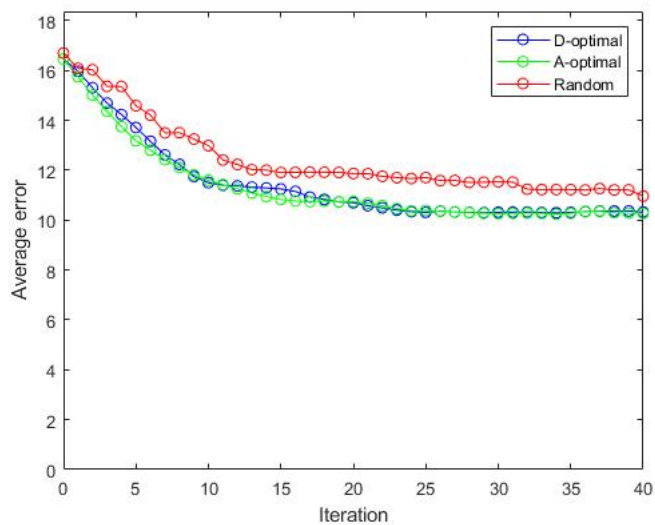


Figure 12. Experiment 3. Mean reconstruction error in ROI by iterations. Reconstruction error is estimated with a sample of 10000 randomized prior distributions. Error analysis is conducted for all the samples and the mean value of all errors is plotted.

Adding obstruction near ROI affects greatly all methods. After all, scanning the image from the obstruction's side (or the opposite site) provides very little information from the ROI. The effect of obstruction can be observed from the target functions in figure 10. Both methods show a flat part in the middle, which indicates the range of completely useless angles. Besides the flat part we have usable angles, but they provide less information than a non obstructive angle. We can also observe a staircase like structure in the target

functions which is most likely due to having the ability to place more line integrals and sensors when moving away from the middle part which is covered by the obstruction.

A- and D-optimal designs are still almost equal, but the mean errors have increased with both methods. Both methods are still linear up to the 10th iteration, but the mean values for almost every iteration have gotten higher. The 15th iteration mean value for example has risen from 6 to 11 compared to experiment two. D-optimality still has slightly worse values than A-optimality at the start, but it manages to catch up at the 9th iteration in a similar way as in experiment two.

Compared to experiment two the selected projection angles are now more focused on avoiding one part of available angles. Both methods are avoiding angles between 54° and 108° according to histograms in figure 11. These histograms are using 16 projection angles and ten bins to have comparable results with experiment two even if experiment three stabilizes earlier. A-optimality uses even angles while D-optimality still has the "staircase" structure. D-optimality also seems to have unevenly selected angles based on the histogram, but this is simply due to selecting 16 angles. Taking a look at the full projection angle selection reveals that D-optimality selects projection angles evenly, but does this by completing one staircase at a time. Depending on how many angles one selects for the histograms the results can vary a lot. A-optimality on the other hand selects the angles individually and maintains a balance no matter how many angles one selects for the histograms. This is why the histogram of D-optimality is uneven in this test.

6 DISCUSSION

6.1 Current study

Both A- and D-optimal designs are good at optimizing the target area by decreasing errors compared to the random method. A-optimality seems to be the best optimal design from the two tested ones at increasing the accuracy of results. D-optimality is able to catch up, but this takes multiple projection angles and by the time D-optimality is able to catch up to A-optimality, A-optimality is already at a stable state and gains diminishing returns from successive projection angles. At this point both methods are equal and this gives A-optimality a distinctive advantage when compared to D-optimality.

A-optimality is more accurate at the start and later becomes equal with D-optimality. This is due to differences in projection angle selection. D-optimality has a tendency to pick angles close to previous angles which leads to "staircase" like structure on both sides. A-optimality does the opposite and chooses angles between previously chosen angles which leads to even angles even when there is obstruction present. Both methods equalize later, but the even angles of A-optimality give it an advantage at early angles as even angles are able to produce more accurate reconstruction.

6.2 Future work

The future study could include other optimal designs e.g. C-, E- or T-optimality. Parameter sensitivity could also be interesting and we could test if D-optimality is able perform better in other areas than A-optimality. In this study we used high absorption obstruction that prevents projection beams from passing through it. Low absorption obstruction could have different effects when projection beams are able to pass through the obstruction. This would not block projection beams, but projection beams passing through obstruction would have lower intensity and possibly more noise compared to beams which do not pass through obstruction. Pairing the low absorption obstruction with higher noise level could reduce the effectiveness of both methods compared to the random method. Although, low absorption obstruction does not have many practical uses in clinical industry as most obstructions are different metals inside human body. It could be used in factories when dealing with different metals with varying absorption models. These are just some things worthy of consideration, but testing higher noise level is probably one of the more interesting ones.

7 CONCLUSION

The thesis covered A- and D-optimal designs. Small region of interest is able to greatly reduce reconstruction errors due to sequential algorithm having to optimize less pixels. Both methods are good at reducing the errors and increasing the reconstruction accuracy compared to the random method. A-optimal design is superior compared to D-optimal design. This is due to differences in projection angle selection between the methods. A-optimality selects angles between previously chosen angles and D-optimality places angles next to previously chosen angles leading to a "staircase" like structure. Selection method causes A-optimality to choose even angles which leads to increase in accuracy in early iterations compared to D-optimality. Eventually D-optimality is able to catch up when more projection angles are selected, but when D-optimality catches up to A-optimality both methods are equal at that point. Selecting more projection angles leads to diminishing returns for both methods and neither of them is able to outdo the other.

REFERENCES

- [1] Simone Carmignato, Wim Dewulf, and Richard Leach. *Industrial X-ray computed tomography*. Springer, 2018.
- [2] PC Seynaeve and JI Broos. The history of tomography. *Journal belge de radiologie*, 78(5):284–288, 1995.
- [3] Kirstine Smith. On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and the guidance they give towards a proper choice of the distribution of observations. *Biometrika*, 12(1/2):1–85, 1918.
- [4] Anirban DasGupta. 29 review of optimal bayes designs. *Handbook of Statistics*, 13:1099–1147, 1996.
- [5] JD Gergonne. The application of the method of least squares to the interpolation of sequences. *Historia Mathematica*, 1(4):439–447, 1974.
- [6] Jari Kaipio and Erkki Somersalo. *Statistical and computational inverse problems*, volume 160. Springer Science & Business Media, 2006.
- [7] Kathryn Chaloner and Isabella Verdinelli. Bayesian experimental design: A review. *Statistical Science*, pages 273–304, 1995.
- [8] M. Burger, A. Hauptmann, T. Helin, N. Hyvönen, and J.-P. Puska. Sequentially optimized projections in x-ray imaging. <https://arxiv.org/abs/2006.12579>, 2020.
- [9] Ahmed Attia, Alen Alexanderian, and Arvind K Saibaba. Goal-oriented optimal design of experiments for large-scale bayesian linear inverse problems. *Inverse Problems*, 34(9):095009, 2018.
- [10] Alen Alexanderian, Philip J Gloor, Omar Ghattas, et al. On bayesian a-and d-optimal experimental designs in infinite dimensions. *Bayesian Analysis*, 11(3):671–695, 2016.
- [11] DM Titterton. Optimal design: some geometrical aspects of d-optimality. *Biometrika*, 62(2):313–320, 1975.