Optimizing Heating Patterns in Thermal Tomography

Juha-Pekka Puska

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo 17.9.2018

Thesis supervisor:

Prof. Nuutti Hyvönen

Thesis advisor:

D.Sc. (Tech.) Lauri Mustonen

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Juha-Pekka Puska

Title: Optimizing Heating Patterns in Thermal Tomography

Date: 17.9.2018 Language: English Number of pages: 6+50

Department of Mathematics and Systems Analysis

Code of major: SCI3053

Supervisor: Prof. Nuutti Hyvönen

Advisor: D.Sc. (Tech.) Lauri Mustonen

Thermal tomography is a promising method for non-destructive testing of materials based on measuring the boundary temperature of an object that is exposed to a known heat flux. The estimation of the internal structure, i.e. the spatially varying heat capacity and thermal conductivity, is an inverse problem, to which the statistical inversion approach is applied.

The question of optimal experiment design is how to conduct an experiment so that the maximum amount of information is gained. In this thesis, the goal is to optimize the time-dependent heating patterns in thermal tomography.

The actual measurements were numerically simulated using finite element modeling. The resulting parameter reconstructions with the optimized heating patterns were compared to a reference pattern to see if, on average, the optimized patterns resulted in smaller reconstruction errors.

The results indicate that on average the reconstruction error was mostly dependent on the rate of increase in the heating and the total amount of heat transferred. The optimization procedure also consistently resulted in patterns with maximum heating within the given constraints.

Keywords: inverse problem, Bayesian inference, optimal experimental design, finite element method, thermal tomography

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Juha-Pekka Puska

Työn nimi: Lämmityskuvioiden optimointi lämpötomografiassa

Päivämäärä: 17.9.2018 Kieli: Englanti Sivumäärä: 6+50

Matematiikan ja systeemianalyysin laitos

Pääaineen koodi: SCI3053

Työn valvoja: Prof. Nuutti Hyvönen

Työn ohjaaja: TkT Lauri Mustonen

Lämpötomografia on lupaava menetelmä materiaalien rikkomattomaan aineenkoetukseen. Siinä kappaleeseen kohdistetaan tunnettu lämpövuo, ja mitataan kappaleen lämpötilaa sen reunoilla. Kappaleen sisäisen rakenteen estimointi on erääntyyppinen inversio-ongelma, johon käytetään tilastollista lähestymistapaa.
Optimaalisessa koesuunnittelussa tavoitteena on suunnittella suoritettava koe siten, että se tuottaa mahdollisimman paljon informaatiota. Tässä työssä tavoitteena on optimoida lämmityskuvioita.
Työssä mittausta simuloitiin numeerisesti käyttäen elementtimenetelmää. Optimoiduilla lämmityskuvioilla saatuja parametrirekonstruktioita verrattiin verrokkikuvioilla saatuihin arvoihin. Kysymyksenä oli, johtavatko optimoidut lämmityskuviot pienempiin virheisiin.
Tulokset indikoivat että keskimääräisen rekonstruktiovirheen pieneneminen oli enimmäkseen riippuvainen lämmityksen kasvunopeudesta sekä siirretystä kokonaislämmöstä. Optimointimenetelmä myös säännöllisesti tuotti lämmityskuvioita joissa lämmitys oli suurin mahdollinen rajoitusehtojen puitteissa.

Avainsanat: inversio-onglema, bayesiläinen päättely,
optimaalinen koesuunnittelu, elementtimenetelmä, lämpötomografia

Preface

First and foremost I want to thank Nuutti Hyvönen for giving me the chance to work on this project, and for his constant and active help and feedback at each step of the process. I also want to thank my advisor Lauri Mustonen for patiently explaining to me the many technical details involved in the project, as well as for the many interesting discussions we had about mathematics, academia, and flying.

Finally, I want to thank all the people in room M329, most notably Vesa Kaarnioja, for making me feel welcome when I first joined the department.

Otaniemi, 17.9.2018

Juha-Pekka Puska

Contents

Symbols and abbreviations

Symbols

Ω	measurement domain
$\partial\Omega$	boundary of measurement domain
u	temperature
$u\|_{\partial\Omega}$	temperature on the boundary
T	end time of measurement
I	time interval of measurement, $[0, T]$
H_j	heater element
s_j	measurement sensor
a	thermal conductivity
b	heat capacity
c	surface heat conductance
ν	unit exterior normal of measurement domain
f	external heating
χ_j	characteristic function of heater
g	heating function
$\tilde{\chi}_j$	characteristic function of pixel
$\varphi_k(t)$	basis functions for heating
M	Measurement points in time
$\mathbf{a}$	parameter vector of thermal conductivity
$\mathbf{b}$	parameter vector of heat capacity
$\mathbf{y}$	measurement vector
$\pi()$	probability distribution
Γ_{post}	covariance matrix of posterior distribution
Γ_{pr}	covariance matrix of prior distribution
Γ_{noise}	covariance matrix of noise

1 Introduction

The question of determining the internal structure of an object from values measured on its boundary belongs to the class of so called *inverse problems*. Inverse problems are defined as the complement of forward, or well-defined problems [1]. This roughly means a solution might not exist, the solution might not be unique, and the solution does not necessarily depend continuously on the data. The characterizing features of inverse problems are that they are often nonlinear and highly unstable with regards to noise and errors in modeling and measurement. Therefore they require special solution techniques to ensure the solutions make sense, and numerical errors can also easily have a large impact. When solving inverse problems, it is important to use all available prior information about the problem.

Thermal tomography is a method whose goal is to reconstruct the internal structure of an object by applying a known heat flux at certain locations on the boundary of the object and measuring the effect of the heating on the temperature of the object on its boundary [2]. It is similar to the more established electrical impendance tomography [3][4] and optical tomography [5]. The reconstruction is possible since the equations governing heat conduction are well known and can be modeled accurately by numerical methods such as finite elements. The most probable application of thermal tomography is the non-destructive testing of objects and materials that can withstand moderate temperature changes.

In the statistical approach to inverse problems, the unknown parameters are modeled as random variables with an assumed probability distribution, and the measurement data as additional information about those variables. Using Bayesian reasoning, assumptions about the probability distributions of the unknowns, known also as the *prior probability*, can be combined with the data into a *posterior probability*. The posterior probability summarizes all information about the unknowns, given the observations. Unlike conventional inversion techniques, which only produce a single estimate, the posterior probability distribution provided by statistical inversion also includes information about the variance and possible error of the result. In this thesis, the unknown random variables are the thermal conductivity and heat capacity inside the object. In a real world application a region of lower thermal conductivity and heat capacity could be interpreted as a crack or a hole in the object, which could lead to the object fracturing or otherwise failing in its intended use.

Optimal experimental design considers how to design an experiment so that it yields the maximum amount of information given some case-dependent constraints [6]. Design in this context means choosing the control variables, such as sampling times or sample sizes. This is important since in real world applications there are usually limitations on how the measurement can be conducted. We might for example only have a limited number of samples available, or limitations might originate from the physical design of the experimental apparatus. In the case of thermal tomography, possible optimization targets are the shapes, sizes and positions of the heaters and sensors, the measurement times, and the heating patterns. In this thesis, the focus is on the heating patterns. The question that this thesis ultimately aims to answer is whether there is an optimal pattern of heating the object with

regards to information gained about the unknown parameters. Intuitively one could think of several optimal solution candidates such as heating with maximal heat flux throughout the experiment, heating initially with maximal intensity and stopping abruptly, or perhaps even periodic heating. Obviously there are real world limitations on how the heating can be conducted. It is not possible to have infinitely high temperatures, discontinuities in the temperature, or active cooling.

In mathematical terms, the goal of the optimization is to make the posterior distribution of the parameters as localized as possible, effectively reducing the posterior uncertainty of the unknown parameters. The degree of localization is encoded in the covariance matrix of the posterior distribution, so the optimization target is to minimize either the trace or the determinant of the said matrix. These targets correspond to well known concepts of A-optimality and D-optimality.

To test the results of the optimization procedure, we simulate measurement using the optimized heating patterns, solve the inverse problem, and compare the reconstruction to the actual target parameter values. If the optimized patterns do yield more information about the object, the reconstruction errors should be lower than if the measurement were conducted using a standard reference pattern. To account for random variability, this procedure is repeated multiple times, each time drawing a sample parametrization from the given prior distribution.

The layout of this thesis is as follows. In Section 2 thermal tomography is described in more detail, the exact mathematical formulation is presented, and a measurement setup is fixed. The chosen setup roughly corresponds to a known experimental setup [11]. In Section 3 the problem is discretized and the parameters are defined for the numerical simulation. The section also provides a brief overview of the finite element method used to solve the forward problem. Section 4 explains the principles of statistical inversion techniques and how they are used in this context. Then the optimization goal and method is explained in more detail. Section 5 shows the results and analysis of the numerical experiments. Finally, Section 6 contains concluding remarks and analysis of the results in a wider context.

2 Thermal tomography

2.1 Background

Thermal tomography is a nondestructive imaging method where the aim is to recover information about the inner structure of an object using temperature measurements on the boundary of the object [7]. It could provide an inexpensive way to test materials and structures for cracks, air bubbles or other deformations that could alter the structural properties of the object.

In thermal tomography, temperature sensors are placed at multiple locations on the boundary, and then the object is heated from specific source location. As the heat diffuses in the object, the evolution of the temperature is observed at the measurement locations. Since the equations governing heat transfer are well known, the measurements bring information about the material parameters affecting the phenomena, namely the thermal conductivity and heat capacity of the object. Higher thermal conductivity will cause the heat to diffuse faster, whereas a higher heat capacity will mean that the object will take longer to heat up. Finally, the process can be repeated for multiple source locations.

The forward problem of solving the diffusion equation is well-posed and stable. Conversely, the inverse problem, that is, reconstructing the internal structure, is ill-posed and non-linear. Solving the (discretized) forward problem is discussed in Section 3 and the inverse problem in Section 4. As can be expected, more measurement points result in more information and therefore better reconstructions. However, the cost of adding more data is higher computational complexity and/or a need for a more complicated measurement device. This serves as the motivation for this thesis: getting better reconstructions without increasing the amount of data.

The foundations for performing thermal tomography using the diffusion equation were laid out in [8]. Bayesian inversion techniques for performing the inversion were introduced in [9]. Since then, it has been shown numerically that the heat capacity, thermal conductivity, and surface heat transfer coefficient can all be estimated simultaneously [10]. However, in this thesis the surface heat transfer coefficient is assumed to be known. Also, a "proof of concept" experimental device has been manufactured [11].

2.2 Mathematical model

Let $\Omega \subset \mathbb{R}^2$ be the imaged body and let

$$u : \Omega \times I \to \mathbb{R}$$

be the function describing its temperature over the time interval $I = [0, T]$, for some $T > 0$. There are $J \in \mathbb{N}$ heaters and sensors attached to the boundary, denoted by H_j and s_j, respectively. The heaters are identified with open subsets of $\partial\Omega$, whereas the sensors are interpreted as points on the boundary. The model for conductive

heat transfer is described by the following parabolic initial boundary value problem

$$\begin{aligned} b\frac{\partial u}{\partial t} - \nabla \cdot (a\nabla u) &= 0 && \text{in } \Omega \times I, \\ \nu \cdot a\nabla u &= c(f-u) && \text{on } \partial\Omega \times I, \\ u &= 0 && \text{in } \Omega \times \{0\}, \end{aligned} \tag{1}$$

where $a : \Omega \to \mathbb{R}_+,\ b : \Omega \to \mathbb{R}_+,\ c : \partial\Omega \to \mathbb{R}_+$, are the thermal conductivity, heat capacity and surface heat conductance, respectively. Moreover $\nu : \partial\Omega \to \mathbb{R}^2$ is the unit exterior normal of Ω and $f : \Omega \times I \to \mathbb{R}$ describes the external heating of Ω. It is also assumed that the object is at constant temperature to start with and the exterior temperature is zero. The object is not assumed to be insulated from the environment, but the surface heat conductance is assumed to be known.

The measurements are values of the boundary temperature $u|_{\partial\Omega} : \partial\Omega \times I \to \mathbb{R}$ at the measurement points s_j on the boundary and at predetermined measurement times on the interval I. The measurements start at the same time as the heating and end at the same time, i.e., at time T (although this is not necessarily an optimal choice).

Only one of the heaters is active at a time, so the heating function f is described by

$$f_j(x,t) = g_j(t)\chi_j(x),$$

where j is the index of the active heater H_j and χ_j is its characteristic function:

$$\chi_j(x) = \begin{cases} 1 & \text{if } x \in H_j, \\ 0 & \text{otherwise.} \end{cases}$$

2.3 Measurement setup

For simplicity, the object is chosen to be a two-dimensional unit disk centered around the origin. On the boundary of the disk, there are eight heater elements labeled as $H_j,\ j = 1, \ldots, 8$, and eight equally spaced measurement sensors labeled $s_j,\ j = 1, \ldots, 8$. The heater elements are attached to the boundary of the object ($H_j \subset \partial\Omega$) and have angular width $\pi/8$. The point-like measurement sensors are located at the mid-points between the heater elements. The measurement setup is illustrated in Figure 1.

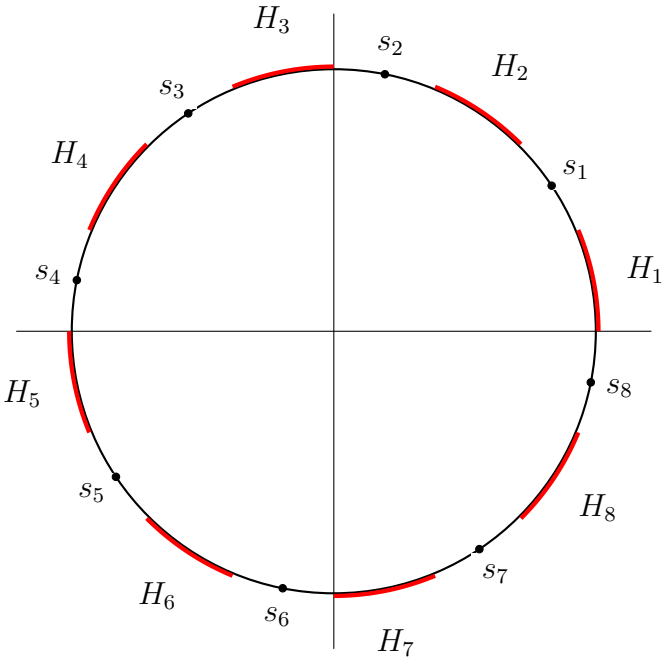

Figure 1: Measurement setup. H_j, $j = 1, \ldots, 8$, are the heaters and s_j, $j = 1, \ldots, 8$, the sensors on the boundary of the imaged disk.

In the measurement process, one of the heaters is turned on and it applies a heat flux according to the chosen heating function g. The temperatures are then measured simultaneously at all the measurement points at the chosen measurement times. The process can then be repeated for the rest of the heaters. Between the measurements the setup is allowed to cool down so that the initial conditions for each measurement are identical.

3 Discretization

3.1 Parametrization

To discretize the heat capacity and thermal conductivity, the domain is divided into non-overlapping pixels that cover the entire domain, as shown in Figure 2. The pixels are chosen so that they all have roughly the same area. Increasing the number of pixels (and thus making each pixel smaller) would in principle make it possible to see smaller features in the object, although the ill-posedness of the inverse problem of thermal tomography sets a lower limit for the reachable resolution. This would also make the inverse problem more demanding computationally.

The parameters a and b are assumed to be constant on each pixel so that

$$a(x) = a_0 + \sum_{l=1}^{L} a_l \tilde{\chi}_l(x), \qquad b(x) = b_0 + \sum_{l=1}^{L} b_l \tilde{\chi}_l(x),$$

where $\tilde{\chi}_l$ is the characteristic function of a pixel in Ω, and a_0 and b_0 the respective base levels around which the parameters vary. In a real world setting, the base level would be set according to the presumed material of the object. Our simulation uses 40 pixels, so the parameter vector is $[\mathbf{a}^T, \mathbf{b}^T]^T \in \mathbb{R}^{80}$, where $\mathbf{a}^T = [a_1, \ldots, a_L]$ and $\mathbf{b}^T = [b_1, \ldots, b_L]$.

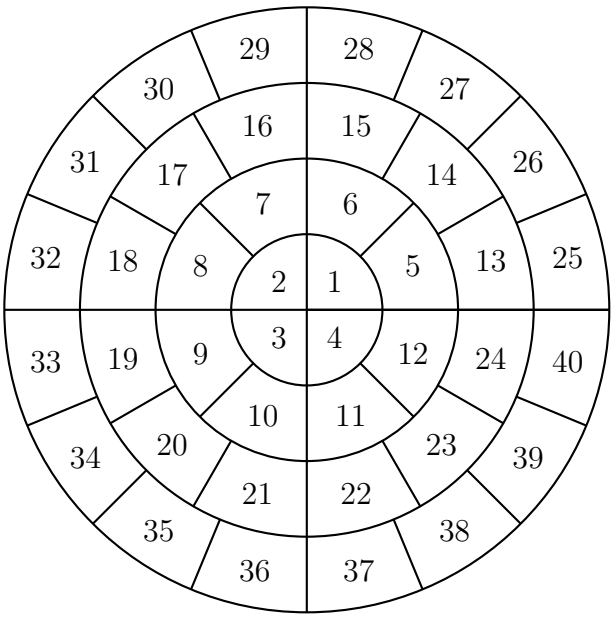

Figure 2: Discretization of heat capacity and thermal conductivity in the unit disk.

The heating pattern is parametrized as

$$g(t) = \sum_{k=1}^{K} g_k \varphi_k(t),$$

where $\varphi_k,\ k = 1, \ldots, K$ are piecewise linear basis functions defined by

$$\varphi_k(t) = \begin{cases} (t - t_{k-1})/\Delta t & t_{k-1} \leq t \leq t_k, \\ (1 - (t - t_k))/\Delta t & t_k \leq t \leq t_{k+1}, \\ 0 & \text{otherwise}, \end{cases} \tag{2}$$

where t_k are equally spaced on the interval $[0, T]$ and Δt is the corresponding offset. This means that our heating function is a piecewise linear function specified by the coefficient vector $\mathbf{g} \in \mathbb{R}^K$. The first heating basis function φ_1 is shown in Figure 3.

In theory, it would be possible that with the multiple heatings, the optimal solution would be different for each heater. However, since the measurement setup has rotational symmetry, this effect is unlikely to be very significant. Also, our numerical experiments support this belief. Therefore, it is assumed that the same heating pattern $g_j = g$ is used for all the heaters.

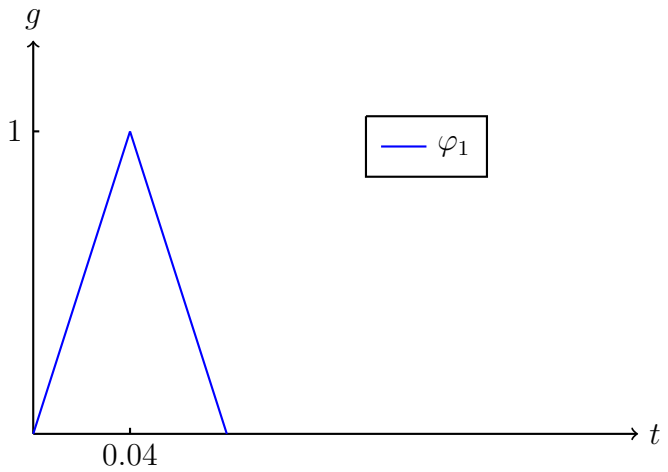

Figure 3: First basis function in (2).

Each sensor measures the boundary temperature at discrete times

$$t_i = \frac{T}{M} i, \quad i = 1, \ldots, M.$$

The total number of measurements for one iteration of the heating process is thus $8M$, and for all eight heaters we get a measurement vector $\mathbf{y} \in \mathbb{R}^{64M}$.

With this measurement setup, we can model the forward problem by the function

$$F : (\mathbf{a}, \mathbf{b}; \mathbf{g}) \mapsto \mathbf{y}$$

that maps the material parameter vectors $\mathbf{a}$ and $\mathbf{b}$, and the chosen heating function described by $\mathbf{g}$, to the temperature measurements on the boundary of the object.

3.2 Solution by FEM and time-stepping

The finite element method (FEM) is a technique for numerically approximating the solution of a partial differential equation. It is favored especially for its ability to

handle complex domains and boundary conditions and it is therefore well suited for various engineering applications [12]. Originally designed mainly for structural engineering, FEM has since been developed by mathematicians into a general purpose framework for partial differential equation problems in various fields. In this thesis, the finite element method is used to solve the forward problem of heat diffusion in the measurement object.

The steps of the finite element method are:

1. Deriving the variational formulation for the given problem.
2. Discretizing the domain.
3. Considering the problem separately in each element.
4. Assembling the elements to form a global system of equations for the problem.
5. Solving the equations to obtain the global solution.

For parabolic problems such as the diffusion equation (1), the most common approach is to use the finite element method to first discretize the problem in the spatial domain. This is called semi-discretization, and it results in a system of ordinary differential equations. Then, the problem is solved forward in time using a finite difference method. The other option, using finite elements to discretize the system in both the spatial and time domains, is also possible but not used here.

The variational or the weak formulation is a way of relaxing the smoothness requirements of the solution and it is central in developing the eventual discrete model for the finite element method. To derive the variational formulation, multiply the first equation of (1) by a test function $v \in H(\Omega)$ and integrate by parts to obtain

$$\begin{aligned} 0 &= \int_\Omega b\frac{\partial u}{\partial t} v \,\mathrm{d}x - \int_\Omega \nabla \cdot (a\nabla u) v \,\mathrm{d}x \\ &= \int_\Omega b\frac{\partial u}{\partial t} v \,\mathrm{d}x + \int_\Omega a\nabla u \cdot \nabla v \,\mathrm{d}x - \int_{\partial\Omega} c(f-u) v \,\mathrm{d}s. \end{aligned}$$

Here we first used integration by parts on the second term to get

$$-\int_\Omega \nabla \cdot (a\nabla u) v \,\mathrm{d}x = \int_\Omega a\nabla u \cdot \nabla v \,\mathrm{d}x - \int_\Omega \nabla \cdot (a\nabla uv) \,\mathrm{d}x,$$

and then using the divergence theorem, deduced

$$\int_\Omega \nabla \cdot (a\nabla uv) \,\mathrm{d}x = \int_{\partial\Omega} a\nabla u \cdot \nu \,\mathrm{d}x.$$

Inserting the boundary condition of (1), one gets the final result, which can be written also as

$$\partial_t (bu, v)_\Omega + (a\nabla u, \nabla v)_\Omega + (cu, v)_{\partial\Omega} = (cf, v)_{\partial\Omega} \qquad \text{for all } v \in H^1(\Omega), \tag{3}$$

where the brackets denote (real) L^2-inner products over respective sets. The weak formulation is accompanied by the initial condition from (1), interpreted in the appropriate weak sense. For (almost) all $t \in [0, T]$, the solution of the problem (3) belongs to the space $H^1(\Omega)$. The question of unique solvability of (3) is discussed for example in [13].

The idea of the spatial discretization is to construct a finite dimensional space V_h that is a subspace of the original solution space $V = H^1(\Omega)$. The solution $u_h \in V_h$ is then defined as a certain best approximation of the original solution. This can be expressed by the *Galerkin orthogonality* property:

$$B(u - u_h, v) = 0 \qquad \text{for all } v \in V_h,$$

where $B(\cdot, \cdot)$ is the bilinear form defined by the left-hand side of (3). In other words,

$$B(u_h, v) = (cf, v) \quad \text{for all } v \in V_h.$$

If we define $\phi_i, i = 1, \ldots, N$ to be a basis for the space V_h, then the solution u_h can be expressed as

$$u_h(t) = \sum_{i=1}^{N} \alpha_i(t)\phi_i, \tag{4}$$

where α_i are time-dependent coefficients that account for the time evolution of the solution.

The finite element method provides a systematic way of constructing the finite dimensional subspace V_h [14]. The domain is first split into triangular elements (though other shapes could be used as well). This triangularization is denoted by $\mathcal{T}_h$, and its single element is denoted by K. In total the mesh contains N_e elements (triangles) and N_p nodes. In practice, to improve the convergence of the method, one should use a finer mesh near irregularities in the shape of the boundary or in the boundary conditions. Conversely, in places where the solution is expected to be smooth, one can use a coarser mesh to save computational resources without affecting the solution considerably. In our problem, the mesh is finer near the heater edges, since at those points the boundary condition is discontinuous. This is shown in Figure 4.

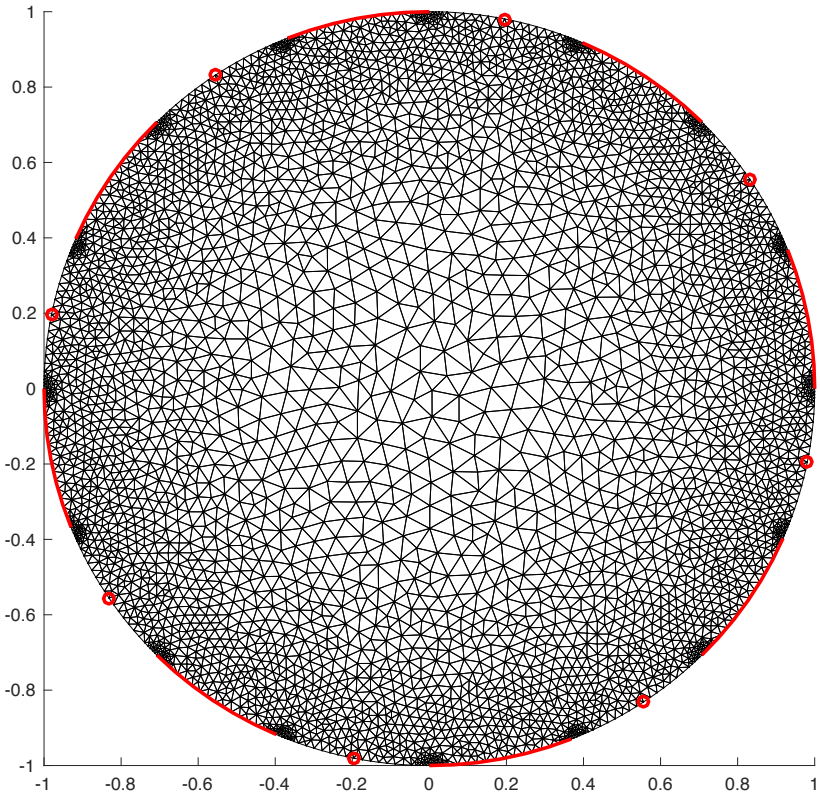

Figure 4: Mesh used for the finite element method in the numerical studies of Section 5.

There are several possible choices for the shapes of the basis functions. The simplest and most common is to set the functions to be piecewise linear:

$$V_h := \{v \in H^1(\Omega) \mid v \in \mathcal{P}^1(K)\ \forall K \in \mathcal{T}_h\},$$

where $\mathcal{P}^1(K)$ is the space of first order polynomials over K. The actual basis is constructed so that there is a basis function associated with each node in the domain, having the value 1 at that node, and 0 at all the other nodes, i.e.,

$$\phi_j(x_i) = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases}$$

where x_i is the i:th node. The finite dimensional weak problem can then be expressed as

$$\partial_t(bu_h, \phi_j)_\Omega + (a\nabla u_h, \nabla\phi_j)_\Omega + (cu_h, \phi_j)_{\partial\Omega} = (cf, \phi_j)_{\partial\Omega} \qquad \text{for all } j \text{ and } t \in I. \quad (5)$$

It is now possible to insert (4) into (5) to form the equations separately on each element. This results in a system of ordinary differential equations that can be expressed in a matrix form as

$$\partial_t B\alpha + (A + C)\alpha = \hat{f}, \quad (6)$$

where $\alpha : I \to \mathbb{R}^N$ contains the time-dependent coefficients from (4).

In finite element terminology, the matrix B is called the *mass matrix* and $A + C$ the *stiffness matrix*. Moreover $\hat{f}$ is a vector containing the coefficients of cf in the chosen finite element discretization. The equation (6) is called the semidiscrete form since so far, the discretization has only been carried out in the spatial domain.

The next step is to use a finite difference method to discretize in the time domain and to solve the equation forward in time. For this, we discretize the interval $[0, T]$ into equally spaced points t_n and let α^n denote $\alpha(t_n)$ and $\hat{f}^n$ denote $\hat{f}(t_n)$. For heat equations the most common time integration technique is the *Crank–Nicolson* (CN) method since it is of second order and unconditionally stable [15]. In the CN method, the time derivative is replaced by a difference quotient and other function evaluations are approximated by interpolated values at the midpoint of the considered time interval

$$B\left(\frac{\alpha^{n+1} - \alpha^n}{\delta t}\right) + \frac{1}{2}\left((A + C)\alpha^{n+1} + (A + C)\alpha^n\right) = \frac{1}{2}(\hat{f}^{n+1} + \hat{f}^n).$$

Here $\delta t > 0$ denotes the used time step. From this, we can approximate the coefficient vector defining the solution at the next time step as

$$\alpha^{n+1} = \left(I + \frac{\delta t}{2} B^{-1}(A + C)\right)^{-1} \left(\frac{\delta t}{2} B^{-1}\left(\hat{f}^{n+1} + \hat{f}^n - (A + C)\alpha^n\right) + \alpha^n\right).$$

where $\alpha^0 = 0$.

4 Bayesian inversion and optimal experimental design

4.1 Bayesian inversion

Bayesian inversion is a statistical paradigm for solving inverse problems. For linear problems where the forward problem is defined by the linear mapping, or matrix,

$$\mathbf{x} \mapsto \mathbf{A}\mathbf{x},$$

solving the inverse problem means finding $\mathbf{x}$ that solves the equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ (in an appropriate sense). In a deterministic approach this roughly means inverting the matrix $\mathbf{A}$ using a regularization method to ensure the solution "makes sense". This produces a single estimate for $\mathbf{x}$ based on the data available. Popular deterministic inversion methods include truncated singular value decomposition, Tikhonov regularization, and Landweber iteration [16].

In the Bayesian framework [1], the unknown parameters are modeled as random variables, and the measurements as additional information about the distribution of those parameters. The result of Bayesian inversion is a probability distribution of the unknowns that incorporates both the prior knowledge and the measurement data. From this posterior distribution we can then deduce information about the unknown parameters, such as their posterior mean and variance.

To perform the inversion, initial knowledge about the unknowns is first encoded into a prior distribution:

$$\pi_{\mathrm{pr}}(\mathbf{a}, \mathbf{b}).$$

Into this distribution we can incorporate knowledge about complex relationships, such as discontinuities and correlations. Forming the prior model can be the most difficult step in the inversion, since the prior knowledge is often qualitative, and transforming it into a probability distribution is typically nontrivial. For more information about prior distributions in tomography methods, see [17] and the references therein.

The next step is forming the *likelihood function* that is the conditional probability of encountering a measurement for given parameters,

$$\pi(\mathbf{y} \,|\, \mathbf{a}, \mathbf{b}; \mathbf{g}).$$

The likelihood function implicitly contains both the forward model as well as information about measurement noise and other uncertainties. To construct it, we need to specify a noise model, which includes possible dependence of the noise on the unknowns.

Using these definitions and the famous Bayes' formula

$$\pi(A|B) = \frac{\pi(B|A)\pi(A)}{\pi(B)},$$

for random events A and B, we can write the posterior distribution of the parameters given the measurement data as

$$\pi(\mathbf{a}, \mathbf{b} \,|\, \mathbf{y}; \mathbf{g}) = \frac{\pi(\mathbf{y} \,|\, \mathbf{a}, \mathbf{b}; \mathbf{g})\pi_{\mathrm{pr}}(\mathbf{a}, \mathbf{b})}{\pi(\mathbf{y}; \mathbf{g})}.$$

The denominator corresponds to the probability of the measurement. However, it only acts as a normalization constant and is usually not important in practice.

Bayesian inversion has several advantages compared to traditional deterministic inversion methods. Firstly, it is easier to incorporate prior information about the unknown quantities in the inversion process, since this information is specified in the prior distribution. A vast amount of research has been devoted to constructing distributions corresponding to many real world phenomena, such as gaps or discontinuities in the object. Secondly, the posterior probability distribution provided by Bayesian inversion gives more information about the unknown than a single point estimate. This makes it possible to analyze reliability of the result and compute error estimates. The downside of Bayesian inversion is its dependence on the accurate modeling of the prior distribution.

In our case, the measurement is modeled by an additive noise model,

$$\mathbf{y} = F(\mathbf{a}, \mathbf{b}; \mathbf{g}) + E,$$

where

$$E \sim N(0, \Gamma_{\text{noise}}).$$

This means that the measurement $\mathbf{y}$ is contaminated with additive, zero-mean Gaussian noise with a positive definite covariance matrix $\Gamma_{\text{noise}} \in \mathbb{R}^{64M \times 64M}$. Since the noise is assumed to be independent of the unknowns, fixing $\mathbf{a}$ and $\mathbf{b}$ does not affect the distribution of E. This means that the conditional density of y conditioned on $\mathbf{a}$ and $\mathbf{b}$ has the distribution of E shifted by $F(\mathbf{a}, \mathbf{b}; \mathbf{g})$, that is,

$$\begin{aligned} \pi(\mathbf{y} \,|\, \mathbf{a}, \mathbf{b}; \mathbf{g}) &= \pi_{\text{noise}}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big) \\ &\propto \exp\left(-\frac{1}{2}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big)^T \Gamma_{\text{noise}}^{-1}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big)\right), \end{aligned} \tag{7}$$

where the constant of proportionality does not depend on the variables of interest.

The mutually independent priors for $\mathbf{a}$ and $\mathbf{b}$ are also chosen to be zero-mean with covariances $\Gamma_a, \Gamma_b \in \mathbb{R}^{L \times L}$, respectively. This corresponds to the prior density

$$\pi_{\text{pr}}(\mathbf{a}, \mathbf{b}) \propto \exp\left(-\frac{1}{2}(\mathbf{a}^T \Gamma_a^{-1} \mathbf{a} + \mathbf{b}^T \Gamma_b^{-1} \mathbf{b})\right). \tag{8}$$

In this work, the covariance matrices of the material parameters are chosen to be of the form

$$\Gamma_{i,j} = \sigma^2 \exp\left(-\frac{\|\hat{x}_i - \hat{x}_j\|_2^2}{2l^2}\right), \qquad i, j = 1, \ldots, L, \tag{9}$$

where $\hat{x}_i \in \Omega$ is the coordinate of the center of the pixel with index i, σ is the pointwise standard deviation parameter and l is the correlation length. This type of covariance assigns a higher correlation to pixels that are close to each other and is a fairly natural choice for a physical material; see, e.g., [18] for usage of such a prior in real-world imaging. Naturally, the values of σ and l depend on the considered application/material.

Now (7) and (8) can be combined to form the posterior distribution:

$$
\begin{aligned}
&\pi(\mathbf{a}, \mathbf{b} \,|\, \mathbf{y}; \mathbf{g}) \qquad (10)\\
&\propto \pi_{\text{noise}}(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y})\pi_{\text{pr}}(\mathbf{a}, \mathbf{b})\\
&\propto \exp\left(-\frac{1}{2}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big)^T \Gamma_{\text{noise}}^{-1}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big)\right)\\
&\qquad \times \exp\left(-\frac{1}{2}(\mathbf{a}^T\Gamma_a^{-1}\mathbf{a} + \mathbf{b}^T\Gamma_b^{-1}\mathbf{b})\right)\\
&= \exp\left(-\frac{1}{2}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big)^T \Gamma_{\text{noise}}^{-1}\big(F(\mathbf{a}, \mathbf{b}; \mathbf{g}) - \mathbf{y}\big) - \frac{1}{2}(\mathbf{a}^T\Gamma_a^{-1}\mathbf{a} + \mathbf{b}^T\Gamma_b^{-1}\mathbf{b})\right),
\end{aligned}
$$

where the constants of proportionality are independent of $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{g}$.

4.2 Optimal experimental design

When performing experiments, in addition to the unknown parameters to be inferred, we often also have a number of *control variables*, or *design variables*, the values of which can be selected beforehand. The design variables do not influence the process of phenomenon being modeled, but have an effect on the outcome of the experiment. In the context of tomography methods, these can include decisions on the spatial and temporal locations of the data points, i.e. where and when we measure. Also, we can appropriately select the input, i.e. the heating in the case of thermal tomography or, e.g., the electrical currents in the case of electrical impendance tomography.

Since different values of the decision variables result in different output distributions of the data, it is natural to assume that some designs are preferable to others by providing more information about the experiment. The goal of optimal experimental design is therefore to improve the statistical inference of the unknown random variables by appropriately selecting the values of these control variables [19].

As its name suggests, finding an optimal design involves solving an optimization problem by finding a design that maximizes a chosen utility function. It is of course possible to improve the inference by simply using more measurement points, but this results in a larger computational load. This can be especially restrictive in the case of complex physical models. Also, sometimes the constraints of the experiment might limit the amount of available data. Therefore it is important to get as informative data as possible.

In related works, optimal experimental design has been applied to electrical impendance tomography to optimize current patterns [20] and electrode positions [21]. Here, the goal is choosing the heating function in a way that will give as much information about the unknown material parameters as possible under certain constraints. These constraints should be chosen so as to roughly correspond to the limitations of the actual physical heating devices.

Formulated using a decision theoretic approach as suggested in [6], finding an

optimal design is equal to maximizing the *expected utility*

$$\begin{aligned} U(\mathbf{g}) &= \int_{\mathcal{Y}} \int_{\mathcal{B}} \int_{\mathcal{A}} u(\mathbf{g}, \mathbf{y}, \mathbf{a}, \mathbf{b})\, \pi(\mathbf{a}, \mathbf{b}, \mathbf{y} \,|\, \mathbf{g})\, \mathrm{d}\mathbf{a}\, \mathrm{d}\mathbf{b}\, \mathrm{d}\mathbf{y} \\ &= \int_{\mathcal{Y}} \int_{\mathcal{B}} \int_{\mathcal{A}} u(\mathbf{g}, \mathbf{y}, \mathbf{a}, \mathbf{b})\, \pi(\mathbf{a}, \mathbf{b} \,|\, \mathbf{y}, \mathbf{g})\, \pi(\mathbf{y} \,|\, \mathbf{g})\, \mathrm{d}\mathbf{a}\, \mathrm{d}\mathbf{b}\, \mathrm{d}\mathbf{y}, \end{aligned} \tag{11}$$

where u is a chosen *utility function* that reflects the usefulness of the experiment performed with the decision variables $\mathbf{g}$, given the values of the parameters $\mathbf{a}$ and $\mathbf{b}$ and the resulting outcome $\mathbf{y}$. Since the specific values of the material parameters and the outcome vector are unknown, they must be marginalized to find the expected outcome. Hence one needs to integrate over $\mathcal{Y}$, the support of $\pi(\mathbf{y}\,|\mathbf{g})$ and $\mathcal{A} \times \mathcal{B}$, that contains the support of $\pi(\mathbf{a}, \mathbf{b} \,|\, \mathbf{y}, \mathbf{g})$. This is equal to finding the expectation of u over the joint distribution of $\mathbf{a}, \mathbf{b}$, and $\mathbf{y}$.

Literature suggests several possible utility functions, each yielding a unique optimization target. An information theoretic approach is maximizing the expected gain in Shannon information. The difference between the information contained in two distributions is measured by the *Kullback–Leibler* (KL) *divergence*

$$D_{\mathrm{KL}}(A \| B) = \int \pi_A(\theta)\, \log \left[\frac{\pi_A(\theta)}{\pi_B(\theta)} \right] \mathrm{d}\theta = \mathbb{E}_A \left[\log \frac{\pi_A(\theta)}{\pi_B(\theta)} \right].$$

For the inverse problem in question, we set the utility function to measure the KL divergence between the posterior and the prior

$$\begin{aligned} u(\mathbf{g}, \mathbf{y}, \mathbf{a}, \mathbf{b}) &= D_{\mathrm{KL}}\big(\pi(\mathbf{a}, \mathbf{b} \,|\, \mathbf{y}; \mathbf{g}) \,||\, \pi_{\mathrm{pr}}(\mathbf{a}, \mathbf{b})\big) \\ &= \int \pi(\mathbf{a}, \mathbf{b} \,|\mathbf{y}; \mathbf{g})\, \log \left[\frac{\pi(\mathbf{a}, \mathbf{b} \,|\mathbf{y}; \mathbf{g})}{\pi_{\mathrm{pr}}(\mathbf{a}, \mathbf{b})} \right] \mathrm{d}\mathbf{a}\, \mathrm{d}\mathbf{b} \\ &= u(\mathbf{g}, \mathbf{y}), \end{aligned} \tag{12}$$

which does not depend on $\mathbf{a}$ or $\mathbf{b}$. A large change in information would mean that the data has provided a high amount of information.

Now we write (11) in the form

$$\begin{aligned} U(\mathbf{g}) &= \int_{\mathcal{Y}} \int_{\mathcal{B}} \int_{\mathcal{A}} u(\mathbf{g}, \mathbf{y}, \mathbf{a}, \mathbf{b})\, \pi(\mathbf{a}, \mathbf{b} \,|\, \mathbf{y}, \mathbf{g})\, \pi(\mathbf{y} \,|\, \mathbf{g})\, \mathrm{d}\mathbf{a}\, \mathrm{d}\mathbf{b}\, \mathrm{d}\mathbf{y} \\ &= \int_{\mathcal{Y}} \int_{\mathcal{B}} \int_{\mathcal{A}} u(\mathbf{g}, \mathbf{y})\, \pi(\mathbf{y}, \mathbf{a}, \mathbf{b} \,|\, \mathbf{g}) \mathrm{d}\mathbf{a}\, \mathrm{d}\mathbf{b}\, \mathrm{d}\mathbf{y} \\ &= \int_{\mathcal{Y}} u(\mathbf{g}, \mathbf{y})\, \pi(\mathbf{y}|\, \mathbf{g}) \mathrm{d}\mathbf{y}. \end{aligned} \tag{13}$$

By inserting the utility function (12) into (13) and noting that by the Fubini's

theorem and Bayes's formula the term

$$\begin{aligned}
&\int_{\mathcal{Y}}\int_{\mathcal{B}}\int_{\mathcal{A}} \log(\pi_{\mathrm{pr}}(\mathbf{a},\mathbf{b}))\pi(\mathbf{a},\mathbf{b};\mathbf{y},\mathbf{g})\mathrm{d}\mathbf{a}\,\mathrm{d}\mathbf{b}\,\mathrm{d}\mathbf{y} \\
&= \int_{\mathcal{B}}\int_{\mathcal{A}} \left(\int_{\mathcal{Y}} \pi(\mathbf{y}|\mathbf{a},\mathbf{b};\mathbf{g})\mathrm{d}\mathbf{y} \right) \log(\pi_{\mathrm{pr}}(\mathbf{a},\mathbf{b}))\pi_{\mathrm{pr}}(\mathbf{a},\mathbf{b})\mathrm{d}\mathbf{a}\,\mathrm{d}\mathbf{b} \\
&= \int_{\mathcal{B}}\int_{\mathcal{A}} \log(\pi_{\mathrm{pr}}(\mathbf{a},\mathbf{b}))\pi_{\mathrm{pr}}(\mathbf{a},\mathbf{b})\mathrm{d}\mathbf{a}\,\mathrm{d}\mathbf{b},
\end{aligned}$$

does not depend on $\mathbf{g}$, we can drop the prior term from the denominator and get an equivalent expected utility of the form

$$U(\mathbf{g}) = \int_{\mathcal{Y}}\int_{\mathcal{B}}\int_{\mathcal{A}} \log(\pi(\mathbf{a},\mathbf{b}\,|\,\mathbf{y},\mathbf{g}))\pi(\mathbf{y},\mathbf{a},\mathbf{b}\,|\,\mathbf{g})\mathrm{d}\mathbf{a}\,\mathrm{d}\mathbf{b}\,\mathrm{d}\mathbf{g}. \tag{14}$$

Maximizing this functional results in the D-optimality criterion.

Another option is the so called Bayesian A-optimality that uses a quadratic loss function that minimizes the mean squared error between the unknown parameters and their estimates:

$$u(\mathbf{g},\mathbf{y},\mathbf{a},\mathbf{b}) = \left\| \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} - \begin{bmatrix}\hat{\mathbf{a}}(\mathbf{y};\mathbf{g})\\ \hat{\mathbf{b}}(\mathbf{y};\mathbf{g})\end{bmatrix} \right\|^2. \tag{15}$$

Inserting this into expression (11), one gets another expected utility

$$U(\mathbf{g}) = \int_{\mathcal{Y}}\int_{\mathcal{B}}\int_{\mathcal{A}} \left\| \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix} - \begin{bmatrix}\hat{\mathbf{a}}(\mathbf{y};\mathbf{g})\\ \hat{\mathbf{b}}(\mathbf{y};\mathbf{g})\end{bmatrix} \right\|^2 \pi(\mathbf{a},\mathbf{b}\,|\mathbf{y},\mathbf{g})\,\mathrm{d}\mathbf{a}\,\mathrm{d}\mathbf{b}\,\pi(\mathbf{y}\,|\,\mathbf{g})\mathrm{d}\mathbf{y}. \tag{16}$$

This condition is appropriate when one wishes to obtain a point estimate of the parameters.

Both of the utility functions (14) and (16) will be used for the optimization, and the designs produced by them will be used for the eventual reconstructions in Section 5. In should be noted that (14) and (16) take considerably simpler forms if the posterior distribution is of the form (10) *and* the forward map F is linear.

4.3 Linearization of the forward model

The challenge with maximizing (14) or (16) is that each iteration of the employed optimization algorithm requires solving forward problems as well as evaluating high-dimensional integrals (cf. (11)). This is computationally unfeasible even for a fairly simple problem as is considered in this thesis. Therefore we perform a linearization to turn the nonlinear forward problem into a far more tractable linear one:

$$F(\mathbf{a},\mathbf{b};\mathbf{g}) \approx F(\mathbf{a}_0,\mathbf{b}_0;\mathbf{g}) + J_{a,b}(\mathbf{a}_0,\mathbf{b}_0;\mathbf{g}) \begin{bmatrix}\mathbf{a}\\ \mathbf{b}\end{bmatrix}. \tag{17}$$

Here $J_{a,b}$ denotes the Jacobian matrix of F with respect to $\mathbf{a}$ and $\mathbf{b}$ evaluated at $\mathbf{a} = \mathbf{a}_0$ and $\mathbf{b} = \mathbf{b}_0$. An entry of a Jacobian matrix tells us the effect that a small change in a pixel value of the parameter would have on an individual measurement.

It follows from the linearity of (1) with respect to f that

$$J_{a,b} = \sum_{k=1}^{K} g_k J_{\varphi_k},$$

where J_{φ_k} is the Jacobian for the heating basis function φ_k, i.e., $J_{\varphi_k} = J_{a,b}(\mathbf{a_0}, \mathbf{b_0}; \varphi_k)$, with slight abuse of notation. For multiple active heaters, the Jacobians have to be calculated separately for each heater,

$$(J_{a,b})_n = \sum_{k=1}^{K} g_k (J_{\varphi_k})_n,$$

although the symmetry of the measurement setup could be used to ease the computational load. The result is a set of matrices $(J_{\varphi_k})_n \in \mathbb{R}^{JM \times 80}$ where k is the index of the heating function and $1 \leq n \leq 8$ the index of the heater.

We can precalculate the Jacobians for each of the heating basis functions, and then form the Jacobian for any chosen heating pattern by calculating a linear combination. This linearization means that only one forward problem has to be solved for each combination of a heating basis function φ_k and a heater element n.

The actual Jacobians are computed by perturbing a single element in one of the two material parameter vectors and using a difference approximation to compute the numerical derivative. In other words, we solve the forward problem for the parameters $\mathbf{a}_0$ and $\mathbf{b}_0$ for a single heating basis function, add a small perturbance of size $\varepsilon > 0$ to one of the components, recompute the forward solution, and compute the difference of these two matrices. In total, $8(K+1) \cdot 80$ forward problems must be solved. Approximating the Jacobians is computationally a fairly intensive task but it only has to be performed once as long as the background material parameters are not changed.

Inserting the linearization (17) into the posterior distribution (10), it can be straightforwardly shown that the covariance matrix of the posterior for a single active heater is of the form [1]:

$$\begin{aligned}
\Gamma_{\text{post}} &= \left(\Gamma_{\text{pr}}^{-1} + \left(\sum_{k=1}^{K} g_k J_{\varphi_k} \right)^T \Gamma_{\text{noise}}^{-1} \left(\sum_{k=1}^{K} g_k J_{\varphi_k} \right) \right)^{-1} \\
&= \left(\Gamma_{\text{pr}}^{-1} + \mathbf{G}^T \Gamma_{\text{noise}}^{-1} \mathbf{G} \right)^{-1},
\end{aligned}$$

where $\mathbf{G} = \sum_{k=1}^{K} g_k (J_{\varphi_k})$. For multiple heaters the inverse posterior becomes

$$\begin{aligned}
\Gamma_{\text{post}}^{-1} &= \Gamma_{\text{pr}}^{-1} + \begin{bmatrix} \mathbf{G}_1^T & \cdots & \mathbf{G}_8^T \end{bmatrix} \begin{bmatrix} \Gamma_{\text{noise}}^{-1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Gamma_{\text{noise}}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{G}_1 \\ \vdots \\ \mathbf{G}_8 \end{bmatrix} \\
&= \Gamma_{\text{pr}}^{-1} + \sum_{n=1}^{8} \mathbf{G}_n^T \Gamma_{\text{noise}}^{-1} \mathbf{G}_n. \qquad (18)
\end{aligned}$$

Here

$$\mathbf{G}_n = \sum_{k=1}^{K} g_k (J_{\varphi_k})_n,$$

where n is the index of the active heater.

Inserting the linearized likelihood function into the expected utility (14), the D-optimality condition can be simplified into the form (see, e.g., [21])

$$\text{Minimize}: \quad \det(\Gamma_{\text{post}}(\mathbf{g})), \tag{19}$$

where Γ_{post} is the posterior covariance. For the A-optimality (16), we similarly get a computationally feasible optimization target

$$\text{Minimize}: \quad \operatorname{tr}(\Gamma_{\text{post}}(\mathbf{g})), \tag{20}$$

if the considered point estimate $[\hat{\mathbf{a}}(\mathbf{y};\mathbf{g}), \hat{\mathbf{b}}(\mathbf{y};\mathbf{g})]$ in (16) is chosen to be the mean or the maximal value of the "linearized posterior density".

However, due to numerical stability reasons the determinant of the posterior cannot be computed directly, but instead we compute the logarithm of the determinant, since it will have the same extreme points. By resorting to the logarithm, we are able to exploit the fact that the positive definite and symmetric (inverse) posterior covariance has a Cholesky factorization, i.e., $\Gamma_{\text{post}}^{-1} = C^T C$ with C being (upper) triangular. The minimization target is then

$$\begin{aligned}
\log\big(\det(\Gamma_{\text{post}}(\mathbf{g}))\big) &= -\log\big(\det(C^T C)\big) \\
&= -2\log\big(\det(C)\big) \\
&= -2\log\big(\prod_i c_{ii}\big) \\
&= -2\sum_i \log(c_{ii}),
\end{aligned} \tag{21}$$

where c_{ii} are the diagonal elements of the aforementioned Cholesky factor. As a consequence, the D-optimality criterion reduces to optimizing the log-sum of the diagonal elements of C.

The actual optimization algorithm in Section 5 is a gradient-based one, so we need to compute the partial derivatives of the target functions given in (19) and (20). For this, we need the following matrix formulas [22]:

$$\begin{aligned}
\frac{d}{d\tau}\det(B(\tau)) &= \det(B(\tau))\operatorname{tr}\big(B^{-1}\frac{d}{d\tau}B(\tau)\big), \\
\frac{d}{d\tau}\operatorname{tr}\big(B(\tau)\big) &= \operatorname{tr}\big(\frac{d}{d\tau}B(\tau)\big), \\
\frac{d}{d\tau}(B(\tau))^{-1} &= -B(\tau)^{-1}\partial B(\tau)B(\tau)^{-1}.
\end{aligned}$$

To simplify the notation, we also set

$$\mathbf{A} = \Gamma_{\text{post}}^{-1} = \big(\Gamma_{\text{pr}}^{-1} + \mathbf{G}^T\Gamma_{\text{noise}}^{-1}\mathbf{G}\big),$$

and denote

$$\partial_i = \frac{\partial}{\partial g_i}.$$

Let us first consider the A-optimality, i.e., minimizing (20). The derivative of the trace of the posterior for a single active heater is

$$\begin{aligned}
&\partial_i\big(\mathrm{tr}(\Gamma_{\mathrm{post}})\big) \\
&\quad = \mathrm{tr}(\partial_i \Gamma_{\mathrm{post}}) \\
&\quad = \mathrm{tr}\big(-\Gamma_{\mathrm{post}}\big(\partial_i \mathbf{A}\big)\Gamma_{\mathrm{post}}\big) \\
&\quad = \mathrm{tr}\Big(-\Gamma_{\mathrm{post}}\big(\underbrace{\partial_i \Gamma_{\mathrm{pr}}^{-1}}_{=0} + \partial_i(\mathbf{G}^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G})\big)\Gamma_{\mathrm{post}}\Big) \\
&\quad = \mathrm{tr}\Big(-\Gamma_{\mathrm{post}}\big((\partial_i \mathbf{G}^T)\Gamma_{\mathrm{noise}}^{-1}\mathbf{G} + \mathbf{G}^T\partial_i(\Gamma_{\mathrm{noise}}^{-1}\mathbf{G})\big)\Gamma_{\mathrm{post}}\Big) \\
&\quad = \mathrm{tr}\Big(-\Gamma_{\mathrm{post}}\big((\partial_i \mathbf{G})^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G} + \mathbf{G}^T(\underbrace{\partial_i \Gamma_{\mathrm{noise}}^{-1}}_{=0}\mathbf{G} + \Gamma_{\mathrm{noise}}^{-1}\partial_i \mathbf{G})\big)\Gamma_{\mathrm{post}}\Big) \\
&\quad = \mathrm{tr}\Big(-\Gamma_{\mathrm{post}}\big((\partial_i \mathbf{G})^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G} + \mathbf{G}^T\Gamma_{\mathrm{noise}}^{-1}\partial_i \mathbf{G}\big)\Gamma_{\mathrm{post}}\Big).
\end{aligned}$$

The derivative of the sum-matrix $\mathbf{G}$ is simply:

$$\partial_i \mathbf{G} = \partial_i\Big(\sum_{k=1}^{K} g_k J_{\varphi_k}\Big) = J_{\varphi_i}.$$

In consequence, we finally get

$$\partial_i\big(\mathrm{tr}(\Gamma_{\mathrm{post}})\big) = \mathrm{tr}\Big(-\Gamma_{\mathrm{post}}\big((J_{\varphi_i})^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G} + \mathbf{G}^T\Gamma_{\mathrm{noise}}^{-1}J_{\varphi_i}\big)\Gamma_{\mathrm{post}}\Big).$$

For multiple active heaters the expression is only slightly more complicated:

$$\partial_i\big(\mathrm{tr}(\Gamma_{\mathrm{post}})\big) = \mathrm{tr}\Big(-\Gamma_{\mathrm{post}}\sum_{n}^{8}\big((J_{\varphi_i})_n^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G}_n + \mathbf{G}_n^T\Gamma_{\mathrm{noise}}^{-1}(J_{\varphi_i})_n\big)\Gamma_{\mathrm{post}}\Big). \tag{22}$$

In a similar manner, the derivative of the logarithm of the determinant for a single heater can be written as (cf. (21))

$$\begin{aligned}
\partial_i\big(\log(\det(\Gamma_{\mathrm{post}})\big) &= \mathrm{tr}(\mathbf{A}\partial_i \mathbf{A}^{-1}) \\
&= \mathrm{tr}\big(\mathbf{A}(-\mathbf{A}^{-1}(\partial_i \mathbf{A})\mathbf{A}^{-1})\big) \\
&= -\mathrm{tr}\big((\partial_i \mathbf{A})\Gamma_{\mathrm{post}}\big) \\
&= -\mathrm{tr}\big(((J_{\varphi_i})^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G} + \mathbf{G}^T\Gamma_{\mathrm{noise}}^{-1}J_{\varphi_i})\Gamma_{\mathrm{post}}\big).
\end{aligned}$$

For multiple heaters, the corresponding formula becomes

$$\partial_i\Big(\log\big(\det(\Gamma_{\mathrm{post}})\big)\Big) = \mathrm{tr}\Big(-\sum_{n}^{8}\big((J_{\varphi_i})_n^T\Gamma_{\mathrm{noise}}^{-1}\mathbf{G}_n + \mathbf{G}_n^T\Gamma_{\mathrm{noise}}^{-1}(J_{\varphi_i})_n\big)\Gamma_{\mathrm{post}}\Big). \tag{23}$$

4.4 Alternative noise model

Thus far, it has been assumed that the measurement noise is independent of the measurements and the level of noise was controlled only by a fixed parameter. An alternative approach would be to define the noise level to be dependent on the magnitude of the measurement, i.e., for example,

$$\Gamma_{\text{noise}} = \Big(\operatorname{diag}\big(F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g})\big)^2 + \varepsilon^2_{\text{noise}}\mathbf{I}\Big)\sigma^2, \tag{24}$$

where $\sigma > 0$ is a positive variance parameter and $F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g})$ is the measurement produced by the heating $\mathbf{g} = \sum_{k=1}^{K} g_k\varphi_k$ when the parameters $\mathbf{a}$ and $\mathbf{b}$ are set to their expected values $\mathbf{a}_0$ and $\mathbf{b}_0$. This approach would mean that heating with a larger intensity increases the amount of noise and correspondingly decreases the amount of information provided by the measurement, effectively penalizing larger values of g. Again, due to the linearity of (1) with respect to f

$$F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g}) = \sum_{k=1}^{K} g_k F(\mathbf{a}_0, \mathbf{b}_0; \varphi_k).$$

Since we have introduced a dependence on $\mathbf{g}$ for the noise model ($\Gamma_{\text{noise}} = \Gamma_{\text{noise}}(\mathbf{g})$), the assumption $\partial_i\Gamma^{-1}_{\text{noise}} = 0$ employed in the previous section no longer holds. Instead,

$$\partial_i\Gamma_{\text{noise}} = 2\sigma^2\operatorname{diag}\big(F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g})\big)\operatorname{diag}\big(F(\mathbf{a}_0, \mathbf{b}_0; \varphi_i)\big)$$

and

$$\begin{aligned}\partial_i\Gamma^{-1}_{\text{noise}} &= -\Gamma^{-1}_{\text{noise}}\partial_i\Gamma_{\text{noise}}\Gamma^{-1}_{\text{noise}} \\ &= -2\sigma^{-2}\operatorname{diag}(F(\mathbf{a}_0, \mathbf{b}_0; \varphi_i)\big)\operatorname{diag}\big(F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g})\big)\big(\operatorname{diag}(F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g})\big)^2 + \varepsilon^2_{\text{noise}}\mathbf{I}\big)^{-2}.\end{aligned}$$

Also, since the heating depends on the heater, for multiple heaters the submatrices of the noise covariance matrix depend on the heater index and the posterior covariance matrix introduced in (18) turns into

$$\Gamma^{-1}_{\text{post}} = \Gamma^{-1}_{\text{pr}} + \begin{bmatrix}\mathbf{G}_1^T & \cdots & \mathbf{G}_8^T\end{bmatrix}\begin{bmatrix}\Gamma^{-1}_{\text{noise},1} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \Gamma^{-1}_{\text{noise},8}\end{bmatrix}\begin{bmatrix}\mathbf{G}_1 \\ \vdots \\ \mathbf{G}_8\end{bmatrix}.$$

Now we can calculate the derivatives of the optimization target functions for the new additive Gaussian noise model defined by (24):

$$\begin{aligned}&\partial_i\big(\operatorname{tr}(\Gamma_{\text{post}})\big) \\ &\quad= \operatorname{tr}\Big(-\Gamma_{\text{post}}\big(\underbrace{\partial_i\Gamma^{-1}_{\text{pr}}}_{=0} + \partial_i(\mathbf{G}^T\Gamma^{-1}_{\text{noise}}\mathbf{G})\big)\Gamma_{\text{post}}\Big) \\ &\quad= \operatorname{tr}\Big(-\Gamma_{\text{post}}\big((\partial_i\mathbf{G})^T\Gamma^{-1}_{\text{noise}}\mathbf{G} + \mathbf{G}^T(\partial_i\Gamma^{-1}_{\text{noise}}\mathbf{G} + \Gamma^{-1}_{\text{noise}}\partial_i\mathbf{G})\big)\Gamma_{\text{post}}\Big) \\ &\quad= \operatorname{tr}\Big(-\Gamma_{\text{post}}\big((\partial_i\mathbf{G})^T\Gamma^{-1}_{\text{noise}}\mathbf{G} + \mathbf{G}^T(\partial_i\Gamma^{-1}_{\text{noise}})\mathbf{G} + \mathbf{G}^T\Gamma^{-1}_{\text{noise}}\partial_i\mathbf{G}\big)\Gamma_{\text{post}}\Big),\end{aligned}$$

and

$$\partial_i\big(\log(\det(\Gamma_{\text{post}})\big) = -\text{tr}\big(((J_{\varphi_i})^T\Gamma_{\text{noise}}^{-1}\mathbf{G} + \mathbf{G}^T(\partial_i\Gamma_{\text{noise}}^{-1})\mathbf{G} + \mathbf{G}^T\Gamma_{\text{noise}}^{-1}J_{\varphi_i})\Gamma_{\text{post}}\big).$$

For multiple heaters, these become

$$\partial_i\big(\text{tr}(\Gamma_{\text{post}})\big) = \text{tr}\bigg(-\Gamma_{\text{post}}\sum_n^8\big((J_{\varphi_i})_n^T\Gamma_{\text{noise}}^{-1}\mathbf{G}_n + \mathbf{G}_n^T(\partial_i\Gamma_{\text{noise,n}}^{-1})\mathbf{G}_n + \mathbf{G}_n^T\Gamma_{\text{noise,n}}^{-1}(J_{\varphi_i})_n\big)\Gamma_{\text{post}}\bigg),$$

and

$$\partial_i\bigg(\log\big(\det(\Gamma_{\text{post}})\big)\bigg) = \text{tr}\bigg(-\sum_n^8\big((J_{\varphi_i})_n^T\Gamma_{\text{noise}}^{-1}\mathbf{G}_n + \mathbf{G}_n^T(\partial_i\Gamma_{\text{noise,n}}^{-1})\mathbf{G}_n + \mathbf{G}_n^T\Gamma_{\text{noise,n}}^{-1}(J_{\varphi_i})_n\big)\Gamma_{\text{post}}\bigg),$$

respectively.

4.5 Reconstruction of parameters

The last step in Bayesian inversion is exploring the posterior probability distribution to gain information about the unknowns. Since the result of Bayesian inversion is a probability distribution, if we wish to obtain a single estimate for the unknown parameters, we must compute a point estimate. In the case of thermal tomography, a point estimate is obviously required, since that corresponds to the desired visualizable reconstruction of the physical parameters.

The most popular point estimate is arguably the *maximum a posteriori* (MAP) estimate due to its simplicity. It is the value of the parameters that maximizes the posterior probability and it is defined as

$$x_{\text{MAP}} = \arg\max_{x\in\mathbb{R}^n}\pi(x\,|\,y),$$

if y is the data and x the unknown. For our problem, the MAP estimate is of the form (cf. (10))

$$(\mathbf{a}^*,\mathbf{b}^*) = \arg\min_{\mathbf{a},\mathbf{b}\in\mathbb{R}^L}\big(\|C_{\text{noise}}(F(\mathbf{a},\mathbf{b};\mathbf{g})-\mathbf{y})\|^2 + \|C_{\text{prior,a}}(\mathbf{a}-\mathbf{a}_0)\|^2 + \|C_{\text{prior,b}}(\mathbf{b}-\mathbf{b}_0)\|^2\big), \tag{25}$$

where C_{prior} and C_{noise} are the Cholesky factors of the inverse covariance matrices, i.e.,

$$\Gamma_{\text{prior}}^{-1} = C_{\text{prior}}^T C_{\text{prior}},$$
$$\Gamma_{\text{noise}}^{-1} = C_{\text{noise}}^T C_{\text{noise}}.$$

Finding the MAP estimate involves solving an optimization problem. This is most commonly approached using some gradient based minimization algorithm, such as the Gauss–Newton method [23]. However, for simplicity, we only consider in this work solving (25) approximately via one-step linearization. Therefore we once again need to perform a linearization of the dependence of F on $\mathbf{a}$ and $\mathbf{b}$. The minimization problem then becomes,

$$(\mathbf{a}^*, \mathbf{b}^*) = \underset{\mathbf{a},\mathbf{b}\in\mathbb{R}^{\mathrm{L}}}{\arg\min} \Bigg(\left\| C_{\mathrm{noise}}\big(J_{\mathrm{a,b}}(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g}) \begin{bmatrix} \mathbf{a}-\mathbf{a}_0 \\ \mathbf{b}-\mathbf{b}_0 \end{bmatrix} + \mathbf{F}(\mathbf{a_0}, \mathbf{a_0}; \mathbf{g}) - \mathbf{y}\big) \right\|^2 + \\ \|C_{\mathrm{prior,a}}(\mathbf{a}-\mathbf{a}_0)\|^2 + \|C_{\mathrm{prior,b}}(\mathbf{b}-\mathbf{b}_0)\|^2 \Big).$$

This is now a linear least squares problem that can be rearranged as

$$(\mathbf{a}^*, \mathbf{b}^*) = \underset{\mathbf{a},\mathbf{b}\in\mathbb{R}^{L}}{\arg\min} \left\| \begin{bmatrix} C_{\mathrm{noise}} \left(J_{a,b} \begin{bmatrix} \mathbf{a}-\mathbf{a}_0 \\ \mathbf{b}-\mathbf{b}_0 \end{bmatrix} - \tilde{\mathbf{y}} \right) \\ C_{\mathrm{prior}} \begin{bmatrix} \mathbf{a}-\mathbf{a}_0 \\ \mathbf{b}-\mathbf{b}_0 \end{bmatrix} \end{bmatrix} \right\|^2 .$$

where $\tilde{\mathbf{y}} = F(\mathbf{a}_0, \mathbf{b}_0; \mathbf{g}) - \mathbf{y}$. If we write

$$\mathbf{A} = \begin{bmatrix} C_{\mathrm{noise}} J_{a,b} \\ C_{\mathrm{prior}} \end{bmatrix}, \qquad \mathbf{z} = \begin{bmatrix} C_{\mathrm{noise}} \tilde{\mathbf{y}} \\ 0 \end{bmatrix},$$

the optimization problem can be written in the simple form

$$\mathbf{d}_* = \underset{\mathbf{d}\in\mathbb{R}^{2L}}{\arg\min} \|\mathbf{A}\mathbf{d} - \mathbf{z}\|^2 . \tag{26}$$

This linear least-squares problem can be solved efficiently using, e.g., the built-in MATLAB backslash operator.

5 Numerical experiments

The parameter values used in the numerical experiments are listed in Table 1.

5.1 Optimizing heating patterns

The first step in the numerical experiments is forming the Jacobian matrices for the linearization as explained in Section 4.2. Solving the $8(K+1) \times 80$ forward problems required in total for building the Jacobians took roughly 30 minutes on a consumer-grade computer with a four core processor. The finite element mesh used in the forward solver (shown in Figure 4) has $N_p = 4189$ nodes, $N_t = 7992$ elements and $N_e = 12180$ edges. The mesh size used for the time stepping was $\delta t = T/60 = 1/30$. The magnitude of the perturbance was $\varepsilon = 0.01$.

An example of the Jacobian matrix $(J_{a,b})_j$ for one of the heaters H_j is shown in Figure 5. The rows correspond to the measurement sensors and measurement times: the first eight rows correspond to the first measurement time, rows 9 to 16 to the second measurement time, and so on. The columns indicate the pixel values for the parameters: the first 40 columns are the thermal conductivities **a** and the next 40 are the heat capacities **b**. The Jacobian gives insight to how changing parameter vectors affect the measurements. The change is smaller in later measurements, which is intuitively reasonable since for them the heat has more time to dissipate evenly. Also, the largest changes are caused by changes in the pixels directly on the heaters, whereas those close to the center of the object have a smaller effect.

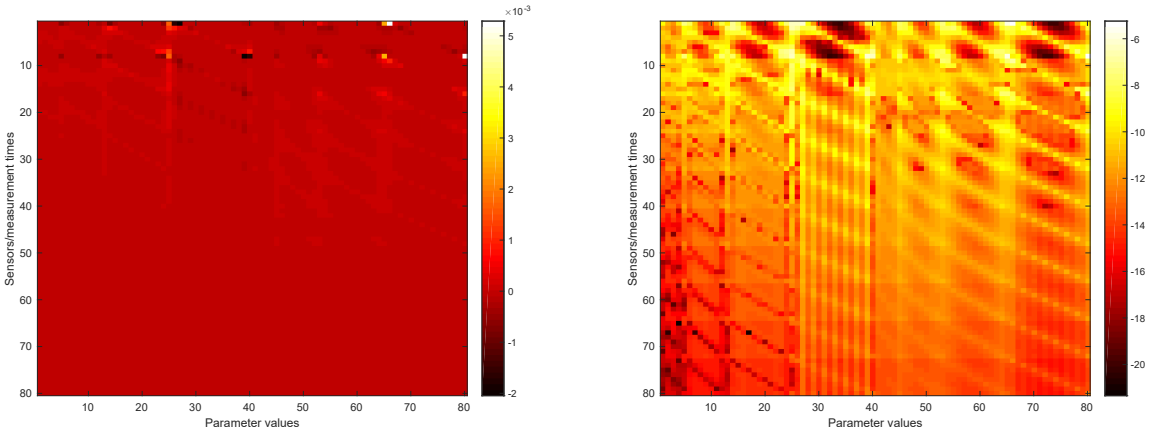

Figure 5: Absolute values and logarithm of absolute values of the Jacobian elements.

Next, we form the functions to be minimized/maximized, that is, the trace of the posterior covariance and the logarithm of the determinant of its inverse. Also the gradients given by formulas (22) and (23) must be computed. The initial guess for the optimization algorithm is the same that is eventually used as a reference solution when comparing the reconstructions. It is a simple linear heating from 0 at time $t = 0$ to 1 at time $t = T$ shown in Figure 6. The presented results correspond to the case of simultaneous optimization of all 8 heating patterns.

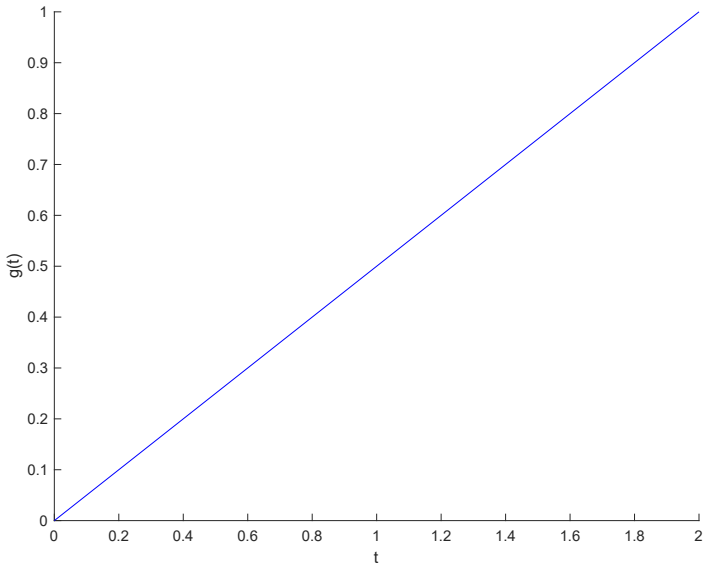

Figure 6: Reference heating pattern.

The initial objective function values accounting only for the prior information are:

$$\begin{aligned} \log\big(\det(\Gamma_{\text{prior}})\big) &= -329.4867, \\ \operatorname{tr}\big(\Gamma_{\text{prior}}\big) &= 80. \end{aligned}$$

After incorporating the measurement data corresponding to the reference heating pattern shown in Figure 6, the values become:

$$\begin{aligned} \log\big(\det(\Gamma_{\text{post}}(\mathbf{g}^{(0)}))\big) &= -365.1262, \\ \operatorname{tr}\big(\Gamma_{\text{post}}(\mathbf{g}^{(0)})\big) &= 16.8010. \end{aligned} \tag{27}$$

Here $\mathbf{g}^{(0)}$ denotes the reference heating pattern. Correspondingly we will use $\mathbf{g}^A$ and $\mathbf{g}^D$ for the heating patterns optimized using the A- and D-optimality criteria respectively. As expected, the values of the targets improve, that is, decrease, since the measurement with the reference heating gives information about the parameters.

The actual optimization is performed using MATLAB's `fmincon` function using the sequential quadratic programming algorithm and with the gradient of the objective function specified by the user. MATLAB also provides other optimization algorithms but they resulted in almost the same patterns.

As an initial condition, the heating pattern is set to be zero at time $t = 0$. Upper and lower bounds are set such that the heating stays in the interval $[-3, 3]$. The first and second derivatives are also given maximum and minimum values to prevent sharp, jagged patterns and fast oscillations, which can be considered nonphysical. All the values of the parameters for these constraints, as well as other parameters used for the simulations, are shown in Table 1.

Table 1: Values of parameters used.

a_0	4	Base level of thermal conductivity
b_0	4	Base level of heat capacity
c_{heater}	20	Surface heat conductance at the heaters
$c_{\text{nonheater}}$	0.05	Surface heat conductance between the heaters
H	8	Number of heaters
s	8	Number of sensors
T	2	End time of measurement
M	10	Measurement points in time
K	50	Heating basis functions
σ	1	Standard deviation of unknown parameters
γ	0.01	Standard deviation of noise
l	0.5	Correlation length
$\max \partial_t g$	1	Constraints for heating
$\min \partial_t g$	-1	
$\max \partial_{tt} g$	5	
$\min \partial_{tt} g$	-5	
$\max g$	3	
$\min g$	-3	
N	1000	Number of samples
N_p	4189	Points in FEM mesh
N_e	12180	Edges in FEM mesh
N_t	7992	Elements in FEM mesh

The optimization itself is very fast, taking only seconds to complete. Without the gradient specified, the optimization process takes slighly longer, but is still completed in under a minute. After running the optimization algorithm, the values of the objective function are

$$\begin{aligned} \log\big(\det(\Gamma_{\text{post}}(\mathbf{g}^D))\big) &= -387.4964, \\ \operatorname{tr}\big(\Gamma_{\text{post}}(\mathbf{g}^A)\big) &= 9.4027. \end{aligned} \tag{28}$$

Comparing these to corresponding values for the unoptimized patterns in (27), we can see that there is a clear difference. This suggests that the optimization should provide clear benefits for the inference of the unknown parameters. The patterns can also be cross evaluated with regards to the objective functions:

$$\begin{aligned} \log\big(\det(\Gamma_{\text{post}}(\mathbf{g}^A))\big) &= -384.5229, \\ \operatorname{tr}\big(\Gamma_{\text{post}}(\mathbf{g}^D)\big) &= 10.0304. \end{aligned} \tag{29}$$

While the values are slightly worse than in (28), they are still significantly better than the unoptimized values of (27).

The heating patterns corresponding to the two optimization criteria are shown in Figure 7. Initially both of the patterns increase linearly, with the A-optimal

solution decreasing near the end of the time interval $[0, 1]$. It seems that the optimal strategy is to increase the temperature of the heater as quickly as possible given the constraints. There is a slight difference in the heating patterns resulting from the two optimality criteria. However, since at the later times the heat does not anymore have much time to dissipate through the object, the effect on the actual result, i.e. the values of the target functions, is small; see (28) and (29).

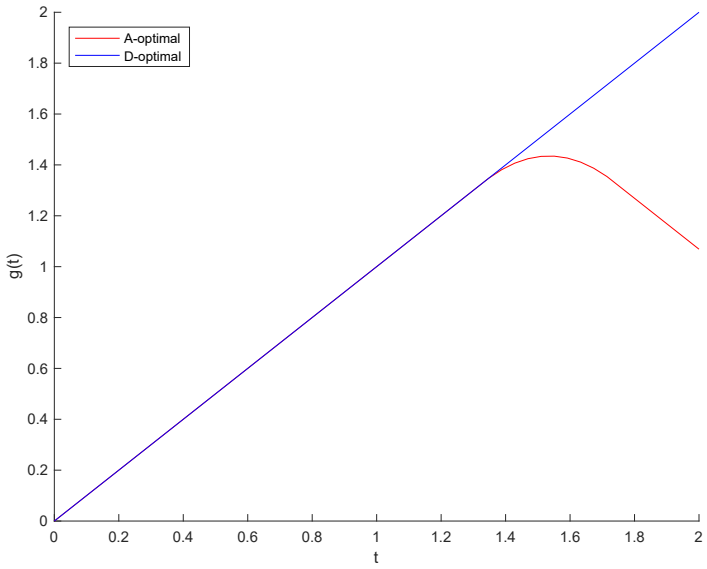

Figure 7: Optimized heating patterns. Blue curve corresponds to D-optimality and red to A-optimality.

5.2 Reconstruction of parameters

To evaluate the accuracy of reconstructions corresponding to different heating patterns, the first step is to draw samples from the prior distributions for the material parameters. The number of samples is $N = 1000$ as also listed in Table 1. As indicated by (8), the parameters are distributed according to a normal distribution, with the means $\mathbf{a}_0$ and $\mathbf{b}_0$, and the covariance defined by (9). The covariance matrix is computed using parameters $\sigma = 1$ and $l = 0.5$, which are the same as in Section 5.1 where the heating patterns were optimized. Figure 8 shows a visualization of the covariance matrix revealing its structure.

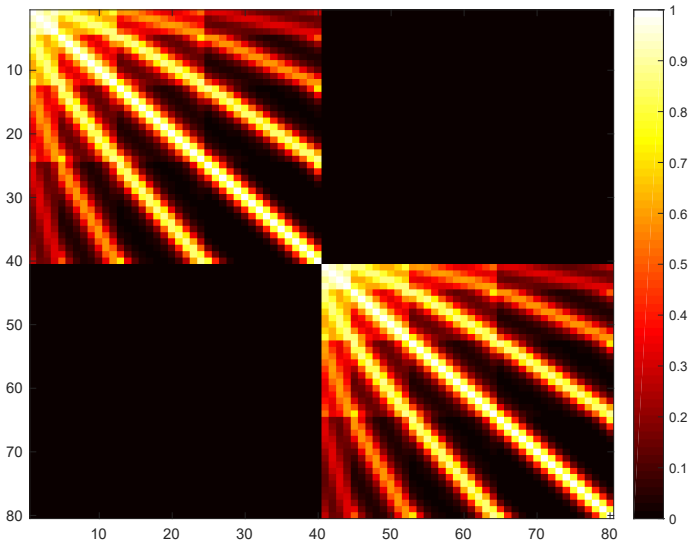

Figure 8: Prior covariance of parameter vector $[\mathbf{a}^T, \mathbf{b}^T]^T$. The axes correspond to the numbering of pixels in Figure 2.

Drawing the $N = 1000$ samples is performed with the built-in MATLAB function `mvnrnd`. Since negative parameter values would be nonphysical, all values smaller than 0.01 are rounded up to 0.01. A visualization of a random sample is shown in Figure 9. Independently and identically distributed Gaussian noise with zero mean and standard deviation $\gamma = 0.01$ is added to the measurements; note that this noise model is also the same as in Section 5.1.

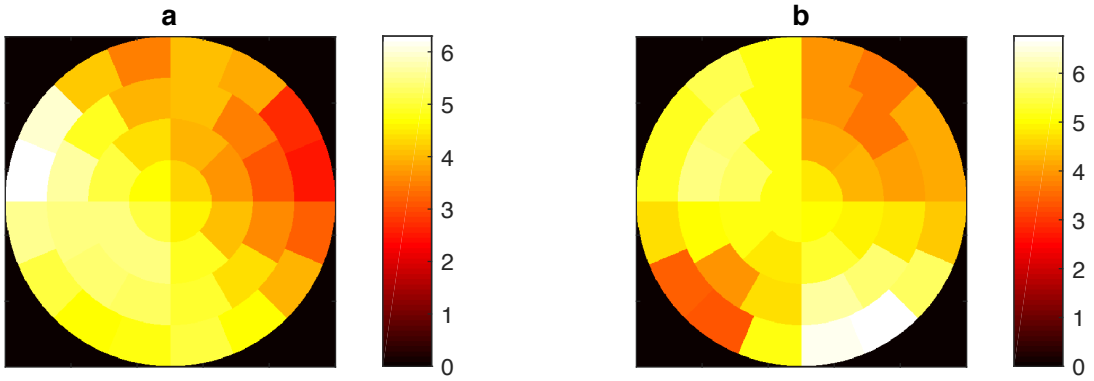

Figure 9: Sample parameters drawn from the prior distribution.

For each sample, a forward solution is computed using both of the optimized heating patterns, as well as the reference pattern. Solving the forward problems took roughly 20 minutes on a consumer grade computer. Figure 10 shows a set of simulated (non-noisy) measurement data corresponding to one of the samples and the reference heating pattern.

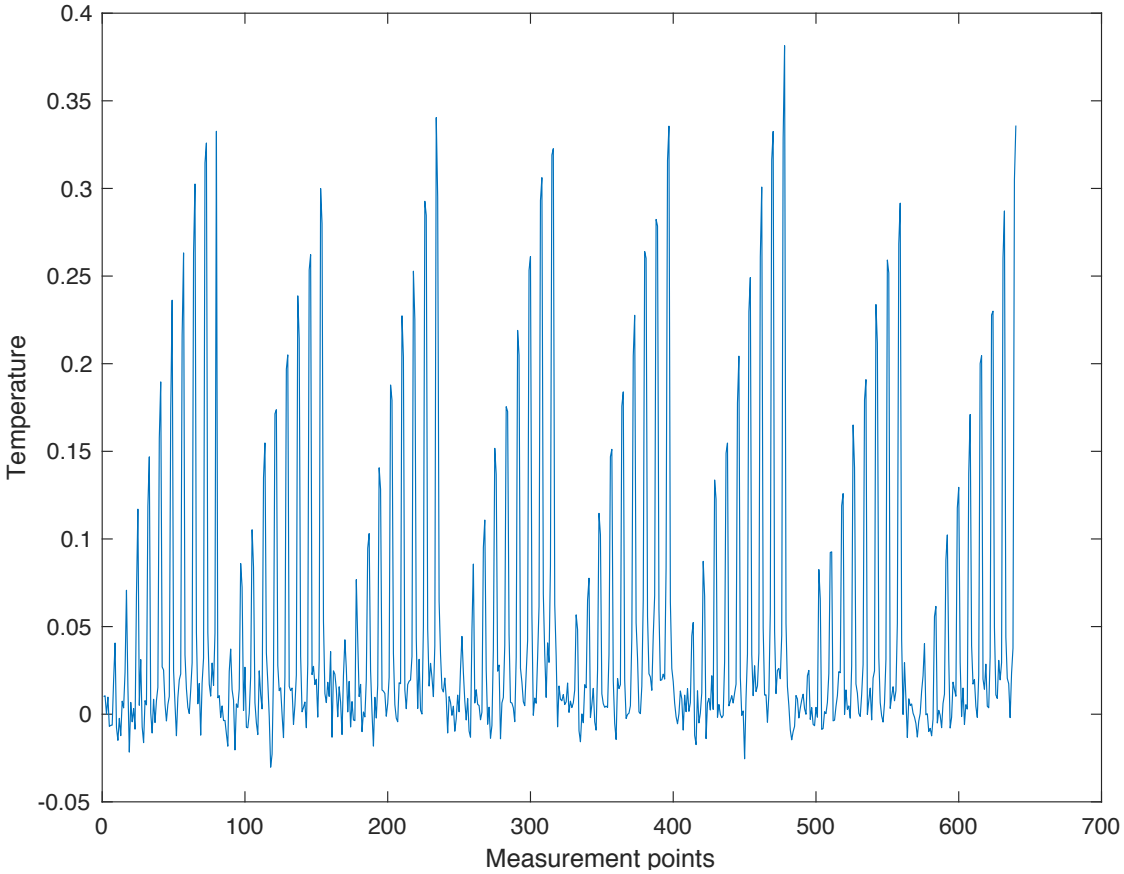

Figure 10: Example of simulated measurement data. The indices for the horizontal axis are obtained by concatenating together the measurement vectors for all 8 heaters.

The reconstruction for each of the $N = 1000$ data sets is computed according to (26). Finally, the last step is to calculate the average squared error for each of the heating patterns,

$$E = \frac{1}{N} \sum_{j=1}^{N} \|(\hat{\mathbf{a}}^{(j)}, \hat{\mathbf{b}}^{(j)}) - (\mathbf{a}^{(j)}, \mathbf{b}^{(j)})\|^2, \tag{30}$$

where $\hat{\mathbf{a}}, \hat{\mathbf{b}}$ are the reconstructed parameters. Note that this approximately corresponds to the squared $L^2(\Omega)$-error because the pixels shown in Figure 2 are approximately of the same size.

The resulting squared average errors for the chosen parameters are given in Table 2.

Table 2: Average reconstruction errors.

Pattern	E
Reference pattern	18.7067
Optimized pattern using trace condition (A-optimality)	12.0336
Optimized pattern using determinant condition (D-optimality)	12.2064

The results show that the optimized patterns do indeed perform better than the reference pattern with a significant improvement on the average reconstruction error. The superior performance of A-optimality was to be expected since the utility function of A-optimality, as shown in (15), is the same squared error that was used to evaluate the reconstructions.

Figure 11 shows the vectorized reconstructions for a single sample for the different heating patterns. Since we have a large number of measurements, the reconstructions are all fairly good. The reconstructions corresponding to the sample in Figure 9 are visualized for the base pattern in Figure 12 and for the optimized patterns in Figures 13 and 14, and their squared errors are listed in Table 3. All the reconstructions clearly identify the region of lower thermal conductance in the upper right quadrant of the object, as well as the areas with higher heat capacity in the upper right and lower left quadrants.

Table 3: Squared error for the example reconstructions in Figure 11.

Pattern	E
Reference pattern	14.1233
Optimized pattern using trace condition (A-optimality)	8.9279
Optimized pattern using determinant condition (D-optimality)	10.4938

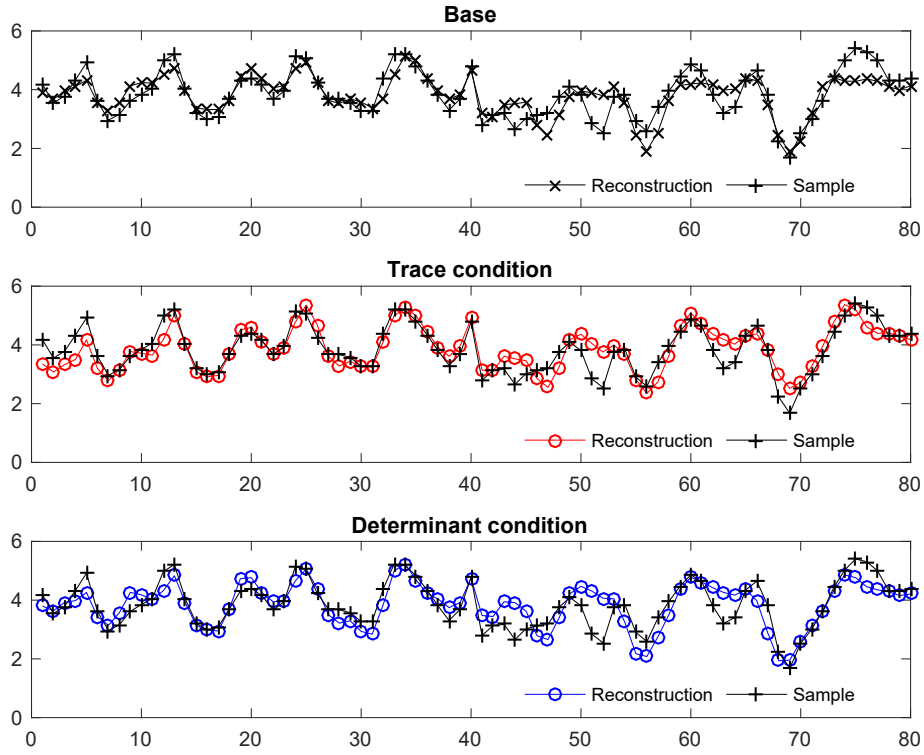

Figure 11: Reconstructions for a sample using each of the heating patterns. The first 40 values correspond to **a** and the last 40 to **b**. The numbering of pixel values is as in Figure 2.

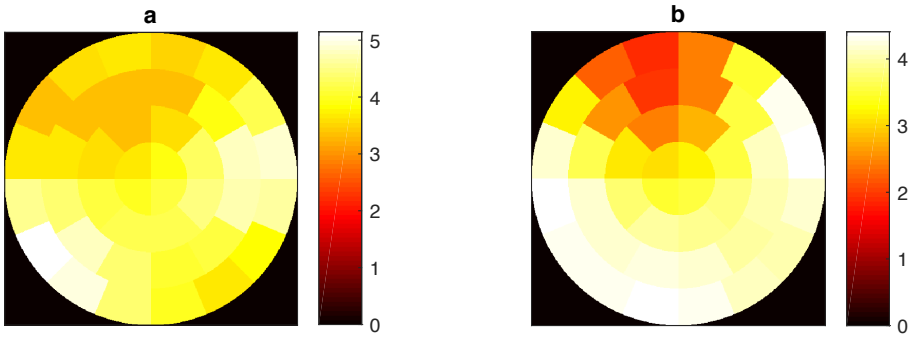

Figure 12: Reconstruction for the samples in Figure 9 using the reference heating pattern.

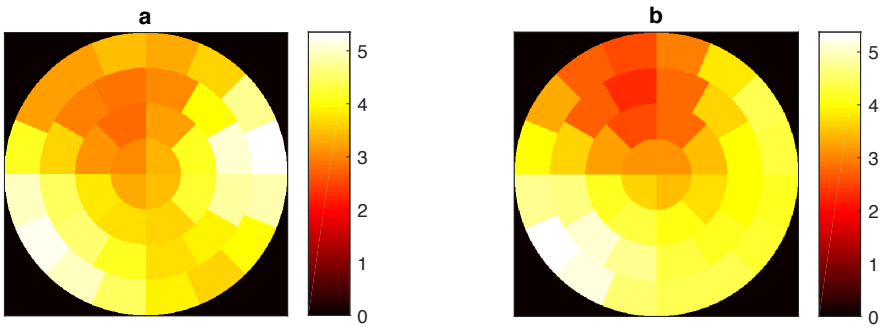

Figure 13: Reconstruction for the samples in Figure 9 using the optimized heating pattern (trace condition).

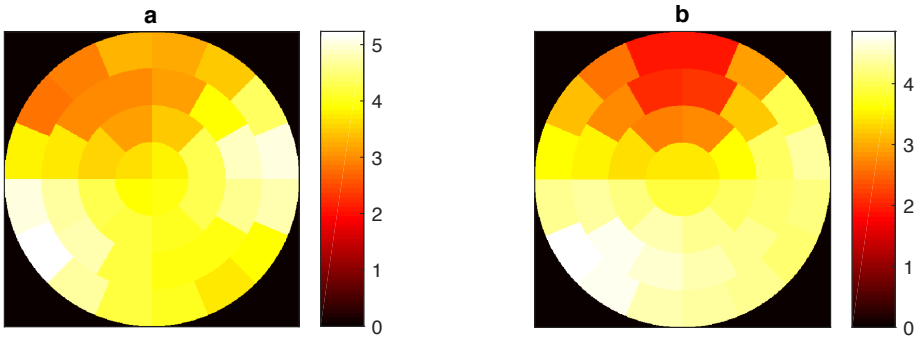

Figure 14: Reconstruction for the samples in Figure 9 using the optimized heating pattern (determinant condition).

The reconstruction errors (30) for the first 50 samples are shown in Figure 15. The reconstructions seem to be fairly consistent with no huge outliers. However, even in this figure we can clearly see that all the largest reconstruction errors are produced by the reference pattern.

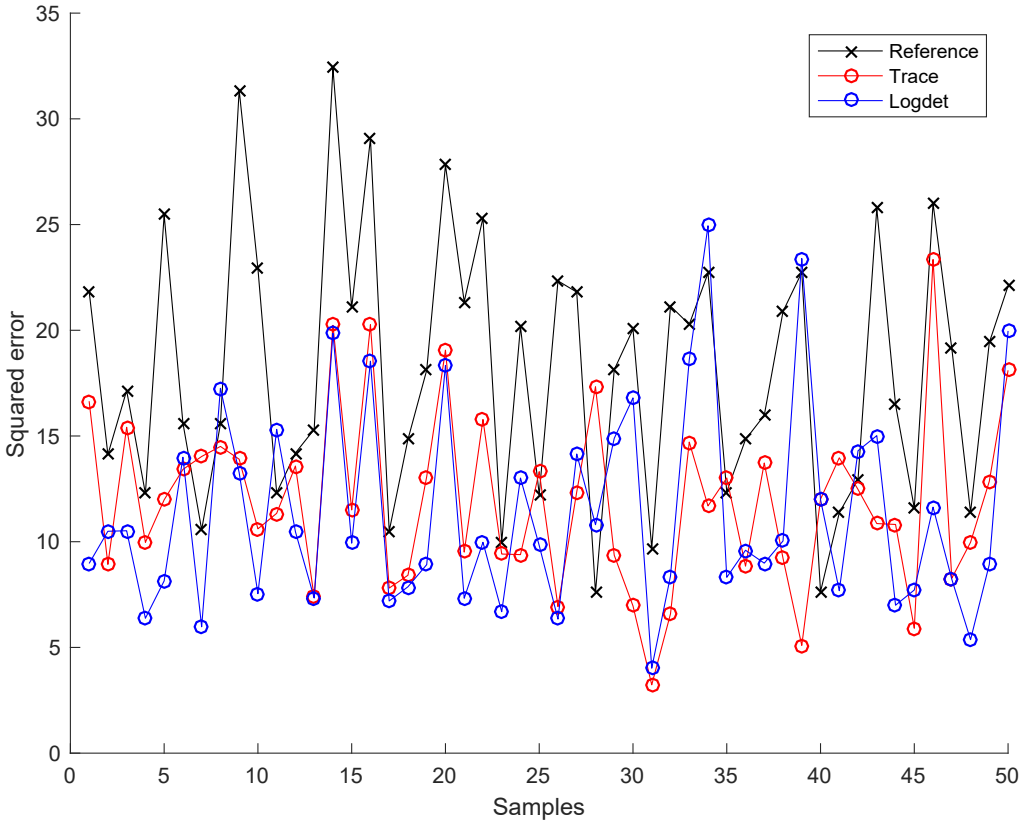

Figure 15: Reconstruction errors for the first 50 samples.

Figure 16 shows the pixelwise mean values over all samples and reconstructions. There seems to be a tendency for the reconstruction algorithm to underestimate the thermal conductivity, especially near the edges of the object, and conversely overestimate the heat capacity, especially near the center. It is worth noting that deviation of the sample average (the plus signs) from $a_0 = b_0 = 4$ is a random property of the drawn sample and is expected to be of the order $1/\sqrt{N} \approx 0.03$. The figures 17, 18 and 19 show the same data in the measurement geometry, further illustrating these findings. Since all the parameters have the same average, ideally these images should be of uniform color, with deviation resulting from the randomness of the samples. Note here the different scaling of the colorbar compared to the previous reconstructions.

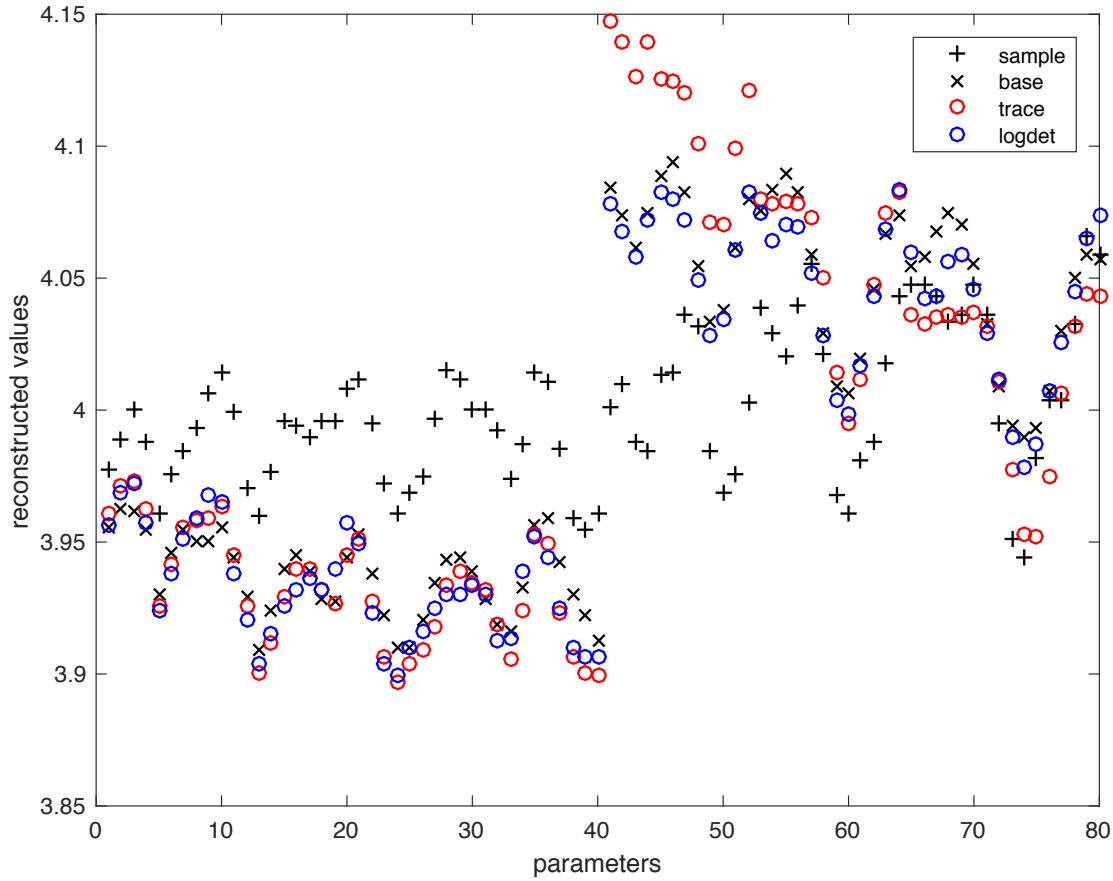

Figure 16: Average reconstruction. The numbering of the parameters is as in Figure 11.

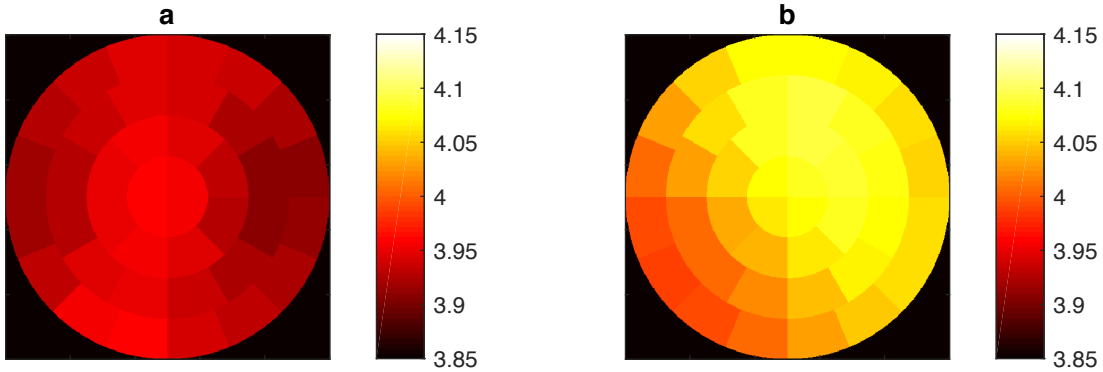

Figure 17: Average reconstruction with the reference heating pattern, shown in the measurement geometry.

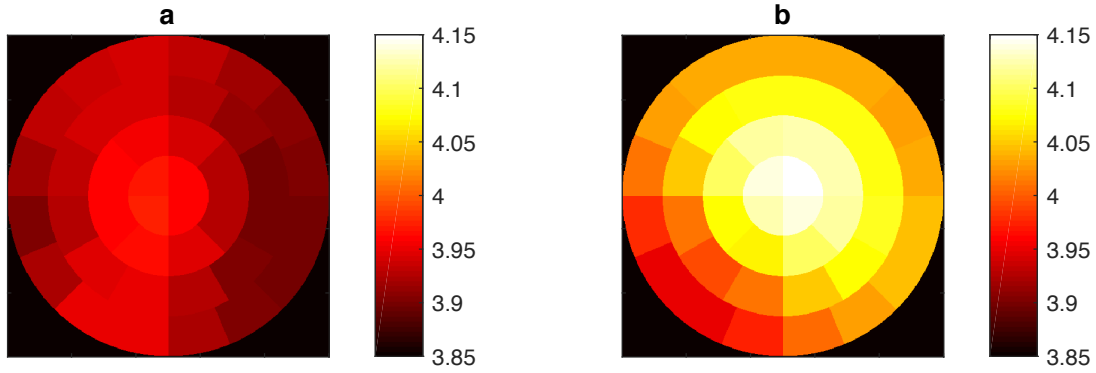

Figure 18: Average reconstruction with the optimized heating heating pattern (trace condition), shown in the measurement geometry.

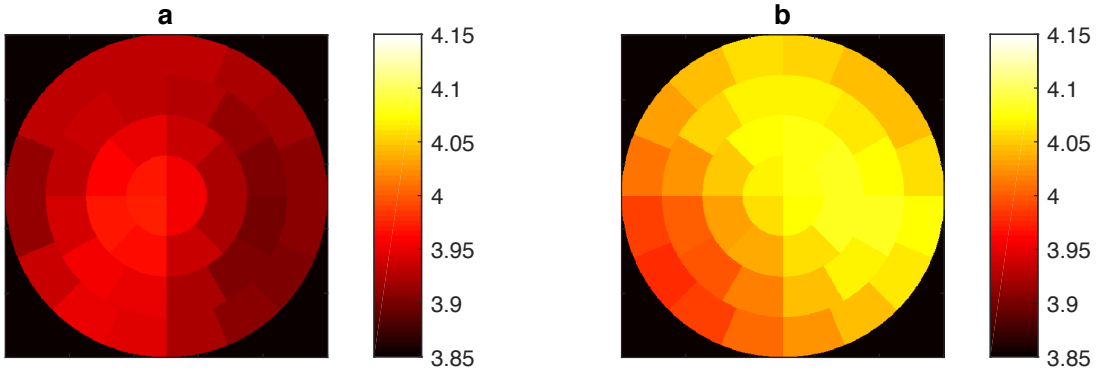

Figure 19: Average reconstruction with the optimized heating pattern (determinant condition), shown in the measurement geometry.

5.3 Optimization with alternate parameters

Since the optimized heating patterns from Section 5.2 were essentially maximal heating constrained by the first derivative, the improvements in reconstructions were probably mostly due to the greater heat flux providing a better signal-to-noise ratio (especially since the noise level is independent of the measurements). To test this hypothesis, the optimization and following reconstructions should be carried out with stricter constraints on the heating pattern.

Figure 20 shows the results of the optimization with the upper and lower bounds for the temperature of the heater set to 1, i.e., the same as the upper limit of the reference pattern. Both of the optimized patterns initially increase linearly, but with a steeper slope than the reference pattern. The results in Table 5 show that the optimized patterns still produce better reconstructions, but with the trace condition, i.e., A-optimality performing significantly worse than before. This would suggest that the optimality criteria favor designs with a large temperature gradient and not only maximal heat flux.

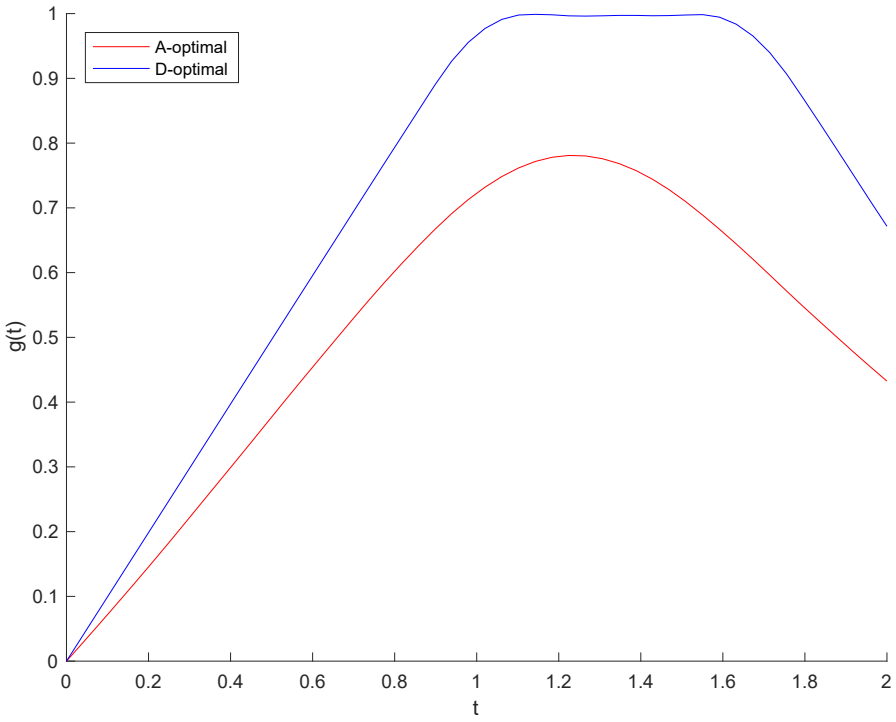

Figure 20: Optimized heating patterns with $\max|g| = 1$.

Table 4: Reconstruction errors for heating patterns with upper bound for the absolute value of g set to 1 instead of 3.

Pattern	E
Reference pattern	19.3493
Optimized pattern using trace condition	16.4756
Optimized patterns using determinant condition	13.8064

If the first derivative constraint is set to be 0.5, i.e. the same as the first derivative of the reference pattern, the optimization with either of the two objective functions results in the same pattern as the reference in Figure 6. Lowering the upper bound to 0.5 for the heating results in the heating patterns shown in Figure 21 and the mean reconstruction errors in Table 5. Now the total heating for the optimized patterns is less than that of the reference pattern. This time the optimized patterns also perform worse than the reference pattern, which is in accordance with our hypothesis about maximal heating gradient leading to the lowest reconstruction error.

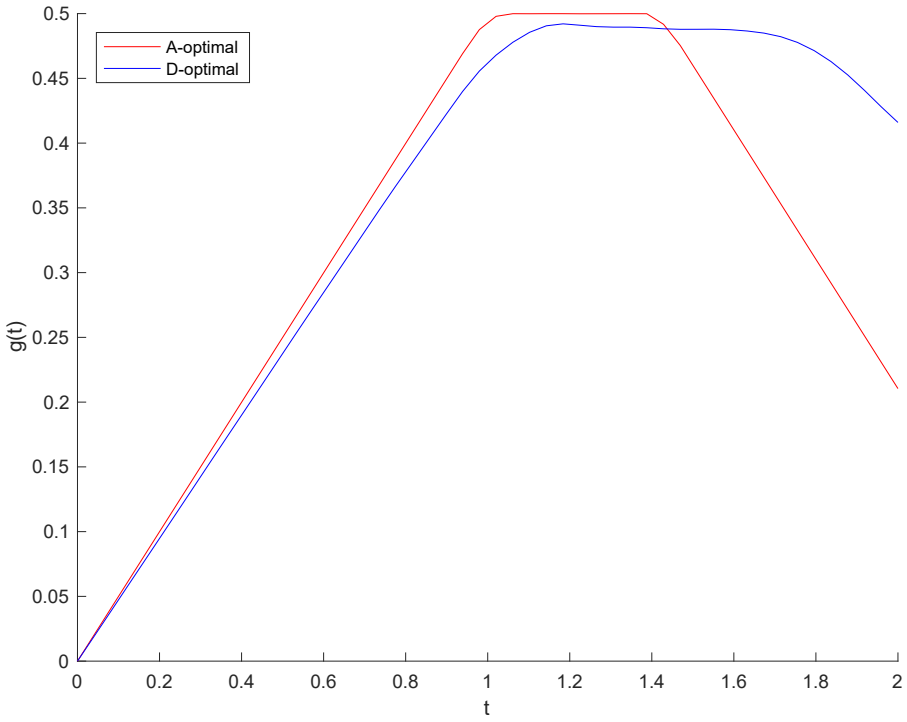

Figure 21: Optimized heating patterns with maximum values of $|g|$ and $|g_t|$ set to 0.5.

Table 5: Reconstructions errors for heating patterns with maximum values for $|g|$ and $|g_t|$ set to 0.5.

Pattern	E
Reference pattern	19.7499
Optimized pattern using trace condition	22.0242
Optimized patterns using determinant condition	23.4989

5.4 Optimization using alternate noise model

In the previous sections, the noise level was fixed beforehand and did not depend on the magnitude of the corresponding measurement. As could be expected, this resulted in optimized heating patterns that tend to the maximal values permitted by the constraints as quickly as possible. In this section, the optimization is performed using the noise model defined in Section 4.3. As shown in equation (24), the noise level is now set (almost) relative to the measurement. Since the magnitude of the measurements is typically less than 0.5, the variance parameter σ is set to 0.05 to make the level of noise similar to the component of noise that is independent of measurements considerd in previous subsections. The parameter controlling the component of noise that is independent of the measurements is set to $\varepsilon_{\text{noise}} = 10\text{e}{-6}$,

i.e. to an extremely low value. The constraints for the heating are the same as in the original experiment (cf. Table 1), with the first derivative being in the interval $[-1, 1]$, and the second derivative in the interval $[-5, 5]$. The initial guess, i.e. the reference pattern, is also the same as before (see Figure 6). With these initial parameters, the target function values are

$$\log\left(\det\left(\Gamma_{\text{post}}(\mathbf{g}^{(0)})\right)\right) = -447.9602,$$
$$\operatorname{tr}\left(\Gamma_{\text{post}}(\mathbf{g}^{(0)})\right) = 2.1570.$$

With the new noise model, evaluating the target function is computationally more intensive. Performing the optimization with the gradient specified by the user takes now roughly a minute for each condition. The target function values are

$$\log\left(\det\left(\Gamma_{\text{post}}(\mathbf{g}^{D})\right)\right) = -450.4670,$$
$$\operatorname{tr}\left(\Gamma_{\text{post}}(\mathbf{g}^{A})\right) = 1.9752.$$

The optimized heating patterns cross-evaluated with regards to their respective target functions values result in

$$\log\left(\det\left(\Gamma_{\text{post}}(\mathbf{g}^{(A)})\right)\right) = -450.5318,$$
$$\operatorname{tr}\left(\Gamma_{\text{post}}(\mathbf{g}^{(D)})\right) = 2.0093.$$

The resulting patterns are shown in Figure 21. Both of the patterns initially increase linearly, with the A-optimal solution leveling out at around time $t = 0.8$. The slight oscillation in the pattern is likely caused by numerical errors. Unlike for the noise model in Section 5.1, this time also the D-optimal heating pattern starts to decrease but only at time $t = 1.4$.

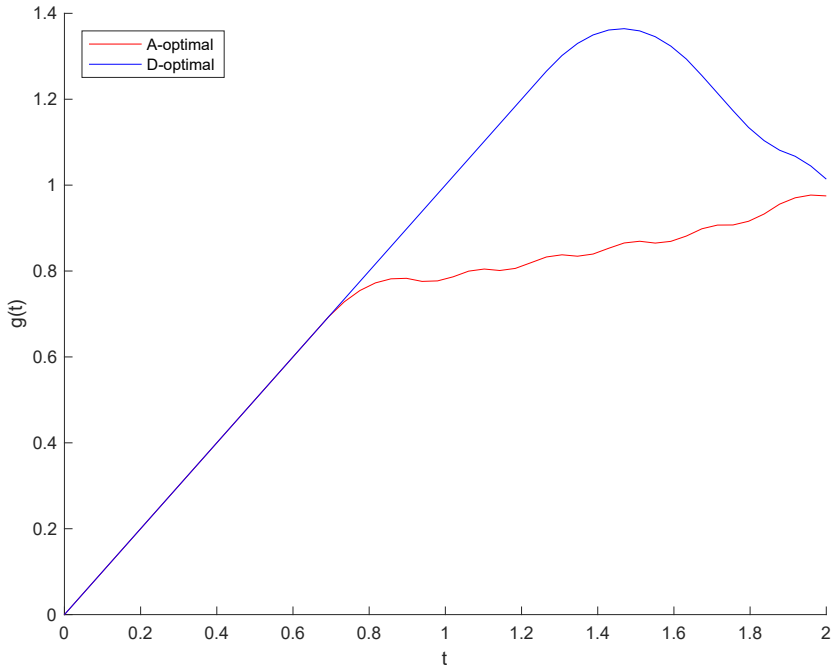

Figure 22: Optimized heating patterns with the alternative noise model.

Experimenting with the constraints yields similar results as previously. Setting the upper bound to 1 causes the D-optimal solution to first increase linearly until the upper bound is reached, and then stay there (almost) until the end of the heating (Figure 23). This is different from the designs of the previous section, where the optimal solutions seemed to favor large gradients in the heating and not just maximal heat flux. The A-optimal solution was not constrained by the bound so it does not change from the previous test. The target function values for these parameters are

$$\log\left(\det\left(\Gamma_{\text{post}}(\mathbf{g}^D)\right)\right) = -450.5383,$$
$$\operatorname{tr}\left(\Gamma_{\text{post}}(\mathbf{g}^A)\right) = 1.9752.$$

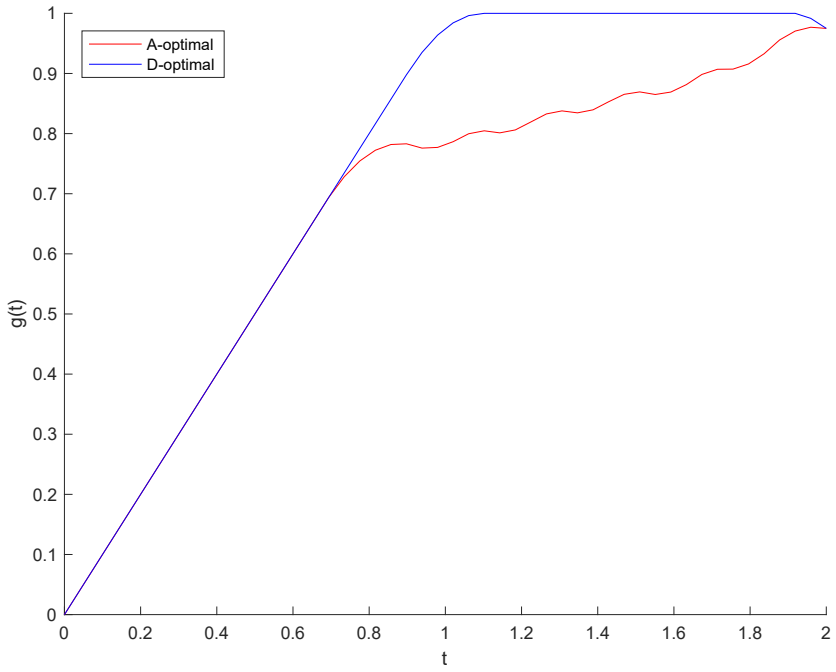

Figure 23: Optimized heating patterns with the alternative noise model and upper bound 1 for $|g|$.

Setting the bound for the absolute value of the first derivative to be the same as in the reference pattern, that is 0.5, yields nearly identical patterns for both of the conditions with both initially increasing linearly (Figure 24). Near the end of the time interval, the A-optimal solution decreases but as previously mentioned, this has very little impact on the measurements. The target function values for these parameters are

$$\log\left(\det\left(\Gamma_{\text{post}}(\mathbf{g}^D)\right)\right) = -447.9602,$$
$$\operatorname{tr}\left(\Gamma_{\text{post}}(\mathbf{g}^A)\right) = 2.1531.$$

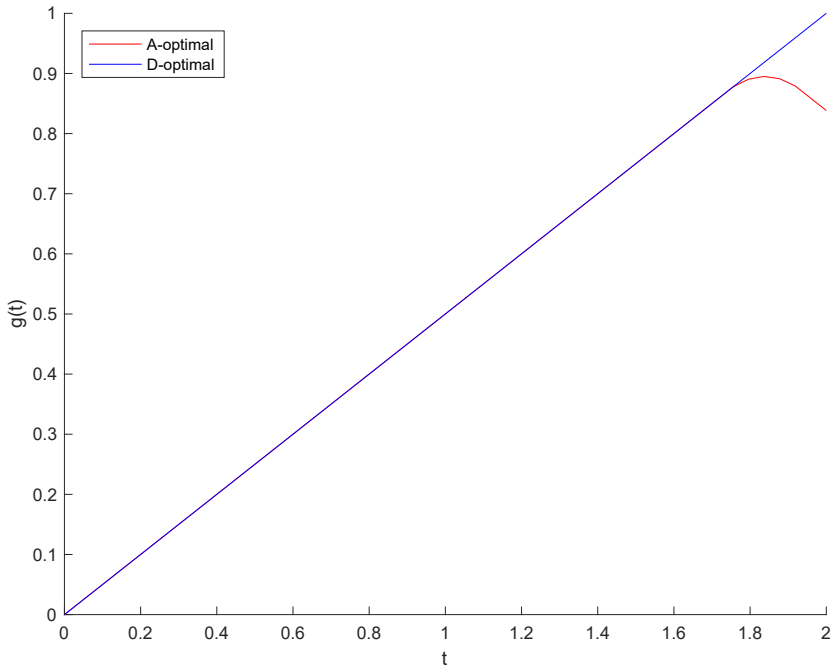

Figure 24: Optimized heating patterns with the alternative noise model and bound 0.5 for $|g_t|$.

With the upper bound for $|g|$ set to 0.5 and for $|g_t|$ to 0.5, both solutions increase linearly until the upper bound is reached, and stay there almost until the end of the heating (Figure 25). The target function values are

$$\log\left(\det\left(\Gamma_{\text{post}}(\mathbf{g}^D)\right)\right) = -448.2688,$$
$$\operatorname{tr}\left(\Gamma_{\text{post}}(\mathbf{g}^A)\right) = 2.0822.$$

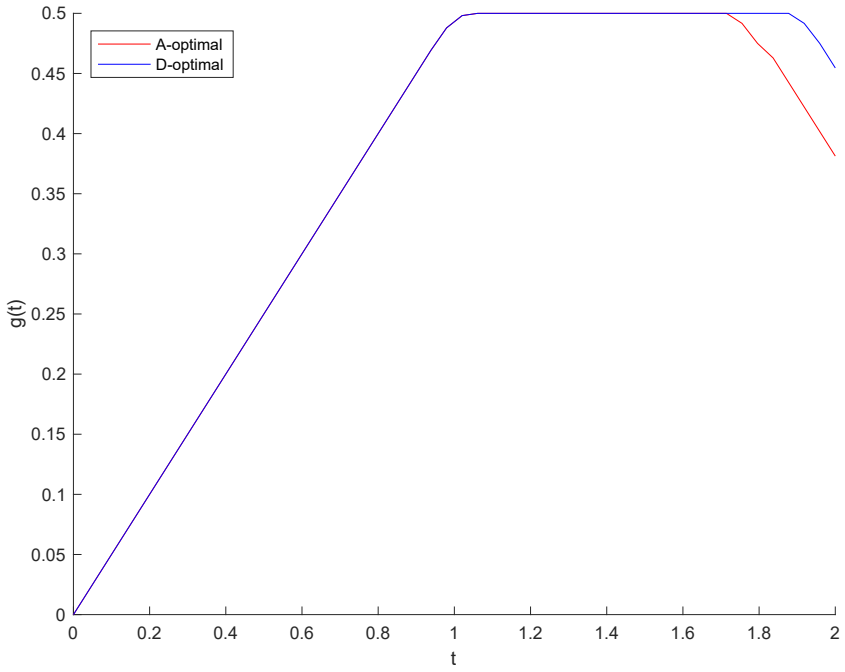

Figure 25: Optimized heating patterns with the alternative noise model and bound 0.5 for both $|g|$ and $|g_t|$.

In contrast to computing reconstructions with the independent noise model, computing reconstructions with the alternate noise model suffers from huge variability in the quality of the reconstructions. As can be seen in Figure 26, a small fraction of the reconstructions are significant outliers with regards to the squared error, and therefore throw off the average error for the samples. The average errors, listed in Table 6, are an order of magnitude larger than with the independent noise model. Also, this time the reference pattern performed on average better than the optimized patterns, possibly due to the smaller magnitude of heating.

Table 6: Reconstruction errors for heating patterns of Figure 22 with $\varepsilon_{\text{noise}} = 10\text{e}{-6}$.

Pattern	E
Reference pattern	206.2594
Optimized pattern using trace condition	265.4402
Optimized patterns using determinant condition	294.5021

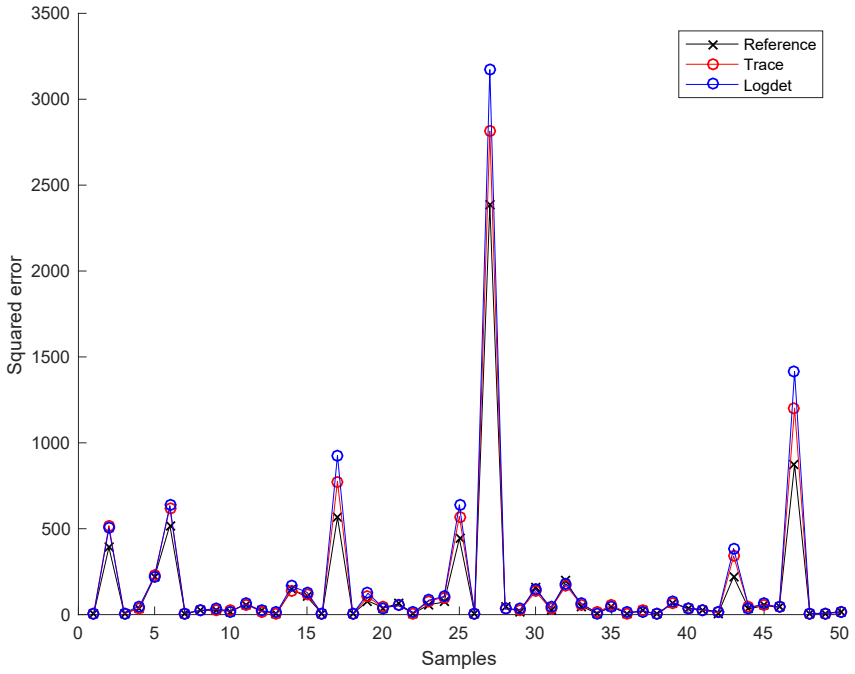

Figure 26: Errors for first 100 reconstructions with heating patterns of Figure 22, with $\varepsilon_{\text{noise}} = 10\text{e}{-6}$.

This variability is most likely caused by numerical errors within the computation. For measurement times and locations where the magnitude of the measurement is small, the small value chosen for the independent noise results in small values in the noise covariance matrix, and conversely large values in the inverse. This could lead to under- and overflow problems. Figures 27 and 28 show an example of a reconstruction with a large error. In this example, the reconstruction of the thermal conductivity has roughly correct form, i.e., the areas with lower thermal conductivity correspond to those in the sample, but the values are off by a constant factor. The values of the heat capacity seemingly bear no resemblance to the sample.

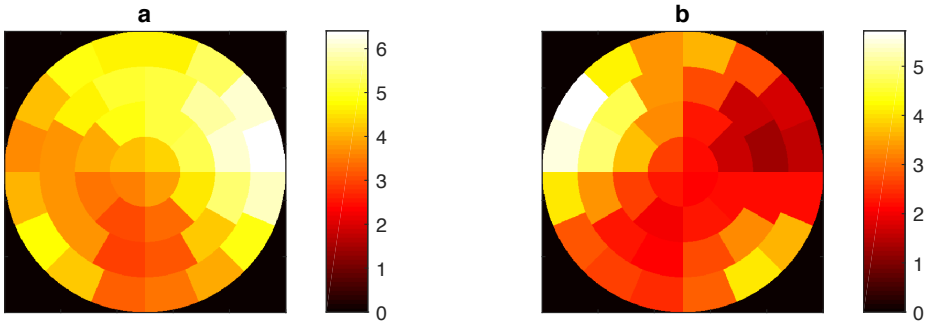

Figure 27: Sample of parameters from the prior distribution.

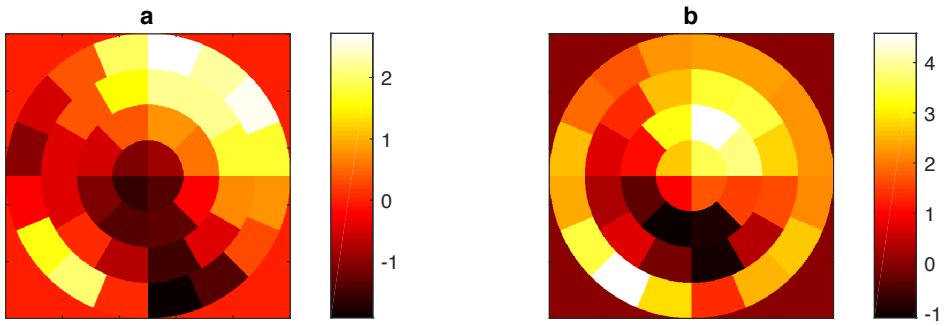

Figure 28: Reconstruction corresponding to sample in Figure 27 with reference heating pattern.

Also the average reconstructions, shown in Figures 29, 30, and 31, are very different from those in Figures 17, 18, and 19. It can be seen that the reconstruction algorithm now has a tendency to dramatically underestimate the thermal conductivity in the center of the object, as well as the heat capacity near, but not on the boundary. The ring structure seen here is also a lot unlike the samples drawn from the prior, which have smoothly varying areas of higher and lower thermal conductivity and heat capacity.

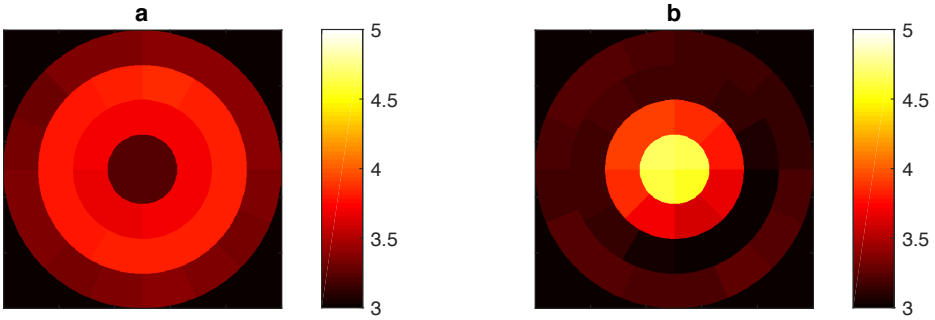

Figure 29: Average reconstruction with reference pattern and measurement dependent noise model.

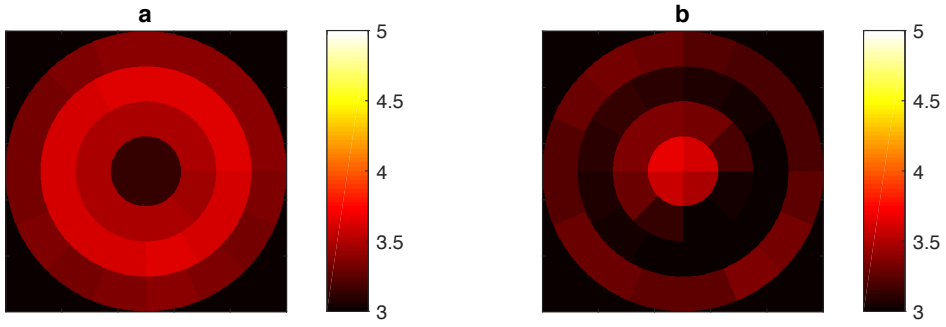

Figure 30: Average reconstruction with the A-optimal pattern in Figure 22 and measurement dependent noise model.

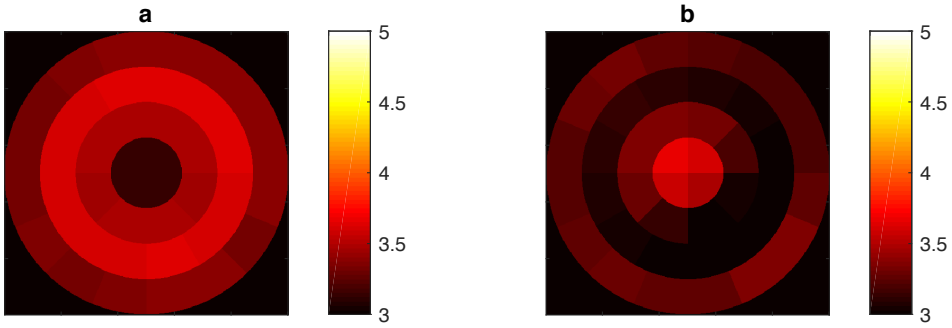

Figure 31: Average reconstruction with the D-optimal pattern in Figure 22 and measurement dependent noise model.

To test the hypothesis that the errors are caused by the small level of background noise, we increase the value of the parameter controlling the independent component of the noise to $\varepsilon_{\text{noise}} = 0.01$ and rerun the optimization and inversion computations. The errors, listed in Table 7, are reduced to a similar magnitude as with the measurement-independent noise model.

Table 7: Reconstruction errors for heating patterns of Figure 22 with $\varepsilon_{\text{noise}} = 0.01$.

Pattern	E
Reference pattern	10.6091
Optimized pattern using trace condition	11.8048
Optimized patterns using determinant condition	11.8674

Figure 32 shows that there are still outliers with significantly higher than average reconstruction errors, but they are less frequent and the errors are also smaller in magnitude.

It is quite unexpected that this time the reference pattern produces the lowest average reconstruction error. Investigation into the cause of this phenomenon should be included in any further work on the subject.

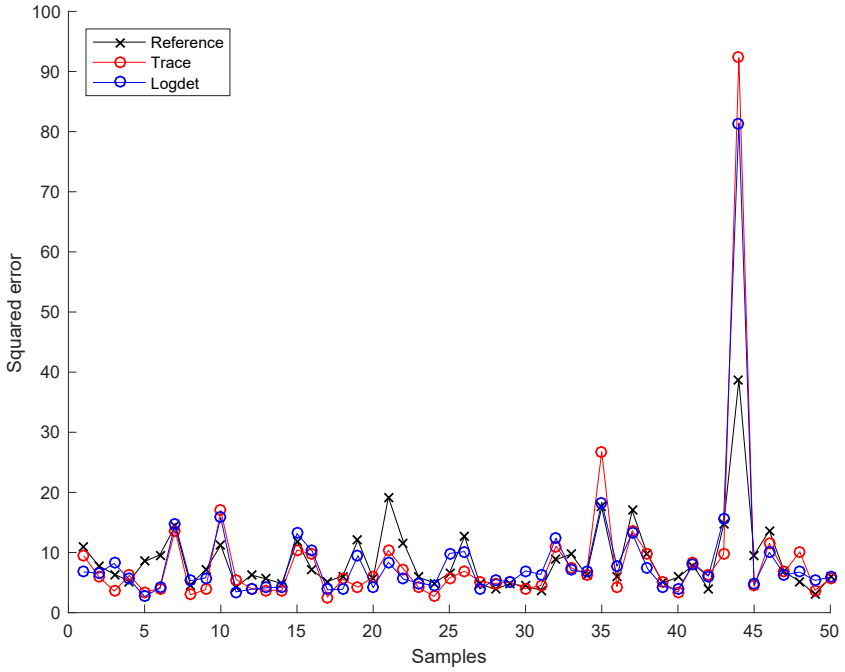

Figure 32: Reconstruction errors for heating patterns of Figure 22 with $\varepsilon_{\text{noise}} = 0.01$.

6 Concluding remarks

This thesis investigated the feasibility of applying optimal experimental design techniques to thermal tomography, and more specifically to the selection of the heating patterns. Optimizing the heating patterns requires no change to the experiment apparatus and therefore could provide better results with very little additional effort and no additional cost, with the only constraints being the physical limitations of the heaters.

The experiment was modeled as a heat diffusion problem in two dimensions, with heaters and sensors attached to the boundary of the object. The object itself was a unit disk with the thermal conductivity and heat capacity discretized into a finite number of pixels. The measurement data was obtained by numerically simulating the experiment using the finite element method. The inverse problem was solved using Bayesian techniques with a single linearization step to make the problem computationally feasible.

Two optimization criteria were considered, corresponding to Bayesian A- and D-optimality. The goal was to make the posterior distribution as localized as possible, with the degree of localization measured by the trace or the determinant of the covariance matrix of the posterior distribution. If heating with the optimized pattern does provide more information about the object, this should result in better reconstructions of the unknown parameters. This was tested by drawing random samples from the prior, simulating the measurement data, computing the reconstructions by solving the inverse problem, and finally computing the average errors for each of the reconstructions.

The results of the optimization showed that the optimized designs mostly tended to first increase the heating within the constraint on the first derivative. If the highest allowable temperature were reached, the heating would not stay at this upper bound, but began decreasing. Relaxing the constraints for the first and second derivative of the heating pattern tended to result in designs with more oscillations. Both optimization targets resulted in same or similar designs.

Performing the reconstructions using the optimized heating patterns did indeed initially result in better reconstructions as theory would predict. The improvements were significant, with the average error decreasing over 30% relative to a reference pattern. This means that optimizing the heating patterns does indeed provide a benefit and could improve the performance of the method in a real world setting. However, when the experiment was repeated with stricter constraints for the heating, the optimized designs performed worse. It seems that the time derivative and the total amount of heat transferred are important factors in determining the amount of information gained by thermal tomography: better reconstructions can presumably be obtained by simply increasing the heat flux applied to the object.

With a fixed noise level it is not surprising that the optimized designs favoured maximal heating since it would provide the best possible signal-to-noise ratio. However, setting the noise level to be (linearly) dependent on the magnitude of the measurements still resulted in similar optimized patterns as with fixed noise, which is not an obvious outcome.

To progress further towards an actual application of thermal tomography, this study should be extended into three dimensions and one should also consider physically realistic parameters. Especially investigation into what are the physically realistic constraints for actual heaters would be an important further question, since the constraints of the optimization problem have a significant effect on the optimal heating designs. Another possibility would be to use optimal experimental design principles on other optimization targets, such as the positions of the heaters. In addition, one should investigate how the optimal designs affect nonlinear, iterative reconstruction algorithms that can be considered the state-of-the-art reconstruction methods for thermal tomography.

References

[1] Kaipio, J., Somersalo, E. *Statistical and Computational Inverse Problems*, Springer, New York, 2005.

[2] Hyvönen, N., Mustonen, L. *Thermal tomography with unknown boundary*, SIAM J. Sci. Comp., 40:B663-B683 (2018).

[3] Somersalo, E., Cheney, M., Isaacson, D. *Existence and uniqueness for electrode models for electric current computed tomography.* SIAM J. Appl. Math. 52:1023-1040 (1992).

[4] Uhlmann, G. *Electrical impendance tomography and Calderóns problem.* Inverse problems 25:123011 (2009).

[5] Arridge, S. R. *Optical tomography in medical imaging* Inverse problems 15, R41-R93 (1999).

[6] Huan, X., Marzouk, Y. *Simulation-based optimal Bayesian experimental design for nonlinear systems.* J. Comput. Phys. 232:288-317 (2013).

[7] Toivanen, J. *Thermal tomography*, Doctoral thesis, University of Eastern Finland, Kuopio, 2016.

[8] Bakirov, V. F., Kline, R. A. *Diffusion-based thermal tomography.* J. Heat Transfer 127:1276-1279 (2005).

[9] Kolehmainen, V., Kaipio, J. P., Orlande, H. R. B. *Reconstruction of thermal conductivity and heat capacity using a tomographic approach.* Int. J. Heat Mass Transfer 51:1866-1876 (2008).

[10] Toivanen, J. M., et al. *Simultaneous estimation of spatially distributed thermal conductivity, heat capacity and surface heat transfer coefficient in thermal tomography.* Int. J. Heat Mass Transfer 55:7958-7968 (2012).

[11] Toivanen, J. M., et al. *3D thermal tomography with experimental measurement data.* Int. J. Heat Mass Transfer 78:1126-1134 (2014).

[12] Johnson, C. *Numerical Solution of Partial Differential Equations by the Finite Element Method* Dover Books on Mathematics Series, Courier Corporation, 2012.

[13] Lions, J. L., Dautray, R. *Evolution Problems I. Mathematical Analysis and Numerical Methods for Science and Technology vol. 5.*, Springer-Verlag, 2000.

[14] Brenner, S., Ridgway, S. *The mathematical theory of finite element methods.* Vol. 15. Springer Science & Business Media, 2007.

[15] Crank, J., Nicolson, P. *A practical method for numerical evaluation of solutions of partial differential equations of the heat conduction type.* Proc. Camb. Phil. Soc. 43:50-67 (1947).

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

[17] Kaipio, J. P., Kolehmainen, V., Somersalo, E., Vauhkonen, M. *Statistical inversion and Monte Carlo sampling methods in electrical impendance tomography.* Inverse problems 16:1487-1522 (2000).

[18] Dardé, J., Hyvönen, N., Seppänen, A., Staboulis, S. *Simultaneous recovery of admittivity and body shape in electrical impendance tomography: an experimental evaluation.* Inverse Problems 28:085004 (2013).

[19] Chaloner, K., Isabella, V. *Bayesian experimental design: A review.* Statistical Science 10:273-304 (1995).

[20] Kaipio, J. P., et al. *Posterior covariance related optimal current patterns in electrical impedance tomography.* Inverse Problems 20:919-936 (2004).

[21] Hyvönen, N., Seppänen, A., Staboulis, S. *Optimizing electrode positions in electrical impedance tomography.* SIAM J. Appl. Math. 74:1831-1851 (2014).

[22] Petersen, K. B., Petersen, M. S. *The matrix cookbook.* Technical University of Denmark, vol. 7, 2008.

[23] Nocedal, J., Wright, S. *Numerical optimization*, Springer-Verlag, 1999.