

Optimisation of the measurement setting in parallel beam X-ray imaging

Aatu Selkee

School of Science

Bachelor's thesis
Espoo 23.8.2024

Supervisor

Prof. Nuutti Hyvönen

Advisor

Prof. Nuutti Hyvönen

Copyright © 2024 Aatu Selkee

The document can be stored and made available to the public on the open internet pages of Aalto University.
All other rights are reserved.



Author Aatu Selkee		
Title Optimisation of the measurement setting in parallel beam X-ray imaging		
Degree programme Bachelor's programme in science and technology		
Major Mathematics and system analytics		Code of major SCI3029
Teacher in charge Prof. Nuutti Hyvönen		
Advisor Prof. Nuutti Hyvönen		
Date 23.8.2024	Number of pages 35	Language English

Abstract

In this thesis, parallel beam X-ray tomography is considered. The imaging setup is discretised and modelled as a linear inverse problem, which is solved using Bayesian methods. The goal is to optimise the measurement setting with respect to an optimality criterion given by Bayesian optimal experimental design. We present ways to find the posterior distribution with both a Gaussian prior and a total variation prior using A-optimality. In addition, we present the derivative of the A-optimality target function with respect to the variance of the noise in the model. These results are verified numerically with different experiments.

Two algorithms are considered in this thesis. One of them is an "offline" algorithm, which uses a Gaussian prior. This algorithm is independent of the measurement data and can thus be run before the data is gathered. The other one is an "online" algorithm, which uses the total variation prior. The total variation prior is more complex than the Gaussian prior. As a result, there is no closed-form formula for the posterior distribution using the total variation prior. Consequently, lagged diffusivity iteration is employed to approximate the posterior distribution using Gaussian distributions.

In the numerical experiments, we conduct three tests, where we attempt to show the capabilities of the algorithms in different circumstances. The first experiment aims to demonstrate how the algorithm with the Gaussian prior focuses the majority of radiation on a predefined region of interest. The other experiments aim to show the reconstructions of two target absorption distributions in both high noise and low noise conditions. In the low noise scenario, the algorithm with the total variation prior performed better due to its more realistic assumptions. However, in high noise conditions, the total variation prior performed worse than the Gaussian prior. Overall, both algorithms performed well in the considered reconstruction tasks.

Keywords Parallel beam X-ray tomography, Bayesian optimal experimental design, A-optimality, Gaussian prior, total variation prior, lagged diffusivity iteration



Tekijä Aatu Selkee

Työn nimi Mittausasetelman optimointi yhdensuuntaisten säteiden
röntgenkuvantamisessa

Koulutusohjelma Teknistieteellinen kandidaattiohjelma

Pääaine Matematiikka ja systeemitieteet

Pääaineen koodi SCI3029

Vastuuopettaja Prof. Nuutti Hyvönen

Työn ohjaaja Prof. Nuutti Hyvönen

Päivämäärä 23.8.2024

Sivumäärä 35

Kieli Englanti

Tiivistelmä

Tässä tutkielmassa käsitellään yhdensuuntaisten säteiden röntgenkuvantamista. Kuvantamisjärjestelmä diskretoidaan ja mallinnetaan lineaarisena inversio-ongelmana, joka ratkaistaan bayesiläisin menetelmin. Tavoitteena on optimoida mittausasetelma bayesiläisen optimaalisen koesuunnittelun antaman optimaalisuuskriteerin mukaisesti. Tutkielmassa esitellään menetelmiä posteriorijakauman löytämiseksi sekä Gaussin priorilla että kokonaisvaihtelupriorilla (engl. total variation) käyttäen A-optimaalisuuskriteeriä. Lisäksi esittelemme A-optimaalisuuskriteerin derivaatan mittausmallin kohinan varianssin suhteen. Näitä tutkimustuloksia sovelletaan numeerisissa kokeissa erilaisille koeasetelmille.

Tutkielmassa käsitellään kahta algoritmia. Algoritmeista ensimmäinen on "offline-algoritmi, jossa käytetään gaussista prioria. Tämä algoritmi ei riipu kuvantamisesta saatavasta datasta, joten sitä voidaan käyttää jo ennen datan keräämistä. Toinen algoritmeista on "online-algoritmi, jossa käytetään kokonaisvaihteluprioria. Kokonaisvaihtelupriori on monimutkaisempi kuin gaussinen prior, minkä takia kaavaa posteriorijakaumalle ei ole. Tästä syystä viivytetyn diffuusion iteraatiota käytetään posteriorijakauman estimoimiseksi gaussisten jakaumien avulla.

Numeerisissa kokeissa tehdään kolme erilaista testiä, joissa esitellään algoritmien kykyjä erilaisissa olosuhteissa. Ensimmäinen koeasetelma pyrkii kuvaamaan, kuinka gaussisen priorin algoritmi keskittää suurimman osan säteilystä mielenkiintoiseen alueeseen. Muut kokeet pyrkivät näyttämään kahden tavoitejakauman rekonstruktioita sekä suurella että pienellä mittauskohinan määrällä absorptiodataan nähden. Pienen kohinan testissä kokonaisvaihtelu priorin algoritmi suoriutuu gaussista algoritmia paremmin sen realistisempien oletusten takia. Kuitenkin suuren kohinan testissä kokonaisvaihtelu priorin algoritmi suoriutuu huonommin kuin gaussinen prior. Kokonaisuudessaan molemmat algoritmit toimivat hyvin rekonstruktio tehtävissä.

Avainsanat Yhdensuuntaisten säteiden röntgenkuvantaminen, bayesiläinen optimaalinen kokeiden suunnittelu, A-optimaalisuus, gaussinen prior, kokonaisvaihtelu prior, viivytetyn diffuusion iteraatio

Contents

Abstract	3
Abstract (in Finnish)	4
Contents	5
1 Introduction	6
2 Literature review	7
3 Parallel beam X-ray imaging	8
3.1 Imaging setup	8
3.2 The forward model	9
3.3 Problem formulation	10
4 Bayesian inverse problems	11
4.1 Regularisation techniques	11
4.2 Bayesian inversion	11
5 Prior models	13
5.1 Gaussian prior	13
5.2 Total variation prior and lagged diffusivity iteration	14
6 Optimal experimental design	16
7 Algorithms and their implementations	17
7.1 Gradient descent	17
7.2 Golden section line search	20
7.3 Algorithm with a Gaussian prior	21
7.4 Algorithm with a TV prior and lagged diffusivity iteration	22
8 Numerical experiments	23
8.1 Circular region of interest	24
8.2 Open ring reconstruction	25
8.3 Shepp-Logan phantom reconstruction	30
9 Conclusions	32

1 Introduction

Inverse problems are a widely researched topic in modern mathematics [7, 12]. They offer a wide range of applications, such as determining an initial heat distribution inside an object based on the current distribution [7], or restoring the original image from a blurred version [7, 12]. In these examples, the goal is to determine the initial parameters using observations, which is typical for inverse problems. Conversely, in a forward problem, one tries to determine the observations based on a set of initial parameters.

A subset of inverse problems called linear inverse problems involves determining initial parameters, when the relationship between the observation and the parameters can be expressed as a linear transformation. Linear inverse problems have an advantage that they are often easier to solve compared to non-linear inverse problems due to modern computers being well optimised for large matrix operations. Although nonlinear inverse problems can represent more complex phenomena, they often result in more ill-posed problems, which means that either the existence, uniqueness or stability of a solution is more severely compromised [15]. In this thesis, we strictly consider finite-dimensional linear inverse problems, particularly in the field of parallel beam X-ray tomography.

Usually, inverse problems are not considered without taking noise into account. This makes determining the initial parameters harder due to the information being distorted [18]. However, taking noise into account in our models often yields better outcomes in practice, as applications necessarily include some noise [3]. In this thesis, we assume the noise to be distributed according to a multivariate normal distribution. This results in some nice properties for the posterior distribution assuming that the prior distribution is also Gaussian [10].

There are two main solution methodologies for solving linear inverse problems [7, 10]. The first is a classical approach using regularisation methods and least squares techniques. This approach yields a point estimate for the optimal parameters, which is not ideal for inference, giving motivation for another method, a probabilistic approach using Bayes' rule and Bayesian statistics. By applying Bayes' rule, we obtain a probability distribution for the solution of the inverse problem, enabling more versatile inference compared to the classical approach [7, 10].

Bayesian optimal experimental design (OED) is a way of finding the optimal parameters for the considered measurement setup [17]. In Bayesian OED, we aim to maximize the "utility" received from a specific experiment [10]. There are many ways to define "utility" [17], however, in this thesis we consider the optimal design parameters to minimise the so-called A-optimality target function. In other words, the optimal design parameters minimise the sum of the variances of the unknown pixelwise absorptions [17], which works well in imaging tasks.

The goal of this thesis is to implement an algorithm similar to the approach described by Helin et al. [5] for edge-promoting X-ray imaging. The main difference is that we use a non-sequential algorithm instead of a greedy, sequential one to optimise the intensities of the employed X-rays simultaneously. This algorithm employs a total variation (TV) prior, which promotes piecewise constant functions [5, 10] and

better preserves edges in the target. In addition, we implement an algorithm that uses a Gaussian prior, to compare different prior distributions. Unlike the TV prior, the considered Gaussian prior tends to smooth out the target.

The studied algorithms are gradient-based and hence suffer from local minima. Although the exhaustive search used in [3] partially mitigates this issue, the gradient-based algorithms are computationally more efficient, and thus preferable. The implemented algorithms are expected to yield noticeably better reconstructions inside the pre-assigned region of interest (ROI) compared to the background of the image. Specifically, the algorithm should ensure that the posterior variance of pixel values within the ROI is lower than the variance of pixel values in the background. Overall, the approach in this thesis is similar to the algorithm used by Maaninen [10] in magnetorelaxometry imaging.

To maintain the scope of this thesis focused, many reasonable topics will be left out. Firstly, other types of OED metrics apart from A-optimality will not be experimented with. D-optimality is another commonly used optimality metric, which attempts to maximize the information gain from the posterior compared to the prior [3]. According to Wanyonyi et al. [17], A-optimality performs better than D-optimality in screening tasks. Burger et al. [3] numerically demonstrate that using the A-optimality criterion yields results closer to the global optimum in a simple X-ray imaging task compared to using D-optimal designs. Therefore, we will strictly consider A-optimal designs in this thesis. Secondly, we will not experiment with any other priors than the Gaussian prior and the TV prior. For example, Maaninen [10] also considers the Perona-Malik prior, which will not be included in this thesis as Maaninen observes it to perform slightly worse than the TV prior in magnetorelaxometry.

The following outlines the structure of this thesis. In section 2, we will briefly consider a few important sources for this thesis. In section 3, we will outline the basic setup of X-ray imaging. In sections 4 and 5, we will present the basics of the Bayesian method for solving inverse problems. Section 6 will define what it means for parameters defining a measurement setting to be optimal. Every algorithm that will be used in the numerical experiments in section 8 will be presented in section 7. Finally, we will present the conclusions in section 9.

2 Literature review

The field of X-ray tomography has attracted a significant amount of research prior to this thesis. However, there appear to be only a few works addressing Bayesian OED in this field [3, 5, 13]. Burger et al. [3] and Helin et al. [5] use a sequential algorithm instead of a non-sequential one, which they argue is more tractable, since the objective function is not convex. On the other hand, Ruthotto et al. [13] use a gradient based optimisation algorithm to find a minimum for an X-ray tomography problem that closely resembles the one in [3]. Contrary to [13], we consider two design variables, the angle and the offset of the X-ray source, similar to [3] and [5]. Considering also the offset allows the algorithm to choose the importance of each

ray more freely, but increases the computational complexity of the problem.

Bayesian OED has been previously applied to many imaging techniques [16], with the consideration of Maaninen in [10] being arguably the closest to the setting of this thesis. Maaninen [10] tests a gradient-based optimisation algorithm in magnetorelaxometry imaging, which is very similar to the one presented in this thesis. When using a Gaussian prior Maaninen applies gradient descent simultaneously to all design parameters as is the case in this thesis. With edge-promoting priors, such as the TV prior, Maaninen uses a sequential algorithm similar to the ones used in [3] and [5].

3 Parallel beam X-ray imaging

In this section, the fundamentals of parallel beam X-ray imaging are presented. Section 3.1 covers the basic setup for the imaging process. We present the forward model that is used to define the corresponding inverse problem in section 3.2. In section 3.3, the main inverse problem of this thesis is defined. The optimisation problem related to the OED is also discussed in section 3.3.

3.1 Imaging setup

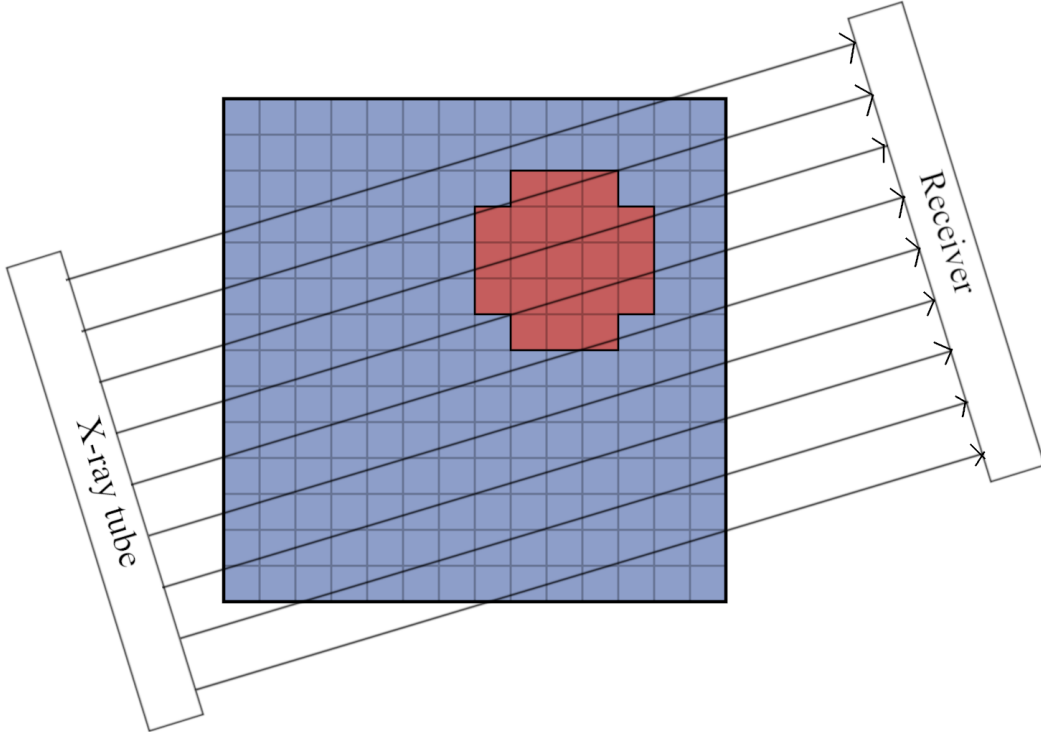


Figure 1: A picture of the basic imaging setup. The red area represents the ROI and the arrows represent the individual X-rays. The rays are emitted from the X-ray tube and detected at the receiver.

Parallel beam X-ray tomography is an imaging technique usually used to obtain a reconstruction of the absorption inside a target [7]. Since the work of Radon in 1917, it has been known that the reconstruction of the human body can be constructed based on 2D slices of the human body. In X-ray tomography, these slices are obtained by emitting X-rays of a certain intensity through the body from many angles and offsets and detecting the intensity data after passing through the body [7].

The imaging setup in parallel beam X-ray tomography consist of three main parts: the X-ray tube, the detector and the gantry. The gantry is often circular and lays around the target object, usually a human. Inside the gantry, moving parts hold the X-ray tube and the detector. These parts allow the X-ray tube to shoot many parallel X-rays from many angles and offsets, which yield data on the attenuation of intensity on each line [9]. A picture of the X-ray tube shooting parallel rays from a specific angle is shown in figure 1. Nowadays, parallel beam X-ray tomography has widely been replaced by other methods, which are similar to parallel beam X-ray tomography, but use a circular cone of X-rays instead of parallel rays.

3.2 The forward model

Using the setup provided in section 3.1, the attenuation of intensity I on a single ray L of an infinitesimal length dl is calculated using the following equation [7]:

$$dI = -Iu(x)dl(x), \quad (1)$$

where $u: D \rightarrow \mathbb{R}_+$ is a function that maps a point in the target object D to the corresponding absorption value [3]. Equation (1) yields the following formula for the log-intensity I at the receiver [7]:

$$\log(I) - \log(I_0) = \log\left(\frac{I}{I_0}\right) = \int_{I_0}^I \frac{dI}{I} = - \int_L u(x)dl(x). \quad (2)$$

Based on (2), the log-intensity data on the receiver is the value of the line integral along a path of some X-ray. The problem can then be expressed as, "reconstruct the distribution $u(x)$ based on values of line integrals passing through the domain D of $u(x)$ ".

Equation (2) gives us an easy way to define the forward model by discretising the target domain D into $n = N^2$ pixels. The measurement is made by computing the inner product of a vector containing the lengths $dl(x)$ of a ray passing through a pixel x and a vector containing the corresponding absorption values $u(x)$ [3]. Since it is inefficient to compute separate dot products for every ray that we emit, we organise the lengths of each intersection into a matrix $R_k \in \mathbb{R}^{m \times n}$, where m is the number of offsets of the ray and k indicates that we are considering the k -th illumination angle. Since we consider images from every direction simultaneously, matrices R_k can again be stacked on top of one another resulting in the following system matrix:

$$R = \begin{bmatrix} R_1 \\ R_2 \\ \vdots \\ R_k \end{bmatrix} \in \mathbb{R}^{km \times n}.$$

This gives rise to the following forward model [3]:

$$y = Rx + N, \quad (3)$$

where $x \in \mathbb{R}^n$ is the vector of pixel wise absorption values and $N \in \mathbb{R}^{km}$ is a vector containing noise values for each ray. Then $y \in \mathbb{R}^{km}$ is the measurement of the experiment containing information about the attenuation along each X-ray.

3.3 Problem formulation

The main problem addressed in this thesis is to optimise the measurement setup related to (3) from the standpoint of limited radiation dose. We will aim to implement an algorithm that optimises design parameters based on A-optimality with the goal of concentrating the allowed radiation dose inside the target area. We will also aim to reconstruct the absorption distribution inside the ROI.

The ROI outlines the specific area under investigation. In practise, the ROI might represent, e.g., a damaged organ or a particular body part of interest. It seems logical to concentrate the X-rays towards only a single area, if other body parts can be saved from the collateral radiation. This motivates the need for an algorithm that is able to automatically find the X-rays that carry the most information about the ROI.

For efficiently solving the inverse problem corresponding to the forward model discussed in section 3.2, we aim to optimise certain design parameters at the measurement setup. These design parameters d relate to the intensities of the X-rays going through the target. In this thesis, we adopt a model, where the dose of a single ray is inversely proportional to the variance of the corresponding component of the noise vector:

$$\text{dose}_i = \frac{c}{(d_i + \epsilon)^2}, \quad (4)$$

where $c > 0$ and $\epsilon > 0$ are constants with ϵ usually being small. Since radiation is harmful to humans, we want to restrict the total radiation dose. In this thesis, this is not done by reducing the number of angles like in [3], but the total dose is capped in the optimisation process. We consider images from every angle at once and limit the dose by varying the intensities of individual X-rays. This should yield results, where the uninteresting X-rays have lower intensities compared to the ones yielding more information about the ROI. This restriction results in the following constrained optimisation problem:

$$\begin{cases} d^* = \underset{d \in \mathbb{R}^{km}}{\operatorname{argmin}} \Phi_A(d), \\ \sum_{i=1}^{km} \frac{1}{(d_i + \epsilon)^2} \leq C, \end{cases} \quad (5)$$

where C a positive constant and $\Phi_A(d)$ is the A-optimality target function that is described in more detail in section 6.

The unfeasible set $\{d \mid \sum_{i=1}^{km} \frac{1}{d_i^2} > C\}$ complicates the optimisation problem significantly, since the feasible set is not convex. This property makes most gradient based

optimisations methods harder to use, since these methods are prone to converging to a local minimum. Line search methods are also nontrivial to use, since it is possible to hit the unfeasible set, even if both endpoints of the line search are feasible. These optimisation methods will be discussed in more detail in sections 7.1 and 7.2.

4 Bayesian inverse problems

In this section, we consider the Bayesian approach to inverse problems. In section 4.1, we quickly review regularisation methods that are used to solve linear inverse problems. In addition, we discuss the flaws of these classical methods. In section 4.2, we present the Bayesian approach for solving inverse problems. In other words, we display the needed formulas for the posterior distribution.

4.1 Regularisation techniques

Traditionally, inverse problems have been solved using regularisation methods, such as Tikhonov regularisation [7]. Since many inverse problems have solutions that exhibit irregular behaviour as the data-fit gets better, it is useful to solve a linear inverse problem $Ax = y$ by tackling the following minimisation problem:

$$\operatorname{argmin}_{x \in X} F_\delta(x) = \|Ax - y\|^2 + \mathcal{J}(x),$$

where $\mathcal{J}: X \rightarrow [0, \infty)$ is a regularisation function that increases as x increases in an appropriate sense [1]. In the basic form of Tikhonov regularisation, the regularisation function is chosen to be $\mathcal{J}(x) = \delta \|x\|^2$, where $\delta > 0$ is a parameter that guarantees the unique existence of the minimiser [1, 7].

The problem with the traditional approach is that we only get a point estimate of the correct solution without information about its reliability. For this reason, it is useful to aim at obtaining a probability distribution as a solution to an inverse problem. With the Bayesian approach, it is also easier to take prior knowledge explicitly into account in terms of a prior distribution, since with traditional methods, the prior is implicitly incorporated into the regularisation function \mathcal{J} or the parameters of some other deterministic regularisation technique [1].

4.2 Bayesian inversion

With Bayesian approach to inverse problems, we assume that every variable is sampled from some probability distribution, including the parameters we want to solve for. The solution to the inverse problem is then achieved by forming the posterior distribution for the parameters of interest [7]. Consider the linear forward model described in (3), i.e., the model

$$Y = RX + N, \tag{6}$$

where Y , X and N are now random variables. Also let $d \in \mathbb{R}^{km}$ be the vector of design parameters that enters (6) via $N = N(d)$ and $y_{\text{observed}} \in \mathbb{R}^{km}$ the observed

projection images. In addition, assume that the random variables X and N are independent of one another. This model is called an additive noise model [10]. Using Bayes' formula, we get the following equation for the conditional probability density of x given some measurement y_{observed} :

$$\pi(x|y_{\text{observed}}; d) = \frac{\pi(y_{\text{observed}}|x; d)\pi_{\text{pr}}(x)}{\pi(y_{\text{observed}}; d)}, \quad (7)$$

where $\pi : \mathbb{R}^{km} \rightarrow \mathbb{R}_+$ is the likelihood function [3].

It is common to assume that the random variables X and N follow multivariate Gaussian distributions

$$X \sim \mathcal{N}(x_{\text{pr}}, \Gamma_{\text{pr}}) \text{ and } N \sim \mathcal{N}(0, \Gamma_{\text{noise}}),$$

where $x_{\text{pr}} \in \mathbb{R}^n$ is the prior mean vector for the image data, $\Gamma_{\text{pr}} \in \mathbb{R}^{n \times n}$ the corresponding positive definite prior covariance matrix and $\Gamma_{\text{noise}}(d) \in \mathbb{R}^{km \times km}$ the positive definite covariance matrix for the noise. In addition, we assume that the individual components of the noise are independent, which implies that $\Gamma_{\text{noise}}(d)$ is diagonal. More precisely, the noise covariance matrix is defined as

$$\Gamma_{\text{noise}}(d) = \text{diag}(d_1^2, d_2^2, \dots, d_{km}^2) + \epsilon^2 I, \quad (8)$$

where $\epsilon > 0$ is a small constant that could e.g., represent digital noise, which is impossible to remove. Mathematically ϵ guarantees that $\Gamma_{\text{noise}}(d)$ is positive definite. Defining the noise covariance this way allows us to interpret the dose of radiation in (4) to be proportional to the variance of the noise. This may not be entirely accurate physically, however such model is intuitively acceptable and has a suitable behaviour for the problem. That is, more radiation induces lower noise in the measurement, since more photons hit the detector.

Based on these assumptions, the posterior distribution in (7) is also a multivariate Gaussian distribution with the covariance matrix [3, 7]

$$\Gamma_{\text{post}}(d) = (\Gamma_{\text{pr}}^{-1} + R^T \Gamma_{\text{noise}}(d)^{-1} R)^{-1} \quad (9)$$

and the mean

$$x_{\text{post}}(d) = \Gamma_{\text{post}}(\Gamma_{\text{pr}}^{-1} x_{\text{pr}} + R^T \Gamma_{\text{noise}}(d)^{-1} y_{\text{observed}}). \quad (10)$$

However, these can be very time consuming to calculate, since (9) requires matrix inversions of sizes $n \times n$, $n \times n$ and $km \times km$, while (10) requires inversions of sizes $n \times n$ and $km \times km$. Each inversion can be done in $\mathcal{O}(n^3)$ or $\mathcal{O}((km)^3)$ operations. Therefore it is beneficial to apply the Woodbury matrix identity to (9) and (10) to write them as [2]

$$\Gamma_{\text{post}}(d) = \Gamma_{\text{pr}} - \Gamma_{\text{pr}} R^T (R \Gamma_{\text{pr}} R^T + \Gamma_{\text{noise}}(d))^{-1} R \Gamma_{\text{pr}} \quad (11)$$

and

$$x_{\text{post}}(d) = x_{\text{pr}} + \Gamma_{\text{pr}} R^T (R \Gamma_{\text{pr}} R^T + \Gamma_{\text{noise}}(d))^{-1} (y_{\text{observed}} - R x_{\text{pr}}), \quad (12)$$

respectively [3, 10]. Both of these calculations only require the inversion of the matrix $R\Gamma_{\text{pr}}R^T + \Gamma_{\text{noise}}(d) \in \mathbb{R}^{km \times km}$, which is much better computationally. In both cases we require an inversion of a $km \times km$ matrix, but in (11) and (12) we do not have to compute the additional inversions of the $n \times n$ matrices. Note that since we expect the number of pixels to be much larger than other quantities, i.e. $km \ll n$, the matrix inversion in (11) and (12) is expected to be faster than in (9) and (10). Also note that $R\Gamma_{\text{pr}}R^T + \Gamma_{\text{noise}}(d)$ is guaranteed to be invertible, since $\Gamma_{\text{noise}}(d)$ is positive definite and $R\Gamma_{\text{pr}}R^T$ is positive semi definite.

It should be noted that (11) and (12) can only be applied when the prior distribution is Gaussian. If this is not the case, the posterior distribution is no longer Gaussian, which significantly complicates the calculations [10]. Consequently, the posterior mean and covariance cannot be explicitly represented. When the prior distribution is not Gaussian, some other methods, such as lagged diffusivity iteration in section 5.2 must be used to approximate the parameters of the posterior distribution.

5 Prior models

The assumptions on the prior distributions have a considerable impact on the results of the algorithm. Mainly the characteristics of the reconstruction depend greatly on the chosen prior covariance function. In this section we will look into two common priors that each have different advantages.

5.1 Gaussian prior

In X-ray imaging, photons are used to capture data on the absorption inside the imaged body [7]. Classically, these photons arriving at the detector are assumed to be Poisson-distributed, with exponential inter-arrival times [7]. Since the Poisson distribution resembles a Gaussian distribution given the expected value is high enough, we use a Gaussian distribution to model the amount of noise in the data. In addition, a Gaussian prior is used as the prior distribution of the unknown, as it allows for an explicit representation of the posterior distribution, although the assumptions of the Gaussian prior do not entirely capture the characteristics of the unknown distribution.

The characteristics of a Gaussian prior are defined by a covariance function that determines the correlation between any two components of the unknown absorption. Intuitively two values are correlated, if the corresponding locations are close to one another with respect to the Euclidean norm. A Gaussian prior with such a property can be defined based on the following formula for the covariance matrix [4]:

$$(\Gamma_{\text{pr}})_{ij} = \sigma^2 e^{-\frac{\|x_i - x_j\|_2^2}{2l^2}}, \quad i, j \in \{1, 2, \dots, n\}, \quad (13)$$

where $\sigma, l \in \mathbb{R}$ are the point wise standard deviation and the correlation length respectively, and x_i and x_j are the coordinates of the midpoints of pixels with indices

i and j in the unknown absorption. Equation (13) demonstrates that the prior covariance is symmetric, due to the symmetry of the Euclidean norm. Hence, it can be shown that (13) yields a positive definite covariance matrix. We also see that as the correlation length grows, the exponent grows as well, which results in larger correlations for components that are further apart. Similarly if the correlation length decreases, the correlation of two far away components approaches quickly zero. The parameter σ defines the overall level of variance for any component.

A Gaussian prior offers some considerable benefits in the context of linear inverse problems. Firstly, if we assume that the prior and the noise are independent and follow a Gaussian distribution in (3), the posterior will also follow a Gaussian distribution with the mean vector and covariance matrix described in section 4.2. This allows for easy computational methods in optimisation algorithms. Another advantage of Gaussian priors is that the design parameters can be optimised non-sequentially [10]. This means that one can optimise for every parameter at once for example using gradient descent. A non-sequential approach is also possible in the setup of [3], however, the non-sequential approach requires evaluating a very large number of design parameter combinations, which is infeasible if the number of images is large. In this thesis, we use the non-sequential algorithm, since it is possible for us to calculate the explicit derivatives of the target function described in section 6. This makes it feasible to perform the optimisation non-sequentially.

One aspect of using a Gaussian prior distribution is that the posterior covariance can be calculated "offline". This means that the measurements do not affect the posterior covariance, but only the posterior mean. Therefore the posterior covariance can be calculated without knowing the measurements [3]. Hence, in an OED application the computational complexity does not matter as much, since we can do the optimisation computations before even meeting the "patient".

5.2 Total variation prior and lagged diffusivity iteration

A total variation prior is a prior distribution that promotes piecewise constant functions. Specifically, the TV prior promotes functions, where large regions of the domain remain constant with sharp edges between each region. Intuitively, the TV prior should be more accurate for X-ray imaging compared to a Gaussian prior in section 5.1, since the contents of the ROI are assumed to have piecewise constant absorption values. For instance, this distribution more accurately models the absorption values within the human body due to the unique absorption characteristics of each organ.

The TV prior distribution for the absorption X is defined as

$$\pi(x) \propto \exp(-\gamma\Phi(x)), \quad \text{with } \Phi(x) = \int_D \varphi(\|\nabla x(t)\|_2) dt, \quad (14)$$

where $\gamma > 0$ is a free parameter and $\varphi(t) = \sqrt{t^2 + T^2} \approx |t|$ is a differentiable approximation of the absolute value function with a small constant $T > 0$ [5]. We use the approximation to guarantee differentiability, which is needed for the lagged diffusivity iteration. The variable γ controls how strongly different pixels

correlate with each other [6]. With this prior, we can not use (11) and (12) as mentioned in section 4.2. This motivates the use of lagged diffusivity iteration from [5] to approximate the posterior corresponding to $\Phi(x)$, which is otherwise difficult to calculate. The idea is to iteratively estimate the log-posterior using quadratic approximations that allow unique minimisers.

We achieve the method by first calculating the gradient $\nabla_x \Phi(x)$. From (14), we get [5]

$$\nabla_x \Phi(x) = H(x)x, \quad (15)$$

where $H(x)$ is element wise defined as

$$H_{i,j}(x) = \int_D \frac{1}{\sqrt{\|\nabla_t x(t)\|_2^2 + T^2}} \nabla \phi_i(t) \cdot \nabla \phi_j(t) dt. \quad (16)$$

Here $\{\phi_i(t)\}_{i=1}^n$ is a Lagrangian finite element basis of D , where $\phi_i(t)$ evaluates to one at the midpoint of the i :th pixel and is zero otherwise. Moreover, $x \in \mathbb{R}^n$ is identified with the function $x(t) = \sum_{i=1}^n x_i \phi_i(t)$. As Helin et al. [5] point out, $H(x)$ is a stiffness matrix of an elliptic partial differential operator with a homogeneous Dirichlet boundary condition. This implies that $H(x)$ is positive definite [5] and thus invertible, which is necessary for the iterative algorithm.

In the iterative process of lagged diffusivity, we define the j :th approximation of the posterior mean $x_{\text{post}}^{(j)}$ as

$$x_{\text{post}}^{(j)} = \Gamma_{\text{post}}^{(j-1)} R^T (R \Gamma_{\text{post}}^{(j-1)} R^T + \Gamma_{\text{noise}})^{-1} y_{\text{observed}}, \quad (17)$$

where $\Gamma_{\text{post}}^{(j-1)} = \gamma^{-1} H(x_{\text{post}}^{(j-1)})^{-1}$ can be interpreted as a covariance matrix for a zero-mean Gaussian distribution [6]. In the lagged diffusivity iteration, $x_{\text{post}}^{(j)}$ is repeatedly calculated until a stopping criterion is satisfied. An approximate posterior covariance matrix for the absorption values can then be calculated using the posterior mean by the following equation [6]:

$$\Gamma_{\text{post}} = \Gamma_{\text{post}}^{(j-1)} - \Gamma_{\text{post}}^{(j-1)} R^T (R \Gamma_{\text{post}}^{(j-1)} R^T + \Gamma_{\text{noise}})^{-1} R \Gamma_{\text{post}}^{(j-1)}. \quad (18)$$

However, it is not necessary to calculate Γ_{post} during the iterative process. It is sufficient to calculate Γ_{post} when the stopping criterion is satisfied. In the end, we also set $x_{\text{post}} = x_{\text{post}}^{(J)}$, where J is the index when the stopping criterion was satisfied. The full algorithm will be presented in section 7.4.

As mentioned in section 5.1, the posterior covariance can be calculated "offline" when using a Gaussian prior. In contrast, with the TV prior, it is not possible to calculate the posterior covariance without knowing the results of the measurement. Therefore, lagged diffusivity iteration is employed as an "online" algorithm, which updates the posterior distribution as new measurement data becomes available. Overall, the more realistic assumptions of the TV prior makes it preferable compared to the Gaussian prior.

6 Optimal experimental design

Optimal experimental design (OED) in the context of X-ray imaging aims to choose the design parameters of a measurement setup so that information on the unknown absorptions is maximised. In this thesis, we will consider only A-optimality, although there exist many other types of OED as well. Every type of Bayesian OED is designed to minimise or maximise some statistical quantity. These commonly include the expected squared distance of the unknown from a point estimate as in A-optimality or the information gain in the posterior distribution compared to the prior distribution as in D-optimality [3].

A-optimality criterion can be interpreted as minimising the posterior variance of the pixel wise absorption values X inside the ROI in (6) [8]. This is intuitive, since it is logical to aim for maximum certainty about the absorption values inside the ROI when solving the inverse problem. The pixel wise absorption values define the target object, and therefore it makes sense to use A-optimality.

Often A-optimality is considered to be the best metric for optimisation in screening tasks [8]. Other types of optimality, such as D-optimality may perform worse than A-optimality in imaging, since the A-optimality criterion focuses on concrete aspects of the problem. A-optimality criterion also performs well in terms of other types of optimality criteria [8]. For example, A-optimal designs often perform well with g-efficiency compared to D-optimal designs, which is remarkable, since D-optimality is often credited to do better in terms of many different optimality criteria [8]. A-optimal designs often also perform well in terms of the D-optimality criterion.

Although A-optimality seems to be the best optimality criterion in terms of screening tasks, there are also negative aspects of using A-optimality. Minimisation of the A-optimality target function may lead to only locally optimal solutions [3], although, e.g. D-optimality also suffers from the same problem. This complicates the optimisation process and could benefit from using more advanced optimisation methods.

The A-optimality criterion corresponds to the minimisation of the following expression:

$$\Phi_A(d) = \text{tr}(A\Gamma_{\text{post}}(d)A^T), \quad (19)$$

where $\Gamma_{\text{post}}(d)$ is the posterior covariance matrix described in (11) or (18) and $A \in \mathbb{R}^{n \times n}$ is a weight matrix that can be used to assign different levels of importance for different components of the unknown [3, 10]. In this thesis, A is diagonal and is used to define the ROI. The diagonal of the matrix A is the vectorised version of the ROI, where the pixels inside the ROI are indicated by 1 and pixels outside the ROI are indicated by 0. If we set $A = I \in \mathbb{R}^{n \times n}$ to be the identity matrix, then the A-optimality criterion $\Phi_A(d)$ in (19) yields the expected squared reconstruction error $\mathbb{E}(\|x - x_{\text{post}}\|_2^2)$ with respect to the Euclidean norm [3]. This is a useful metric when trying to maximise the reliability of x_{post} as a reconstruction of the true unknown absorption values. The modified A-optimality target function $\frac{1}{N}\sqrt{\Phi_A}$ can also be used as a performance metric. It corresponds to the expected $L^2(D)$ -reconstruction error, where D is the ROI [3, 10].

7 Algorithms and their implementations

In this section, we describe the algorithms used in the numerical experiments. The basic gradient descent is described in section 7.1. The calculations needed for evaluating the gradient are also presented. Section 7.2 presents an implementation for the line search algorithm that is used in combination with the gradient descent.

In section 7.3, we implement a basic algorithm with a Gaussian prior. In section 7.4, however, we use the TV prior, requiring a different algorithm, since (11) and (12) are not applicable when using the TV prior. The algorithm has to use the lagged diffusivity iteration described in section 5.2. The numerical experiments using these algorithms are presented in section 8.

7.1 Gradient descent

In this thesis, the optimisation for the best design parameters d is executed using gradient descent. Gradient descent is an optimisation algorithm that takes steps towards the direction of fastest decrease in the target function. For gradient descent to be implemented, we need the first partial derivatives of the target function $\Phi_A(d)$ described in section 6 with respect to the design parameters.

Target function gradient

The needed partial derivatives are calculated explicitly in the following way:

$$\frac{\partial \Phi_A(d)}{\partial d_i} = \frac{\partial}{\partial d_i} \text{tr}(A \Gamma_{\text{post}}(d) A^T).$$

Since the trace and the derivative are linear operators, the derivative can be moved to the inside of the trace-function. Also A is independent of d_i , which yields us

$$\begin{aligned} \frac{\partial}{\partial d_i} \text{tr}(A \Gamma_{\text{post}}(d) A^T) &= \text{tr}\left(\frac{\partial}{\partial d_i} A \Gamma_{\text{post}}(d) A^T\right) \\ &= \text{tr}\left(A \frac{\partial \Gamma_{\text{post}}(d)}{\partial d_i} A^T\right). \end{aligned}$$

Let us consider only the derivative of the posterior covariance matrix $\Gamma_{\text{post}}(d)$. Based on (11) and the linearity of the partial derivative operator, we get

$$\begin{aligned} \frac{\partial \Gamma_{\text{post}}(d)}{\partial d_i} &= \frac{\partial}{\partial d_i} (\Gamma_{\text{pr}} - \Gamma_{\text{pr}} R^T (R \Gamma_{\text{pr}} R^T + \Gamma_{\text{noise}}(d))^{-1} R \Gamma_{\text{pr}}) \\ &= \frac{\partial}{\partial d_i} \Gamma_{\text{pr}} - \frac{\partial}{\partial d_i} \Gamma_{\text{pr}} R^T (R \Gamma_{\text{pr}} R^T + \Gamma_{\text{noise}}(d))^{-1} R \Gamma_{\text{pr}}. \end{aligned}$$

Since Γ_{pr} and R are independent of d , we get

$$\begin{aligned} \frac{\partial \Gamma_{\text{post}}(d)}{\partial d_i} &= -\frac{\partial}{\partial d_i} \Gamma_{\text{pr}} R^T (R \Gamma_{\text{pr}} R^T + \Gamma_{\text{noise}}(d))^{-1} R \Gamma_{\text{pr}} \\ &= -\Gamma_{\text{pr}} R^T \frac{\partial}{\partial d_i} ((R \Gamma_{\text{pr}} R^T + \Gamma_{\text{noise}}(d))^{-1}) R \Gamma_{\text{pr}}. \end{aligned}$$

The derivative of an inverse of a matrix can be calculated using the following formula:

$$\frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1},$$

where A is an invertible matrix that depends on x . Let $Z(d) = (R\Gamma_{\text{pr}}R^T + \Gamma_{\text{noise}}(d))^{-1}$. We get the partial derivative of $Z(d)$ as

$$\begin{aligned} \frac{\partial Z(d)}{\partial d_i} &= \frac{\partial}{\partial d_i} (R\Gamma_{\text{pr}}R^T + \Gamma_{\text{noise}}(d))^{-1} \\ &= -Z(d) \left(\frac{\partial}{\partial d_i} (R\Gamma_{\text{pr}}R^T + \Gamma_{\text{noise}}(d)) \right) Z(d). \end{aligned}$$

Since Γ_{pr} and R are independent of d , we get

$$\begin{aligned} \frac{\partial Z(d)}{\partial d_i} &= -Z(d) \left(\frac{\partial}{\partial d_i} (R\Gamma_{\text{pr}}R^T + \Gamma_{\text{noise}}(d)) \right) Z(d) \\ &= -Z(d) \left(\frac{\partial}{\partial d_i} R\Gamma_{\text{pr}}R^T + \frac{\partial}{\partial d_i} \Gamma_{\text{noise}}(d) \right) Z(d) \\ &= -Z(d) \left(\frac{\partial}{\partial d_i} \Gamma_{\text{noise}}(d) \right) Z(d). \end{aligned}$$

Based on section 3.3,

$$\Gamma_{\text{noise}}(d) = \begin{bmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & d_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_{km}^2 \end{bmatrix} + \epsilon^2 I.$$

Hence we get the derivative

$$\begin{aligned} \frac{\partial Z(d)}{\partial d_i} &= -Z(d) \left(\frac{\partial}{\partial d_i} (\Gamma_{\text{noise}}(d)) \right) Z(d) \\ &= -Z(d) L(d_i) Z(d), \end{aligned}$$

where $L(d_i) \in \mathbb{R}^{km \times km}$ is given component wise as

$$L(d_i)_{jl} = \begin{cases} 2d_i & \text{if } j = l = i, \\ 0 & \text{otherwise,} \end{cases}$$

where the only nonzero element is at (i, i) . Substituting back, we get the following formula:

$$\begin{aligned} \frac{\partial \Gamma_{\text{post}}(d)}{\partial d_i} &= -\Gamma_{\text{pr}} R^T \frac{\partial Z(d)}{\partial d_i} R \Gamma_{\text{pr}} \\ &= \Gamma_{\text{pr}} R^T Z(d) L(d_i) Z(d) R \Gamma_{\text{pr}}. \end{aligned}$$

Finally, we obtain the wanted partial derivative

$$\begin{aligned}
\frac{\partial \Phi_A(d)}{\partial d_i} &= \frac{\partial}{\partial d_i} \text{tr}(A\Gamma_{\text{post}}(d)A^T) \\
&= \text{tr}\left(A \frac{\partial \Gamma_{\text{post}}(d)}{\partial d_i} A^T\right) \\
&= \text{tr}(A\Gamma_{\text{pr}}R^T Z(d)L(d_i)Z(d)R\Gamma_{\text{pr}}A^T).
\end{aligned} \tag{20}$$

Barrier function

In addition to the basic gradient descent, we consider a barrier function to mitigate the effect of the feasible set. In this thesis, we use a logarithmic barrier function defined by the equation

$$B(d) = -\theta \ln\left(C - \sum_{j=1}^{km} \frac{1}{(d_j + \epsilon)^2}\right),$$

where $\theta > 0$ is a small constant and C the radiation dose limit. The logarithmic barrier function helps the algorithm to find a minimum inside the feasible set, since the gradient close to the barrier is large and points away from the unfeasible set. The gradient of the logarithmic barrier function is

$$\frac{\partial B(d)}{\partial d_i} = -\frac{2\theta(d_i + \epsilon)^{-3}}{C - \sum_{j=1}^{km} \frac{1}{(d_j + \epsilon)^2}}.$$

The A-optimality criterion with the barrier function is then

$$\Phi_A(d) = \text{tr}(A\Gamma_{\text{post}}(d)A^T) - \theta \ln\left(C - \sum_{j=1}^{km} \frac{1}{(d_j + \epsilon)^2}\right) \tag{21}$$

and the gradient function is

$$\frac{\partial \Phi_A(d)}{\partial d_i} = \text{tr}(A\Gamma_{\text{pr}}R^T Z(d)L(d_i)Z(d)R\Gamma_{\text{pr}}A^T) - \frac{2\theta(d_i + \epsilon)^{-3}}{C - \sum_{j=1}^{km} \frac{1}{(d_j + \epsilon)^2}}. \tag{22}$$

Gradient descent algorithm

Based on the obtained gradient function, we can define the gradient descent algorithm described as algorithm 1.

Algorithm 1 Gradient descent

Choose initial parameter d_0 and the stopping criterion τ .
Set $d = d_0$.
while not converged **do**
 Calculate $\nabla\Phi_A(d)$ using (22).
 if $\|\nabla\Phi_A(d)\|_2 < \tau$ **then**
 break
 end if
 Set d according to algorithm 2.
 Calculate $\Phi_A(d)$ based on (21). ▷ Only for monitoring progress.
end while
return d

7.2 Golden section line search

In this thesis, we apply a line search method to aid the gradient descent to converge faster. Line search methods help gradient based optimisation methods to select the optimal step size, which often results in fewer iterations before convergence. The golden section search is a method, which iteratively computes function values to seek a minimum value along the given line section [11]. However, since the feasible set is not convex, as noted in section 3.3, a modified version of the line search is used.

A modified version of the golden section search helps us ensure that we hit the feasible set when updating the design parameters. First we try to apply the golden section search between the points d and $d - \alpha\nabla\Phi_A(d)$, where α is the initial step size of the algorithm. If we hit the unfeasible set, we try to apply the golden section search again, but using a shorter interval. If we hit the unfeasible set n times, the $n + 1$:th golden section search will be between the points d and $d - \alpha\beta^n\nabla\Phi_A(d)$, where $\beta \in [0, 1)$ is a constant. This guarantees that we hit the feasible set every iteration of the gradient descent. However, in theory we might still "jump over" the feasible set to the other side. In practice, we did not experience such phenomena happen while carrying out the numerical experiments in section 8.

The golden section search was selected instead of other line search methods, since it offers a good balance of accuracy and speed. The golden section search requires one target function evaluation at each iteration, since one of the previous two points can be reused [11]. The golden section line search is an exact line search, which means that it is guaranteed to return a solution d that is at most a length τ away from the optimal solution assuming the target function is convex along the considered line segment, where τ is the stopping criterion [11]. This makes the golden section search efficient for finding good and feasible solutions for the gradient descent algorithm. Note that we did not experiment with the commonly employed Armijo rule, which is an inexact line search method. It may perform better than the golden section search, which we found to be sufficiently effective for our purposes. Our implementation of the modified golden section search algorithm is presented in algorithm 2.

Algorithm 2 Golden section search

Select initial step size α , a reduction constant $\beta \in [0, 1)$, the stopping criterion τ and the initial parameter d_0 .

Initialise the inverse golden ratio $\phi^{-1} = 0.6180339$, the reduction $r = 1$ and the initial parameter $d = d_0$.

while feasible solution not found **do**

 Set $a = d$ and $b = d - \alpha r \nabla \Phi_A(d)$.

 Set $\lambda = a + (1 - \phi^{-1})(b - a)$ and $\mu = a + \phi^{-1}(b - a)$.

 Set $r = \beta r$.

if a, b, λ or μ not in feasible set **then**

continue

end if

while $\|b - a\|_2 > \tau$ **do**

if $\Phi_A(\lambda) > \Phi_A(\mu)$ **then**

 Set $a = \lambda, \lambda = \mu$ and $\mu = a + \phi^{-1}(b - a)$.

if μ not in feasible set **then**

break ▷ Did not find a feasible solution

end if

else

 Set $b = \mu, \mu = \lambda$ and $\lambda = a + (1 - \phi^{-1})(b - a)$.

if λ not in feasible set **then**

break ▷ Did not find a feasible solution

end if

end if

end while

end while

if $\Phi_A(\lambda) < \Phi_A(\mu)$ **then**

 Set $d = a$.

else

 Set $d = b$.

end if

return d

7.3 Algorithm with a Gaussian prior

In this section, we will present an algorithm that we use in section 8 to solve the problem described in section 3.3. The idea of algorithm 3 is to calculate the optimal design parameter d and the posterior covariance matrix Γ_{post} . Based on Γ_{post} we get many figures for inference presented in section 8. Algorithm 3 is based on algorithm 1 by Maaninen [10] with modifications mainly after the optimisation step.

Algorithm 3 is initialised by choosing an appropriate initial design parameter d , a ROI, the radiation dose limits, as well as an initial covariance matrix Γ_0 . In the basic algorithm, Γ_0 is initialised according to 5.1. On the other hand, the radiation dose limits are chosen, so that they can be incrementally increased to make the

Algorithm 3 Offline algorithm using the Gaussian prior

Choose the initial design parameter d_0 , the ROI A and the radiation dose limits L .
 Initialise the covariance matrices Γ_0 and Γ_{noise} according to sections 5.1 and 4.2.
 Set $d = d_0$ and $\Gamma_{\text{pr}} = \Gamma_0$.
 Calculate the system matrix R based on the number of angles k and the number of offsets m .
for C **in** L **do**
 Calculate $d = \underset{d \in \mathbb{R}^{km}}{\operatorname{argmin}} \Phi_A(d)$ according to algorithm 1 with the dose limit C .
 If x_{post} is wanted, conduct the measurement according to (6) and measure y_{observed} .
 Calculate the posterior covariance matrix Γ_{post} and the posterior mean x_{post} according to (11) and (12) respectively.
 Update $\Gamma_{\text{pr}} = \Gamma_{\text{post}}$ and $x_{\text{pr}} = x_{\text{post}}$.
 Plot $\operatorname{diag}(\Gamma_{\text{post}})$, x_{post} and other metrics. ▷ Only for monitoring progress
end for

algorithm perform better over time. This can be done for example using equally spaced points between some minimum and maximum radiation levels. In addition, the system matrix R needs to be calculated based on the chosen grid. After the initialisation step, we optimise the design parameter d with the gradient descent described in algorithm 1. Every iteration, our algorithm updates the prior distribution to the previously calculated posterior distribution, which allows the algorithm to use previously gathered knowledge of our target. Our algorithm is presented as algorithm 3.

As described in section 3.3, we capture projection images from many directions simultaneously. We optimise the intensities of the rays concurrently, making the iterative method described in [3] unnecessary. This approach has the advantage that the optimisation process potentially approaches the global optimum with respect to the A-optimality criterion, which is not guaranteed with the sequential approach [3]. However, local minima may pose a problem in this algorithm, which are not as problematic in [3] due to the brute force search. Nevertheless, Burger et al. [3] found the sequential algorithm to perform well in similar tasks to ones presented in section 8. Comprehensive testing of the algorithm will be conducted in section 8.

7.4 Algorithm with a TV prior and lagged diffusivity iteration

In this section, we present an algorithm using a TV prior as the prior distribution instead of the Gaussian one used in section 7.3. This makes it more difficult to handle the posterior distribution as there is no explicit formula for it. We use lagged diffusivity iteration described in section 5.2 to iteratively approximate the posterior mean. This also allows us to calculate an approximate posterior covariance according to (18).

The algorithm starts by initialising the design parameter d , the ROI via A and

Algorithm 4 Online algorithm using the TV prior [6]

Choose initial design parameter d_0 , the ROI, the radiation dose limits L , the smoothing parameter $T > 0$, the strength parameter γ and the stopping criterion τ .

Initialise covariance matrices Γ_0 and Γ_{noise} according to the section 3.3.

Set $d = d_0$, $\Gamma_{\text{pr}} = \Gamma_0$ and $x_{\text{pr}} = \bar{0}$.

Calculate the system matrix R based on the number of angles k and the number of offsets m .

for C **in** L **do**

 Calculate $d = \underset{d \in \mathbb{R}^{km}}{\text{argmin}} \Phi_A(d)$ according to algorithm 1 with the dose limit C .

 Conduct the measurement according to (6) and measure y_{observed} .

 Set $j = 1$, $x_{\text{post}}^{(0)} = x_{\text{pr}}$ and $\Delta\Phi = \tau + 1$.

while $\Delta\Phi > \tau$ **do**

 Calculate $x_{\text{post}}^{(j)}$ according to (17).

 Calculate $\Delta\tau = |\Phi(x_{\text{post}}^{(j)}) - \Phi(x_{\text{post}}^{(j-1)})| / \Phi(x_{\text{post}}^{(j)})$.

 Set $j = j + 1$.

end while

 Update $x_{\text{pr}} = x_{\text{post}}^{(j)}$ and calculate $\Gamma_{\text{pr}} = \Gamma_{\text{post}}^{(j)}$ according to (18).

 Plot $\text{diag}(\Gamma_{\text{pr}})$, x_{post} and other metrics. ▷ Only for monitoring progress

end for

the radiation limits L in exactly the same way as in section 7.3. Additionally, we must initialise the parameters needed for the lagged diffusivity iteration, including the smoothing parameter $T > 0$, the strength parameter of the prior distribution γ and the stopping criterion τ . Before starting the iterative process, the system matrix R is also calculated.

During the iterative process, we calculate the optimal design parameters in the same way as in algorithm 3. However, calculating the posterior distribution differs from algorithm 3. During the lagged diffusivity iteration, we repeatedly approximate the posterior mean using (17). This process depends on the observed measurement, which makes this an "online" algorithm. After hitting the stopping criterion, the posterior covariance matrix is calculated using (18). Lastly, we begin the new iteration with a higher maximum dose of radiation using the calculated posterior as the prior distribution. This algorithm is presented as algorithm 4.

8 Numerical experiments

In this section, we present a few measurement setups, which demonstrate the capabilities of the algorithms described in section 7. Firstly, in section 8.1, we describe how algorithm 3 uses the most of the radiation to scan the predefined ROI. Secondly, in section 8.2, we illustrate, how algorithms 3 and 4 perform in an absorption reconstruction task. This assessment is conducted using a circular target featuring an open ring. Lastly, in section 8.3, we apply the same algorithms to the well-known

Shepp–Logan phantom to qualitatively analyse their performance.

8.1 Circular region of interest

In this section, we show how algorithm 3 directs the maximum amount of radiation towards the predefined ROI. Algorithm 4 is not considered in this section. Therefore, the calculations are performed "offline", as noted in section 5.1. Since the "offline" algorithm is utilised, no measurement data is required, and thus no target image is considered.

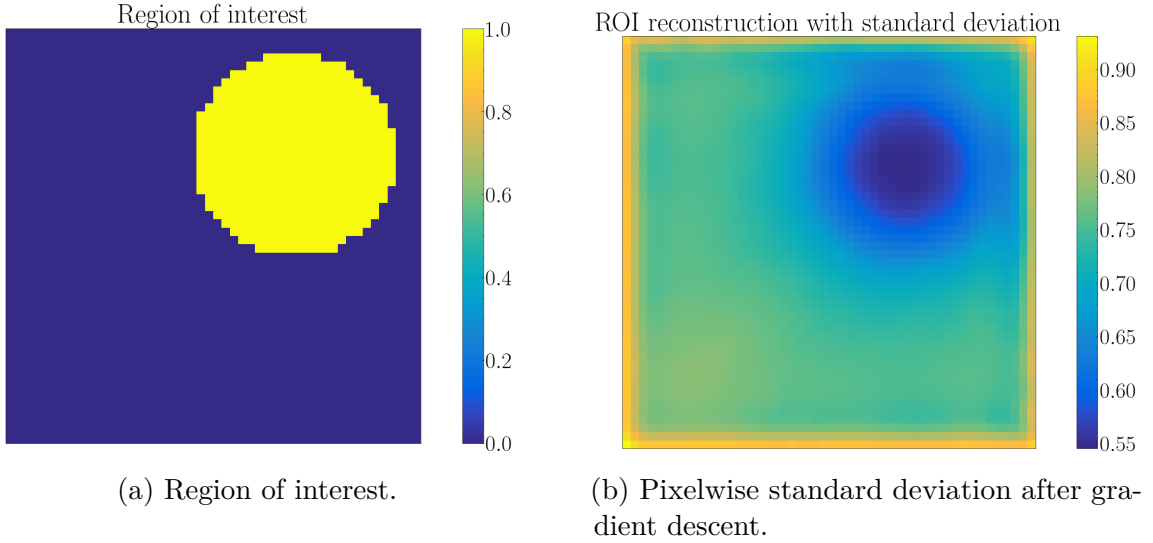


Figure 2: Image of the ROI and the resulting pixelwise variances. We notice that the standard deviation is lower inside the ROI compared to outside the ROI. The edges of the ROI are not prioritised as much as its central parts.

For the experiment, we use a 50×50 grid of pixels. The initial design parameter is initialised as $d_0 = [0.5 \ 0.5 \ \dots \ 0.5]^T$. Additionally, the initial covariance matrix is initialised as in (13), with a small correlation length $l = 0.05$ and the pointwise standard deviation $\sigma = 1$. Finally, the list of radiation dose limits, the number of imaging angles and the number of offsets are initialised as $L = [100000]$, $k = 8$ and $m = 40$ respectively. For the gradient descent, the step size is initialised to $\alpha = 0.1$ and the reduction constant to $\beta = 0.5$. The stopping criteria in gradient descent and golden section search are set to 0.001 and 0.0001 respectively. Finally, the constant θ for the barrier function is set to $\theta = 10^{-5}$. The region of interest is shown in figure 2a.

Figure 2b shows the standard deviation of each pixel after running the gradient descent optimisation. The figure is intuitively acceptable, since the aim of A-optimality is to reduce variance inside the ROI. The decreased standard deviation, and consequently lower variance, enhances our confidence in the contents of the ROI, leading to improved reconstructions of that segment of the absorption image. This is quantitatively demonstrated in sections 8.2 and 8.3. In addition, figure 2b has a small boundary of higher variance. This is due to the non-zero correlation length l . The pixels on the edge have a fewer number of pixels close by, which would correlate

with the given pixel. This results in higher standard deviation for the pixels near the edge.

The posterior distribution in algorithm 4, including the covariance matrix, depends on the measurement. This is not the case with algorithm 3, since the posterior covariance matrix can be calculated using (11). This makes algorithm 4 approximate the target instead of concentrating on the ROI. Therefore the posterior standard deviation using algorithm 4 mimics the underlying target absorption instead of strictly reducing the variance inside the ROI.

8.2 Open ring reconstruction

In this section, we consider both algorithms 3 and 4 in a reconstruction task. As the ROI, we use the whole discretised image and as the target absorption we use a configuration described in figure 3a. In contrast to the previous section, the experiments in this section are conducted "online", which makes it possible to make reconstructions of the target. This means that after each iteration of gradient descent, we take a measurement given by (6), which is used to calculate the posterior mean. This mean is the new reconstruction of the target object.

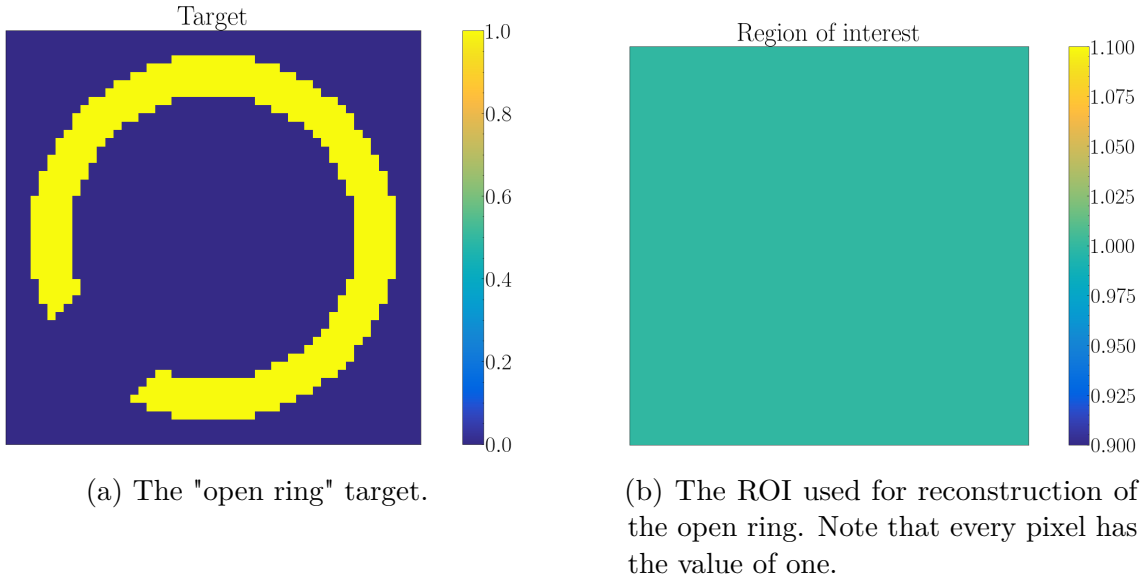


Figure 3: The ROI and the target used in this section.

For both algorithms, we use similar initial parameters. Both images are discretised to consist of 50×50 pixels and both reconstructions are made using $k = 8$ angles and $m = 40$ offsets. The same stopping criteria are used as in section 8.1, and the parameters for the gradient descent and golden section search are also kept the same. However, both algorithms use an initial noise level of $d_0 = [0.1 \ 0.1 \ \dots \ 0.1]^T$, since the noise level of the previous section is too high for convergence and good reconstructions. Additionally, the parameters for the lagged diffusivity iteration are set to $\gamma = 100$ and $T = 10^{-6}$. Finally, the radiation dose limits are chosen to be 10 equally spaced points between 35200 and 100000 for algorithms 3 and 4. This should

allow both algorithms to gradually reduce the noise level in the areas of interest within the reconstruction while keeping the dose of radiation manageable.

Gaussian prior

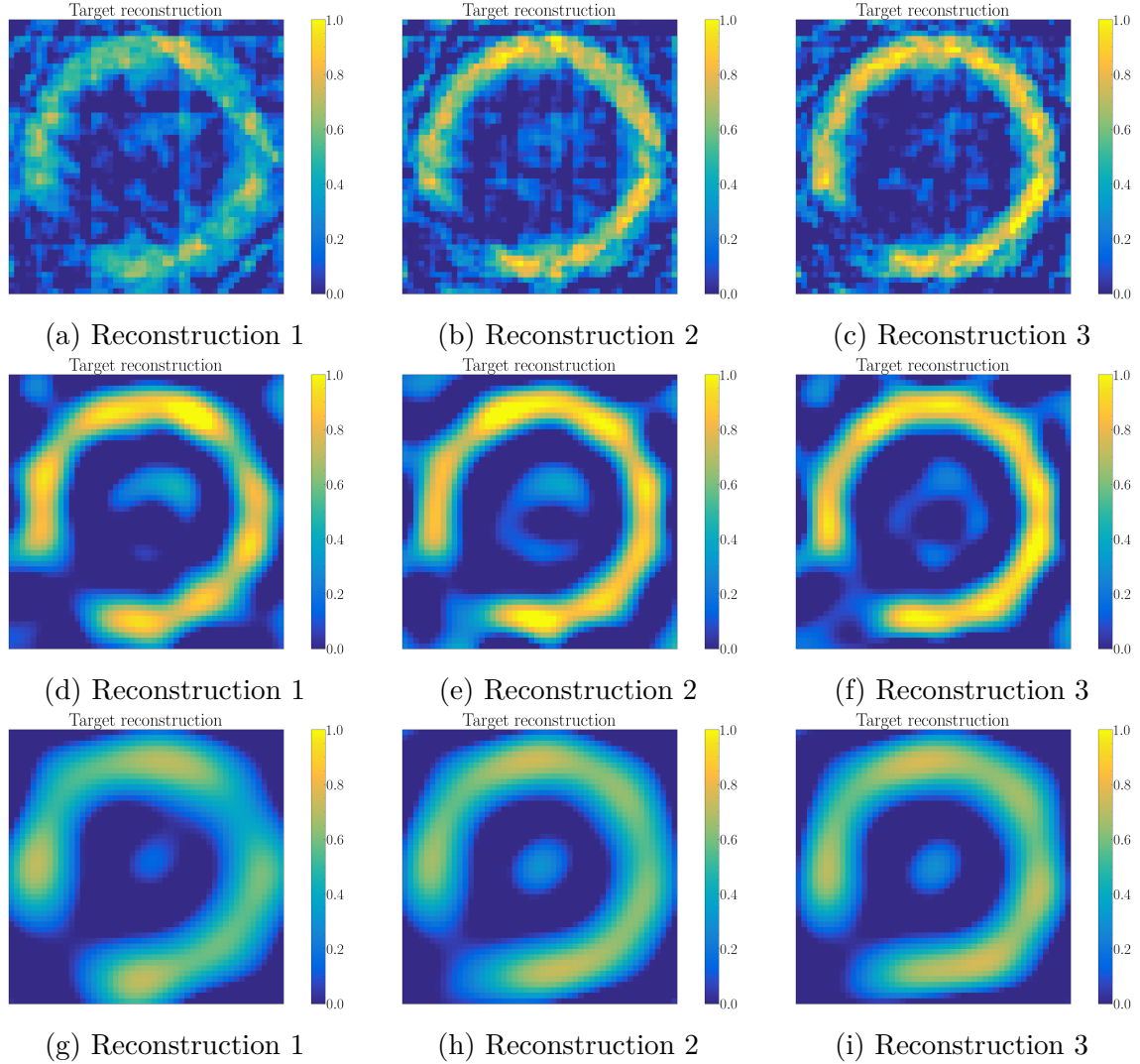
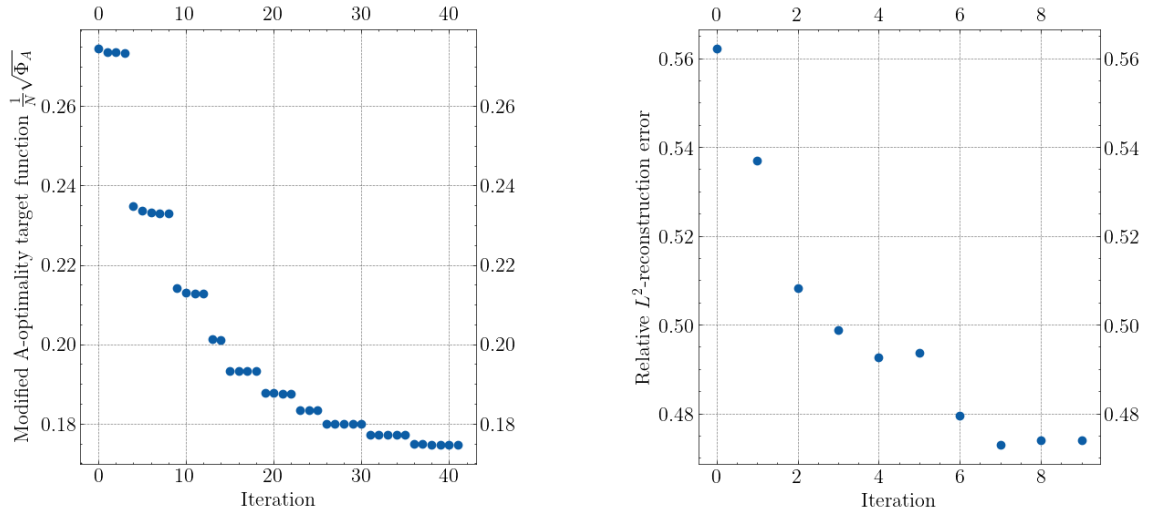


Figure 4: Reconstructions of the target described in figure 3a using algorithm 3 and the Gaussian prior. First three reconstructions of the target are shown using three different correlation lengths for the prior distribution. The rows from top to bottom correspond to correlation lengths of $l = 0.001$, 0.1 and 0.2 , respectively.

The reconstructions with a Gaussian prior of the target 3a can be seen in figure 4. Each row represents one run of the algorithm with a different correlation length for the prior distribution. The rows correspond to correlation lengths of $l = 0.001$, 0.1 and 0.2 from top to bottom. The effect of the correlation length can clearly be seen in the figures. In the top row, the paths of individual X-rays can clearly be seen and the final reconstruction is noisy. The third reconstruction (top right) is, however, an

adequate reconstruction. The reconstructions in the bottom row have the correct shape and the noise is also low, however, the algorithm does not reconstruct the edges separating the target from the background sharply. This results in the ROI appearing blurry, rendering every reconstruction sub-optimal. The middle row performs the best with this target. The width of the target ring is approximately the same as the correlation length l , which makes the prior distribution suitable for this problem. However, there is still a substantial amount of noise in the background.

As a result, the Gaussian prior distribution is unable to completely model the targets examined in this thesis. The employed Gaussian prior promotes smooth edges on the reconstruction, as can be seen in figure 4, which is unrealistic in the parallel beam X-ray imaging context. The edges can be sharpened by lowering the correlation length of the prior distribution. However, good reconstructions with a small correlation length require large amount of X-rays, which often requires more radiation. This is often not feasible in the real world, which motivates the use of the TV prior for the same reconstruction task.



(a) The modified target function. Each group represents a single run of the gradient descent algorithm, with the observed jumps corresponding to updates in the prior distribution.

(b) The relative L^2 -reconstruction error. Each data point represents errors after finding the optimal design parameter with gradient descent.

Figure 5: Two performance metrics as a function of iteration using the Gaussian prior with the correlation length of $l = 0.1$, which visually is the best parameter for the prior distribution.

Figure 5 shows two different performance metrics as a function of iteration corresponding to the middle row of figure 4. Figure 5a shows the value of the modified A-optimality target function over each iteration of the gradient descent. The larger changes correspond to updating the prior distribution. As described in section 6, the modified A-optimality target function represents the expected reconstruction error, which decreases at every iteration. Another performance metric

is shown in figure 5b. It represents the realised relative L^2 -reconstruction error. The relative L^2 -reconstruction error depends on the reconstruction, which in turn depends on the target that has not been drawn from the prior distribution. We clearly see that the shapes of the two plots are anyway rather similar. Overall, the reconstruction gets better as more iterations are included based on both figures.

Total variation prior

Figure 6 shows the reconstruction of the target described in 3a using algorithm 4. In contrast to figure 4, figure 6 does not modify the parameters of the prior distribution. The figure shows six different reconstructions after updating the prior distribution and allowing more radiation. We see that the reconstruction gets better as the radiation limit is increased, which is expected, since the noise for more informative rays is reduced. According to figure 7b, the relative reconstruction error clearly decreases. The best reconstruction according to the Euclidean metric is in figure 6e, which is in line with a visual inspection.

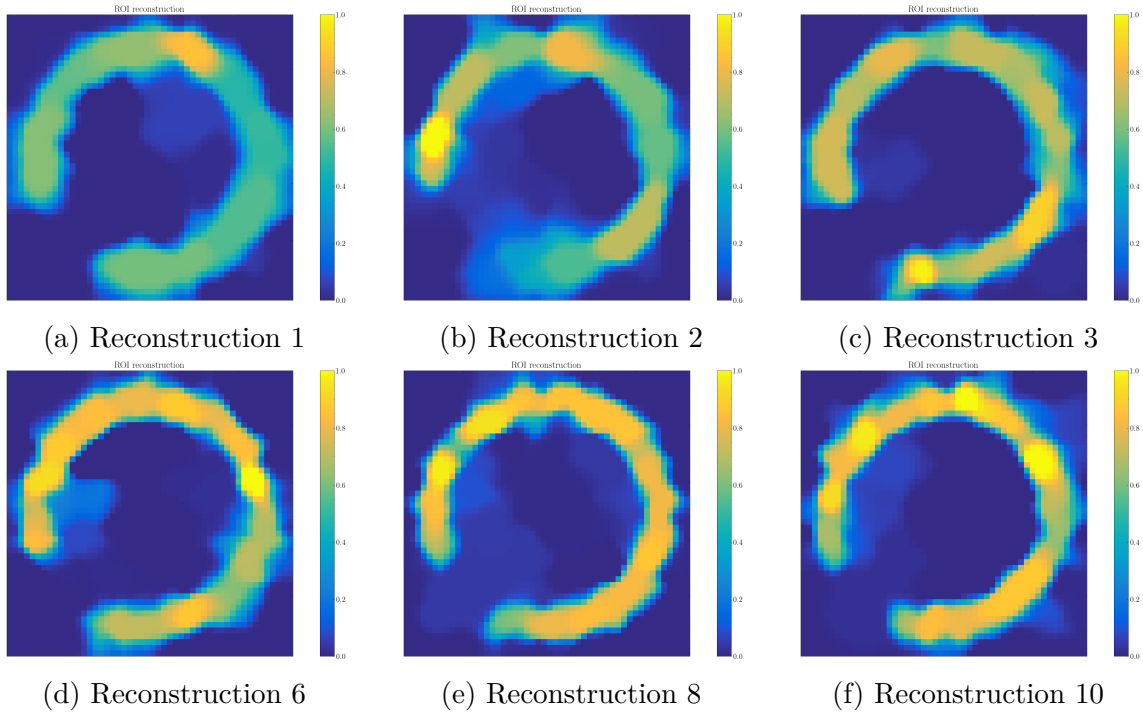
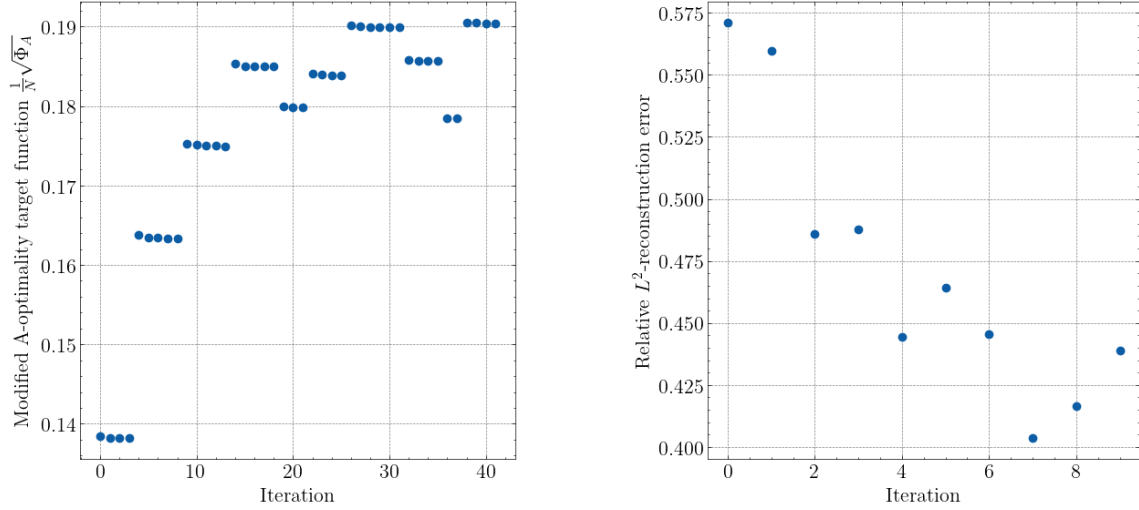


Figure 6: Reconstructions of the "open ring" target using a TV prior. Reconstructions get better as the radiation limit is increased.

In figure 6, it is possible to see the characteristics of the TV prior. The TV prior promotes piecewise constant functions with sharp edges. These patches can easily be recognised for example in figure 6a, where the object consists of a yellow and 2 green patches. In addition, the background also has some patches. The sharp edges can also be seen in the same figure. In the initial reconstructions, these edges exhibit higher variance. However, by the eighth and tenth reconstructions, the algorithm

accurately models the sharp edges. Algorithm 4 is adaptive, which means that as the prior distribution is updated, the algorithm adapts to the contents of the image. Therefore, the new iterations use more radiation to examine the areas with high variance, i.e. the edges.



(a) The modified target function. Each group represents a single run of the gradient descent algorithm, with the observed jumps corresponding to updates in the prior distribution.

(b) The relative L^2 -reconstruction error. Each data point represents errors after finding the optimal design parameter with gradient descent.

Figure 7: Two performance metrics as a function of iteration using the TV prior. The relative L^2 -reconstruction error is lower than using the Gaussian prior.

Based on figure 6, the TV prior performs better compared to the Gaussian prior. Visually the reconstructions are considerably better using the TV prior. This is supported by comparing the relative L^2 -reconstruction errors in figures 5b and 7b for the Gaussian prior and the TV prior respectively. The relative reconstruction error is lower across all iterations using the TV prior compared to the Gaussian prior, which makes the reconstructions better. Similar to the Gaussian prior, the relative reconstruction error using the TV prior decreases across each iteration. However, the relative error has more noise compared to the Gaussian prior. This is likely due to the posterior covariance matrix also depending on the measurement, which is affected by the randomness. On the other hand, the modified A-optimality target function in figure 7a decreases on each run of gradient descent, however, it increases when updating the prior distribution. This happens, since with algorithm 4, the measurement affects the posterior distribution, whereas with algorithm 3, the posterior distribution is independent of the measurement.

8.3 Shepp-Logan phantom reconstruction

The Shepp-Logan phantom is a traditional test image that has been used to test the performance of image reconstruction algorithms. It represents the human head and was devised by Shepp and Logan [14] in 1974. In this thesis, we use the Shepp-Logan phantom to examine the capabilities of algorithms 3 and 4.

As seen in figure 8, the Shepp-Logan phantom consist of 10 ellipses stacked inside one another [14]. The yellow area represents the skull, which is thicker around the forehead area. The five small light blue ellipses represent small tumors and other remaining areas represent other brain organs and tissue. The color values represent the absorption levels of each pixel. The skull is approximately twice as dense as other organs. This model is based on the real anatomy of the brain. The goal of this section is to reconstruct figure 8 as accurately as possible from projection images.

We use a circular ROI around the Shepp-Logan phantom, which should marginally improve the quality of the reconstruction compared to the ROI used in the previous section. Once again, we use a grid of 50×50 pixels to calculate the reconstruction. We use a correlation length of $l = 0.05$ and as the initial design parameter, we use $d_0 = [0.5 \ 0.5 \ \dots \ 0.5]^T$. In addition, for the Gaussian prior, we use six equally spaced dose limits between 1408 and 100000, which allows the algorithm to gradually make the reconstruction better. On the other hand, for the TV prior, we use six dose limits between 51200 and 100000, since with lower initial radiation, the reconstructions of the target remain unrecognisable throughout the experiment. Additionally, for the strength parameter of the TV prior, we use $\gamma = 100^2$ for the optimisation and $\gamma = 10^2$ for the lagged diffusivity iteration. Otherwise, we use similar parameters as in sections 8.1 and 8.2.

The Shepp-Logan phantom is more complex than the target in figure 3a. In addition, for the Gaussian prior, the starting value of noise is higher compared to section 8.2, which makes the reconstruction task considerably harder. This section outlines the process of subsequent reconstructions improving over time. Using both algorithms, the initial reconstructions are expected to be lacking, while the final reconstruction should be significantly more accurate. We do not consider the differences between the Gaussian prior and the TV prior, due to the difference in the initial noise values and radiation dose limits.

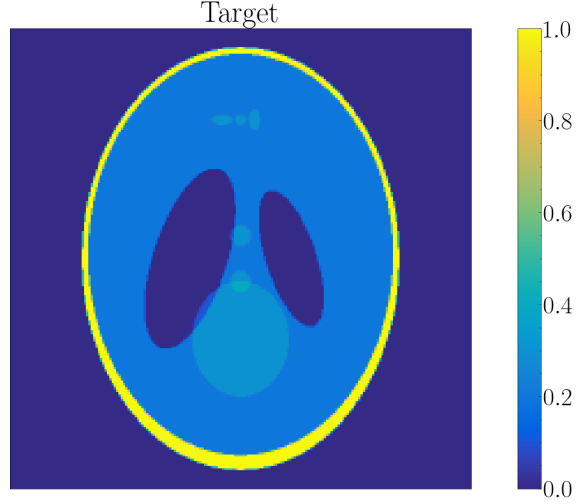


Figure 8: The Shepp-Logan phantom.

Gaussian prior

The reconstructions of the Shepp-Logan phantom in figure 8 using algorithm 3 can be seen in figure 9. The reconstructions using the Gaussian prior start out by being unrecognisable. Gradually, the reconstruction is enhanced with the largest performance improvement being between reconstructions 1 and 2. The reconstruction is adequate, however, the noise in the background is still present. The tumors and the skull are also lost, which decreases the accuracy. The edges of the target are quite blurred, which is expected due to the properties of the Gaussian prior distribution. Overall, the reconstruction has the right shape, but with missing features, noisy background and blurred edges.

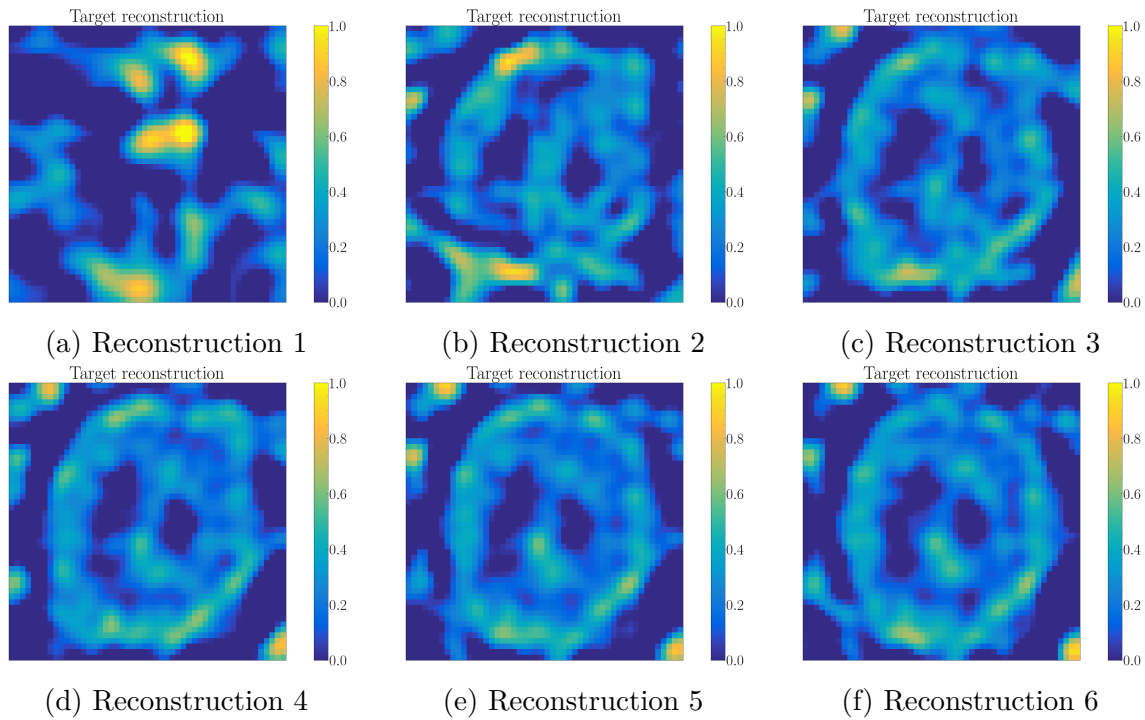


Figure 9: The first six reconstructions of the Shepp-Logan phantom using a Gaussian prior distribution. The accuracy of the reconstructions gets better as the radiation limit is increased. However, many parts of the original target are lost.

Total variation prior

Figure 10 shows the reconstructions of the target shown in figure 8 using algorithm 4. There are six different reconstructions of the target, which visually get marginally better as the algorithm proceeds. Similar to the Gaussian prior, the TV prior does not clearly find the tumors inside the target, and the skull is also lost. The improvement in reconstruction accuracy is primarily attributed to the algorithm focusing on the inner features of the target. The outside edges of the target do not improve as much or even partially get worse on each reconstruction.

Similar to figure 6, the main characteristics of the TV prior can also be seen in figure 10. The reconstructions consist of multiple areas with constant absorption.

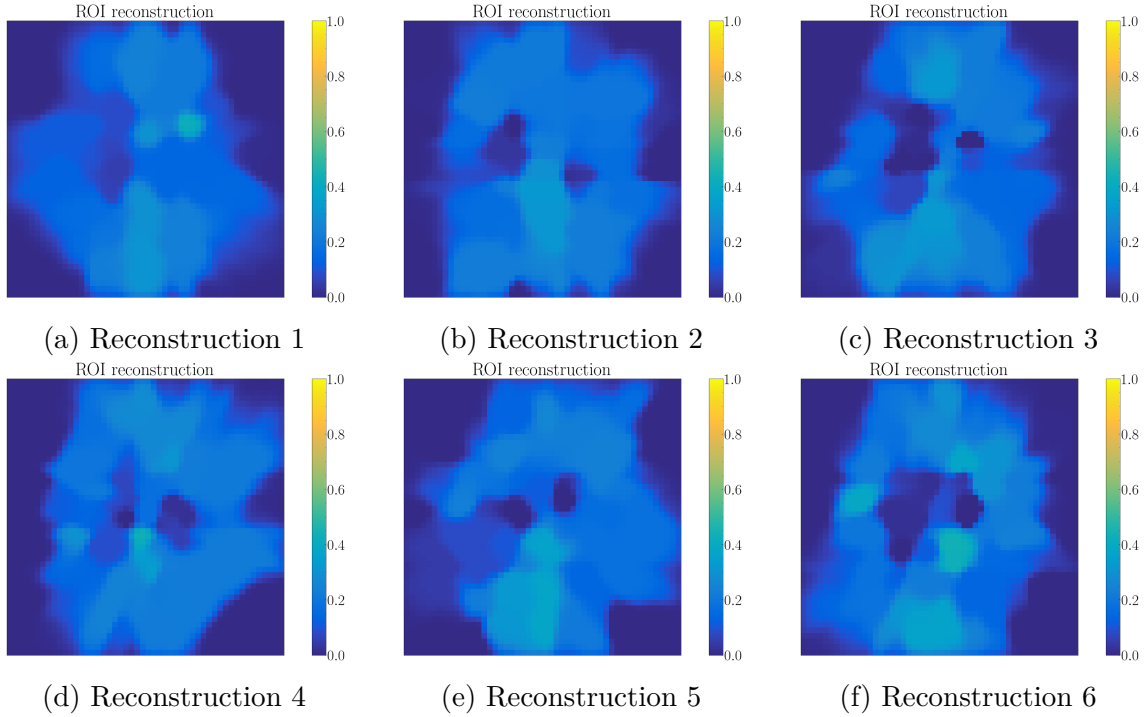


Figure 10: Reconstructions of the Shepp-Logan phantom using the TV prior. Reconstructions marginally improve as the radiation limit is increased.

The edges of each area are sharp, which is also expected. However, due to the more difficult projection geometry, the background is a little more noisy. This makes the reconstruction considerably worse than the reconstruction of the open ring in section 8.2, even though the initial noise level is lower than in section 8.2.

9 Conclusions

The goal of this thesis was to create an algorithm to optimise the measurement setting in parallel beam X-ray tomography. In conclusion, we considered two algorithms for this task. One of the algorithms, namely algorithm 3, utilised a Gaussian prior distribution. This algorithm enables efficient computation of the optimal intensities "offline", facilitating a more streamlined screening process in practise. The second algorithm, namely algorithm 4, utilised a TV prior distribution. The TV prior has more accurate assumptions considering realistic imaging targets. This makes it preferable even though the computations need to be conducted "online". The choice of an appropriate prior distribution is essential, since the final reconstruction exhibits the properties defined by the chosen prior. Therefore multiple prior distributions were evaluated.

In this thesis, a probabilistic approach was chosen. The Bayesian approach is often more stable than traditional regularisation methods under imperfect information, which is common for many screening applications. In addition, the Bayesian approach allows us to easily take prior knowledge into account, which is implicitly incorporated

in traditional methods.

To optimise the measurement setting, the A-optimality criterion was employed. The criterion enables the selection of the optimal measurement setting, so that the information on the unknown absorption values is maximised. Using A-optimality, we aim to minimise the posterior variance of the pixel wise absorption values inside the ROI, which is a natural choice for the information measure. For the optimisation process, gradient descent was utilised. Other methods, such as Newton’s method could also be used, however, we found gradient descent to perform sufficiently well to obtain reasonable results.

Both algorithms performed well in reconstruction tasks. However, algorithm 4 seemed to perform better in many applications due to the more realistic assumptions on the prior distribution. We found algorithm 3 to smooth out the sharp edges in the reconstruction process. However, the reconstructions by algorithm 4 were considerably more precise. The results in section 8.2 demonstrated that both algorithms can easily reconstruct simple targets. In contrast, the algorithms faced greater difficulty reconstructing the more complex target in section 8.3, however, satisfactory reconstructions were still achieved despite high level of initial noise.

Several aspects were omitted to maintain the scope of this thesis. Firstly, besides A-optimality, other types of Bayesian OED were not considered. For example, D-optimality would be a natural extension for further research. Additionally, exploring other types of prior distributions, such as the Perona-Malik prior, or different reconstruction geometries, could help to identify the optimal prior for a given screening setup. Lastly, the same algorithms could be applied to three-dimensional tasks. Helin et al. have conducted experiments using cone-beam tomography [5], however, Bayesian OED in three dimensional parallel-beam X-ray tomography remains unexplored. This approach is more realistic and applicable in real-world scenarios, making it an intriguing subject for future research.

References

- [1] G. S. Alberti, E. De Vito, M. Lassas, L. Ratti, and M. Santacesaria. Learning the optimal tikhonov regularizer for inverse problems. *Advances in Neural Information Processing Systems*, 34:25205–25216, 2021.
- [2] S. Ameli and S. C. Shadden. A singular woodbury and pseudo-determinant matrix identities and application to gaussian process regression. *Applied Mathematics and Computation*, 452:128032, September 2023. ISSN 0096-3003. doi: 10.1016/j.amc.2023.128032. URL <http://dx.doi.org/10.1016/j.amc.2023.128032>.
- [3] M. Burger, A. Hauptmann, T. Helin, N. Hyvönen, and J.-P. Puska. Sequentially optimized projections in x-ray imaging. *Inverse Problems*, 37(7):075006, jun 2021. doi: 10.1088/1361-6420/ac01a4. URL <https://dx.doi.org/10.1088/1361-6420/ac01a4>.

- [4] D. K. Duvenaud. *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge, 2014.
- [5] 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, 2022.
- [6] T. Helin, N. Hyvönen, J. Maaninen, and J.-P. Puska. Bayesian design of measurements for magnetorelaxometry imaging. *Inverse Problems*, 39(12):125020, 2023. URL <https://arxiv.org/abs/2305.19940>.
- [7] E. Somersalo J. Kaipio. *Statistical and Computational Inverse Problems*. New York : Springer, 2004. ISBN 0-387-22073-9.
- [8] B. Jones, K. Allen-Moyer, and P. Goos. A-optimal versus D-optimal design of screening experiments. *Journal of Quality Technology*, 53(4):369–382, 2021. doi: 10.1080/00224065.2020.1757391. URL <https://doi.org/10.1080/00224065.2020.1757391>.
- [9] W. A. Kalender. X-ray computed tomography. *Physics in Medicine Biology*, 51(13):R29–R43, jun 2006. doi: 10.1088/0031-9155/51/13/R03. URL <https://dx.doi.org/10.1088/0031-9155/51/13/R03>.
- [10] J. Maaninen. Bayesian experimental design for magnetorelaxometry imaging. Master’s thesis, Aalto university - School of Science, Espoo, 2023-03-21.
- [11] F. Oliveira. *Nonlinear optimisation*. Aalto university - School of Science, 2022. Modified by N. Hyvönen.
- [12] A. Ribes and F. Schmitt. Linear inverse problems in imaging. *IEEE Signal Processing Magazine*, 25(4):84–99, 2008. doi: 10.1109/MSP.2008.923099.
- [13] 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. doi: 10.1137/17M1143733. URL <https://doi.org/10.1137/17M1143733>.
- [14] L. A. Shepp and B. F. Logan. The fourier reconstruction of a head section. *IEEE Transactions on Nuclear Science*, 21(3):21–43, 1974. doi: 10.1109/TNS.1974.6499235.
- [15] R. Snieder and J. Trampert. *Linear and Nonlinear Inverse Problems*, pages 93–164. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000. ISBN 978-3-540-45597-4. doi: 10.1007/3-540-45597-3_3. URL https://doi.org/10.1007/3-540-45597-3_3.
- [16] S. Tassiopoulou, G. Koukiou, and V. Anastassopoulos. Algorithms in tomography and related inverse problems—a review. *Algorithms*, 17(2):71, 2024. doi: <https://doi.org/10.3390/a17020071>.

- [17] S. Wanyonyi, A. Okango, and J. Koech. Exploration of D-, A-, I- and G-optimality criteria in mixture modeling. *Asian Journal of Probability and Statistics*, 12(4):15—28, May 2021. doi: 10.9734/ajpas/2021/v12i430292. URL <https://journalajpas.com/index.php/AJPAS/article/view/225>.
- [18] L. Zhang and H. Schaeffer. Stability and error estimates of bv solutions to the abel inverse problem. *Inverse Problems*, 34(10):105003, aug 2018. doi: 10.1088/1361-6420/aad1c7. URL <https://dx.doi.org/10.1088/1361-6420/aad1c7>.