of the measurement setting

Gradient based optimisation in parallel beam x-ray imaging

Aatu Selkee

School of Science

Bachelor's thesis
Espoo 31.8.2018 KORJAA OIKEAKSI

Supervisor

Prof. Nuutti Hyvönen

Advisor

Prof. Nuutti Hyvönen

Aalto University
School of Science

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Aatu Selkee

Title Gradient based optimisation in parallel beam x-ray imaging

Degree programme Teknistieteellinen kandidaattiohjelma

Major Matematiikka ja systeemitieteet **Code of major** SCI3029

Teacher in charge Prof. Nuutti Hyvönen

Advisor Prof. Nuutti Hyvönen

Date 31.8.2018 KORJAA OIKEAKSI **Number of pages** 21 **Language** English

Abstract

Tiivistelmässä on lyhyt selvitys kirjoituksen tärkeimmästä sisällöstä: mitä ja miten on tutkittu, sekä mitä tuloksia on saatu.

Tämän opinnäytteen tiivistelmäteksti kirjoitetaan opinnäytteen luettavan osan lomakkeen lisäksi myös pdf-tiedoston metadataan \thesisabstract-makron avulla (katso yllä). Kirjoita tähän luettavaan tiivistelmälomakkeeseen menevä teksti. Tässä saa olla erikoismerkkejä kuten kreikkalaiset kirjaimet ja rivinvaihto- ja kappaleenjakomerkit. Tämän tekstin on muuten oltava sama kuin metedatatiivistelmän teksti.

Jos tiivistelmäsi ei sisällä erikoismerkkejä eikä kaipaa kappaleenjakoa, voit hyödyntää makroa \abstracttext joka käyttää ylläantamaasi metedata-kelpoista tiivistelmää (katso kommentti alla).

Keywords Avainsanoiksi valitaan kirjoituksen, sisältöä keskeisesti, kuvaavia, käsitteitä

Aalto University
School of Science

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of the bachelor's thesis

Author Aatu Selkee

Title Thesis template

Degree programme Bachelor's Programme in Science and Technology

Major Mathematics and Systems Sciences **Code of major** SCI3029

Teacher in charge Prof. VALVOVAN PROFESSORIN NIMI

Advisor DSc (Tech.) OHJAAJAN NIMI

Date 31.8.2018 KORJAA OIKEAKSI **Number of pages** 21 **Language** English

Abstract

Your abstract in English. Keep the abstract short. The abstract explains your research topic, the methods you have used, and the results you obtained.

Keywords Resistor, resistance, temperature

Contents

1 Introduction

Nowadays, inverse problems are a widely researched topics in modern mathematics [1, 2]. Inverse problems offer a wide range of applications, such as determining an initial heat distribution ~~in a rod based~~ inside an object on the current heat distribution [1] as well as restoring the original image from a blurred version [1, 2]. All of these problems have one thing in common: they are an inverse of some forward problem. A forward problem tries to determine the observations based on some initial parameters. On the other hand, inverse problems try to find the initial parameters using the observations of an experiment.

In particular, a subset of inverse problems called linear inverse problems involves determining initial parameters, where 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 compared to non-linear inverse problems due to modern computers being well optimised for large matrix operations. They are often also easier to interpret. Trivially nonlinear inverse problems can model more complex behaviourLisää lähde tai muokkaa. In this thesis we strictly consider linear inverse problems, particularly in the field of tomography. Problems in tomography are often modelled ~~as~~ finite-dimensional linear inverse problems, since the area we are interested in can be discretised, which results into the following linear measurement model [3];

$$Y = RX + N, \tag{1}$$

where X is a random variable following ~~the~~ a prior distribution and N is a random variable independent ~~of the random variable~~ X representing noise in the experiment. Furthermore Y is the measurement of the conducted experiment and R is a projection matrix that takes a vectorised image and transforms it to some ~~amount~~ number of data-points. ~~This setup can be used~~ One can for example try to sequentially optimise, which projection angle yields the most amount of information or decreases the variance of the initial parameters the most [3]. Tämä vielä voisi olla vasta myöhemmin. → kts. * seuraavalla sivulla

Usually inverse problems are not considered without taking noise into account. This makes determining the initial parameters harder due to the information being "polluted" ~~by the noise~~ [4]. However, taking noise into account in our models often yields better outcomes due to real life applications and data always having some noise present [3]Lisää toinen lähde tai muokkaa. In this thesis we assume the noise to be distributed according to a multivariate normal distribution, which results in some nice properties for the posterior distribution assuming that the prior distribution is also Gaussian [5].

There are two main solution methods for solving linear inverse problems [1, 5]. First of the two methods is a classical approach using regularisation methods and least squares techniques. This method yields a point estimate for the optimal parameters, which is not ideal for performing inference. This gives ~~rise to~~ motivation for another method, a probabilistic approach using Bayes' rule and Bayesian statistics. Using Bayes' rule, we get a probability distribution for the optimal parameters, which allows us to do better inference compared to the classical approach [1, 5]. Hence we will be using the Bayesian approach in this thesis.

for the considered measurement setup.

Bayesian optimal experimental design (OED) is a way of ~~defining~~ finding the optimal parameters [6]. In Bayesian OED we aim to maximize the "utility" received from a specific experiment [5]. There are many ways to define "utility" [6], however, in this thesis we consider the parameters to be optimal if they minimise the so-called A-optimality target function $\Phi_A(x)$. A-optimality tries to minimise the variance of the parameters [6], which works well in our use case.

The aim of this thesis is to implement an algorithm in the field of X-ray tomography, similar to the approach described by Burger et al. [3]. The main differences are that we aim to implement a gradient based optimisation algorithm instead of a brute force search. We make this possible by first fixing the projection matrix R in ~~equation~~ (1). Intuitively this means that our algorithm takes an X-ray image from predetermined directions around the target area. Then we consider a vector d of values related to the intensities of each ray in R. It is possible to compute the derivatives $\frac{\partial \Phi_A}{\partial d_i}$, $\frac{\partial^2 \Phi_A}{\partial d_i^2}$ and $\frac{\partial^2 \Phi_A}{\partial d_i d_j}$ of Φ_A with respect to some parameters d_i and d_j analytically. This makes ~~it possible for us~~ allows to use gradient descent or Newton's method for optimising the intensities more efficiently compared to the brute force search ~~by~~ such as in Burger et al. [3]. This approach may, however, be more susceptible to local minima compared to the brute force search. Overall this approach is similar to the algorithm used by Maaninen [5] to magnetorelaxometry imaging. We apply this algorithm to X-ray tomography.

\eqref{...}

for choosing subsequent projection angles.

This thesis aims to answer whether or not a gradient based optimisation methods can be used in X-ray imaging to ~~get similar results compared to brute force searches.~~ to optimize the radiation dose of different X-rays. The implemented algorithm should produce ~~figures~~ reconstructions that clearly differentiate between the background of the image and the region of interest (ROI). That is, the algorithm should produce pictures, where the variance of the pixels on the ROI is lower compared to the variance of the other pixels. We should also be able to reconstruct the ROI strictly based on the results of the experiment.

The numerical experiments in this thesis will be conducted mainly using a Gaussian prior. We will also ~~try to reconstruct the ROI using~~ employ a total variation (TV) prior, which ~~should make the edges of the ROI sharper~~ promotes edges in the reconstruction [7, 5]. TV prior promotes piece wise constant functions, which should describe ~~the ROI~~ target better compared to the Gaussian priors, which smooth out the edges ~~of the ROI~~.

To maintain the scope of this thesis smaller, many things that would be reasonable to try will be left out. Firstly, other types of OED metrics apart from A-optimality will not be experimented with. According to Wanyonyi et al. [6], A-optimality performs better than D-optimality in screening tasks. D-optimality is another commonly used optimality metric, which tries to maximize the information gain from the posterior compared to the prior [3]. Burger et al. [3] also show that using A-optimality criterion yields results closer to the global optimum compared to using D-optimal designs. Therefore, in this thesis we will strictly consider A-optimal designs. Secondly, we will not experiment with any other priors than the Gaussian prior and the TV prior in this thesis. For example, Maaninen [5] experiments with the Perona-Malik prior, which will not be included in this thesis as Maaninen determines it to perform slightly worse than the TV prior.

The following outlines the structure of this thesis. In section 2 we will take a

[Joko x-ray tai X-ray.]

short look at a few important sources of 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 a solution to an inverse problem to be optimal~~ 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 ~~give an conclude~~ present conclusions in section 9.

2 Literature review

X-ray tomography has ~~been investigated by many people prior to this thesis~~ attracted enormous amounts of research prior to this thesis [3, 7, 8]. However, there seem to be only a few works tackling Bayesian OED for X-ray tomography. Burger et al. [3] and Helin et al. [7] use a sequential algorithm instead of a gradient based algorithm, which they argue produces more optimal results, since the objective function they use is not convex. On the other hand Ruthotto et al. [8] use a gradient based optimisation algorithm to find a minimum to a similar x-ray tomography problem considered in [3]. It is worth to mention that in this thesis we consider two parameters, the angle and the offset of the x-ray source, which is similar to the problem in [3] and [7]. However, Ruthotto et al. [8] only consider the angle of the source. Considering also the offset allows the algorithm to choose the importance of each ray more freely, however, it increases the computational complexity of the problem.

~~Many other tomography problems have also been investigated [9].~~ Bayesian OED has been previously applied to many imaging techniques, with the considerations of Maaninen in [5] arguably the closest to the setting of this thesis. Maaninen [5] demonstrates 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 similar to this thesis. With edge promoting priors, such as a TV prior, Maaninen uses a sequential algorithm similar to the ones used in [3] and [7]. In this thesis we will also use a sequential algorithm when experimenting with the TV prior.

3 Parallel beam X-ray imaging

In this section the basics of X-ray imaging will be presented. In section 3.1, we will present the basic setup for the imaging process. In section 3.2 we will present the forward model that will be used to define the corresponding inverse problem. In section 3.3 the main problem of this thesis will be defined. The optimisation problem related to the OED will be discussed.

3.1 Imaging setup or something

Lisää Enemmän selitystä itse menetelmästä. Esim joku kuva ja vähän tekstiä miten toimii tms. Mainitse miksi parallel beam

X-ray tomography is ~~a~~ an imaging technique usually used to obtain a reconstruction of ~~a~~ the absorption inside a target [1]. Since the work of Radon in 1917, it has been known that the reconstruction of ~~a body~~ the human body can be constructed based on 2D slices of the human body. In x-ray tomography, these slices are obtained by shooting x-rays of a certain intensity

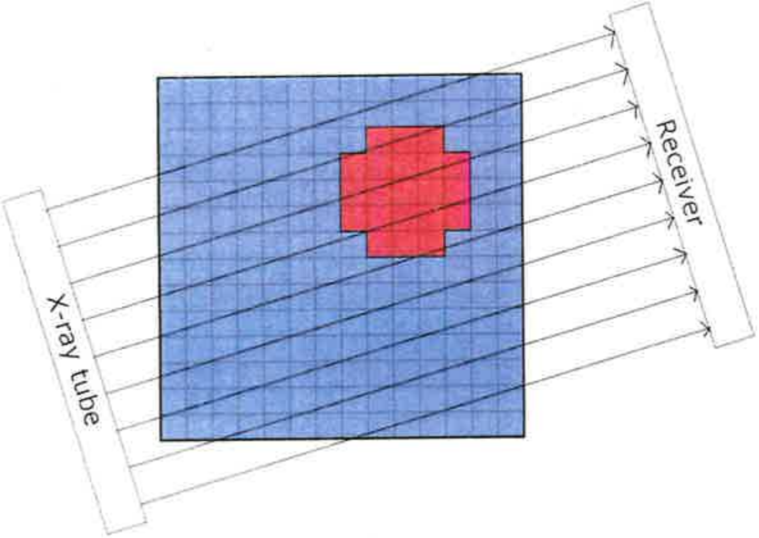

Figure 1: Picture of the basic imaging setup. The red area represents the ROI and the arrows represent the individual x-rays.

through the body from many angles and offsets and detecting the absorption data after passing through the body [1].

The imaging setup in parallel beam x-ray tomography consist of 3 main parts, the x-ray tube, the detector and the gantry [10]. The gantry is a circular object, which lays around the target object, usually a human. Inside the gantry are rotating parts, which hold the x-ray tube and the detector. Tähän vähän lisää ei tarvitse paljaa.

When taking a scan, the x-ray tube sends x-rays towards the detector from many different angles. These

3.2 The forward model

Using the setup provided in section 3.1, the attenuation of intensity I on a single ray L of length dl is given by the equation [1] MAYBE TOO SIMILAR TO THE BOOK ihan OK
an infinitesimal

$$dI = -Iu(x)dl(x), \tag{2}$$

where $u : D \to \mathbb{R}_+$ is a function that maps a point from the target object D and takes it to the corresponding absorption value [3]. Equation (2) yields the following
\eqref

equation for the intensity I at the receiver [1].

$$\log(I) - \log(I_0) = \log\left(\frac{I}{I_0}\right) = \int_{I_0}^{I} \frac{dI}{I} = -\int_L u(x)dl(x) . \tag{3}$$

Equation (3) gives us an easy way to define the forward model by discretising the target domain D into $n = N^2$ pixels and 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 shoot, we organise the lengths of each ray into a matrix $R_k \in \mathbb{R}^{m \times n}$, where m is the number of offsets of the ray. These matrices R_k can again be stacked on top of one another for every k angles of the images resulting in the following matrix.

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

\vdots

Tässä vaiheessa voisi olla vielä pienet (so. 'deterministiset') kirjaimet

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

$$Y = RX + N, \tag{4}$$

where $X \in \mathbb{R}^n$ is the vector of pixel wise absorption values and $N \in \mathbb{R}^{km}$ is a vector containing additive noise values for each ray. Then $Y \in \mathbb{R}^{km}$ is the measurement of the experiment containing ~~absorptions of each ray~~. information on the absorption along each ray.

Tell what equation 4 represents and reference it.

3.3 Problem formulation

The main problem in this thesis is ~~to solve the inverse problem corresponding to the forward equation 4~~ to optimize the measurement setup related to (4) from the standpoint of radiation dose. We will aim to implement an algorithm that optimises design parameters based on A-optimality with the goal of concentrating on the ~~ROI inside the target area~~ allowed radiation dose inside the target area. We will also aim to reconstruct the image inside the ROI.

Delete ~~When~~ For efficiently solving the inverse problem corresponding to the forward model discussed in section 3.2, we aim to ~~determine the parameters that produced the current output~~ optimize certain design parameters of the measurement setup. These design parameters d relate to the intensities of the X-rays going through the target. In this thesis, we ~~define the radiation dose from a single image by the formula~~ adopt a model where the dose of a single X-ray is inversely proportional to the variance of the corresponding component of the noise vector,

$$\text{dose}_i = \frac{c}{d_i^2}, \tag{5}$$

where c is a positive constant. Since radiation is harmful to humans, we want to restrict the total ~~dose of the~~ radiation dose. This results into the following constrained optimisation problem

$$\begin{cases} d^* = \underset{d \in \mathbb{R}^{km}}{\operatorname{argmin}} \Phi_A(d), \\ \sum_{i=1}^{km} \frac{1}{d_i^2} \leq D \end{cases} \tag{6}$$

{\displaystyle }

where D 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 $\sum_{i=1}^{km} \frac{1}{d_i^2} > D$ complicates the optimisation problem significantly, since the feasible set is ~~no longer~~ 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 ~~hard~~ nontrivial to use, since it is possible to hit the unfeasible set. These optimisation methods and the challenges implementing them will be discussed in more detail in sections 7.1 and 7.2.

KIRJOITA MIKÄ ON PÄÄ KYSYMYS MIHIN VASTATAAN?

4 Bayesian inverse problems

Traditionally, inverse problems have been solved using regularisation methods, such as Tikhonov regularisation [1]. Regularisation, and other point estimate methods generally perform well in many problems. Since many inverse problems ~~have solutions that blow up as the solution gets better~~ exhibit irregular behaviour when the data fit gets better, it is useful to solve ~~the~~ an inverse problem $Ax = y$ by solving the following minimisation problem:

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

where $J : X \to [0, \infty)$ is an increasing regularisation function [11]. In ~~the~~ Tikhonov regularisation, the regularisation function is chosen to be $J(x) = \delta\|x\|^2$, where $\delta > 0$ is a parameter that guarantees the existence of the minimum [1, 11]. However, the problem with this approach is that we ~~get strictly~~ only get a point estimate of the correct solution.* ~~Often~~ For this reason, it is useful to get a probability distribution as a solution to an inverse problem. ~~since it is easier to use for inference compared to only a point estimate.~~ It also easier to take prior knowledge ~~of the distributions into account~~ into account in the form of a prior distribution, since with traditional methods, the prior is ~~incorporated~~ implicitly into the regularisation function J [11]. Therefore in this thesis we use Bayesian inversion to solve the problem described ~~in~~ section 3.3.MUISTA LISÄTÄ MISSÄ KAPPALEESSA ONGLEMA KUVAILLAAN

* without information about its reliability.

or some other determinist regularization technique.

4.1 Bayesian inversion or Bayes' rule

With Bayesian approach to inverse problems, we assume that every variable is sampled from a certain distribution, including the parameters we want to solve for. The solution to the inverse problem is then achieved by calculating the posterior distributions of the parameters [1]. ~~Now~~ Consider the linear forward model described in chapter 3.2OR IN EQUATION 4, i.e. the model

$$Y = RX + N. \quad (+)$$

that enters (+) via $z = R(d)z$ and

Also let $d \in \mathbb{R}^{km}$ be the vector of design parameters, $y_{\text{observed}} \in \mathbb{R}^{km}$ the observed projection image ~~and $x \in \mathbb{R}^n$ the real data sampled from the prior distribution $\pi_{pr} : \mathbb{R}^n \to \mathbb{R}_+$~~. In addition, assume that the random variables X and N are independent of one another. This model is called an additive noise model, since first

we apply a transformation to X and then add the independent random variables together [5]. ~~Then~~ Using the Bayes formula we get the following equation for the conditional probability x given some measurement y_{observed}

RX and N

density of

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

where $\pi : \mathbb{R}^{km} \to \mathbb{R}_+$ is the likelihood function [3]. It is common to assume that the random variables X and N follow multivariate Gaussian distributions

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

for Γ_{noise} see below

where $x_{pr} \in \mathbb{R}^n$ is the prior mean vector for the image data, $\Gamma_{pr} \in \mathbb{R}^{n\times n}$ the corresponding positive definite prior covariance matrix and $\Gamma_{\text{noise}} \in \mathbb{R}^{km\times km}$ 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

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

where $\epsilon > 0$ is a small constant to guarantee that $\Gamma_{\text{noise}}(d)$ is positive definite. Defining the noise covariance this way allows us to interpret the dose of radiation in ~~equation~~ (5) as ~~proportional to the variance of the noise~~ inversely. This may not ~~physically~~ be entirely accurate, however ~~the~~ such model has suitable behaviour for the problem. That is, more radiation induces lower noise in the measurement.

physically; is intuitively acceptable and

Based on these assumptions we ~~get~~ deduce that the posterior distribution in ~~equation~~ (7) is also a multivariate Gaussian distribution with the covariance matrix [1, 3]

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

and the mean EHKÄ TURHA VIITTAUS KOSKA SAMA KUIN YLLÄ [1, 3] [then ok]

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

However, these can be very time consuming to calculate, since in ~~equation~~ (9) we require matrix inversions of matrices of shapes $n \times n$, $n \times n$ and $km \times km$ and in equation (10) inversions of shapes $n \times n$ and $km \times km$, each of which takes $O(n^3)$ or $O((km)^3)$ time complexity. Therefore it is beneficial to apply the Woodbury matrix identity to ~~equations~~ (9) and (10) to write them as

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

and

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

respectively [3, 5]. Both of these calculations only require the inversion of the matrix $R\Gamma_{pr}R^T + \Gamma_{\text{noise}}(d) \in \mathbb{R}^{km\times km}$, which is much better computationally, since in both regret

cases we require an inversion of a $km \times km$ matrix, but in ~~equations~~ (11) and (12) we do not have to compute the additional inversions of the $n \times n$ matrices. Note also that often $km << n = N^2$.

It should be noted that ~~equations~~ (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 [5], and, e.g., the posterior mean and covariance do not allow explicit representations. When the prior distribution is not Gaussian some other methods, such as lagged diffusivity iteration in section 5.2 must be used to ~~obtain the posterior distributions~~ approximate 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 given a target ROI depend greatly on the chosen prior covariance function. In this section we will look into two common priors that each have ~~differing~~ different advantages.

5.1 Gaussian prior

In X-ray imaging, photons are used to capture the ~~absorption data~~ data on the absorption inside the imaged body [1]. Classically, these photons arriving at the detector are assumed to be Poisson distributed, with exponential inter arrival times [1]. This Poisson distribution can often be approximated as a Gaussian distribution assuming that the number of photons, i.e. the intensity of the x-ray is high enough. Therefore it is reasonable to assume that the ~~prior distribution is Gaussian with good accuracy.~~ The noise distribution in ~~equation (4) can also be assumed to follow a Gaussian distribution [4]~~ (4) is Gaussian. Often it is also reasonable to assume the prior is Gaussian as well.

A Gaussian prior is ~~a covariance function that determines the correlation between any two data points~~ with such a property that determines the correlation between any two components of the unknown. Intuitively two ~~points~~ values are correlated, if they ~~are close to one another~~ correspond to locations that are close to one another with respect to the Euclidean norm. ~~This is seen from the~~ A Gaussian prior covariance can be ~~is~~ defined based on the following formula [12]

$$K_{ij} = \sigma^2 e^{-\frac{\|\tilde{p}_i - \tilde{p}_j\|}{2l^2}}, \quad i,j = 1,\dots,n, \tag{13}$$

where $\sigma, l \in \mathbb{R}$ are the ~~variance~~ pointwise standard deviation and the correlation length respectively, and $\tilde{p}_i$ and $\tilde{p}_j$ are the coordinates of the pixel corresponding to the components x_i and x_j in the unknown absorption. Equation (13) defines how different ~~data points~~ components of the unknown are correlated with one another. From the equation we see that the prior ~~is~~ covariance is symmetric, since the Euclidean norm is also symmetric. Therefore ~~equation~~ (13) ~~will~~ yields a symmetric positive definite covariance matrix. We also see that as the correlation length grows, the exponent grows as well, which results in larger correlations for points that are further apart. Similarly if the correlation length decreases, the correlation of two far away data points approaches quickly zero. The parameter sigma defines the overall level of correlation between any two data points.

Based on the covariance function in ~~equation~~ (13) it is possible to define the prior covariance matrix Γ_{pr}. The prior covariance matrix is element wise defined in the following way

$$(\Gamma_{pr})_{ij} = K(r_i, r_j)$$

[Tämän osuuden voi liittää jo kaavan (13) yhteyteen.]

where r_i and r_j are the coordinates of the midpoints of pixels i and j in the discretised image.

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 equation~~ (4) the posterior will also follow a Gaussian distribution with the mean vector and covariance matrix described in section 4.1. This allows for easy computational methods in optimisation algorithms.

Another advantage of Gaussian priors is that the parameters can be optimised non-sequentially [5] ONKO TÄSSÄ TAPAUKSESSA, SILLÄ VAIKUTTAA, ETTÄ Bayesian design of measurements for magnetorelaxometry imagingissa oleva pseudokoodi algoritmi ei tässä tapauksessa olisi sequential???. This means that one can optimise for every parameter at once for example using gradient descent. Onko oikea määritelmä? This speeds up the optimisation algorithm, since when optimising sequentially, one must calculate the optimal solutions many times for a few parameters at a time. mainitaan pelkästään, että hidastaa, koska systeemimatriisi suuri, mutta R vakio kokonaisuudessaan tässä tutkielmassa [Keskustellaan tästä uudestaan]

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

~~Another advantage of the Gaussian priors is that the posterior covariance does not depend on the measurements...~~

5.2 Total variation prior and lagged diffusivity iteration

6 Optimal experimental design

Optimal experimental design (OED) ~~is a way to define what it means for something to be the optimal solution to a problem~~ aims to choose the design parameters of a measurement setup so that the information on the unknown is maximized. 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 type of statistical quantity.~~ These commonly include ~~the squared norm to some point estimate as~~ the expected distance of the unknown from a point estimate in A-optimality or the information gain as in D-optimality [3].

A-optimality criterion can be interpreted as minimizing the posterior variance of the pixel wise absorption values X inside the ROI in equation 4 [13]. This is intuitive, since it is logical to aim for maximum certainty about the absorption values inside the ROI when ~~calculating the optimal result~~ 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 [13]. Other types of optimality, such as D-optimality perform worse than A-optimality in imaging, since A-optimality criterion focuses on more important aspects of the problem. A-optimality criterion also performs ~~good~~ well with other types of optimality criteria [13]. 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 [13]. 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. ~~A-optimality sometimes suffers from the solutions being locally optimal, however global optimality is not always ensured [?].~~ This complicates the optimization process and could benefit from using more advanced optimization methods, such as ADAM, which may help to avoid local minima [14]. Ehkä ei kannata mainita ADAM:ia

Minimization of the A-optimality target may lead to only locally optimal solutions.

A-optimality criterion can be calculated using the following equation REMEMBER TO DEFINE THE BARRIER FUNCTION SOMEWHERE

$$\Phi_A(d) = \mathrm{tr}(A\Gamma_{\mathrm{post}}(d)A^T), \tag{14}$$

where $\Gamma_{\mathrm{post}}(d)$ is the posterior covariance matrix described in equation 11 and $A \in \mathbb{R}^{n\times n}$ is the diagonal measurement matrix that defines the ROI [3, 5]. 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 ~~equation~~ (14) yields the expected reconstruction error $\mathbb{E}(\|x - x_{\mathrm{post}}\|_2^2)$ with respect to the euclidean norm [3]. This is a useful metric when trying to ~~measure the effectiveness of the optimisation procedure.~~ maximize the reliability of x_{post} as a reconstruction of the true unknown

7 Algorithms and their implementations

In this section we will describe the basic algorithms used in the numerical experiments. The first algorithm presented in section 7.1 presents the needed gradient calculations as well as the basic gradient descent. Next, ~~up in~~ section 7.2 will present an implementation for the line search that is used in combination with the gradient descent.

The algorithm in section 7.3 will implement a basic algorithm with a Gaussian prior. In section 7.4 we will implement an algorithm using a TV prior, which can not be done using the same algorithm as in section 7.3, since ~~the equations~~ (1) and (2) ~~do~~ are not valid ~~not apply~~ when using a TV prior. The algorithm has to use lagged diffusivity iteration described in section 5.2. The numerical experiments using these algorithms and the conclusions will be presented in sections 8 and 9 respectively.

7.1 Gradient descent

The optimisation for the best design parameters d will be executed using ~~the~~ gradient descent in this thesis. Gradient descent is an optimisation algorithm that takes steps towards the direction of largest decrease. 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. The needed partial derivatives are calculated explicitly in the following way: REMEMBER THE BARRIER FUNCTION AND

IT'S DERIVATIVE

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

Since trace and 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

$$\frac{\partial}{\partial d_i}\mathrm{tr}(A\Gamma_{\mathrm{post}}(d)A^T) = \mathrm{tr}\Bigl(\frac{\partial}{\partial d_i}A\Gamma_{\mathrm{post}}(d)A^T\Bigr)$$
$$= \mathrm{tr}\Bigl(A\frac{\partial \Gamma_{\mathrm{post}}(d)}{\partial d_i}A^T\Bigr)$$

Let us ~~Next up~~ we consider only the derivative of the posterior covariance matrix $\Gamma_{\mathrm{post}}(d)$. ~~Consider equation (1)~~ Based on ~~the equation~~ (11) and the linearity of the partial derivative operator we get

$$\frac{\partial \Gamma_{\mathrm{post}}(d)}{\partial d_i} = \frac{\partial}{\partial d_i}(\Gamma_{\mathrm{pr}} - \Gamma_{\mathrm{pr}}R^T(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}R\Gamma_{\mathrm{pr}})$$
$$= \frac{\partial}{\partial d_i}\Gamma_{\mathrm{pr}} - \frac{\partial}{\partial d_i}\Gamma_{\mathrm{pr}}R^T(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}R\Gamma_{\mathrm{pr}}$$

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

$$\frac{\partial \Gamma_{\mathrm{post}}(d)}{\partial d_i} = -\frac{\partial}{\partial d_i}\Gamma_{\mathrm{pr}}R^T(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}R\Gamma_{\mathrm{pr}}$$
$$= -\Gamma_{\mathrm{pr}}R^T\frac{\partial}{\partial d_i}((R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1})R\Gamma_{\mathrm{pr}}$$

The derivative of the inverse of the 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. Next let $Z(d) = (R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}$. We get the partial derivative of Z as

$$\frac{\partial Z(d)}{\partial d_i} = \frac{\partial}{\partial d_i}(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}$$
$$= -(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}\Bigl(\frac{\partial}{\partial d_i}(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))\Bigr)(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}$$

Since Γ_{pr} and R are independent of d and $Z(d) = (R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}$, we get

$$\frac{\partial Z(d)}{\partial d_i} = -(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}\Bigl(\frac{\partial}{\partial d_i}(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))\Bigr)(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))^{-1}$$
$$= -Z(d)\Bigl(\frac{\partial}{\partial d_i}(R\Gamma_{\mathrm{pr}}R^T + \Gamma_{\mathrm{noise}}(d))\Bigr)Z(d)$$
$$= -Z(d)\Bigl(\frac{\partial}{\partial d_i}(\Gamma_{\mathrm{noise}}(d))\Bigr)Z(d).$$

Based on section 3.3, we have that

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

Hence we get the derivative Miten fiksusti esittää matriisi, missä paljon nollia

Big Big

$$\frac{\partial Z(d)}{\partial d_i} = -Z(d)\Big(\frac{\partial}{\partial d_i}(\Gamma_{\text{noise}}(d))\Big)Z(d)$$

$$= -Z(d) \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 2d_i & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix} Z(d).$$

$L(d_i)$,

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

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

~~Let $L(d_i)$ represent the previous matrix, i.e.~~

$$L(d_i) = \begin{bmatrix} 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 2d_i & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 \end{bmatrix}$$

Substituting back we get the following formula

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

Finally, we obtain the wanted partial derivative

$$\frac{\partial \Phi_A(d)}{\partial d_i} = \frac{\partial}{\partial d_i} \text{tr}(A \Gamma_{\text{post}}(d) A^T)$$
$$= \text{tr}(A \frac{\partial \Gamma_{\text{post}}(d)}{\partial d_i} A^T)$$
$$\frac{\partial \Phi_A(d)}{\partial d_i} = \text{tr}(A \Gamma_{\text{pr}} R^T Z(d) L(d_i) Z(d) R \Gamma_{\text{pr}} A^T). \tag{15}$$

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

7.2 Golden section line search

In this thesis, we apply a line search method to aid the gradient descent to converge faster. The golden section search is a method, which iteratively computes function

Algorithm 1 Gradient descent

1: Choose initial parameter d_0 and the stopping criterion τ.
2: Set $d = d_0$.
3: **while** not converged **do**
4: Calculate $\nabla\Phi_A(d)$ using to equation 15.
5: **if** $\|\nabla\Phi_A(d)\|_2 < \tau$ **then**
6: **break**
7: **end if**
8: Set d according to algorithm 2.
9: Calculate $\Phi_A(d)$. ▷ Only for monitoring progress.
10: **end while**
11: **return** d

values to seek a minimum value along the given line [15]. However, since the feasible set is not convex as noted in section 3.3, a modified version of the line search must be used. This modified version helps us to 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 step size of the algorithm. If we hit the unfeasible set, we try to apply the golden section search again, but using a shorter length. If we hit the unfeasible set n times, the $n+1$:th golden section search will be between the points d and $d - \alpha r^n \nabla\Phi_A(d)$, where $r \in [0, 1)$ 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 have not seen such cases happen while doing 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 each iteration, since one of the previous two points can be reused [15]. 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 [15]. 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.

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

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 $\phi^{-1} = 0.618033988749895$, the reduction $r = 1$ and $d = d_0$.
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 $\mu = a + \phi^{-1}(b - a)$.
  if a, b, λ or μ not in feasible set **then**
    continue
  end if
  Set $r = \beta r$.
  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

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 [5] with modifications mainly after the optimisation step.

Algorithm 3 is initialised by choosing an appropriate initial parameter d, a ROI we are interested in, as well as choosing an initial covariance matrix Γ_0. The system matrix R also 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 section 1.

As described in section 3.3, we capture images from every direction simultaneously. We optimise for the intensities of those 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 A-optimality (unless it gets stuck at a local minimum), which is not guaranteed with the sequential approach [3]. However, Burger et al. [3] found the sequential algorithm to perform well in similar tasks to ones presented in section 8. Our algorithm is presented as algorithm 3.

Algorithm 3 Basic algorithm

1. Choose initial design parameter d_0 and the ROI A.
2. Initialise covariance matrices Γ_0 and Γ_{noise} according to the section 3.3.
3. Set $d = d_0$ and $\Gamma_{\text{pr}} = \Gamma_0$.
4. Calculate the system matrix R based on the number of angles k and the number of offsets m.
5. Calculate $d = \underset{d \in \mathbb{R}^{km}}{\operatorname{argmin}} \Phi_A(d)$ according to algorithm 1.
6. Calculate the posterior covariance matrix Γ_{post} and the posterior mean x_{post} according to equations 11 and 12 respectively.
7. **return** d, Γ_{post}, x_{post}

7.4 Algorithm with a TV prior and lagged diffusivity iteration

8 Numerical experiments

9 Conclusions

References

[1] E. Somersalo Jari Kaipio. *Statistical and Computational Inverse Problems*. Springer, 2004.

[2] Alejandro Ribes and Francis Schmitt. Linear inverse problems in imaging. *IEEE Signal Processing Magazine*, 25(4):84–99, 2008. doi: 10.1109/MSP.2008.923099.

[3] Martin Burger, Andreas Hauptmann, Tapio Helin, Nuutti Hyvönen, and Juha-Pekka Puska. Sequentially optimized projections in x-ray imaging. 06 2020.

[4] Linan Zhang and Hayden 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.

[5] J. Maaninen. Bayesian experimental design for magnetorelaxometry imaging. Master's thesis, School of Science, Espoo, 01 2023.

[6] Samson Wanyonyi, Ayubu Okango, and Julius Koech. Exploration of d-, a-, i-and g-optimality criteria in mixture modeling. *Asian Journal of Mathematics Statistics*, 12:15–28, 05 2021. doi: 10.9734/AJPAS/2021/v12i430292.

[7] Tapio Helin, Nuutti Hyvönen, and Juha-Pekka Puska. Edge-promoting adaptive bayesian experimental design for x-ray imaging. 2021.

[8] Lars Ruthotto, Julianne Chung, and Matthias 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.

[9] Styliani Tassiopoulou, Georgia Koukiou, and Vassilis Anastassopoulos. Algorithms in tomography and related inverse problems a review. *MDPI algorithms*, 2024.

[10] Paula R. Patel and Orlando De Jesus. *CT Scan*. StatPearls Publishing, Treasure Island (FL), 2023. URL http://europepmc.org/books/NBK567796.

[11] Giovanni S. Alberti, Ernesto De Vito, Matti Lassas, Luca Ratti, and Matteo Santacesaria. Learning the optimal tikhonov regularizer for inverse problems. 2021.

[12] David Kristjanson Duvenaud. *Automatic Model Construction with Gaussian Processes*. PhD thesis, University of Cambridge, 2014.

[13] Katherine Allen-Moyer Bradley Jones and Peter 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.

[14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. 2017. URL https://arxiv.org/abs/1412.6980.

[15] Fabricio Oliveira and modified by Nuutti Hyvönen. *Nonlinear optimisation*. 2022.