Zuse Institute Berlin

Takustrasse 7
D-14195 Berlin-Dahlem
Germany

YVONNE FREYTAG

Optimal Experimental Design to Estimate the Time of Death in a Bayesian Context

ZIB Report 17-14 (March 2017)

Zuse Institute Berlin
Takustrasse 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Optimal Experimental Design to Estimate the Time of Death in a Bayesian Context

Yvonne Freytag

March 28, 2017

Abstract

This thesis is devoted to the interdisciplinary work between mathematicians and forensic experts: the modeling of the human body cooling process after death laying the foundation for the estimation of the time of death. An inverse problem needs to be solved. In this thesis the inverse problem computes the time of death given the measured body temperature and the Forward Model that simulates the body cooling process. The Forward Model is based on the heat equation established by Fourier. This differential equation is numerically solved by the discretization over space by the *Finite Element Method* and the discretization over time by the *Implicit Euler Method*. The applications in this thesis demand a fast computation time. A model reduction is achieved by the *Proper Orthogonal Decomposition* in combination with the *Galerkin Method*. For reasons of simplification the computations and the measurements are restricted to a cylindrical phantom that is made out of homogeneous polyethylene.
The estimate of the time of death is accompanied by an uncertainty. The inverse problem is incorporated by Bayesian inference to interpret the quality of the estimate and the efficiency of the experiment. The uncertainty of the estimate of the time of death is minimized by approaching the *Optimal Design of the Experiment*. An objective function measures the certainty of the data and lays the foundation of the optimization problem. Solving the optimization problem is successfully done by relaxing the complex discrete NP-hard problem and applying a gradient-based method.
The results of this thesis clearly show that the design of an experiment has a great influence on the outcome of the quality of the estimate. The comparison of the estimate and its properties based on different designs and conditions reveals the efficiency of the *Design of Experiment* in the context of the estimation of the time of death.

Zusammenfassung

Diese Masterarbeit widmet sich, im Rahmen eines interdisziplinären Forschungsprojekts von Mathematikern und Rechtsmedizinern, der Schätzung des Todeszeitpunktes. Das inverse Problem bestimmt ein Schätzer des Todeszeitpunktes basierend auf Messungen der Körpertemperatur an einer Leiche und der Simulation des Abkühlungsprozesses des menschlichen Körpers nach Eintritt des Todes. Die Wärmeleitungsgleichung nach Fourier dient als Vorwärtsmodel. Diese partielle Differentialgleichung wird numerisch mit der Ortsdiskretisierung durch die *Finiten–Elemente-Methode* und mit der Zeitdiskretisierung durch das *Implizite Euler-Verfahren* gelöst. Um den numerischen Aufwand gering zu halten, wird das Model mit Hilfe der *Proper Orthogonal Decomposition* mit *Galerkin-Ansatz* reduziert. Weiterhin werden als vereinfachte Grundlage die numerischen Rechnungen an einem Zylinder aus Polyethylen durchgeführt.
Das invserse Problem wird durch den Bayessche Ansatz gelöst, wobei dem Schätzer eine Unsicherheit beigemessen wird. Die Minimierung der Unsicherheit dient als Grundlage zur Bestimmung des *Optimalen Designs eines Experimentes*. Das daraus resultierende NP-schwere Optimierungsproblem wird relaxiert, so dass dies in dieser Arbeit durch das Gradientenverfahrens gelöst werden kann.
Es wird gezeigt, dass das Design eines Experimentes einen großen Einfluss auf die Qualität der Schätzung besitzt.

Contents

1 Introduction

1.1 Motivation

The estimation of an accurate time of death of a human body is of significant importance in many fields, e.g., finding the right suspect during a criminal investigation or by confirming the evidence of innocence based on an alibi during the estimated time of death. Estimating the time of death is not only important for homicide cases. It is relevant for medical and scientific aspects as well as in civil matters. Determining the effectivity of an insurance policy might depend on the time factor. The realization of a will depends on the justness of the probate, whether the wife or husband died first.
To develop a practical model interdisciplinary work between forensic medical experts and mathematicians is necessary. Several different approaches have evolved concerning the question of estimating the time of death as accurate as possible, whereby the origin of the work focuses on the temperature based method which analyzes the heat loss of the body after death. The practicability of the method depends on specific conditions and parameters, e.g., the outside temperature. Approximately for 24 hours after the cooling process started the temperature based method could be applied.
The time of death is evaluated based on the measured body temperature at certain points of measurement. Post-mortem temperature-based models are necessary as they provide a curve of the body temperature at a point of measurement with respect to the time that has passed since death.

1.2 Current State of the Art

The empirical model according to Henssge is one option to determine the time of death with the help of a widespread designed database of post-mortem cooling cases under specific conditions [36]. Experimental post-mortem cooling curves that include the related conditioned factors and circumstances (e.g., gender, weight, height, outside temperature) are recorded in a database. Complications concerning the data ascertainment occur since the experiments on corpses are difficult to realize due to the lack of sampling and in addition for ethical and moral reasons. Furthermore, the number of considered conditions is limited. On one hand, certain relevant conditions are not taken into account. On the other hand, the relevance of inspected conditions is controversial, e.g., the body mass index that does not distinguish between the relevant amount of different tissues.
The alternative approach by Mall [33] avails oneself of the *Heat Flow Finite Element Method* relying on laws of nature. The dynamic process is displayed by a differential equation, the heat equation by Fourier. The resulting computational model simulates the body cooling process that is influenced by complex and to some extend unknown factors. The different tissue compartments vary among their thermal properties. Individual anatomical features are taken into account. The method quickly develops into a high-dimensional model that considers only approximations of thermodynamic relevant structures.
While observing the corpse, forensic experts perform measurements of the body temperature at certain points in space and certain time intervals. Ordinarily, the collection of

measurements is limited over space to one rectal point. The data ascertainment provides the basis for the back-calculation to provide an estimate of the time of death of the corpse.

1.3 Problem Statement

This thesis focuses on the modeling of the body cooling process after death. Identifying the parameters and designing the experiments are two important aspects that have to be considered to evaluate a reliable estimate for the time of death. The foundation for this analysis is the computational model which has to be validated such that the results of the simulation coincide with the reality. By assuming that the mathematical formulation of the cooling process is perfectly known the usability of the model relies on the assumption of the unknown input parameters. Parameters are not completely predictable and underlay certain changes during the cooling process. Considering the muscle tissue it holds that different bodies, or even different parts of the muscle tissue of one person, have a unique structure and various properties. Another example considers the outside temperature during the cooling process that can only be determined precisely to a certain extent. These so called uncertainties have to be taken into account.
The input parameters and the data that emanates from the experimental setting influence the usability of the model. The experimental setting is chosen in such way that reliable estimates are computed while considering cost and practicability at the same time. The question evolves in which setting the experiment should take place to provide good results that obtain the maximum amount of information under given constraints. By systematically planning the design of the setting the resulting model shall be optimized regarding the accuracy and the stability, while keeping an eye on the experimental effort.
The experimental design includes the definition of control variables. These variables might include the sample size, the selection of points of measurement, the prior accuracy, and the assumptions respectively the knowledge considering the thermal parameters. Certain decisions have to be made prior or during the data ascertainment accordingly as the experiment consists of more than one series.
Considering experiments that cannot be performed more than once, the approach of designing the optimal setting is unalterable, since the data collection is restricted. Prior available data has to be valued. The experiment is put into a statistical context. The degree of optimization is measured by statistical inference of the quantities of interest, therefore we aim for estimates with maximal statistical reliability. An estimate of the time of death with a small variance is a good intuitive example for a desired target in terms of an optimal design of the experiment. The question to be answered is if an optimal experimental setting leads to significant better estimates in comparison to estimates of alternative experimental settings.

1.4 Structure of the Thesis

The current chapter is followed by Chapter 2 which covers the computation of the *Forward Model* that simulates the cooling process of the body over time and space. The computation is based on the *Finite Element Method*. The reduction of the high-dimensional model

by the *Proper Orthogonal Decomposition* application in combination with the *Galerkin Method* is introduced.
The estimation of the time of death results in an *Inverse Problem*. Chapter 3 treats the parameter identification, whereby it is stated in a Bayesian statistical context. Furthermore, the non-linearity of the model is examined and the local linearization is analyzed.
Chapter 4 approaches the topic of *Optimal Design of Experiments*. An optimization problem results, whose solution yields certain challenges. Former numerical results will be illustrated to examine the outcome of an optimal design. The last chapter summarizes the results of this thesis and gives an outlook on open issues and problems.

2 Forward Model

2.1 Heat Equation

This thesis uses a post-mortem Forward Model to provide the temperature in any point of the body with respect to the time that has passed since death [33]. The forward problem is given by the heat equation by Fourier [52, 17]. This partial differential equation (PDE) describes the distribution and transfer of the heat by conduction, convection, and radiation over space $D \subset \mathbb{R}^3$ and time $t \in [0, T]$, whereby $t = 0$ defines the starting point of the cooling process and $T > 0$. The temperature u is displayed as

$$u = u(x, t),\ x \in D,\ t \in [0, T].$$

The differential form of Fourier's law of thermal conduction,

$$c\rho \frac{\partial u}{\partial t} = \nabla \cdot (\kappa\ \nabla u) + q \text{ in } D \times (0, T], \tag{2.1}$$

involves the heat capacity $c(x)$, the mass density ρ, and the thermal conductivity $\kappa(x)$. The heat capacity and the thermal conductivity distinguish between the different tissues, nevertheless for simplification reasons we write $c = c(x)$, $\kappa = \kappa(x)$. The mass density ρ is fixed in this model for all points of measurement. Additional heat sources are accumulated in $q(x), x \in D$, for instance heat as a result of continuing metabolism after death. Here we assume that the model does not include additional heat sources, i.e. $q = 0$.

The convection interacts with the outside at the boundary ∂D. Convection transfers heat by the movement of molecules. The thermal radiation concerns the electromagnetic radiation emitted towards and from the body over time. It affects the environment and not only the immediate surrounding of the body. Therefore, the two distinct temperatures, u_{env} and u_{out} are considered. The convection and radiation are represented in the boundary condition,

$$n^T(\kappa\ \nabla u) = \gamma(u_{out} - u) + \epsilon\sigma(u_{env}^4 - u^4) \text{ in } \partial D \times (0, T). \tag{2.2}$$

The boundary equation involves the heat transfer coefficient γ, the emissivity ϵ, and the Stefan-Boltzmann constant σ.
Considering the application we can simplify the heat equation as follows. Since the temperature difference $(u_{env} - u)$ of the radiation term is small the term approximates to a linear relationship which allows us to write (2.2) in the form,

$$n^T(\kappa\ \nabla u) = \gamma(u_{out} - u) + 4u_{env}^3\ \epsilon\sigma(u_{env} - u) \text{ in } \partial D \times (0, T). \tag{2.3}$$

We assume the outside temperature and environmental temperature to be identical and constant over time. The setup is further simplified; no heat factors are included. Therefore, the temperature is no subject to variation such that (2.3) simplifies to,

$$n^T(\kappa\ \nabla u) = (\gamma + 4u_{out}^3\ \epsilon\sigma)(u_{out} - u) \text{ in } \partial D \times (0, T). \tag{2.4}$$

The emissivity and Stefan-Boltzmann constant concerning (2.4) are fixed. An initial condition $u(x, 0)$ is included that states the body temperature at the starting point of the cooling process,

$$u(x, 0) = u_0(x) \text{ in } D. \tag{2.5}$$

The initial condition is known based on the average body temperature of the human body controlled by the thermoregulatory center.

Definition 1. *The considered computational Forward Model of the cooling process includes* (2.1)*,* (2.4)*, and* (2.5)*, and is defined as,*

$$F(p; x, t) = u(x, t), \tag{2.6}$$

with the state variable as a mapping,

$$u : D \times [0, T] \to \mathbb{R}, \tag{2.7}$$

and the parameter set,

$$p(x) := [c(x), \kappa(x), \gamma, u_{out}],\ x \in D. \tag{2.8}$$

Remark. *The objection that the parameters in* $p(x)$ (2.8) *are temperature-dependent is justified. Due to the lack of scientific knowledge and research in the area of temperature-dependent parameters with respect to the human tissue, these parameters will be inspected independently from the current body temperature. Nevertheless, the literature suggests different parameter ranges, whereby the discrepancies concerning the parameter determination will be addressed and apprehended in Section 3.4.2.*

The solution of the heat equation with its boundary condition and initial value cannot be obtained analytically, but numerically e.g. applying the *Finite Element Method*.

2.2 Finite Element Method

The computation of the solution of the Forward Model is based on the *Finite Element Method (FEM)*. Theoretical background can be found in Chapter 4 of *Numerische Mathematik 3* by Deuflhard and Weiser [14]. Further details are found in [21, 43, 51, 14]. An overview of the FEM follows in this section by applying the theory to the heat equation. The FEM approximates the solution of (2.1), (2.4), and (2.5), by a linear combination of so-called basis functions.

The first step involves the reformulation of the original problem to approach the FEM. We consider a simple stationary heat equation,

$$-\nabla \cdot (\kappa\ \nabla u) = q \text{ in } D \subset \mathbb{R}^3. \tag{2.9}$$

The constant term regarding the emissivity and Stefan-Boltzmann constant is embedded for simplification in γ. The Robin boundary conditions are stated with,

$$n^T \kappa\ \nabla u + \gamma(u - u_{out}) = 0 \text{ on } \partial D. \tag{2.10}$$

The second order PDE (2.9) requires a twice differentiable function $u \in C^2(D) \cap \bar{C}(\partial D)$ as the solution. Since this requirement is too strong in most cases, the so-called strong formulation is transformed to a weak formulation by means of the integration against a test function $v \in C^1(D)$ and the application of Green's theorem [14, p. 397]. This transformation results in a variational statement,

$$\int_D \kappa \ \nabla u^T \nabla v \ dx - \int_{\partial D} \kappa \ v n^T \nabla u \ ds = \int_D -\nabla \cdot (\kappa \ \nabla u) \ v \ dx. \tag{2.11}$$

If u is a solution of (2.9) it follows that,

$$\int_D \kappa \ \nabla u^T \nabla v \ dx - \int_{\partial D} \kappa \ v n^T \nabla u \ ds = \int_D qv \ dx. \tag{2.12}$$

A weak formulation evolves by inserting the boundary condition (2.10) with $\beta := \gamma \ u_{out}$,

$$\int_D \kappa \ \nabla u^T \nabla v \ dx + \int_{\partial D} \gamma uv \ ds = \int_D qv \ dx + \int_{\partial D} \beta v \ ds. \tag{2.13}$$

The Sobolev space $H^1(D)$ is introduced to analyze the solution.

Definition 2. *With $m \geq 0$ and $1 < p < \infty$ the Banach space*

$$W^{m,p}(D) = \{u \in L^p(D) : \ \nabla u \in W^{m-1,p}(D)\} \text{ and } W^{0,p}(D) = L^p(D)$$

is called a Sobolev space. The corresponding norm is defined as,

$$||u||^p_{W^{m,p}(D)} = ||\nabla u||^p_{W^{m-1,p}(D)} + ||u||^p_{L^p(D)},$$

whereby $L^p(D) = \{D \to \mathbb{R}| \ ||u||^p_{L^p} = \int_D u^p dx < \infty\}$ defines the Lebesgue space. If $p = 2$ the Sobolev space is an Hilbert space,

$$H^m(D) := W^{m,2}(D).$$

Further details are found in [14, pp. 403 ff.]. It holds that every solution of the strong formulation is a solution of the weak formulation.

Definition 3. *A function in the Sobolev space $u \in H^1(D)$ is called a weak solution of* (2.9) *and* (2.10) *if the function u solves the weak formulation* (2.13) *$\forall v \in H^1(D)$.*

The left-hand side of (2.13) states a symmetric bilinear operator,

$$a(u, v) := \int_D \kappa \ \nabla u^T \nabla v \ dx + \int_{\partial D} \gamma uv \ ds, \ u, v \in H^1(D).$$

The right-hand side of (2.13) displays a linear function $b \in H^1(D)$ with $\langle b, v \rangle$ denoting the dual pairing of a function and a vector,

$$\langle b, v \rangle := \int_D qv \ dx + \int_{\partial D} \beta v \ ds. \tag{2.14}$$

Hence we end up with a generalized formulation,

$$a(u, v) = \langle b, v \rangle, \ \forall v \in H^1(D). \tag{2.15}$$

2.2.1 The Galerkin Approach

The *Galerkin Method* approximates the infinite-dimensional function space $H^1(D)$ by a finite-dimensional function space $V_h \subset H^1(D)$, whereby the subspace is defined as $V_h = span\{\varphi_1, \ldots, \varphi_N\},\ N \in \mathbb{N}^+$. The discretization of the solution is then given by a linear combination of basis functions φ_i $(i = 1, \ldots, N)$ with the weight vector $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)^T$,

$$u_h(x) = \sum_i \alpha_i\ \varphi_i(x). \tag{2.16}$$

The *Galerkin discretization*,

$$a(u_h, v_h) = \langle b, v_h \rangle,\ \forall\ v_h \in V_h \subset H^1(D), \tag{2.17}$$

inserts the basis functions φ_i as test functions v_h in (2.17),

$$\sum_i \alpha_i\ a(\varphi_i, \varphi_j) = \langle b, \varphi_j \rangle,\ \forall j = 1, \ldots, N. \tag{2.18}$$

By defining the vector

$$\boldsymbol{f_h} := (\langle b, \varphi_1 \rangle, \ldots, \langle b, \varphi_N \rangle)^T,$$

and the symmetric positive semi-definite matrix

$$(A_h)_{ij} := a(\varphi_i, \varphi_j) \in \mathbb{R}^{N \times N},$$

a linear system of equations evolves,

$$A_h\ \boldsymbol{\alpha} = \boldsymbol{f_h}. \tag{2.19}$$

The theorem by Lax-Milgram states the existence of a unique solution of the weak formulation [14, pp. 404-406].

Concerning the L^2-inner product, the *Galerkin Method* computes the discretized solution u_h which is closest to the strong solution u. The proof can be found in [14, p. 95].

After setting the groundwork for the FEM the domain is subdivided into disjoint elements,

$$D = D^1 \cup \cdots \cup D^m.$$

Functions on the domain D such as the solution u are represented by a linear combination of basis functions, $\varphi_i,\ i = 1, \ldots, N$. In the case of $D \subset \mathbb{R}^3$ the elements D^i can be chosen as tetrahedra. Globally continuous, piecewise polynomial functions with local support are the preferred choice of FE-basis functions. Due to the local support of the basis functions the so-called stiffness matrix A_h computes as a sparse matrix. The stiffness matrix incorporates the parameters of the stationary model, the thermal conductivity, and the heat transfer coefficient,

$$(A_h(\kappa, \gamma))_{ij} = \kappa \int_D \nabla \varphi_i^T \nabla \varphi_j\ dx + \gamma \int_{\partial D} \varphi_i \varphi_j\ ds. \tag{2.20}$$

This sparsity is an advantage of the FEM and the linear system (2.19) can be solved by direct or iterative solvers [14, pp. 149 ff.].

As a next step, we apply the *Galerkin Method* on the time-dependent heat equation with the Robin boundary conditions (2.4). The mass density ρ is fixed, w.l.o.g. $\rho = 1$. Therefore (2.9) is expanded with the instationary term yielding the PDE,

$$c\frac{\partial u}{\partial t} - \nabla \cdot (\kappa \ \nabla u) = q \text{ in } D \times (0, T], \tag{2.21}$$

with the initial condition (2.5). The weights $\boldsymbol{\alpha}_i$ from previous discretization (2.16) now incorporate the time-dependence,

$$u_h(x,t) = \sum_i \boldsymbol{\alpha_i}(t) \ \varphi_i(x), \tag{2.22}$$

$$\Rightarrow \dot{u}_h(x,t) = \sum_i \dot{\boldsymbol{\alpha}}_{\boldsymbol{i}}(t) \ \varphi_i(x). \tag{2.23}$$

The variational statement evolves to,

$$\int_D \kappa \ \nabla u^T \nabla v \ dx + \int_{\partial D} \gamma u \ v \ ds = \int_D qv \ dx + c\int_D -\dot{u}v \ dx + \int_{\partial D} \beta v \ ds. \tag{2.24}$$

The *Galerkin Method* is applied with the left-hand side being identical to the stationary case (2.13), stating the stiffness matrix A_h (2.20). The right-hand side yields in addition to (2.14) the positive semi-definite mass matrix,

$$(M_h(c))_{ij} := c\int_D -\varphi_i\varphi_j \ dx, \tag{2.25}$$

such that the ordinary differential equation (ODE) of order one results,

$$\begin{aligned} &A_h(\kappa,\gamma) \ \boldsymbol{\alpha} = M_h(c) \ \dot{\boldsymbol{\alpha}} + \boldsymbol{f_h}, \\ &M_h(c), \ A_h(\kappa,\gamma) \in \mathbb{R}^{N\times N}. \end{aligned} \tag{2.26}$$

The ODE is solved by the Backward Euler Method [13]. A system of linear equations with $\boldsymbol{\alpha^j} := \boldsymbol{\alpha}(t_j)$ is solved by discretizing the time interval equidistantly with step size τ, whereas $\boldsymbol{\alpha^1} = \boldsymbol{\alpha}(0)$ and $\boldsymbol{\alpha^{k+1}} = \boldsymbol{\alpha}(T)$,

$$A_h \ \boldsymbol{\alpha^{j+1}} = M_h \left(\frac{\boldsymbol{\alpha^{j+1}} - \boldsymbol{\alpha^j}}{\tau}\right) + \boldsymbol{f_h}, \ j = 1, \ldots, k, \tag{2.27}$$

with the initial condition defined as

$$u_0(x) = \sum_i \alpha_i(0) \ \varphi_i(x). \tag{2.28}$$

2.3 Computation of the Forward Model

The parameters ρ, ϵ, and σ are fixed and embedded in a constant matrix which will be apprehended in (2.32). The other parameters (2.8) represent the input parameters of the computed Forward Model.
The domain D is the digital corpse based on a computed tomography scan. The segmentation is done by the ZIB Amira software [48, 54] that identifies the three main thermodynamic tissue types: fat, bones, and muscles. Details of the concept of 3D Medical Imaging by segmentation and its different techniques can be found in [39, pp. 467 ff.].
The body temperature u is computed for any point with respect to the time that has passed since death. The noise e is added to incorporate accumulation of variables that are not observable or of no interest. Changes in the initial body temperature and parameters vanish over time by converging to the outside temperature, therefore the final temperature is unbiased.
Further information of the Heat Flow Model regarding the human body can be found e.g. in papers by Mall and Eisenmenger [36, 33].

2.4 Model Reduction

The large-scale dynamic system and the resulting intricacy of the designed Forward Model are accompanied by a demanding computing time. For several applications considered in this thesis a faster computation is necessary. The model reduction will be achieved by the *Proper Orthogonal Decomposition (POD)* application in combination with the the *Galerkin Method*. An overview of the POD theory is embedded in the following section. The theory of the POD is based on the notes of Volkwein [50]. The theory of the *Galerkin Method* follows the previous section. The reference [45] describes an example how to reduce a model using the *POD Galerkin Method*.
The foundation for the model reduction are potential solutions so-called snapshots computed with the Forward Model. Furthermore, the mass matrix and stiffness matrix, as well as the initial condition, of the forward problem, are provided and build the foundation of the model reduction by the *POD Galerkin Method*.

2.4.1 Proper Orthogonal Decomposition

The key property of the *POD Method* is to produce an optimal orthogonal basis for a given data set, where the main properties of the solution $u \in V$ are represented by the POD basis of a low-dimensional space V_h which is a subspace to V. The dynamics of the system, i.e. the heat flow of the body, is captured by snapshots stating an ensemble of potential solutions and providing the foundation for the POD modes that define the POD basis. In our context the snapshots are computed by the high-dimensional Forward Model consisting of n different potential values of the thermal parameters and m discrete time steps, $0 = t_1 < t_2 < \cdots < t_m = T$. By considering a set of parameter samples,

$$\Xi_n = \{p^1, \ldots, p^n\},\ n \in \mathbb{N}^+ \tag{2.29}$$

the solution matrix S is constructed by snapshots $F(p^i) \in \mathbb{R}^{N\times m}$ such that,

$$S := [F(p^1), \dots, F(p^n)] \in \mathbb{R}^{N\times M},\ M = mn. \tag{2.30}$$

The dimension M includes the solution over discrete points in time and different parameter sets, while the dimension N represents the number of mesh nodes modeling the discretization over space given by the FEM.

The POD basis is evaluated by applying the *Singular Value Decomposition (SVD)* to the solution matrix. The main information of the solution matrix will then be represented by a small number of vectors corresponding to the greatest singular values, as will be shown.

Theorem 1. *Let $S \in \mathbb{R}^{N\times M}$ and $rank(S) = d \leq min(N, M)$. The factorization,*

$$S = W\Sigma V^T,$$

exists with the orthogonal matrices

$$W \in \mathbb{R}^{N\times N} \text{ and } V \in \mathbb{R}^{M\times M},$$

and the diagonal matrix,

$$\Sigma = diag(\sigma_1, \dots, \sigma_d, 0, \dots 0) \in \mathbb{R}^{N\times M},$$

that consists of the singular values in descending order,

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_d \geq 0.$$

Proof. The proof can be found in [12, p. 146]. □

The magnitude of a singular value agrees with the significance and influence on the representation of S. A reduction of the computational effort to evaluate a representative of the matrix S can be achieved by regarding only a bounded number of the greatest singular values σ_i with their corresponding vectors w_i and v_i.

Definition 4. *The Truncated SVD of dimension l states the representation of the matrix $S \in \mathbb{R}^{N\times M}$ by the rank l matrix,*

$$\widetilde{S}_l := W_l\Sigma_l V_l^T$$

in order that only the $l \leq d$ greatest singular values with their corresponding vectors w_i and v_i are regarded,

$$\begin{aligned} W_l &= [w_1, \dots, w_l] \in \mathbb{R}^{N\times l}, \\ \Sigma_l &= diag(\sigma_1, \dots, \sigma_l) \in \mathbb{R}^{l\times l}, \\ V_l &= [v_1, \dots, v_l] \in \mathbb{R}^{M\times l}. \end{aligned}$$

The matrix $\widetilde{S}_l$ is the best rank l approximation to S considering the Frobenius– and the 2-norm [49].

In the context of model reduction the singular values σ_i usually decline rapidly and the *Truncated SVD* can be applied to evaluate the POD basis.

Theorem 2. *The vectors of the orthogonal matrix W of the SVD form the span of the column space of the solution matrix S. For $l \in \{1, \dots, d\}$ the vectors $\{w_i\}_{i=1}^{l}$ build the POD basis of rank l. For every $l \leq d$ the approximation of the columns of S by the corresponding set $\{w_i\}_{i=1}^{l}$ is optimal among all rank l approximations to the columns of S.*

Proof. The proof can be found in [50, p. 6 ff.]. □

Remark. *With given $\epsilon > 0$ the $l \leq d$ greatest singular values $\sigma_i > \epsilon$ with their corresponding vectors w_i are regarded and define an adequate representation of the solution matrix such that a certain error is not exceeded. Therefore, the cost of the computation is reduced. The choice of ϵ, which estimates the dimension l of the POD basis and therefore the arising error, is of central importance, though no general a-priori rules are available concerning the application of the SVD truncation. With the introduced method of snapshots it is not clear to what extent one might neglect certain singular values before analyzing the result of the model reduction. Though the approximation quality by the matrix $\widetilde{S}_l$ of rank l is bounded by $\sum_{i=1}^{l} \sigma_i$ [50, pp. 6 ff.].*

2.4.2 POD Galerkin Method

The model reduction is achieved by substituting the applied basis vectors of the *Galerkin Method* with the POD modes (Theorem 2),

$$\varphi_i = w_i, \ i = 1, \dots l.$$

The *POD Galerkin Method* reduces the dynamical system by an orthogonal projection onto the subspace of the POD modes. The projection of the mass matrix and stiffness matrix of the ODE (2.26) result in the following transformations with $\langle \cdot, \cdot \rangle$ defining the scalar product,

$$\begin{aligned}
&\langle M \sum_i \dot{\alpha}_i \ \varphi_i + f_h, \varphi_j \rangle = \langle A \sum_i \alpha_i \ \varphi_i, \varphi_j \rangle \quad \forall j = 1, \dots, l, \\
\Leftrightarrow\ &\varphi_j^T M \ \sum_i \dot{\alpha}_i \ \varphi_i + \varphi_j^T f_h = \varphi_j^T A \ \sum_i \alpha_i \ \varphi_i \quad \forall j = 1, \dots, l, \\
\Leftrightarrow\ &\sum_i (\varphi_j^T M \ \varphi_i) \ \dot{\alpha}_i + \varphi_j^T f_h = \sum_i (\varphi_j^T A \ \varphi_i) \ \alpha_i \ \forall j = 1, \dots, l.
\end{aligned}$$

By projection we obtain,

$$\begin{aligned}
(\widetilde{M})_{ij} &:= \varphi_j^T M \varphi_i \ \forall i, j = 1, \dots, l, \\
(\widetilde{A})_{ij} &:= \varphi_j^T A \varphi_i \ \ \forall i, j = 1, \dots, l, \\
(f_p)_j &:= \varphi_j^T f_h \quad \forall j = 1, \dots, l, \\
(\alpha_p)_j &:= \varphi_j^T \alpha \quad \forall j = 1, \dots, l,
\end{aligned} \tag{2.31}$$

such that a reduced order ODE of dimension $l < N$ evolves,

$$\widetilde{M}\ \dot{\boldsymbol{\alpha}}_{\boldsymbol{p}} + \boldsymbol{f}_{\boldsymbol{p}} = \widetilde{A}\ \boldsymbol{\alpha},$$

with

$$\widetilde{M}, \widetilde{A} \in \mathbb{R}^{l \times l} \text{ and } \boldsymbol{\alpha}_{\boldsymbol{p}}, \boldsymbol{f}_{\boldsymbol{p}} \in \mathbb{R}^{l},$$

and the initial condition,

$$u_0(x) = \sum_j (\alpha_p)_i(0)\ \varphi_i(x).$$

The definitions (2.20) and (2.25) permit the partitioning of the mass matrix and stiffness matrix into constant *initial matrices*,

$$M_0,\ A_0 \in \mathbb{R}^{N \times N}, \tag{2.32}$$

and *correction matrices*,

$$\Delta M_c,\ \Delta A_\kappa,\ \Delta A_\gamma \in \mathbb{R}^{N \times N}, \tag{2.33}$$

that are individually scalable by the distinct parameter,

$$\begin{aligned} M(c) &= M_0 + c\ \Delta M_c, \\ A(\kappa, \gamma) &= A_0 + \kappa\ \Delta A_\kappa + \gamma\ \Delta A_\gamma. \end{aligned} \tag{2.34}$$

The *POD projection* can be applied individually on the distinct parameter matrices due to linearity reasons (2.34),

$$\begin{aligned} (\widetilde{M})_{ij} &= (\varphi_i^T M \varphi_j) = (\varphi_i^T M_0 \varphi_j + \varphi_i^T \Delta M_c \varphi_j) && \forall i, j = 1, \ldots, l, \\ (\widetilde{A})_{ij} &= (\varphi_i^T A \varphi_j) = (\varphi_i^T A_0 \varphi_j + \varphi_i^T \Delta A_\kappa \varphi_j + \varphi_i^T \Delta A_\gamma \varphi_j) && \forall i, j = 1, \ldots, l. \end{aligned}$$

We introduce the abbreviated notation of the individual matrices,

$$\widetilde{M}_0,\ \Delta\widetilde{M}_c,\ \widetilde{A}_0,\ \Delta\widetilde{A}_\kappa,\ \Delta\widetilde{A}_\gamma \in \mathbb{R}^{l \times l},$$

such that,

$$\begin{aligned} \widetilde{M} &:= \widetilde{M}_0 + \Delta\widetilde{M}_c, \\ \widetilde{A} &:= \widetilde{A}_0 + \Delta\widetilde{A}_\kappa + \Delta\widetilde{A}_\gamma. \end{aligned}$$

The *POD Galerkin Method* is efficient in terms of a model reduction if the resulting dimension of the application of the *Truncated SVD* is largely smaller than the number of grid points and a certain error due to the reduction is not exceeded. Therefore, the output of the reduced-order model is analyzed.

2.5 Application and Numerical Results

The *POD Galerkin Method* is applied to reduce the high-dimensional Forward Model. For practical reasons a phantom that provides the foundation for the computation is developed. The error resulting from the model reduction as well as the time performance of the reduced model are analyzed.

2.5.1 Phantom

Experiments are performed on phantom models made out of different material. The validation is realized step by step regarding the complexity of the model and approaching the anatomical properties of a human body. In this thesis the cylindrical phantom is observed with a radius of 10.25cm and a height of 29.5cm which is made out of homogeneous polyethylene, as seen in Figure 1. For one, the evaluated experimental measurements are realizable repeatedly. For another, the phantom's geometry is well-defined and the unknown parameters are assessable. The next step involves a phantom made out of beeswax, brine solution, and bone tissue. This choice of material holds as a good representative of the main human tissues: fat, bones, and muscles.
The phantom is heated in an incubator at an internal ambient temperature of approximately 40°C. After 30 hours the phantom obtains a constant initial temperature of 39.4°C. Afterwards, the phantom is placed in a climatic chamber providing a uniform temperature field over the period of the cooling process. The temperature measurement probe is placed at a point of measurement 15cm deep located near the geometrical center [47]. In this thesis we will only consider computations and experimental measurements based on the polyethylene phantom.

Figure 1: The polyethylene phantom model

The first step to evaluate an accurate reduced model of the cooling process of the polyethylene phantom involves the parametrization of the system, i.e. the definition of the set of parameters that characterize the system. The model is made out of one material. We

assume a constant ambient temperature of 25.1°C. The heat capacity, the thermal conductivity, and the heat transfer coefficient of polyethylene, and primarily the starting time of the cooling process, represent the unknown parameters of the model.

2.5.2 Computation of the Reduced-Order Model

At first, the solution matrix is computed consisting of snapshots supplied by the high-dimensional Forward Model. The grid of the phantom is adjusted to $N = 21155$ nodes. Let $m = 61$ be the number of time steps over the time interval $[0, 72.000]$ seconds, therefore a step size of 1200 seconds is provided. In every time step a solution is computed by the high-dimensional Forward Model. The computations are performed for different parameter combinations considering the material of the phantom, whereby the parameters are chosen in a plausible interval [25],

- the heat capacity of polyethylene: $c \in [1800, 1880]\ \frac{J}{kg\ ^\circ\mathrm{C}}$,
- the thermal conductivity of polyethylene: $\kappa \in [0.36, 0.44]\ \frac{\mathrm{W}}{\mathrm{m}\ ^\circ\mathrm{C}}$,
- the heat transfer coefficient of polyethylene: $\gamma \in [3.0, 3.8]\ \frac{\mathrm{W}}{\mathrm{m}^2\ ^\circ\mathrm{C}}$.

Information of the thermal properties are found in literature or can be evaluated by experimental measurements. Details concerning the choice of the parameters is apprehended in Section 3.4.2. Nevertheless, the properties are considered as unknowns that include certain prior information, e.g. the expectation value and the range of the parameters. The $M = 7625$ potential solutions are computed and form the solution matrix,

$$S \in \mathbb{R}^{21155 \times 7625}.$$

After evaluating the solution matrix, the *Truncated SVD* is applied as introduced in Definition 4. The corresponding POD modes of the l greatest singular values form the basis of the orthogonal projection. For further applications an efficient model reduction accompanied by a small maximum error is required.

2.5.3 Error Analysis of the Reduced-Order Model

The reduction of the model is accompanied by an error measured on the basis of the difference of the temperature output of the two models in Kelvin. A reasonable compromise of the computation time and the accuracy of the reduced-order model has to be found. For one, the accuracy depends on the size and choice of the snapshots. For another, the accuracy depends on the dimension of the POD basis and the number of discretization time steps.
To measure the quality and efficiency of the model reduction we define the maximum error.

Definition 5. *The maximum error of two matrices* $S,\ \widetilde{S}_l \in \mathbb{R}^{N \times M}$ *is defined as,*

$$e_{max}(S, \widetilde{S}) = ||S - \widetilde{S}||_{max}, \tag{2.35}$$

with the maximum norm,

$$||S||_{max} = \max_{i,j} |s_{ij}|, \ \forall i, j.$$

The application of the SVD truncation regarding the chosen dimension l of Definition 4 highly depends on the magnitude and regression of the singular values. The greatest value of the solution matrix is $\sigma_1 \geq 3e6$, and the singular values decline rapidly with $\sigma_8 \leq 1e2$ and $\sigma_{13} \leq 1e0$ as displayed in Figure 2.

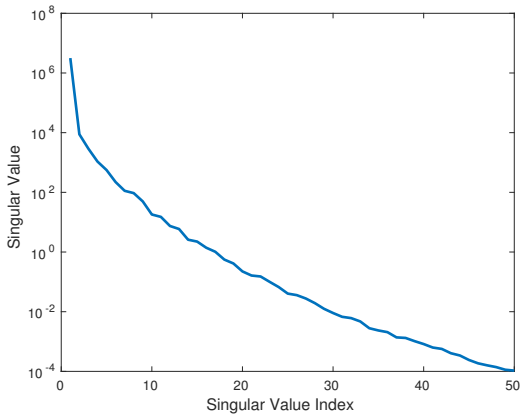


Figure 2: Singular Values: 1:50

The maximum error of the solution matrix S and the truncated solution matrix of rank l stated in Definition 4 is evaluated with Definition 5,

$$e_{max}(l) = ||S - \widetilde{S}_l||_{max}. \tag{2.36}$$

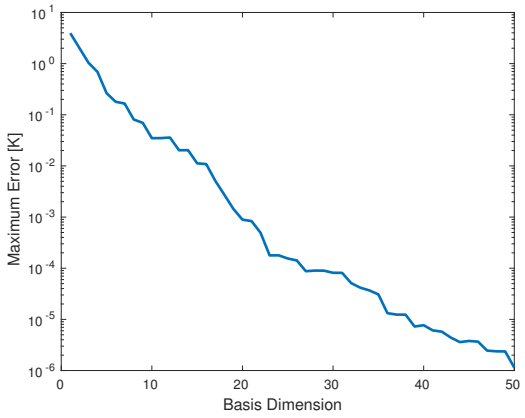


Figure 3: The maximum error considering the solution matrix and the truncated solution matrix of rank l declines rapidly by choosing a greater basis dimension l.

Figure 3 shows the error depending on the dimensional truncation stated by $e_{max}(l)$ in (2.36). Clearly, the error declines rapidly with increasing basis dimension l and one can

assume that a basis of dimension $l \geq 8$ results in a proper approximation of the solution matrix such that,

$$e_{max}(l) \leq 1\text{e}{-1}.$$

A selection of an efficient dimension of the POD Basis must be ensured. To justify the representation of the model by the low-dimensional model, the output of the low-dimensional model and the output of the high-dimensional model is computed and interrelated by the maximum norm stating the error of the modeling process that results from the model reduction.

Definition 6. *Set u_c as the computed solution of the high-dimensional model and $u_r(l)$ as the computed solution of the low-dimensional model based on the l-dimensional POD basis. The maximum error resulting by the model reduction over space and time is defined as,*

$$e_{max}(l) = ||u_c - u_r(l)||_{max} = \max_{x,t} |u_c(x,t) - u_r(x,t;l)|, \ x \in D, \ t \in [0,T], \ l \leq d. \quad (2.37)$$

The following figures show the maximum error considering different dimensions of the POD basis. The quality of the low-dimensional model concerning the chosen POD basis dimension is specified by means of Definition 6.

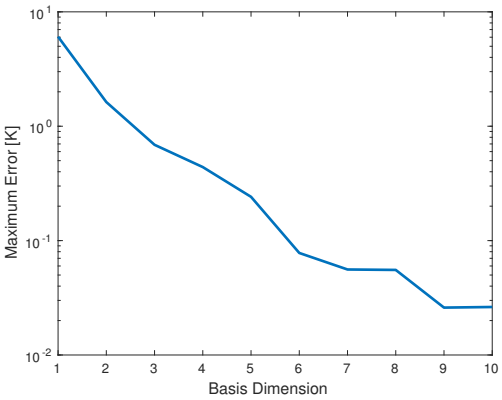


Figure 4: The maximum error of the high-dimensional model and the low-dimensional model converges to zero by increasing the dimension of the POD basis.

The error vanishes and converges to zero by increasing basis dimension as seen in Figure 4. Choosing a basis of dimension $l \geq 8$ provides sufficient results with,

$$e_{max} \leq 3\text{e}{-2}.$$

The basis dimension $l = 8$ is fixed and the step size of the Backward Euler method is optimized [12]. Both, the maximum model reduction error by Definition 6, and the computational time are considered. A compromise solution of the two mentioned aspects is to be found.

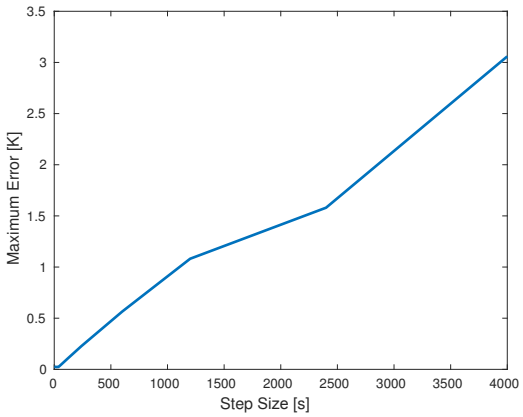


Figure 5: The maximum error depending on the step size of the Euler method

The maximum error that is displayed in Figure 5 varies in dependence of the step size of the Backward Euler method. The error increases almost linearly with respect to the step size. We observe the time performance to make a justified decision.

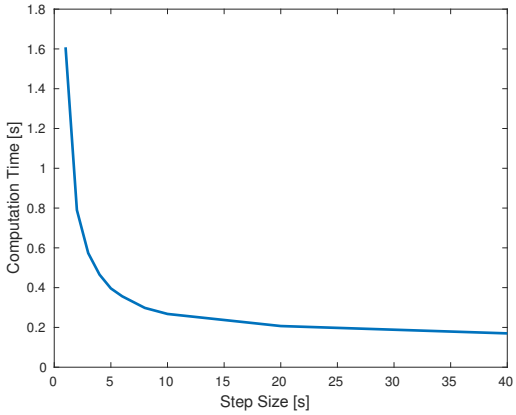


Figure 6: The computational time depending on the step size of the Euler method

As seen in Figure 6 a step size $\tau \geq 20$ seconds leads to a reasonable time performance and converges to the computational time that contains the POD Galerkin projection (2.31). On the other hand, a computation with a step size of $\tau \leq 40$ seconds is accompanied by a maximum error of $\leq 3\text{e}{-2}°\text{K}$. Therefore, it is justified to choose a step size $20 \leq \tau \leq 40$ seconds.
The basis dimension $l = 8$ and step size $\tau = 40$ seconds are fixed. The body temperature curve of the high-dimensional model and the low-dimensional model at a given point of measurement $x \in D$ is displayed in Figure 7. The absolute error is computed by

$$e_{abs}(t) = |u_c(x,t) - u_r(x,t)| \ \forall t.$$

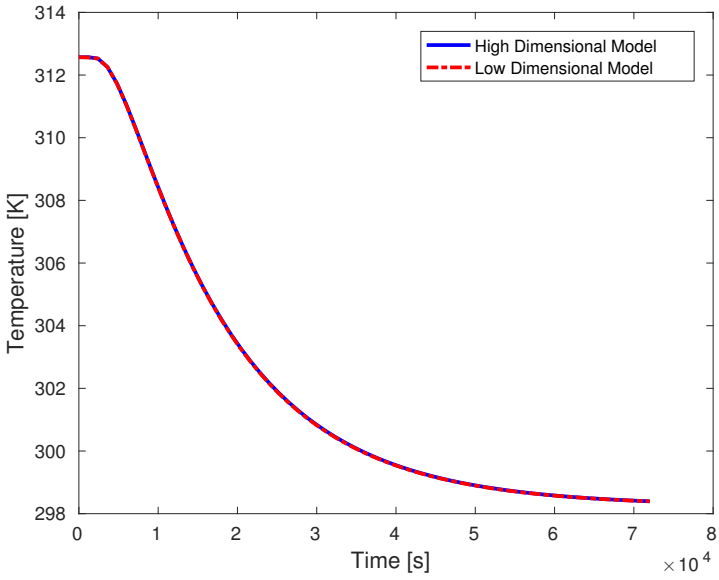


Figure 7: The body temperature curve at the point of measurement over time

The absolute error over time regarding the fixed point of measurement as seen in Figure 8 declines while converging to the outside temperature.

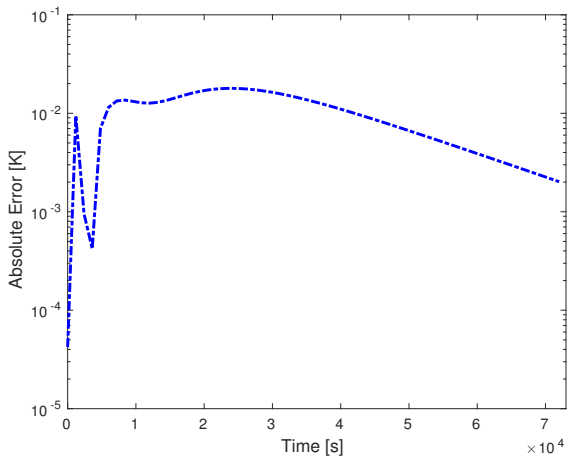


Figure 8: Absolute error over time between the high-dimensional and low-dimensional model at one point of measurement

A computational time reduction is achieved by the *Proper Orthogonal Decomposition* application in combination with the the *Galerkin Method*. The high-dimensional model goes along with a computation time of more than 600 seconds. The low-dimensional model computes a solution in less than 1 second.

Remark. *One has to keep in mind that the implementation of the low-dimensional model is based on the solution matrix computed by the high-dimensional Forward Model. The former computations have to be considered while comparing the computation time and performance. For one thing, the computation time of the groundwork considers the time required for evaluating the snapshots. Since the snapshots can be computed in parallel, we have to consider one-time-only the computation time of a solution output by the high-dimensional model to evaluate the solution matrix S. For another, the computation time of the Singular Value Decomposition has to be considered evaluating the POD basis.*
The quality of the output of the low-dimensional model depends on the assumed parameter range that is included in the solution matrix. If the cooling process relies on a different range of parameters a newly computed solution matrix S and POD basis are required.
In this thesis we rely on the computed low-dimensional model based on a snapshot matrix that is generated from the stated parameter range (Section 2.5.2).

3 Inverse Problem

The Forward Model by Definition 1 provides the body temperature based on the given set of thermal parameters, the initial condition and the time that has passed since the cooling process started. Henceforth, the unknown parameters p of (2.8) include the time of death $t \in [0, T]$. Here, the outside temperature is considered to be known. It holds that the body temperature, the parameters and the Forward Model always refer to distinct points of measurement $x \in D$ and the Forward Model in Definition 1 is simplified.

Definition 7. *In the context of the Inverse Problem the Forward Model is defined as,*

$$F(p) = u, \tag{3.1}$$

with the parameters,

$$p := [t, c, \kappa, \gamma]. \tag{3.2}$$

Definition 8. *The forward problem provides the body temperature of the corpse u at certain points of measurement based on the Forward Model of Definition 7.*

The forward problem is deterministic, i.e. a unique solution is computed with a given set of parameters (3.2). In practice the forensic medical expert evaluates the body temperature at certain points of measurement u_m, intending to estimate the time of death, and as the circumstances require the thermal parameters of the Forward Model. For practical applications the forward problem is inverted [31, 26]. The computation of the parameters depends sensitively upon the measured data and a noise e is considered such that,

$$F(p, e) = u_m, \tag{3.3}$$

and the unknown parameters cannot be determined analytically by,

$$F^{-1}(u_m) = p.$$

An optimization problem rises that evaluates the least-square solution [7],

$$\min_p ||F(p) - u_m||^2. \tag{3.4}$$

The problem of estimating the time of death is ill-posed. While the forward problem is deterministic the back-calculation on the contrary does not have a unique solution. Furthermore, changes in the time of death do not have a great impact on the body temperature while converging to the outside temperature in the end. It proceeds that significantly different initial conditions and parameters could have produced the measured body temperature. Therefore, the so-called ill-posed inverse problem has to be approached by certain numerical regularization methods [26].

3.1 Tikhonov Regularization

To estimate the time of death the inverse problem is addressed by the Tikhonov regularization [26].
In this thesis we denote the weighted norm considering $x \in \mathbb{R}^n,\ A \in \mathbb{R}^{n \times n}$ by

$$||x||_A^2 = x^T A\ x. \tag{3.5}$$

Definition 9. *Based on measurements of the body temperature the inverse problem,*

$$\min_p \left|\left|F(p) - u_M\right|\right|_B^2 + \alpha \left|\left|p - \hat{p}\right|\right|_{L^T L}^2, \tag{3.6}$$

evaluates the time of death and the thermal parameters by solving the optimization problem. The Tikhonov regularization includes prior information such that preferences to certain parameters are given by defining the two matrices B and L, the regularization parameter $\alpha > 0$, and the prior assumption of the unknown $\hat{p}$.

The regularization parameters have to be defined reasonably, which might lead to additional difficulties. The generalized Tikhonov regularization uses $B = I$ and $L = I$.

3.2 Parameter Identification

Temperature measurements are performed on a phantom made out of polyethylene, as introduced in Section 2.5.1. At a constant outside temperature of $u_{out} = 25.1°\text{C}$ over discrete points in time $t \in [0, 72000]$ seconds at one point of measurement $x \in D$ [47] the temperature measurements are evaluated. The focus is laid on the evaluation of the unknown thermal parameters, since the time component is given by the supervised experiment. The results of the Forward Model stating the simulated data $u(x,t)$ are compared to the experimental data $u_m(x,t)$. The Tikhonov regularization evaluates the parameters,

$$c = 1840\frac{J}{kg°\text{C}},\ \kappa = 0.4\frac{W}{m°\text{C}},\ \gamma = 3.32\frac{W}{m^2\ °\text{C}}. \tag{3.7}$$

In practice the time component is one of the biggest unknowns. For the experiment it holds that the time component is known and well-founded prior knowledge of the thermal parameters $\hat{p}$ (Section 3.4.2) is given, therefore the generalized Tikhonov regularization with $B = I,\ L = I$ and $\alpha = 1$ performs well.
The temperature of the simulated and the experimental data at one point of measurement is displayed in Figure 9 with the body temperature converging to the outside temperature (298.25°K).

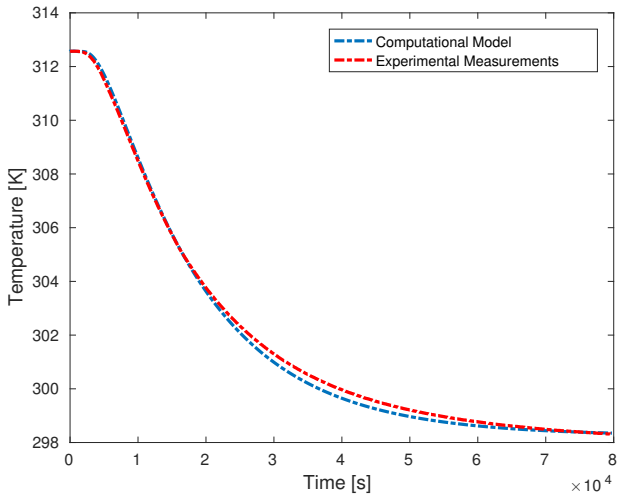


Figure 9: The temperature curve at the point of measurement

The absolute error of the simulated and the experimental data,

$$e_{abs}(t) = |u(x,t) - u_m(x,t)| \; x \in D, \; \forall t,$$

is shown in Figure 10 with $e_{abs}(t) < 3.5\text{e}{-1} \; \forall t$.

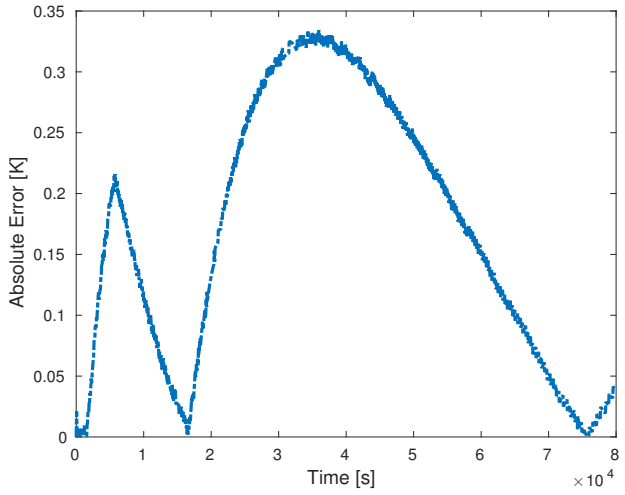


Figure 10: The absolute error at the point of measurement

3.3 Bayesian Inverse Problem

The theory of the following section is based on the book *Statistical and Computational Inverse Problems* by Kaipio and Somersalo [26]. The statistical approach of the inverse problem treats variables of the model (3.3) as random variables. The Bayesian Forward

Model considers the experimental data U_m, the noise E, and the unknown parameters P as random variables,

$$F(P, E) = U_m. \tag{3.8}$$

By assuming independence between the noise E and the unknown parameters P the additive model results,

$$F(P) + E = U_m. \tag{3.9}$$

Definition 10. *The prior information of the unknown parameters P is coded in the* ***prior probability distribution****,*

$$\pi_{pr}(p).$$

Definition 11. *The* ***likelihood probability distribution*** *defines the plausibility of the experimental data U_m conditioned on $P = p$,*

$$\pi(u_m|p).$$

Definition 12. *The* ***noise probability distribution*** *defines the plausibility of the noise E of the model,*

$$\pi_{noise}(e).$$

Theorem 3. *Considering the additive model* (3.9) *the likelihood probability distribution is identical to the noise probability distribution,*

$$\pi(u_m|p) = \pi_{noise}(e).$$

Proof. For the additive model it holds that P and E are mutually independent random variables. The probability distribution of E is unaltered conditioned on $P = p$ with P being fixed. U_m conditioned on $P = p$ is distributed like E with the probability density being translated by $F(p)$,

$$\begin{aligned} &\pi(u_m|p) = \pi_{noise}(u_m - F(p)), \\ \Leftrightarrow\ &\pi(u_m|p) = \pi_{noise}(e). \end{aligned}$$

□

Definition 13. *The solution of the inverse problem is the* ***posterior probability distribution****,*

$$\pi_{post}(p) = \pi(p|u_m),$$

that takes all measurements into account by evaluating the certainty of the unknown parameters $P = p$ conditioned on $U_m = u_m$.

Theorem 4. *The **posterior probability distribution** is defined by,*

$$\pi_{post}(p) \propto \pi_{pr}(p)\ \pi(u_m|p). \tag{3.10}$$

Considering the additive model (3.9) *it holds that,*

$$\pi_{post}(p) \propto \pi_{pr}(p)\ \pi_{noise}(e). \tag{3.11}$$

Proof. With $\pi(u_m) > 0$ and $\pi(p) > 0$, the conditioned probability distributions are computed by applying Bayes' formula [9] with,

$$\pi(p|u_m) = \frac{\pi(u_m, p)}{\pi(u_m)} \text{ and } \pi(u_m|p) = \frac{\pi(u_m, p)}{\pi_{pr}(p)},$$

it holds that,

$$\pi(p|u_m) = \frac{\pi_{pr}(p)\ \pi(u_m|p)}{\pi(u_m)} \propto \pi_{pr}(p)\ \pi(u_m|p).$$

Considering the additive model (3.9) and Theorem 3,

$$\pi(p|u_m) = \frac{\pi_{pr}(p)\ \pi(u_m|p)}{\pi(u_m)} \propto \pi_{pr}(p)\ \pi_{noise}(e).$$

□

For practical reasons the solution of the inverse problem provided by the posterior distribution is not sufficient. Instead, the unknown parameters are ascribed a value by evaluating a point estimator. Different estimates of the unknown parameters can be evaluated based on the posterior probability distribution. One possible point estimator is the *Maximum A Posteriori Estimator (MAP).*

Definition 14. *The **Maximum A Posteriori Estimators** are points of the parameter space at which the posterior probability distribution is maximized,*

$$p_{MAP} = arg\max_{p} \pi_{post}(p).$$

Remark. *Evaluating the MAP results in an optimization problem that might either not have a solution at all, or not provide a unique solution. The computational difficulties are similar to the ones we approach by classic regularization techniques, e.g. the introduced Tikhonov Regularization* (3.6)*. The correlation between the Bayesian Model and the Tikhonov Regularization will be apprehended in Theorem 7.*

3.3.1 Bayesian Normal Linear Model

Definition 15. *A **multivariate Normal (Gaussian) random variable** $X \in \mathbb{R}^n$ with the mean $x_0 \in \mathbb{R}^n$ and the symmetric positive definite covariance matrix $\Gamma \in \mathbb{R}^{n\times n}$ holds the probability density,*

$$\pi(x) = \left(\frac{1}{2\pi|\Gamma|}\right)^{n/2} exp\left(-\frac{1}{2}(x-x_0)^T\Gamma^{-1}(x-x_0)\right) \text{ with } |\Gamma| = det(\Gamma).$$

We write $X \sim \mathcal{N}(x_0, \Gamma)$.

Henceforth, we consider the additive model (3.9), and the prior random variable, as well as the noise random variable holding a Normal probability distribution (Definition 15). In Section 3.4.2 values are assigned to the model parameters. The covariance matrix represents the spread of the random variable. A diagonal entry of the covariance matrix refers to the corresponding marginal variance of one component of the multivariate random variable. The independence of the parameters among themselves holds if a diagonal structure of the covariance matrix is ensured.

Definition 16. *The **Normal prior probability distribution** $p \sim \mathcal{N}(p_0, \Gamma_{pr})$ is defined as,*

$$\pi_{pr}(p) \propto exp\big(-\frac{1}{2}(p-p_0)^T\Gamma_{pr}^{-1}(p-p_0)\big), \tag{3.12}$$

with the mean $p_0 \in \mathbb{R}^n$ and the symmetric positive definite covariance matrix $\Gamma_{pr} \in \mathbb{R}^{n\times n}$.

Definition 17. *The **Normal noise probability distribution** $e \sim \mathcal{N}(e_0, \Gamma_{noise})$ is defined as,*

$$\pi_{noise}(e) \propto exp\big(-\frac{1}{2}\ (e-e_0)^T\ \Gamma_{noise}^{-1}\ (e-e_0)\big), \tag{3.13}$$

with the mean $e_0 \in \mathbb{R}^m$ and the symmetric positive definite covariance matrix $\Gamma_{noise} \in \mathbb{R}^{m\times m}$.

Theorem 5. *The posterior distribution of the additive model based on the Normal prior probability distribution*

$$p \sim \mathcal{N}(p_0, \Gamma_{pr}),$$

and the Normal noise probability distribution

$$e \sim \mathcal{N}(e_0, \Gamma_{noise}),$$

is defined by,

$$\pi_{post}(p) \propto exp\big(-\frac{1}{2}(p-p_0)^T\Gamma_{pr}^{-1}(p-p_0)-\frac{1}{2}(F(p)-u_m-e_0)^T\ \Gamma_{noise}^{-1}(F(p)-u_m-e_0)\big). \tag{3.14}$$

Proof. Considering the additive model (3.9) the posterior distribution can be evaluated by (3.11) of Theorem 4. □

Theorem 6. *The Normal posterior distribution is analytically computable for the linear additive model,*

$$F : \mathbb{R}^n \to \mathbb{R}^m,$$

with the given matrix $A \in \mathbb{R}^{m\times n}$,

$$U_m = A(P) + E,$$

and the Gaussian random variables,

$$P : \Omega \to \mathbb{R}^n \text{ and } E, U_m : \Omega \to \mathbb{R}^m$$

such that,

$$p \sim \mathcal{N}(p_0, \Gamma_{pr}),$$
$$e \sim \mathcal{N}(e_0, \Gamma_{noise}).$$

The Normal posterior distribution evolves characterized by the posterior mean and the posterior covariance matrix,

$$\pi_{post}(p) \propto exp\big(-\frac{1}{2}(p - p_{post})^{\mathsf{T}}\Gamma_{post}^{-1}(p - p_{post})\big), \tag{3.15}$$

$$p_{post} = \big(\Gamma_{pr}^{-1} + A^{\mathsf{T}}\Gamma_{noise}^{-1}A\big)^{-1}\big(A^{\mathsf{T}}\Gamma_{noise}^{-1}(u_m - e_0) + \Gamma_{pr}^{-1}p_0\big), \tag{3.16}$$

$$\Gamma_{post} = \big(\Gamma_{pr}^{-1} + A^{\mathsf{T}}\Gamma_{noise}^{-1}A\big)^{-1}. \tag{3.17}$$

The posterior covariance matrix is a symmetric positive definite matrix [26, pp. 74-76].

Proof. Let

$$\Gamma = \begin{bmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{bmatrix} \in \mathbb{R}^{(n+m)\times(n+m)}$$

be a positive definite symmetric matrix, and $\Gamma_{11} \in \mathbb{R}^{n\times n}$, $\Gamma_{22} \in \mathbb{R}^{m\times m}$. Obviously, it holds that $(\Gamma_{12})^{\mathsf{T}} = \Gamma_{21}$. The (positive definite) Schur complements are well defined by

$$S_{22} = \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21} \text{ and } \; S_{11} = \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}. \tag{3.18}$$

The inverse of Γ is computed by means of the Gaussian elimination [26, pp. 74-75],

$$\Gamma^{-1} = \begin{bmatrix} S_{22}^{-1} & -S_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1}\\ S_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1} & S_{11}^{-1}\end{bmatrix}. \tag{3.19}$$

By setting the off-diagonal blocks of Γ^{-1} equal up to transpose it holds,

$$-\Gamma_{22}^{-1}\Gamma_{21}S_{22}^{-1} = S_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1}. \tag{3.20}$$

Consider the two Gaussian random variables $P : \Omega \to \mathbb{R}^n$ and $U_m : \Omega \to \mathbb{R}^m$ whose joint density $\pi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+$ is defined as,

$$\pi(p, u_m) \propto exp\bigg(-\frac{1}{2}\begin{bmatrix}p - p_0\\ u_m - u_0\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}\Gamma_{11} & \Gamma_{12}\\ \Gamma_{21} & \Gamma_{22}\end{bmatrix}^{-1}\begin{bmatrix}p - p_0\\ u_m - u_0\end{bmatrix}\bigg).$$

Since Bayes' formula gives $\pi(p|u_m) \propto \pi(p, u_m)$, the joint density is considered as a function of p. By shifting the coordinate origin such that $p_0 = 0$ and $u_0 = 0$, and taking (3.20) into account we get,

$$\pi(p, u_m) \propto exp\bigg(-\frac{1}{2}\big(p^{\mathsf{T}}S_{22}^{-1}\; p - 2p^{\mathsf{T}}S_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1}u_m + u_m^{\mathsf{T}}S_{11}^{-1}u_m\big)\bigg). \tag{3.21}$$

By completing the quadratic form in the exponential into squares with respect to p it holds that,

$$\pi(p, u_m) \propto exp\Big(-\frac{1}{2}\big((p - \Gamma_{12}\Gamma_{22}^{-1}u_m)^\mathsf{T} S_{22}^{-1}\big(p - \Gamma_{12}\Gamma_{22}^{-1}u_m\big) + c\big)\Big),$$

where c corresponds to a term that is mutually independent of $P = p$, therefore can be left out of consideration.
The probability distribution of P conditioned on $U_m = u_m$ evolves,

$$\pi(p|u_m) : \mathbb{R}^n \to \mathbb{R}^+, \tag{3.22}$$

$$\pi(p|u_m) \propto exp\Big(-\frac{1}{2}(p - \bar{p})^\mathsf{T} S_{22}^{-1}(p - \bar{p})\Big), \tag{3.23}$$

$$\begin{aligned} \text{with } \ \bar{p} &= p_0 + \Gamma_{12}\Gamma_{22}^{-1}(u_m - u_0), \\ S_{22} &= \Gamma_{11} - \Gamma_{12}\Gamma_{22}^{-1}\Gamma_{21}. \end{aligned} \tag{3.24}$$

The marginal density of P is evaluated by integrating the joint probability density with respect to u_m,

$$\pi(p) = \int_{\mathbb{R}^m} \pi(p, u_m)\ du_m.$$

Setting $L = \Gamma^{-1}$ as in (3.19),

$$L = \begin{bmatrix} S_{22}^{-1} & -S_{22}^{-1}\Gamma_{12}\Gamma_{22}^{-1} \\ S_{11}^{-1}\Gamma_{21}\Gamma_{11}^{-1} & S_{11}^{-1} \end{bmatrix} = \begin{bmatrix} L_{11} & L_{12} \\ L_{21} & L_{22} \end{bmatrix}, \tag{3.25}$$

and rearranging the exponential term to squares with respect to u_m in (3.21) gives,

$$\begin{aligned} \pi(p, u_m) &\propto exp\Big(-\frac{1}{2}\big(p^\mathsf{T} L_{11}\ p - 2p^\mathsf{T} L_{12} u_m + u_m^\mathsf{T} L_{22} u_m\big)\Big) \\ &\propto exp\Big(-\frac{1}{2}\Big(\big(u_m + L_{22}^{-1}L_{21}p\big)^\mathsf{T} L_{22}\big(u_m + L_{22}^{-1}L_{21}p\big) + p^\mathsf{T}\big(L_{11} - L_{12}L_{22}^{-1}L_{21}\big)p\Big)\Big). \end{aligned}$$

We integrate with respect to u_m and refer to

$$T_{22} = L_{11} - L_{12}L_{22}^{-1}L_{21},$$

as the corresponding Schur complement of L such that,

$$\pi(p) \propto exp\Big(-\frac{1}{2}p^\mathsf{T} T_{22}\ p\Big).$$

Since (3.19) defines

$$T_{22}^{-1} = (L^{-1})_{11} = \Gamma_{11},$$

the marginal density of P is

$$\pi(p) \propto exp\Big(-\frac{1}{2}(p - p_0)^\mathsf{T}\Gamma_{11}^{-1}(p - p_0)\Big).$$

Analogously, the marginal density of U_m is

$$\pi(u_m) \propto exp\bigg(-\frac{1}{2}(u_m - u_0)^{\mathsf{T}}\Gamma_{22}^{-1}(u_m - u_0)\bigg).$$

The elaborated theory is applied to the linear inverse model, whereby Bayes' formula gives,

$$\begin{aligned}\pi(p|u_m) &= \pi_{pr}(p)\ \pi_{noise}(u_m - A(p)),\\ &\propto exp\bigg(-\frac{1}{2}(p-p_0)^{\mathsf{T}}\Gamma_{pr}^{-1}(p-p_0) - \frac{1}{2}(u_m - A(p) - e_0)^{\mathsf{T}}(\Gamma_{noise})^{-1}(u_m - A(p) - e_0)\bigg).\end{aligned}$$

Since P and E are Gaussian independent random variables and we consider a linear model such that U_m results as a Gaussian random variable, it holds that,

$$\begin{aligned}&E\left\{\begin{bmatrix}P\\U_m\end{bmatrix}\right\} = \begin{bmatrix}p_0\\u_0\end{bmatrix},\ u_0 = A(p_0) + e_0,\\ &E\left\{(P-p_0)(P-p_0)^{\mathsf{T}}\right\} = \Gamma_{pr},\\ &E\left\{(U_m - u_0)(U_m - u_0)^{\mathsf{T}}\right\} = \Big\{\big(A(P-p_0) + (E-e_0)\big)\ \big(A(P-p_0) + (E-e_0)\big)^{\mathsf{T}}\Big\}\\ &\qquad\qquad = A\ \Gamma_{pr}A^{\mathsf{T}} + \Gamma_{noise},\\ &E\left\{(P-p_0)(U_m-u_0)^{\mathsf{T}}\right\} = E\left\{(P-p_0)(A(P-p_0) + (E-e_0))^{\mathsf{T}}\right\} = \Gamma_{pr}A^{\mathsf{T}},\end{aligned}$$

such that the covariance matrix of the joint probability density evolves

$$cov\begin{bmatrix}P\\U_m\end{bmatrix} = E\left\{\begin{bmatrix}P-p_0\\U_m-u_0\end{bmatrix}\begin{bmatrix}P-p_0\\U_m-u_0\end{bmatrix}^{\mathsf{T}}\right\} = \begin{bmatrix}\Gamma_{pr} & \Gamma_{pr}A^{\mathsf{T}}\\ A\ \Gamma_{pr} & A\ \Gamma_{pr}A^{\mathsf{T}} + \Gamma_{noise}\end{bmatrix}.$$

By (3.23) and (3.24) it follows that

$$p_{post} = p_0 + \Gamma_{pr}A^{\mathsf{T}}(A\ \Gamma_{pr}A^{\mathsf{T}} + \Gamma_{noise})^{-1}(u_m - F(p_0) - e_0), \tag{3.26}$$

$$\Gamma_{post} = \Gamma_{pr} - \Gamma_{pr}A^{\mathsf{T}}(A\ \Gamma_{pr}A^{\mathsf{T}} + \Gamma_{noise})^{-1}A\ \Gamma_{pr}. \tag{3.27}$$

Theorem 6 holds with consideration of the Woodbury Matrix Identity [53, p. 4] that rearranges (3.27) to,

$$\Gamma_{post} = \big(\Gamma_{pr}^{-1} + A^{\mathsf{T}}\Gamma_{noise}^{-1}A\big)^{-1}. \tag{3.28}$$

Since $(A^{\mathsf{T}}\Gamma_{noise}^{-1}A)$ defines a symmetric positive definite matrix it follows that the posterior covariance matrix is a symmetric positive definite matrix. □

Remark. *It holds that $\Gamma_{post} \leq \Gamma_{pr}$, with (3.27) of Theorem 6. The uncertainty of the model after adding measurements (information) is decreased. In the case of the introduced Normal Linear model the expectation value p_{post} of the posterior distribution is equivalent to the MAP (Definition 14). For the non-linear case the posterior distribution with its properties cannot be computed analytically.*

Theorem 7. *Maximizing the posterior distribution* (3.14) *yields the MAP. Evaluating the MAP is equivalent to minimizing the functional* $J(p)$*,*

$$\begin{aligned} J(p) &:= ||F(p) - u_m||^2_{\Gamma_{noise}} + ||p - p_0||^2_{\Gamma_{pr}}, \\ p_{MAP} &= \arg\min_p \ J(p). \end{aligned} \tag{3.29}$$

Proof. Maximizing the posterior distribution defined by an exponential function with a negative exponent (3.14) yields the following minimization problem,

$$\begin{aligned} &\Leftrightarrow \arg\min_p \ \big((p - p_0)^T \Gamma_{pr}^{-1}(p - p_0) + (F(p) - u_m)^T \ \Gamma_{noise}^{-1}(F(p) - u_m)\big), \\ &\Leftrightarrow \arg\min_p \ ||F(p) - u_m||^2_{\Gamma_{noise}} + ||p - p_0||^2_{\Gamma_{pr}}. \end{aligned}$$

□

In the context of the Bayesian Normal model the evaluation of the *MAP* by (3.29) is similar to the Tikhonov regularization (3.6). The statistical approach declares the Tikhonov regularization parameters in a meaningful way.

Remark. *Choosing a Gaussian distribution for the parameters of the model is debatable. It holds that the time of death and the thermal parameters cannot be ascribed a negative value, since all thermal parameters are demonstrably positive given by the law of nature and a non-positive time would imply that the corpse would still be alive. We know for a fact that the parameters do not follow a purely Gaussian distribution. Nevertheless, the use of Gaussian probability distributions have established successfully in the field of statistical inversion theory. The behavior of many real-world incidents is modeled by Gaussian distributions, since Gaussian distributions are easy to construct and to compute. The Gaussian distribution is entirely outlined by its expectation value and its variance. Further, the central limit theorem states that a non-Gaussian distribution approaches a Gaussian distribution by increasing sample size. Details of the central limit theorem can be found in [9, p. 315].*

3.3.2 Linearized Model

To apply the theory of the *Bayesian Normal Linear Model* (Section 3.3.1) we consider a linear model that is evaluated by the local approximation of the Forward Model. Therefore, the non-linear posterior distribution is linearized at the *MAP* with $\hat{p} = p_{MAP}$ by first-order Taylor approximation,

$$F(p) \approx F(\hat{p}) + F'(\hat{p})(p - \hat{p}). \tag{3.30}$$

The linearized model can now be applied in the context of the theory of the additive Bayesian Normal Linear Model, as introduced in Section 3.3.1, to evaluate the posterior distribution and its characteristics by Theorem 6.

3.4 Posterior Distribution

The linearization simplifies the non-linear model. The non-Gaussian part of the posterior distribution that originates from the non-linear Forward Model cannot be neglected. Therefore, we compare the model of the original posterior distribution to the locally approximated linear one given in (3.30).

3.4.1 Sampling

We utilize the *Metropolis-Hastings algorithm* to explore the posterior probability distribution (3.14) by generating samplings of the posterior distribution,

$$d = \{d_1, \ldots, d_n\},\ n \in \mathbb{N}^+. \tag{3.31}$$

Metropolis et al. proposed the algorithm in 1953 [35]. Hastings extended the algorithm in 1970 [23]. The *MH-algorithm* is a well-established technique to earn a profound insight and understanding of the behavior of a given probability density π referred to as the target density. The target density is defined on a state space D and computable up to a multiplying constant. The algorithm is based on the mathematical model of a Markov chain. The current state of the chain (the random variable) depends on the previously computed states. Hence, the algorithm generates correlated variables. The Markov chain is ergodic stating that the random process is nearly independent of its initial state and it is stationary regarding the target density, i.e. the algorithm converges to the stationary state, if $D(t) \to \pi(d)$, then $D(t+1) \to \pi(d)$. The states converge towards the target density. A proposal density proposes the candidates for the data set. Chapter 6 of *Introducing Monte Carlo Methods with R* by Robert and Casella [44] gives a well-founded overview of the algorithm and the theory of Markov chains. The publication of Hasting [23] summarizes the idea of Metropolis and provides the theory of the method accompanied by further techniques and examples.
By means of the *MH-algorithm*, the Gaussianity of the posterior distribution is analyzed. After the sampling the mean and the variance of the proceeding data set (3.31) can be evaluated.
For the purpose of this thesis, the characteristics of a Gaussian distribution are reviewed to confirm the assumption of a local Gaussian behavior of the posterior distribution [9, 15]. Further, the sample-based density of (3.14) is compared to the analytically computed density of the linearized model (3.15) from Theorem 6. The goal is to verify that the posterior distribution of the linearized model follows in large part the original (to some extent non-linear) posterior distribution. Thus, further computations of properties of the non-linear posterior distribution do not rely on the expensive sample-based method.
Following, the basic steps of the *Metropolis Hastings algorithm* are outlined.

Algorithm 1 Metropolis-Hastings Algorithm

Input: Target density $\pi(\cdot)$, proposal density $q(\cdot,\cdot)$, number of iterations $n \in \mathbb{N}$, initial parameter d_1

Output: Data set $d = \{d_1, \ldots, d_n\}$

Set $k \leftarrow 1$
while $k \leq n$ **do**
  $d_c \sim q(d_k, d_c)$ ▷ Draw a candidate from the proposal distribution
  $\alpha(d_k, d_c) = \min\left(1, \frac{\pi(d_c)q(d_c,d_k)}{\pi(d_k)q(d_k,d_c)}\right)$ ▷ Evaluate acceptance ratio.
  $U \sim Uniform(0,1)$ ▷ Generate U from uniform distribution in $[0,1]$
  if $U < d_c$ **then** $d_{k+1} \leftarrow d_c$ ▷ Accept candidate
  else $d_{k+1} \leftarrow d_k$ ▷ Neglect candidate
  end if
  Set $k \leftarrow k+1$
end while

For simplification reasons the proposal distribution is defined by a uniform distribution. In the case of a symmetric proposal distribution,

$$q(d_k, d_c) = q(d_c, d_k),$$

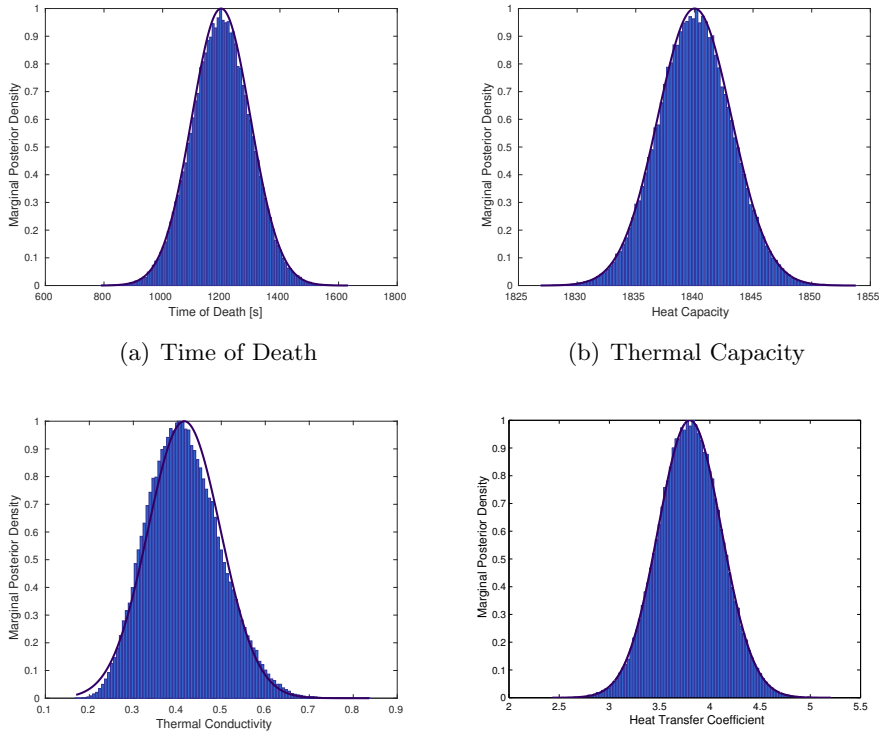
the acceptance probability simplifies to

$$\alpha(d_k, d_c) = \min\left(1, \frac{\pi(d_c)}{\pi(d_k)}\right).$$

For a multivariate target distribution the candidates can be updated component-wise following the steps of the algorithm. During one step a distinct component is updated by the proposal distribution, whereby the other components are fixed, and the acceptance rate is evaluated. It is sufficient to choose the components sequentially and not in a random order [46, 23].

Remark. *Drawn candidates that involve a negative time, a negative heat capacity, a negative thermal conductivity, or a negative heat transfer coefficient, are allocated a probability of zero. Therefore, these candidates are never accepted. The allocation is based on reasons given before (Remark on page 34).*

A data set $d = \{d_1, \ldots, d_n\}$ ($d_i \in \mathbb{R}^4$) of $n =$ 1e6 random variables of the posterior distribution is drawn by the *MH-algorithm.* The quantity assures that the set holds as a profound basis to uncover the underlying structure of the distribution. Starting, the histogram of the data set is constructed by observing the marginal parameters individually. The histograms shown in Figures 11(a), 11(b) and 11(d) give reason to expect good results of the marginal posterior distributions following a Gaussian density. The histogram in Figure 11(c) suggests that the marginal posterior distribution rather follows a Log-normal distribution.



(a) Time of Death (b) Thermal Capacity

(c) Thermal Conductivity (d) Heat Transfer Coefficient

Figure 11: Histogram of the Marginal Random Variables of the Sampled Based Posterior Distribution

The data is mapped into characteristics features [9, 15].

Definition 18. *The **mean** $\bar{d}$ is the average of the data set,*

$$\bar{d} = \frac{1}{n} \sum_{n}^{i=1} d_i.$$

Definition 19. *The **median** d_m is the point below which and above which half of the data lays,*

$$P(D \leq d_m) = P(D \geq d_m) = 0.5.$$

Definition 20. *The **mode** is the data point which most frequently occurs. On a histogram it appears as the highest bar.*

For the Gaussian case it holds that the mean, the median, and the mode are identical. The histogram visually confirms that the data set holds approximately the same value for the mean, the median, and the mode.

Definition 21. *The **variance** measures how far the data set is spread out,*

$$s^2 = \frac{1}{n} \sum_{n}^{i=1} (d_i - \bar{d})^2.$$

On the one hand, we compute the introduced statistical properties of the data set d. On the other hand, the posterior covariance matrix for the Normal linearized model is analytically evaluated by (3.17). By comparing the variance of the distinct parameters of the sampled data $s_i^2,\ i = 1, \dots, 4$, to the marginal variance of the linearized model $\sigma_i,\ i = 1, \dots, 4$, the second important aspect of the validation process is outlined: the two distributions provide an approximately similar marginal variance. The relative difference $r_i = \frac{s_i^2 - \sigma_i^2}{s_i^2}$ gives,

$$|r_t| \leq 4\text{e}{-3},\ |r_c| \leq 8\text{e}{-3},\ |r_\gamma| \leq 2\text{e}{-3},\ |r_\kappa| \leq 2\text{e}{-1},$$

where the relative error of the conductivity, as already assumed in Figure 11(c), stands out.

Definition 22. *The **skewness** is a statistical measure of the symmetry or precisely the amount of asymmetry of a probability distribution about its mean,*

$$g = \frac{1}{n} \sum_{n}^{i=1} \frac{(d_i - \bar{d})^3}{s^3}.$$

A Gaussian distribution is symmetric and possesses a single peak. Therefore, the mean, the mode, and the median are at the centre of the distribution and the tails are exact mirror images of each other. The skewness defines the extent to which a distribution differs from a Gaussian distribution. Hence, the skewness of a perfect Gaussian distribution is zero. The skewness computes to nearly zero for all parameters of the sampled data,

$$|g_t| \leq 5\text{e}{-3},\ |g_c| \leq 2\text{e}{-2},\ |g_\gamma| \leq 6\text{e}{-3},\ |g_\kappa| \leq 4\text{e}{-1},$$

where the value for the conductivity stands out again.

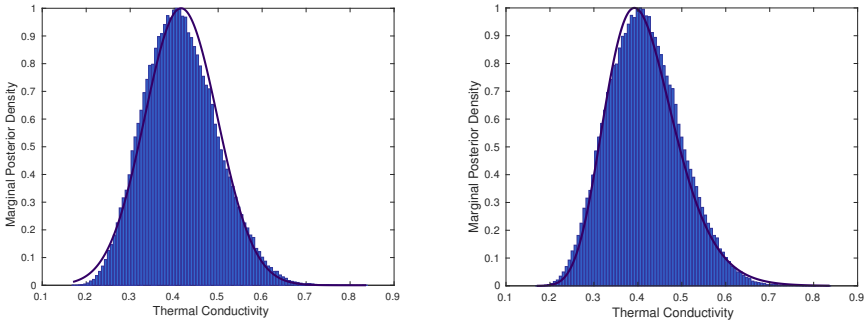
Definition 23. *The **kurtosis** describes the shape of a probability distribution regarding the tail,*

$$k = \frac{1}{n} \sum_{n}^{i=1} \frac{(d_i - \bar{d})^4}{s^4}.$$

The kurtosis of a Gaussian distribution is 3. A kurtosis greater than 3 implies that the distribution produces more extreme values than a Gaussian distribution. A small kurtosis goes along with a data set scattering evenly with few outliers. The data set is analyzed and computes a kurtosis of approximately 3 for all parameters,

$$|k_t - 3| \leq 1\text{e}{-2},\ |k_c - 3| \leq 4\text{e}{-2},\ |k_\gamma - 3| \leq 4\text{e}{-2},\ |k_\kappa - 3| \leq 8\text{e}{-2}.$$

The distribution of the conductivity lets us assume a Log-normal distribution in accordance to the statement given to Figure 11(c), though the other three parameters let us assume a Gaussian distribution. The nuances that differentiate a Normal and a Log-normal distribution are minimal, as seen in Figure 12.



(a) Normal Density (b) LogNormal Density

Figure 12: Histogram of the Thermal Conductivity

In conclusion, the sampling of the posterior distribution mostly confirms the assumption of the posterior distribution following locally a multivariate Gaussian distribution. Further purposes permit an analytic computation of the local statistical properties of the Normal linearized model as representatives of the posterior distribution (3.14).

3.4.2 Determining the Model Parameters

References in the literature and experiments determine prior knowledge. Information of the thermal parameters is accessible from the retailer [25], such that the expectation value of the heat capacity: $\bar{c} = 1840\frac{J}{kg°\mathrm{C}}$, the expectation value of the thermal conductivity: $\bar{\kappa} = 0.4\frac{W}{m°\mathrm{C}}$, and the expectation value of the heat transfer coefficient: $\bar{\gamma} = 3.8\frac{W}{m^2\,°\mathrm{C}}$ are defined. At the crime scene the forensical expert evaluates a first guess of the time of death t_0 in seconds. A prior uncertainty exists and the potential temperature dependence of the thermal parameters is embedded in the variance (the spread) of the parameters [34]. The marginal variance of the thermal parameters is minor in comparison to the marginal variance of the time of death, the greatest unknown of the model. The prior covariance matrix is a diagonal matrix that features the independence between the parameters,

$$\Gamma_{pr} = diag\big(\sigma_t^2, \sigma_c^2, \sigma_\kappa^2, \sigma_\gamma^2\big) \in \mathbb{R}^{4\times 4}.$$

The likelihood distribution quantifies the information that enters the model in the form of temperature measurements at certain points of measurement. Since Theorem 3 holds, the information is coded in the noise distribution. We assume the error mean of $e_0 = 0$. The diagonal entries of the noise covariance matrix, the marginal noise variance of every distinct point of measurement x_i, may differ due to different reliability of the information based on the location over space and time and certain measurement techniques,

$$\Gamma_{noise} = diag\big(\sigma_{x_1}^2, \sigma_{x_2}^2, \ldots, \sigma_{x_m}^2\big) \in \mathbb{R}^{m\times m}.$$

Parameter p_i		Unit	Mean	Range	Variance σ_i^2
Time of Death	t	sec	t_0	$[0, 1e9]$	$1e9$
Heat Conductivity of Polyethylene	κ	$\frac{W}{m°\mathrm{C}}$	0.4	$[0.33, 0.45]$	0.1
Heat Capacity of Polyethylene	c	$\frac{J}{kg°\mathrm{C}}$	1840	$[1680, 2000]$	100
Heat Transfer Coefficient of Polyethylene	γ	$\frac{W}{m^2\ °\mathrm{C}}$	3.8	$[3.0, 4.6]$	1.0
Noise Level Low	e	°C	0	$[0, 0.01]$	0.01
Noise Level Medium	e	°C	0	$[0,\ 0.1]$	0.1
Noise Level High	e	°C	0	$[0,\ 1.0]$	1.0

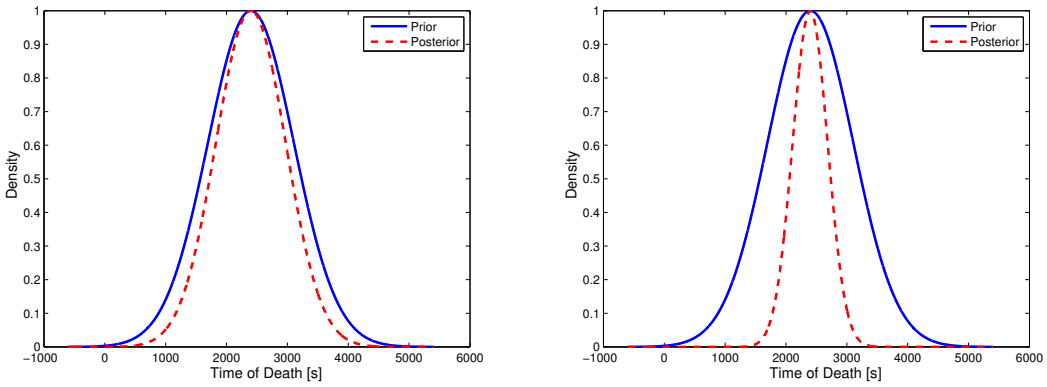
Choosing the right properties establishes the basis for the development of a design in consideration of the allocation of points of measurement. This aspect is fundamental to successfully approach the topic of the *Optimal Design of Experiments* in Section 4.

3.4.3 Computation of the Posterior Distribution

Different posterior distribution arise depending on a distinct selection of points of measurement. The sets of regarded points of measurement differ by the quantity and the quality of the chosen points of measurement. The *MAP* is evaluated, and the model is locally linearized. The posterior covariance matrix that displays the uncertainty of the model is computed by (3.17). Since we focus upon one unknown: the time of death, the marginal distribution of the time of death is plotted and analyzed. The following plots and figures refer to the normalized density.

The Prior versus the Posterior Distribution

The posterior distribution is evaluated by adding information to the model by means of the likelihood information. The uncertainty of the prior distribution decreases evaluating the posterior distribution. The quality of the data is represented by the noise distribution. In Figure 13 the posterior distribution is evaluated based on two different data sets that differ by the magnitude of the noise level. The quality of the data has a great impact on the certainty of the posterior distribution.



(a) Medium Noise Level: 0.1

(b) Low Noise Level: 0.01

Figure 13: The prior distribution versus the posterior distribution regarding different data quality

By considering a lower noise level, representing a good quality of the data, the posterior distribution emerges more slender resulting from a smaller uncertainty. Figure 14 confirms the assumption by comparing different posterior distributions based on different noise levels. The better the quality of the data, the greater the certainty of the posterior distribution evolves.

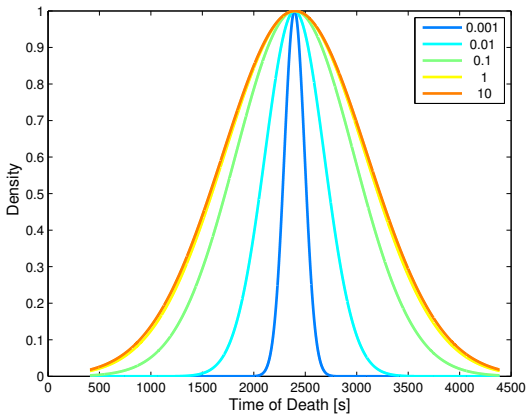


Figure 14: The uncertainty of the posterior distribution decreases with declining noise level.

The Quantity of Measurements

Collecting information has a down-scaling effect on the uncertainty of the model. The certainty of the MAP increases with the number of measurements. Further measurements are added over time, whereby the point of measurement is locally fixed.

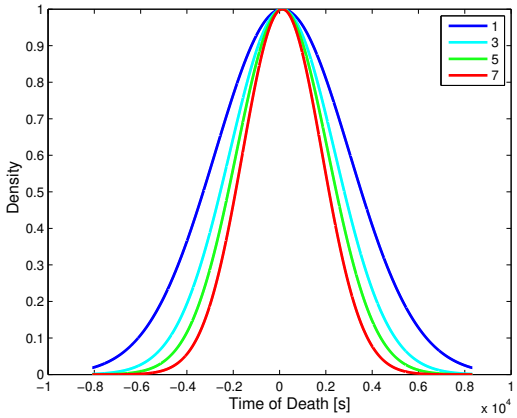
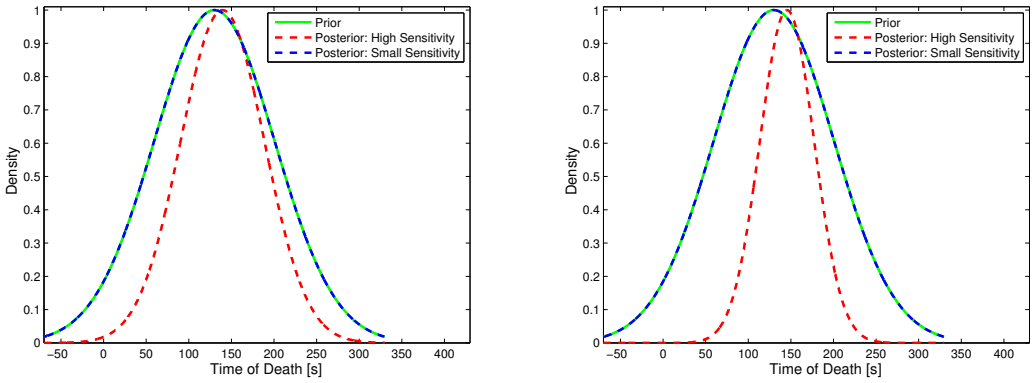


Figure 15: The uncertainty of the posterior distribution decreases with the increase of number of measurements.

Figure 15 displays that the marginal variance of the posterior distribution decreases with the incorporation of further measurements. The observed data integrates information into the model, such that better knowledge of the model is obtained. A similar behavior is observed when comparing the prior distribution to the posterior distribution as displayed in Figure 13. Invariably it holds that data ascertainment increases the certainty as apprehended in Theorem 6.

The Location of Measurements

We analyze the certainty of the posterior distribution regarding different locations of the points of measurement. The number of measurements is constant. Furthermore, the point in time of the measurements is fixed. By considering points of measurement with a high sensitivity the uncertainty of the model is reduced. Figure 16(a) and Figure 16(b) illustrate the posterior distribution that considers points of measurement with a very low sensitivity as shown by the blue line. The posterior distribution is nearly unchanged with respect to the prior distribution (green). Furthermore, the posterior distribution that considers very sensitive points of measurement is displayed by the red line. In marked contrast the posterior distribution deviates clearly from the prior distribution. The effect intensifies by increasing number of measurements as seen by comparing Figure 16(a) to Figure 16(b).



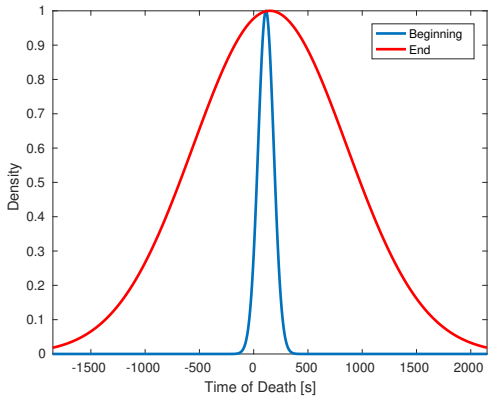
(a) One Measurement

(b) Five Measurements

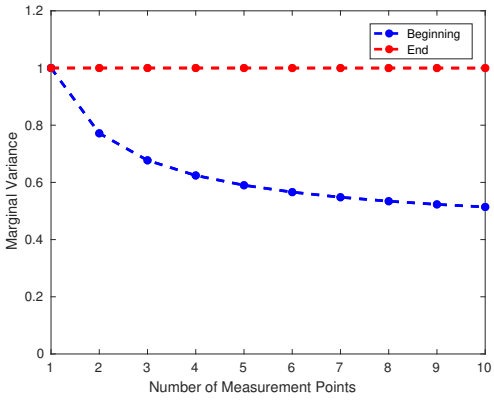
Figure 16: The uncertainty of the posterior distribution varies by incorporating sensitive or non-sensitive points of measurement.

The Instance of Time for Measurements

The influence of the time component of the measurement process regarding the uncertainty of the model is analyzed. Therefore, the number of measurements is fixed and the location of the points of measurement in space is fixed. A range of different moments for measurements is realized, either at the beginning of the cooling process, or at the end of the cooling process. Figure 17(a) lets us hypothesize that measuring at the beginning is beneficial in terms of the certainty of the model. Figure 17(b) clearly shows however that adding measurements at the end of the cooling process does not improve the certainty of the model in contradistinction to adding measurements at the beginning of the cooling process. Apparently the blue line, representing the certainty of the model by taking measurements at the end of the cooling process, does not decline even when increasing the number of measurements. The red line, representing the certainty of the model when taking measurements at the beginning of the cooling process, declines clearly as information is added.



(a) Posterior Distribution: 1 Measurement



(b) Marginal Posterior Variance

Figure 17: The Posterior Distribution regarding different points of measurement in time

After laying the groundwork in Section 2 and Section 3 for the computation of posterior distributions in an exemplary manner demonstrates the importance of a well-conceived experimental setting. Since the uncertainty displays the reliability of the estimate a desired target is a small uncertainty. The former computations give insight into the importance of an optimal design. Optimizing the design of the experiment on a global level is of utmost importance.

4 Optimal Design of Experiments

The quality of the model relies on a proper estimate of the unknown parameters of Definition 7. A well-established parameter identification correlates with the quality of the experimental data. Therefore, one has to focus on a feasible design of the experiment that is adjusted in terms of a statistical criterion. The accuracy, the efficiency, and the stability of the model need to be optimized. The parameters of the model are uncertain and vary among human bodies. Considering e.g. the muscle tissue, it holds that different bodies, or even different parts of the muscle tissue, have an individual structure [5, pp. 48] that might result in distinct thermal properties.
The time of death is determined by back-calculation based on the measured data evaluated by the inverse problem stated in Definition 9. The process of data ascertainment is often expensive, time-consuming or restricted. Especially in the case of the temperature measurements on a corpse, the experiment cannot be repeated under the same condition. Measuring the body temperature is limited to a certain number of allowed measurements and furthermore the placement of the measurements underlay certain – legal and anatomical – restrictions. Two important aspects are examined: When and where shall the measurements of the temperature take place?
In 1993 the topic of the *Optimal Design of Experiments (DOE)* has been treated successfully by Pukelsheim among others [41]. Linear models are examined and the design problem is exclusively approached with the help of statistic tools in combination with linear algebra and convex analysis. The approach of stating an experimental design in this thesis follows the work of Körkel [6, 29]. The experiment is embedded in a statistical context. Different criteria for the measurement of the statistical quality are introduced [4, 42], where the uncertainty is expressed as a functional that depends on the posterior covariance matrix. Therefore, the posterior covariance matrix assesses the quality of the measurement process and is adjusted by allocating weights to the available points of measurement. By minimizing the uncertainty an optimization problem evolves which is approached by means of recent work of Ghattas [2, 1]. The innovative work of Ghattas provides valuable insights in terms of the *Optimal Experimental Design to Estimate the Time of Death in a Bayesian Context*. The optimization problem is NP-hard[1], which is why an heuristic approach is introduced.

[1]The class of *Non-deterministic polynomial-time (NP)* problems are avoided due to their high complexity. NP-hard problems are at least as hard as the hardest problems in NP.

4.1 Experimental Design

The output of the Forward Model depends on the chosen experimental design.

Definition 24. *Considering u as the solution of the Forward Model by Definition 7 an* ***experiment*** *is defined by a mapping,*

$$M : u \in \mathbb{R}^q \to m \in \mathbb{R}^q,$$

stating the available points of measurement such that $q = N_\tau \cdot N_s$ and,

$$\begin{aligned} &m_j := u(x_{i_j}, t_{k_j}),\ i_j \in \{x_1, \dots, x_{N_s}\},\ k_j \in \{t_1, \dots, t_{N_\tau}\}, \\ &\text{over } N_s \in \mathbb{N}^+ \text{ points in space: } \{x_1, \dots, x_{N_s}\} \subset D, \\ &\text{and } N_\tau \in \mathbb{N}^+ \text{ points in time: } \{t_1, \dots, t_{N_\tau}\} \subset [0, T]. \end{aligned} \tag{4.1}$$

Definition 25. *A* ***design*** *based on an experiment from Definition 24 is defined by the weight vector of the available points of measurement,*

$$\boldsymbol{w} = [w_1, \dots, w_q] \in \{0, 1\}^q.$$

The weight $w_j \in \{0, 1\}$ denotes the (non-)occurrence of a measurement at the corresponding point of measurement m_j, where $w_j = 1$ represents an occurring measurement, respectively $w_j = 0$ the opposite.

For reasons of simplification we first examine the problem with respect to space before integrating the time later. Therefore, the set of available points of measurement

$$m := [x_1, \dots, x_{N_s}] \in \mathbb{R}^{N_s}$$

is given with a corresponding weight $w_i \in \{0, 1\}$ assigned to each point of measurement x_i

$$\boldsymbol{w} = [w_1, \dots, w_{N_s}] \in \{0, 1\}^{N_s}.$$

Furthermore, the number of allowed measurements $l \in \mathbb{N}^+$ is limited by

$$\sum_{i=1}^{N_s} w_i \leq l. \tag{4.2}$$

The goal is to choose the optimal subset of all available points of measurement.

4.2 Design of a Bayesian Normal Linear Model

The experimental design is embedded in the theory of the Bayesian Normal Linear model of Section 3.3.1. The statistical quantities of the model depend on the allocated design displayed by the binary weight vector $\boldsymbol{w} \in \{0, 1\}^{N_s}$.

Definition 26. *The **weight-noise distribution** conditioned on a design characterized by*

$$W = diag(w_1, \dots, w_{N_s}) \in \{0,1\}^{N_s \times N_s},$$

with the noise distribution $e \sim N(0, \Gamma_{noise})$ is defined as

$$\pi_{noise}(e|\boldsymbol{w}) \propto exp\big(-\frac{1}{2} e^T W^{1/2} \Gamma_{noise}^{-1} W^{1/2} e\big).$$

Definition 27. *The **weight-noise inverse covariance matrix***

$$W_\sigma \in \mathbb{R}^{N_s \times N_s}$$

is defined as

$$W_\sigma = W^{1/2}\, \Gamma_{noise}^{-1} W^{1/2} = diag\Big(w_1/\sigma_1^2, \dots, w_{N_s}/\sigma_{N_s}^2\Big). \tag{4.3}$$

Remark. *Evidently, the weight-noise inverse covariance matrix W_σ highlights that the components of the posterior covariance matrix are maximized by collecting no information and minimized by collecting information at every available point of measurement x_i. Considering a weight $w_i = 1$ the corresponding diagonal element of W_σ is unaltered with respect to Γ_{noise}^{-1}. On the contrary, a weight $w_i = 0$ shifts the corresponding diagonal element of W_σ to zero. In this case no information is gathered, such that the element of the posterior variance approaches infinity.*

Theorem 8. *The **design posterior distribution** of the linear additive model (see Theorem 6),*

$$F : \mathbb{R}^n \to \mathbb{R}^q,$$

with the given matrix $A \in \mathbb{R}^{q \times n}$,

$$U_m = A(P) + E,$$

and the Gaussian random variables,

$$P : \Omega \to \mathbb{R}^n \text{ and } E, U_m : \Omega \to \mathbb{R}^q,$$

conditioned on a design characterized by $W = diag(w_1, \dots, w_{N_s}) \in \{0,1\}^{N_s \times N_s}$ states

$$\begin{aligned} \pi(p|\ u_m, \boldsymbol{w}) &\propto exp\big(-\frac{1}{2}(p - p_{post})^T \Gamma_{post}^{-1} (p - p_{post})\big) \\ \text{with } \Gamma_{post} &= (\Gamma_{pr}^{-1} + A^T W_\sigma A)^{-1}, \qquad (4.4) \\ p_{post} &= \big(\Gamma_{pr}^{-1} + A^T W_\sigma A\big)^{-1} \big(A^T W_\sigma (u_m) + \Gamma_{pr}^{-1} p_{pr}\big). \qquad (4.5) \end{aligned}$$

Definition 28. *The **Maximum A Posterior Estimator** conditioned on a design characterized by $W = diag(w_1, \dots, w_{N_s}) \in \{0,1\}^{N_s \times N_s}$ is defined as*

$$p_{MAP}(\boldsymbol{w}) = arg \max_{p \in \mathbb{R}^n} \pi(p|u_m, \boldsymbol{w}).$$

Definition 29. *For the time-dependent problem the* ***weight-noise inverse covariance matrix*** $W_\sigma \in \mathbb{R}^{q\times q}$ *is defined with* $q = N_s \cdot N_\tau$. *The matrix holds the weights for all available points of measurement over space and time. The matrix* W_σ *is a diagonal matrix that holds* $W_j \in \mathbb{R}^{N_s\times N_s}$ *as block entries,*

$$W_\sigma = \begin{bmatrix} W_1 & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & 0 & \ddots & 0 \\ \vdots & \ddots & W_j & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & W_{N_\tau} \end{bmatrix} \in \mathbb{R}^{q\times q}. \tag{4.6}$$

Remark. *Adjusting the weight-noise distribution considering all q points of measurement over time and space is trivial. The design posterior distribution in dependence of a certain design over time and space considering* (4.6) *is computed analogously to* (4.4) *and* (4.5).

4.3 Statistical Quality of a Model

Measuring the quality of the experiment is based on the evaluation of the uncertainty of the model. A common ground is to analyze the posterior covariance matrix since it includes the spread and extent of reliability of the unknown parameters. Different posterior covariance matrices emerge depending on the chosen design.
The Normal Linear model provides a closed form of the posterior covariance matrix. Hence, the posterior covariance matrix is computed analytically and independently of the experimental data. The importance of a valid linearized local model is emphasized once more. In terms of non-linear models the problem proceeds that for one, the posterior covariance matrix cannot be evaluated by a closed form, and for another, the posterior covariance matrix depends on the experimental data. Consequently, for strongly non-linear problems the given approach does not evaluate a meaningful optimal design prior to the experimental performance. One might consider a model based on the posterior covariance matrix over all potential experimental data. Alternatively, the evaluation of the posterior covariance matrix is performed via sampling methods, e.g. as introduced in Section 3.4.1. Evaluating a good estimator for non-linear models is often accompanied with a high computation time [1].

4.3.1 Criteria for Measuring the Uncertainty

The uncertainty of the model is assessed by a functional,

$$\phi : \mathbb{R}^{q\times q} \to \mathbb{R}.$$

The functional outlines a chosen criterion according to the focus of the problem. Different criteria have established in the statistical context of measuring the quality of a model [42].

- The A-optimality criterion observes the average of the marginal variances of all parameters,

$$\phi_A(\Gamma) = \frac{1}{n}\, tr(\Gamma).$$

- The D-optimality criterion observes the determinant of the covariance matrix,

$$\phi_D(\Gamma) = det(\Gamma).$$

- The E-optimality criterion observes the eigenvalues of the covariance matrix,

$$\phi_E(\Gamma) = \lambda_{max}(\Gamma).$$

- The optimality criterion observes a distinct selection of particular parameters of interest. Therefore, we obtain a restricted case of the A-optimality criterion by choosing a vector $c \in \mathbb{R}^n$ laying focus on certain unknown parameters,

$$\phi_C(\Gamma) = c^T \Gamma c.$$

Considering the minimization of the average variance of all unknowns by means of an A-optimal design has established as a popular criterion,

$$\phi_A\big(\Gamma_{post}(\boldsymbol{w})\big) = \frac{1}{n}\, tr\big(\Gamma_{post}(\boldsymbol{w})\big). \tag{4.7}$$

Since we are interested in the time of death of the parameter set in Definition 7, a suitable criterion is to suppress the uncertainty of only one unknown. The other parameters are of no major concern. We choose the C-optimal design to lay focus on the marginal variance of the time of death by defining $c = e_1 \in \mathbb{R}^4$ as the corresponding unity vector,

$$\phi_C\big(\Gamma_{post}(\boldsymbol{w})\big) = c^T\ \Gamma_{post}(\boldsymbol{w})\ c. \tag{4.8}$$

4.4 Optimization Problem

Finding the optimal experimental design is done by solving an optimization problem composed of an objective function and suitable constraints. The constraints here ensure a binary design and a restricted number of allowed measurements that are henceforth stated by $l \in \mathbb{N}^+$.

$$\begin{aligned} \min_{w}\ & \phi\big(\Gamma_{post}(\boldsymbol{w})\big) \\ \text{s.t. }\ & \sum_{i=1}^{q} w_i \leq l \\ & w \in \{0,1\}^q \end{aligned} \tag{4.9}$$

The combinatorial optimization problem (4.9) transforms into the problem of the *Sparse Approximate Solutions to Linear Systems*.

Definition 30. *The Sparse Approximate Solutions to Linear Systems (SAS) treats the computation of the vector $w \in \mathbb{R}^q$ with the minimal number of non-zero entries, such that given the matrix $A \in \mathbb{R}^{k \times q}$, the vector $b \in \mathbb{R}^k$, and the scalar $\epsilon \geq 0$ it holds that*

$$||Aw - b||_2 \leq \epsilon.$$

The *SAS*-problem has been addressed as a fundamental problem in matrix computations and states a NP-hard problem [37]. For binary matrices the problem refers to the *Minimum Weight Solution*. The NP-hardness of the *Minimum Weight Solution* has been treated by Gallager in the context of error corrective coding [19, pp. 196]. This thesis will not proof the NP-hardness of the problem or go further into detail. The background of the theory and the history of NP-completeness, as well as examples of NP-hard problems can be found i.a. by Garey and Johnson [20].

4.4.1 Relaxed Optimization Problem

The NP-hard optimization problem is relaxed by solving (4.9) over the range of continuous values instead of binary values, while $l \in \mathbb{N}^+$ invariably refers to the number of allowed measurements. As from now $\vec{\mathbf{0}} \in \mathbb{R}^q$ refers to the vector of all zeros, respectively $\vec{\mathbf{1}} \in \mathbb{R}^q$ to the vector of all ones.

$$\begin{aligned} \min_{w} \ & \phi\big(\Gamma_{post}(\boldsymbol{w})\big) \\ \text{s.t.} \ & \sum_i w_i \leq l \\ & \vec{\mathbf{0}} \leq \boldsymbol{w} \leq \vec{\mathbf{1}} \end{aligned} \tag{4.10}$$

The constraints are embedded in the objective function by considering a penalty function,

$$p(w) : \mathbb{R}^q \to \mathbb{R}^+,$$

and a multiplier $\lambda \geq 0$ that controls the sparsity of the design. We ensure that the number of allowed measurements is not exceeded in the experiment with choosing a proper λ.

$$\begin{aligned} \min_{w} \ & \phi\big(\Gamma_{post}(\boldsymbol{w})\big) + \lambda p(\boldsymbol{w}), \\ \text{s.t.} \ & \vec{\mathbf{0}} \leq \boldsymbol{w} \leq \vec{\mathbf{1}}. \end{aligned} \tag{4.11}$$

The l_p-norm is chosen as a penalty function,

$$l_p(\boldsymbol{w}) = \Big(\sum_i (w_i)^p \Big)^{1/p}.$$

The l_0-norm defines here the number of non-zero components of the vector. The optimization problem (4.11) including the l_0-norm as means of a penalty function leads us back to the NP-hard problem [37] (see Definition 30). The case of a sparse binary solution will be discussed later in Section 4.4.3. To work around the indicated difficulties and avoid the NP-hardness, the optimization problem refers to (4.11) with the continuous solution

$$\vec{\mathbf{0}} \leq \boldsymbol{w} \leq \vec{\mathbf{1}},$$

and the l_1-regularization as a penalty function,

$$p(w) = l_1(w) = \sum_i |w_i|. \tag{4.12}$$

Theorem 9. *If the penalty function $p(w)$ is convex the optimization problem* (4.11) *has a solution and the solution is unique.*

Proof. The objective function is defined as

$$\phi(\Gamma_{post}(\boldsymbol{w})) + \lambda p(\boldsymbol{w}).$$

The index function or trace function $\phi(\cdot)$ is linear. The weight vector enters linearly in $\Gamma_{post}(\boldsymbol{w})$ by (4.3) and (4.4). The posterior covariance matrix and its inverse are symmetric positive definite matrices by Theorem 6. The strict convexity is ensured since $X \to \phi(X^{-1})$ is a strict convex function on the cone of symmetric positive definite matrices [38, p. 82]. With $p(\boldsymbol{w})$ also being a convex function the unique solution exists [24]. □

4.4.2 Gradient Descent

The optimizaion problem is numerically solved by the Gradient descent. In each iteration the first-order optimization algorithm computes the search direction given by the negative gradient of the functional,

$$d = -\nabla\big(\phi\big(\Gamma_{post}(\boldsymbol{w}) + p(\boldsymbol{w})\big).$$

The penalty term as the l_1-norm is a non-differentiable convex function. The subgradient is evaluated as a generalized gradient and stated in [8]. Considering a step size $\gamma \geq 0$ and the sequence,

$$\boldsymbol{w^0},\ \boldsymbol{w^1},\ \boldsymbol{w^2}, \ldots, \boldsymbol{w^n},$$

the Gradient descent computes,

$$\boldsymbol{w^{n+1}} = \boldsymbol{w^n} - \gamma \cdot \nabla\phi\big(\Gamma_{post}(\boldsymbol{w^n}) + \lambda p(\boldsymbol{w^n})\big).$$

The algorithm converges to the desired global minimum (Theorem 9),

$$\phi\big(\Gamma_{post}(\boldsymbol{w^0})\big) + p(\boldsymbol{w^0}) \geq \phi\big(\Gamma_{post}(\boldsymbol{w^1})\big) + p(\boldsymbol{w^1}) \geq \cdots \geq \phi\big(\Gamma_{post}(\boldsymbol{w^{n+1}})\big) + p(\boldsymbol{w^{n+1}}).$$

Details of the Gradient descent can be found in [24, 18]. Both functionals, (4.7) and (4.8), depend on the posterior covariance design-matrix (4.4). Therefore, the derivative of the posterior covariance design-matrix will be evaluated in the following.

Theorem 10. *Consider the matrix $B(x) \in \mathbb{R}^{n\times n}$ whose elements are functions of the scalar parameter x. Let $B(x)$ be a non-singular matrix for all $x \in \mathbb{R}$. The derivative of the inverse $B^{-1} = B(x)^{-1}$ with respect to x is in [40] defined as*

$$\frac{\partial}{\partial x}B^{-1} = -B^{-1}\frac{\partial B(x)}{\partial x}B^{-1}. \tag{4.13}$$

Proof. Let $b_{ij}(x)$ be the components of $B(x)$ and $\widetilde{b}_{ij}(x)$ be the components of $B^{-1}(x)$,

$$\begin{aligned}
&\sum_{j=1}^{n} b_{ij}(x)\widetilde{b}_{jk}(x) = \delta_{ik} \quad \text{with } \delta_{ik} = \begin{cases} 1, & i = k \\ 0, & i \neq k \end{cases} \\
\Leftrightarrow\ & \frac{\partial}{\partial x}\Big(\sum_{j=1}^{n} b_{ij}(x)\widetilde{b}_{jk}(x)\Big) = 0 \\
\Leftrightarrow\ & \sum_{j=1}^{n}\Big(\frac{\partial b_{ij}(x)}{\partial x}\tilde{b}_{ij}(x) + b_{ij}(x)\frac{\partial \tilde{b}_{ij}(x)}{\partial x}\Big) = 0 \\
\Leftrightarrow\ & \frac{\partial B}{\partial x}B^{-1} = -B\frac{\partial B^{-1}}{\partial x}.
\end{aligned}$$

□

Theorem 11. *The inverse of the posterior covariance design-matrix* (4.4) *is written as a linear combination of independent matrices such that,*

$$\begin{aligned}
&\big(\Gamma_{post}(\boldsymbol{w})\big)^{-1} = \big(\Gamma_{pr}^{-1} + \widetilde{w}_1\ A_{w_1} + \cdots + \widetilde{w}_q\ A_{w_i}\big), \\
&\widetilde{w}_i \in \mathbb{R},\ A_{w_i} \in \mathbb{R}^{q\times q}.
\end{aligned} \tag{4.14}$$

Proof.

With $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \ddots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{q1} & \cdots & a_{qn} \end{bmatrix} \in \mathbb{R}^{q\times n}$, $W_\sigma = \begin{bmatrix} \frac{w_1}{\sigma_1^2} & 0 & \cdots & 0 \\ 0 & \frac{w_2}{\sigma_2^2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{w_q}{\sigma_q^2} \end{bmatrix} \in \mathbb{R}^{q\times q}$,

we set $A_{w_i} = \begin{bmatrix} a_{i1}a_{i1} & \cdots & a_{i1}a_{in} \\ \vdots & \ddots & \vdots \\ a_{in}a_{i1} & \cdots & a_{in}a_{in} \end{bmatrix} \in \mathbb{R}^{n\times n}$, $\widetilde{w}_i = \frac{w_i}{\sigma_i^2} \in \mathbb{R}$,

such that $A^T\ W_\sigma\ A = \begin{bmatrix} \sum_i a_{i1}\widetilde{w}_i a_{i1} & \cdots & \sum_i a_{i1}\widetilde{w}_i a_{in} \\ \vdots & \ddots & \vdots \\ \sum_i a_{in}\widetilde{w}_i a_{i1} & \cdots & \sum_i a_{in}\widetilde{w}_i a_{in} \end{bmatrix} = \sum_{i=1}^{q} \widetilde{w}_i \begin{bmatrix} a_{i1}a_{i1} & \cdots & a_{i1}a_{in} \\ \vdots & \ddots & \vdots \\ a_{in}a_{i1} & \cdots & a_{in}a_{in} \end{bmatrix}$,

and $\Gamma_{post}(\boldsymbol{w}) = \big(\Gamma_{pr}^{-1} + A^T W_\sigma A\big)^{-1} = \big(\Gamma_{pr}^{-1} + \widetilde{w}_1\ A_{w_1} + \cdots + \widetilde{w}_q\ A_{w_i}\big)^{-1}$.

□

Corollary 1. *The derivative of the posterior covariance design-matrix with respect to* $w_i \in \mathbb{R}$ *is defined as*

$$\frac{\partial}{\partial w_i}\Gamma_{post}(\boldsymbol{w}) = -\frac{1}{\sigma_i^2}\big(\Gamma_{post}(\boldsymbol{w})\ A_{w_i}(\boldsymbol{w})\ \Gamma_{post}(\boldsymbol{w})\big).$$

Proof. Applying (4.13) and (4.14) gives,

$$\begin{aligned}\frac{\partial}{\partial w_i}\Gamma_{post}(\boldsymbol{w}) &= \frac{\partial}{\partial w_i}\big(\Gamma_{pr}^{-1} + A^T W_\sigma A\big)^{-1} \\ &= -\Gamma_{post}(\boldsymbol{w})\ \frac{\partial}{\partial w_i}\Gamma_{post}(\boldsymbol{w})^{-1}\ \Gamma_{post}(\boldsymbol{w}) \\ &= -\Gamma_{post}(\boldsymbol{w})\ \frac{\partial}{\partial w_i}\big(\Gamma_{pr}^{-1} + \frac{w_1}{\sigma_1^2}\ A_{w_1} + \cdots + \frac{w_q}{\sigma_q^2}\ A_{w_q}\big)\ \Gamma_{post}(\boldsymbol{w}) \\ &= -\frac{1}{\sigma_i^2}\big(\Gamma_{post}(\boldsymbol{w})\ A_{w_i}(\boldsymbol{w})\ \Gamma_{post}(\boldsymbol{w})\big).\end{aligned}$$

□

Corollary 2. *The gradient of the scalar objective function $\phi(\boldsymbol{w})$ with respect to $\boldsymbol{w} \in \mathbb{R}^q$ is evaluated by,*

$$\nabla\phi\big(\Gamma_{post}(\boldsymbol{w})\big) = \Big[\tfrac{\partial\phi(\boldsymbol{w})}{\partial w_1} \quad \tfrac{\partial\phi(\boldsymbol{w})}{\partial w_2} \quad \cdots \quad \tfrac{\partial\phi(\boldsymbol{w})}{\partial w_q}\Big]^T.$$

By considering the trace function or the index function, the gradient of the objective function is defined as,

$$\begin{aligned}(4.8):\ \frac{\partial\phi_C(\boldsymbol{w})}{\partial w_i} &= \frac{\partial}{\partial w_i}\big(c^T\Gamma_{post}(\boldsymbol{w})c\big) = c^T\big(\frac{\partial}{\partial w_i}\Gamma_{post}(\boldsymbol{w})\big)c \\ &= -\frac{1}{\sigma_i^2}\ c^T\big(\Gamma_{post}(\boldsymbol{w})\ A_{w_i}(\boldsymbol{w})\ \Gamma_{post}(\boldsymbol{w})\big)c \qquad (4.15) \\ (4.7):\ \frac{\partial\phi_A(\boldsymbol{w})}{\partial w_i} &= \frac{\partial}{\partial w_i}\bigg(\frac{1}{n}\ tr\big(\Gamma_{post}(\boldsymbol{w})\big)\bigg) = \frac{1}{n}\ tr\big(\frac{\partial}{\partial w_i}\Gamma_{post}(\boldsymbol{w})\big) \\ &= -\frac{1}{n\cdot\sigma_i^2}\ tr\ \Big(\Gamma_{post}(\boldsymbol{w})\ A_{w_i}(\boldsymbol{w})\ \Gamma_{post}(\boldsymbol{w})\Big). \qquad (4.16)\end{aligned}$$

The algorithm that evaluates the optimal design by (4.11) is outlined in Algorithm 3 in the Appendix.

4.4.3 Integer Optimization Problem

Clearly, the relaxed optimization problem (4.11) permits a greater solution set than the original combinatorial optimization problem (4.9). In the context of an experimental design by Definition 25 a strategy has to be evaluated to ascribe an importance and a meaning to the continuous variables, since the original binary assertion $w_i \in \{0,1\}$ cannot be taken for granted. In the case of finding a binary optimal experimental design corresponding to the relaxed optimization problem (4.11) an integer programming problem arises,

$$\begin{aligned}&\min_w\ \phi(\Gamma_{post}(\boldsymbol{w})) + \lambda p(\boldsymbol{w}) \\ &\quad \text{s.t.}\quad w \in \{0,1\}^q.\end{aligned} \qquad (4.17)$$

This integer programming problem cannot be solved in polynomial time. To evaluate a binary design the continuous approach is embedded in the *Branch and Bound algorithm*

(B&B). The functionality of the algorithm is outlined in [10], whereas the mathematical background and theory is described in detail in [30, 16].
An heuristic approach, detached from the B&B algorithm, is introduced that determines for our computational purposes sufficiently satisfactory results affirmed by the efficiency (Definition 33) applied in Section 4.5.

Definition 31. *The solution of the relaxed optimization problem* (4.11) *is defined as* $\boldsymbol{w^*}$

$$\vec{\mathbf{0}} \leq \boldsymbol{w^*} \leq \vec{\mathbf{1}}$$

satisfying

$$l = \sum_i w_i^*.$$

Definition 32. *The heuristic solution*

$$\boldsymbol{w^d} \in \{0,1\}^q$$

of the continuous solution $\boldsymbol{w}^*$ *and the corresponding variable* l *of Definition 31 is defined by setting the components* w_i^d *such that the* l *locations with largest values in* $\boldsymbol{w}^*$ *are rounded to one:* $w_i^d = 1$*, and the remaining locations to zero,* $w_i^d = 0$*, i.e. these points of measurement will be neglected.*

The heuristic solution of Definition 32 and its corresponding continuous solution of Definition 31 refer to the same instance since

$$\sum_i w_i^* = \sum_i w_i^d.$$

Definition 33. *The efficiency of the heuristic solution* $\boldsymbol{w^d}$ *with respect to the optimal continuous solution* $\boldsymbol{w^*}$ *of the optimization problem* (4.11) *including the objective function* $\phi(\boldsymbol{w})$ *is evaluated through,*

$$\mathit{eff}(\boldsymbol{w^d}) = \frac{\phi(\boldsymbol{w^*}) - \phi(\vec{\mathbf{0}})}{\phi(\boldsymbol{w^d}) - \phi(\vec{\mathbf{0}})}.$$

The efficiency rate provides a reference value for the quality of the heuristic solution. Though one has to keep in mind that a small efficiency (< 0.9) does not go along with a poorly chosen heuristic as such, since $\phi(\boldsymbol{w}^*)$ holds merely as a lower bound of the integer programming problem. However, an efficiency ≥ 0.9 is desirable and speaks for the quality of the heuristic solution.

4.4.4 Branch and Bound Algorithm

The Branch and Bound algorithm is of fundamental importance in the field of finding a global solution of an integer programming problem such as (4.17). The enumerative

method divides the initial relaxed non-integer problem P_0 as given by (4.11) into subproblems. The complete solution space $W(P_0)$ is bounded such that a subsolution set $W(P_i)$ is defined that contains additional binary constraints. In the case of an experimental design as introduced in Section 4.2 two subproblems emerge by fixing one component of the weight vector $\boldsymbol{w}$ with either $w_j = 0$ or $w_j = 1$. The branching contrives a tree structure as seen in Figure 18, where each node represents a partially fixed design containing the evolving entities including their solution sets. The root of the tree corresponds to the original continuous optimization problem $P_0(\boldsymbol{w})$. The tree structure displays that the children P_{i0}, P_{i1} of a node P_i represent the optimization problem with the solution subset such that,

$$W(P_{i0}) \dot{\cup} W(P_{i1}) = W(P_i).$$

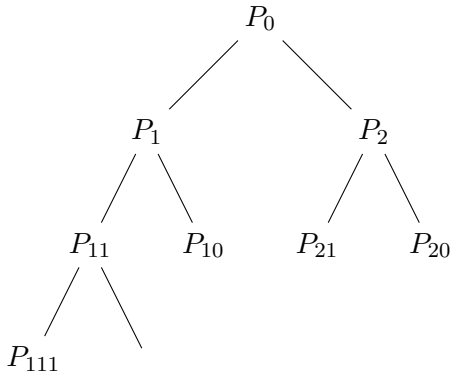


Figure 18: The Binary Tree of Subproblems

The *B&B* is based on the so-called active set A which contains all entities that are still taken into consideration of holding the optimal solution of the integer problem (4.17). The process to trace suboptimal allocations at an early stage results from the so-called bounding and branching steps of the algorithm. We determine the lower bound

$$l_b \leq \phi(\boldsymbol{w}) \; \forall P_i(\boldsymbol{w}) \in A. \tag{4.18}$$

Naturally, the lower bound of the relaxed optimization problem (4.11) holds as a global lower bound of the integer programming problem (4.17). The upper bound $u_b \in \mathbb{R}$ gives the best known integer solution heretofore. The iterations of the algorithm are outlined in Algorithm 2.

Algorithm 2 Branch and Bound Algorithm

Input: Initial problem P_0

Output: Optimal integer solution $\boldsymbol{w}^*$ with $\phi(\boldsymbol{w^*}) \leq \phi(\boldsymbol{w}),\ \forall \boldsymbol{w} \in \{0,1\}^q$

$A \leftarrow P_0$ ▷ Initial active set
$l_b \leftarrow \phi(\boldsymbol{w^0})$ ▷ Initial lower bound
$u_b \leftarrow \infty$ ▷ Initial upper bound
while $A \neq \emptyset$ **do**
    Pick $P_i \in A$ ▷ The bounding step: Find solution of P_i
    $l_b^* = \phi(\boldsymbol{w^*}) \leq \phi(\boldsymbol{w^i}),\ \forall \boldsymbol{w^*},\ \boldsymbol{w^i} \in W(P_i)$
    if $u_b \leq l_b^*$ **then** ▷ P_i is suboptimal and discard.
        $A \leftarrow \{A \setminus P_i\}$
    else ▷ The branching step: Consider subproblems of P_i
        $A \leftarrow \{A \setminus P_i\} \cup \{P_{i_1},\ P_{i_0}\}$
        if $w^* \in \{0,1\}$ **then** ▷ Best known integer solution w_i^*
            $u_b \leftarrow u_b^*$ with $u_b^* = \phi(w^*)$
        end if
    end if
end while

The worst case scenario lists all possible integer solutions. Therefore, the exponential complexity of the original problem is not avoided. To optimize the efficiency of the algorithm the focus is laid on the identification of suboptimality at an early stage. Different techniques evolved, e.g. based on the sequence of the considered nodes: breadth-first search vs. depth-first search [10].
In this thesis the heuristic approach of Definition 32 will be applied for the computations in the following Section 4.5.

4.5 Numerical Experiments and Results

The following experiments in this section are all optimized considering the C-optimal design (4.8). The uncertainty states the marginal variance of the time of death. Heretofore we refer to $\phi(\boldsymbol{w})$ as the uncertainty. In this section, the efficiency of the heuristic solution from Definition 32 is analyzed. The following experiments refer to Definition 24 and provide the basis for choosing a certain design given by Definition 25. Furthermore, different designs are analyzed and compared to each other. The computational results should reveal the meaningful application of an optimal design in the context of the estimation of the time of death.

4.5.1 Experiment 1: Efficiency of the Heuristic Solution

Firstly, we analyze the efficiency of the experiment with $N_s = 15$ points of measurement in space and one point in time, $N_\tau = 1$. With $\boldsymbol{w_l^*}$ we refer to the optimal design considering the restriction of

$$l \in \{1, \ldots, 15\}$$

allowed measurements. The uncertainty of the optimal solution $\boldsymbol{w}_l^*$ is compared to the uncertainty of the corresponding heuristic solution $\boldsymbol{w}_l^{\boldsymbol{d}}$. The uncertainty of the optimal solution holds as a lower bound of the uncertainty of the heuristic solution,

$$\phi(\boldsymbol{w}_l^*) \leq \phi(\boldsymbol{w}_l^{\boldsymbol{d}}),\ \forall l.$$

Figure 19 displays the two uncertainties. The uncertainties are similar, whereas the solution of the optimal design given by the blue line lays marginally below the uncertainty of the heuristic design given by the red line. We assume that in this case the heuristic of Definition 32 performs well. The efficiency rate by Definition 33 confirms the assumption,

$$0.999 \leq \mathit{eff}\,(\boldsymbol{w}_l^{\boldsymbol{d}}) \leq 1,\ \forall l.$$

Since $\phi(\boldsymbol{w}^*) \ll \phi(\vec{\mathbf{0}})$, we evaluate the absolute efficiency,

$$\mathit{eff}_{abs}(\boldsymbol{w}_l^{\boldsymbol{d}}) := \frac{\phi(\boldsymbol{w}_l^*)}{\phi(\boldsymbol{w}_l^{\boldsymbol{d}})}. \tag{4.19}$$

Figure 20 displays the absolute efficiency,

$$0.97 \leq \mathit{eff}_{abs}(\boldsymbol{w}_l^{\boldsymbol{d}}) \leq 1,\ \forall l.$$

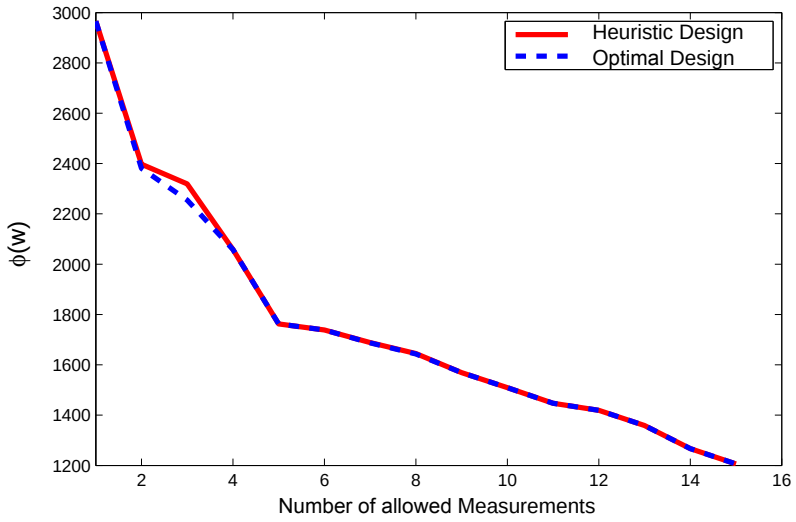


Figure 19: The uncertainty of the optimal solution lays marginally below the uncertainty of the heuristic solution.

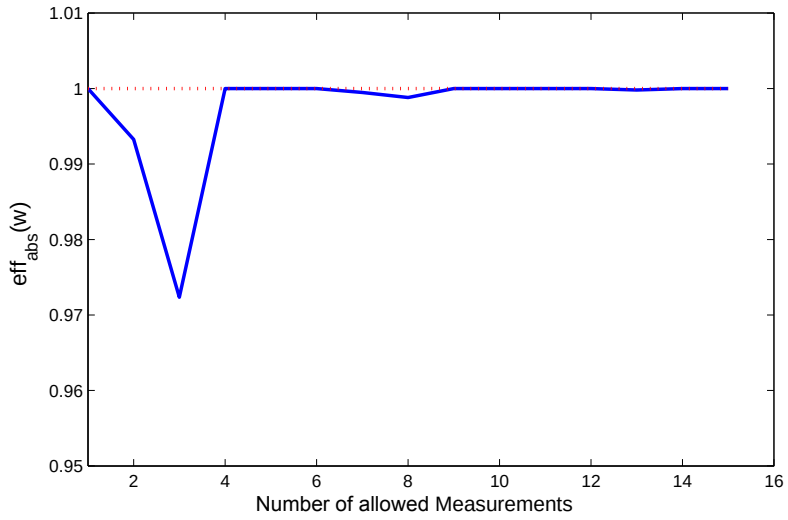


Figure 20: The absolute efficiency of the heuristic solution approximates 1.

Secondly, we analyze the efficiency of more complex experiments that consider a set of $N_s = 187$ points of measurement ($N_\tau = 1$). The optimal design is computed for

$$l \in \{1, \ldots, 150\}$$

allowed measurements. Figure 21 displays that the uncertainty of the heuristic solution is almost identical to the uncertainty of the continuous solution $\forall l$.

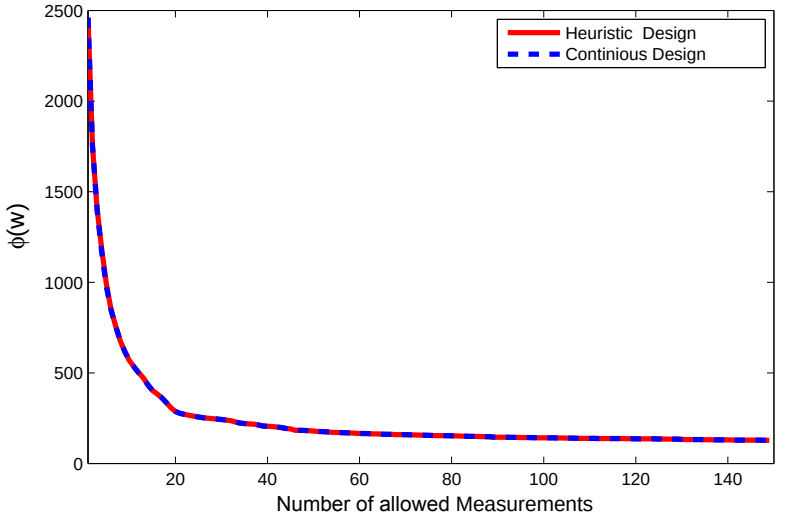


Figure 21: The uncertainty of the continuous solution is almost identical to the uncertainty of the heuristic solution.

Figure 22 refers to the absolute efficiency (4.19). The absolute efficiency converges to 1 for

the experimental designs with a greater number of allowed measurements.

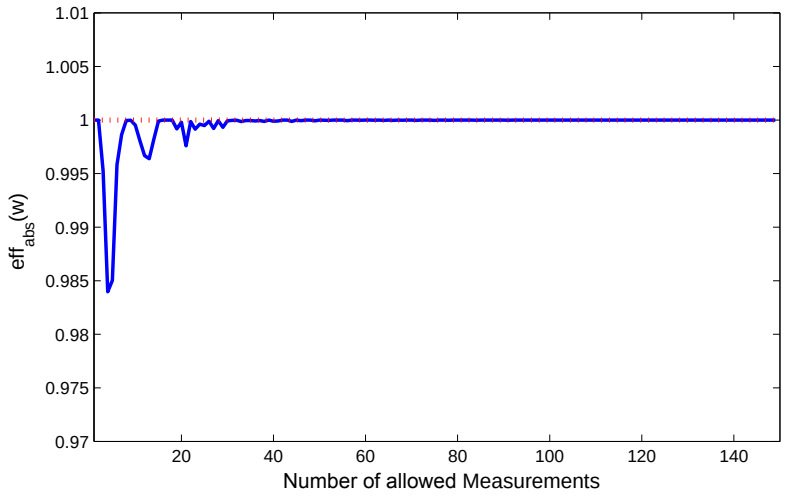


Figure 22: The absolute efficiency of the heuristic solution converges to 1.

Here, the heuristic solution performs nearly as good as the optimal solution. In practice, the efficiency has to be computed to be certain that the heuristic solution gives sufficient results.

4.5.2 Experiment 2: Number of Points of Measurement

We analyze three independent experiments that each hold a certain number of available points of measurement (4.1). For each experiment:

$$e_i,\ i = 1, 2, 3,$$

the optimal design is evaluated. The points of measurement in space are identical for all experiments,

$$N_s^{e_i} = 14, \forall i.$$

The experiments differ by the number of points of measurement in time,

$$N_\tau^{e_1} = 1,\ N_\tau^{e_2} = 2,\ N_\tau^{e_3} = 6,$$

such that the following points in time (in seconds) are considered,

$$t_{e_1} = 100,\ t_{e_2} \in \{100, 1300\},\ t_{e_3} \in \{100, 1300, 2500, 3700, 4900, 6100\}.$$

The total number of points of measurement computes as,

$$q_{e_1} = 14,\ q_{e_2} = 28,\ q_{e_3} = 84.$$

The experiments are embedded in each other, therefore it holds for the set of available points of measurement that,

$$m_{e_1} \subset m_{e_2} \subset m_{e_3}.$$

We analyze to what extent the number of available points of measurement q_{e_i} affect the uncertainty of the optimal design while the number of allowed measurements l stays fixed in all instances. Additionally to the stated aspect, we analyze the effect of the number of allowed measurements while the number of available points of measurement q_{e_i} stays fixed in all instances, i.e. we vary l,

$$1 \in \{1, \ldots, 14\}.$$

Henceforth for simplification reasons the uncertainty of the optimal design with q_{e_i} available points of measurement and l as the number of allowed measurements is defined by

$$\phi(\boldsymbol{w_l^i}) \ \forall i, l.$$

We analyze if the effort of regarding a more complex experiment pays off by observing a significant smaller uncertainty. Figure 23 displays the uncertainty considering the different optimal designs evaluated on the basis of different experiments. We consider experiments that differ by the number of available points of measurement shown by: the red line corresponding to e_1, the blue line corresponding to e_2, and the green line corresponding to e_3. Furthermore, the experiments differ by the number of allowed measurements l.

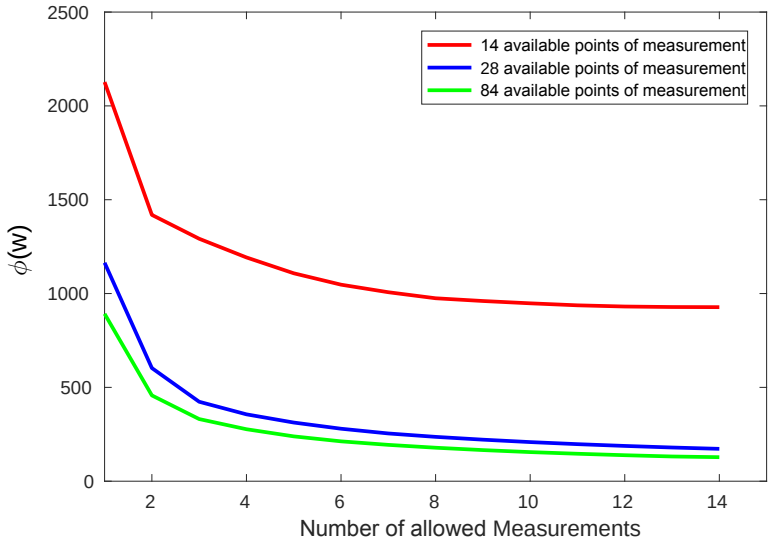


Figure 23: The uncertainty decreases when considering a broader number of available points of measurement and increasing the number of allowed measurements.

Figure 23 shows clearly that the uncertainty is smaller if a greater selection of points of measurement is available. As computed it holds that,

$$\phi(\boldsymbol{w_l^1}) > \phi(\boldsymbol{w_l^2}) > \phi(\boldsymbol{w_l^3}), \ \forall l.$$

Furthermore, the uncertainty decreases when allowing more measurements,

$$\phi(\boldsymbol{w_{(l-1)}^i}) > \phi(\boldsymbol{w_l^i}), \ \forall i, \ l = 2, \ldots 14.$$

These observations are intuitive since an experiment with less restrictions must perform at least as well as an experiment with higher restrictions.
Firstly, we examine the impact regarding the number of available points of measurement. Experiment 1 in comparison to Experiment 2 holds half the number of available points of measurement and the uncertainty of Experiment 1 decreases approximately by half (with constant l) which is displayed by comparing each point of the red line with the corresponding point of the blue line. Nevertheless, the effect does not stay the same when increasing the number of available points of measurement which illustrates through the comparison of Experiment 3 with Experiment 2. Although Experiment 3 considers three times the number of available points of measurement than Experiment 2 the difference of the computed uncertainties is minor as seen by comparing the blue line to the green line.
Secondly, we analyze the effect of the number of allowed measurements l. The uncertainty of Experiment 1, Experiment 2, or Experiment 3 decreases when allowing to perform more measurements. The uncertainty drops significantly in between $1 \leq l \leq 4$, though the effect extenuates for greater l. The benefit of allowing $5 \leq l \leq 14$ measurements might not be worth the effort or costs.
In conclusion it holds that a greater number of available points of measurement and a greater number of allowed measurements is beneficial, though we have to keep in mind that the positive effect vanishes. At first the output changes significantly, while after a certain point no distinct improvement of the uncertainty seems to occur. Therefore, it is important to evaluate the uncertainty of the optimal design for different experiments to evaluate a compromise between a small uncertainty and the effort that comes along with the chosen experiment.

4.5.3 Experiment 3: Instance of Time for Measurements

We consider experiments with $N_s = 25$ available points of measurement in space and $N_\tau = 1$ available points of measurement in time. Five experiments,

$$e_i,\ i = 1, 2, 3, 4, 5,$$

are performed for different points in time in seconds,

$$t_{e_1} = 0,\ t_{e_2} = 3000,\ t_{e_3} = 6000,\ t_{e_4} = 9000,\ t_{e_5} = 12000.$$

In each experiment $l = 3$ measurements are allowed. The optimal design is computed for all five experiments, with the corresponding uncertainty defined as

$$\phi\big(\boldsymbol{w}_{\boldsymbol{l}}^{\boldsymbol{i}}\big), \forall i,\ l = 3.$$

The uncertainty of the optimal design is shown with the red line in Figure 24. In addition, we compute the uncertainty of a design that is optimized in $t_1 = 0$ seconds. Thenceforward the design stays fixed and is not further rearranged. Therefore, a non-optimal design might emerge for $t > t_1$. The uncertainty of the non-optimal design is shown with the blue line in Figure 24.

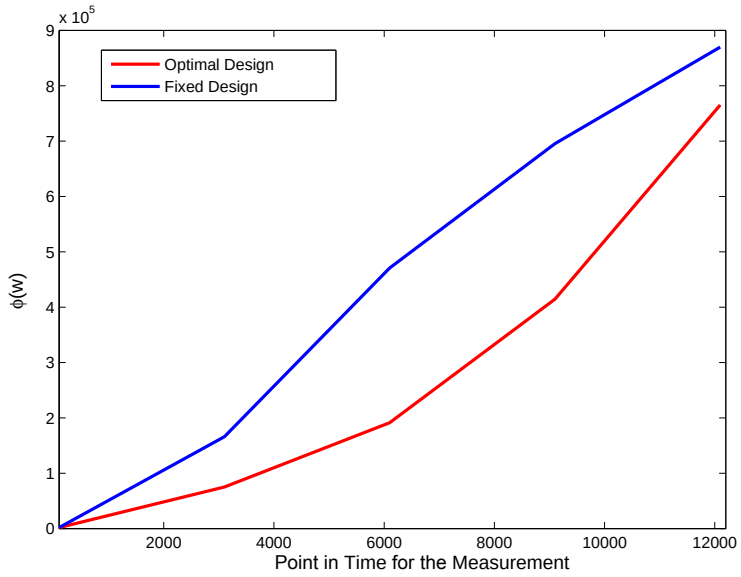


Figure 24: The uncertainty of the optimal design versus the uncertainty of the fixed design

Even by optimizing once in the beginning the certainty in the future differs severely compared to the certainty of the optimal design. Figure 24 shows that the marginal variance increases monotonically in both cases. For the chosen experiment it is important to place measurements at the beginning of the cooling process. Clearly, the uncertainty for the fixed design (blue) is higher compared to the uncertainty of the optimal design (red). The significance of the optimization process and the importance of adapting to new experiments, stated here by the new point in time, is highlighted. The points of measurement of the optimal design shift over time. The computations are performed on the cylinder. The following two illustrations serve as a simplification to display the dynamic of the optimal design over time and space. The allocation of the optimal points of measurement $\boldsymbol{w}_i^* > 0$ (Definition 31) are displayed in Figure 25 in red, respectively the allocation of the points of measurement of the corresponding heuristic solution $\boldsymbol{w}_i^{\boldsymbol{d}} = 1$ (Definition 32) are displayed in Figure 26. Both Figures show that the optimal points distribute widely over the available points of measurement. Additionally, the results of section 4.5.1 are confirmed since the optimal solution approaches a sparse solution.

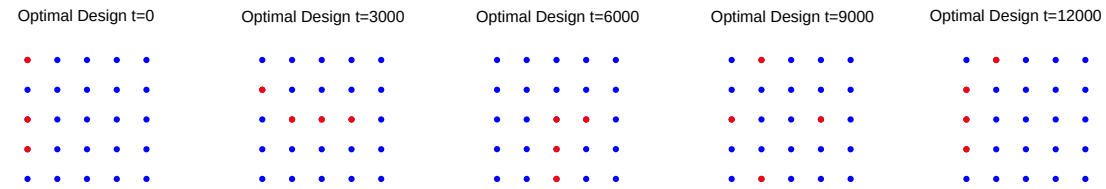


Figure 25: The optimal design over different points in time

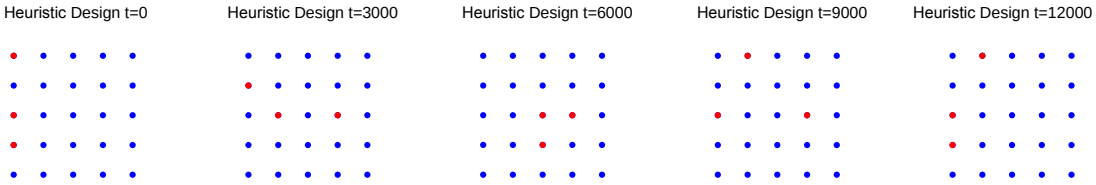


Figure 26: The heuristic design over different points in time

4.5.4 Experiment 4: Optimal Design versus Uniform and Ordinary Design

The importance of choosing the optimal design is clarified by comparing the uncertainty of the optimal design to the uncertainty of the chosen ordinary or uniform design. The experiments,

$$e_i,\ i = 1, \ldots, 6,$$

are set by $N_s = 60$ points of measurement in space, $N_\tau = 1$ points of measurement in time, and $l = 3$ as the number of allowed measurements. The experiments differ by the point in time of the measurements,

$$t_{e_1} = 100,\ t_{e_2} = 1300,\ t_{e_3} = 2500,\ t_{e_4} = 3700,\ t_{e_5} = 4900,\ t_{e_5} = 6100 \text{ seconds.} \quad (4.20)$$

The cylindrical phantom with its available points of measurement over space is shown in Figure 27.

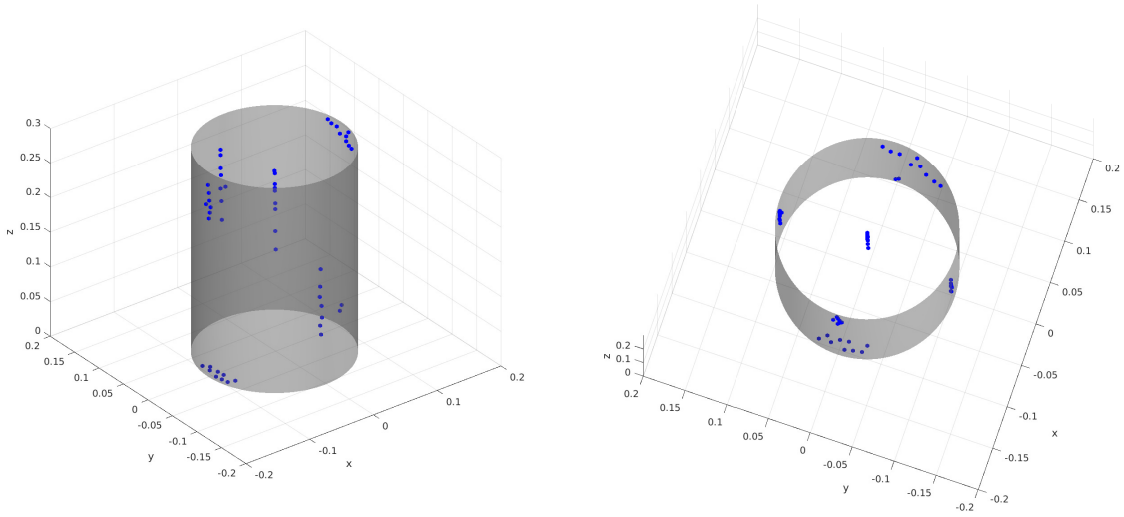
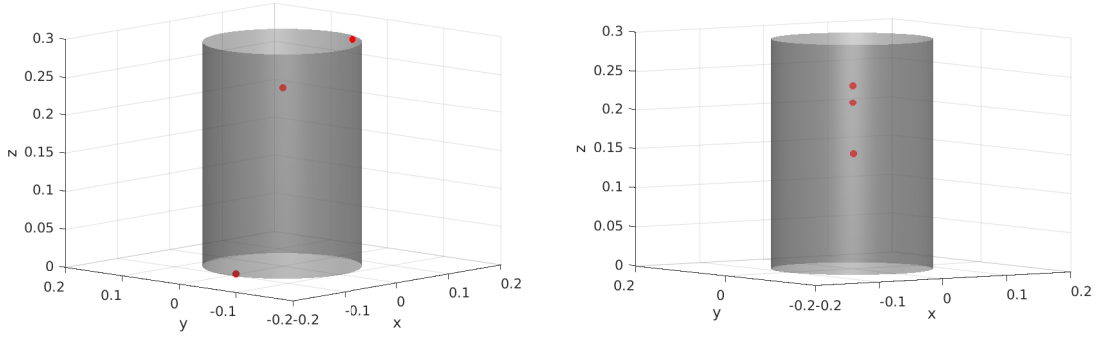


Figure 27: The cylindrical phantom and its 60 available points of measurement

The uniform design and the ordinary design of the cylindrical phantom are illustrated in Figure 28(a) and Figure 28(b). The uniform design considers one point of measurement on the top, one on the bottom and one on the inside of the phantom. The ordinary design is declared by three points of measurement situated at the drilling hole on the inside of the phantom. The ordinary design represents the current state of the measurement process at a corpse. The rectal measurement is performed without considering a different design. The allocation of these points of measurement for both designs is identical for every point in time (4.20).



(a) Uniform Design

(b) Ordinary Design

The optimal design with its corresponding uncertainty, as well as the uncertainties for the ordinary and the uniform design, are evaluated for every point in time (4.20), such that the results can be compared in the following Figure 28. Clearly, the optimal design computes a significantly better certainty than the two alternatives. Furthermore, we observe that the performance of the uniform design is superior to the performance of the ordinary design.

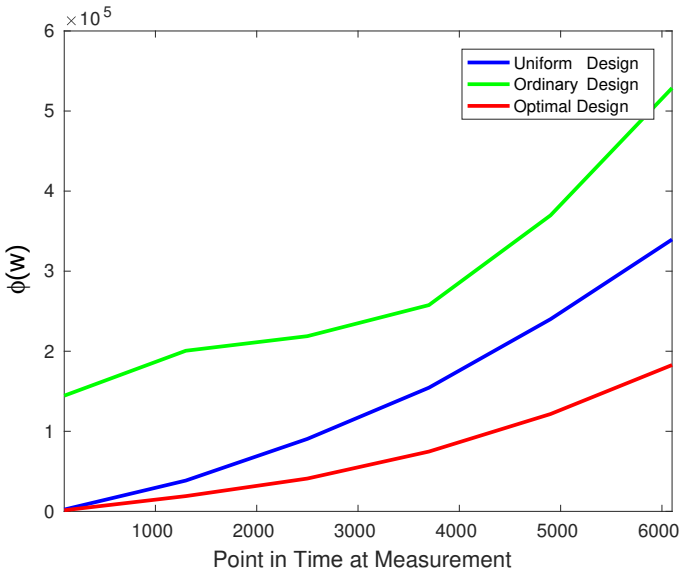


Figure 28: The uncertainty of the three different designs over time

Figure 29 states the results of the optimal design for Experiment 1 given by the point in time of $t_1 = 100$ seconds. One measurement is performed on the top, one on the bottom and one on the surface of the cylinder. The optimal points of measurement change over time as listed in detail in Table 1.

Table 1: Optimal position of the three points of measurement

.

Time in seconds	$t_{e_1} = 100$	$t_{e_2} = 1300$	$t_{e_3} = 2500$	$t_{e_4} = 3700$	$t_{e_5} = 4900$	$t_{e_6} = 6100$
Position	Top	Surface	Surface	Surface	Surface	Surface
	Surface	Inside	Inside	Inside	Surface	Surface
	Bottom	Inside	Inside	Inside	Inside	Inside

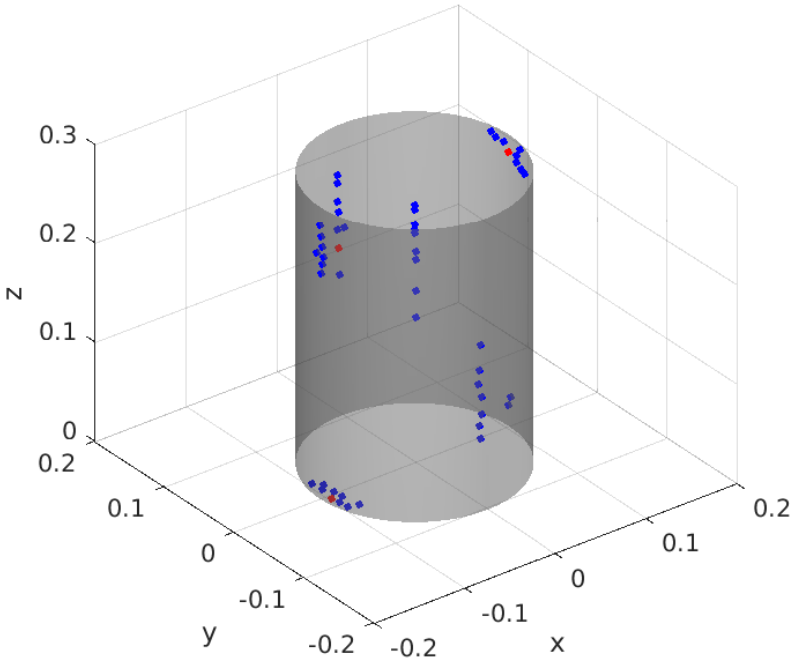
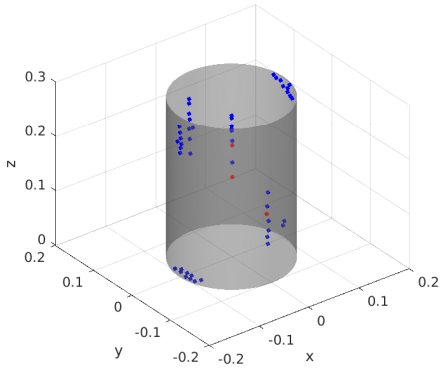
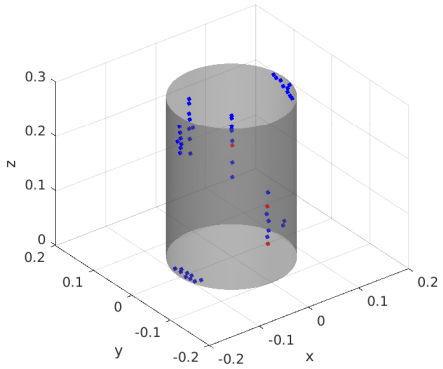


Figure 29: The optimal design for $t = 100$ seconds

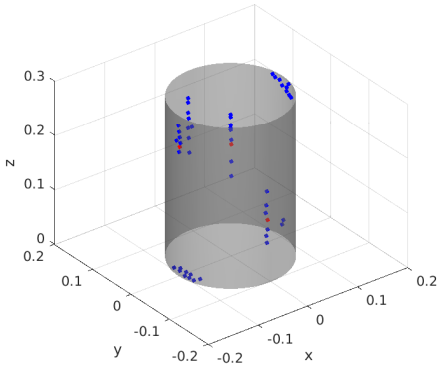
In the beginning of the cooling process the points of measurement on the top and on the bottom of the phantom are favored, while the points of measurement on the inside are unfavored. The points on the inside are rather important later in the cooling process. Figure 30 shows the phantom with the optimal points of measurement for certain points in time corresponding to Table 1.



(a) $t = 1300$ seconds



(b) $t = 4900$ seconds



(c) $t = 6100$ seconds

Figure 30: Different optimal points of measurement over time

We conclude from the experiment that the optimal points of measurement change over time which has significant impact on reducing the uncertainty. The computations are performed for a more complex design of $N_s = 187$ available points of measurement and $l \in \{1, 3, 7\}$. The computational results confirm the former result. It holds that in the beginning of the cooling process the measurements are performed on the top and on the bottom of the phantom. Over time the optimal design changes. For the interested reader the positions of the points of measurement are listed in detail in Table 2 in the Appendix.

5 Conclusion and Outlook

This thesis provides insights how the design of the temperature measurement process impacts the quality of the estimate significantly. The computations in Section 3 and Section 4 are performed on the cylindrical phantom as given in Section 2.5.1. Different designs result in different estimates that are accompanied by a particular uncertainty. The uncertainty varies and can be minimized efficiently as shown in this thesis. By simulating the cooling process and performing experimental measurements the estimate, in this case the *Maximum A Posteriori Estimator (MAP)*, is computed. To evaluate a reliable estimate the optimization of the design of the experiment can now be applied.

The results as part of the *Optimal Design of Experiments* of the simplified model, the cylinder, are generally applicable to other objects, in particular to the human body. The evaluation of the *MAP* and the optimal design should be performed by the forensic experts at the crime scene. The forensic experts evaluate the optimal design, the position of the points of measurement, based on the given circumstances and constraints on-site. Since the cooling process is an ongoing and fast moving process, the design has to be evaluated quickly.

The approach in this thesis focuses on linear or linearized models. The theory for strongly non-linear models needs to be approached differently. The optimal design of the experiment depends on the experimental data that provides the foundation of the estimate of the parameter identification. The planing of an optimal experiment results in a serial process of parameter estimation and experimental designing. The non-linear experimental design considering the parameter identification yields optimization problems embedded in the validation process. New measurements and information give a more reliable *MAP*. One should consider the current *MAP* for the local optimization design process of the linearized model.

Currently only the rectal point of measurement is considered as a point of measurement over space in the field of forensic legal medicine in Germany [11]. This restriction is based on practical and legal purposes, since the method is non-invasive, the point of measurement is reproductively accessible, and the temperature is close to the blood heat. Besides the rectal point of measurement it reveals that measuring the temperature near the digestive system or the esophagus is medically sensible. Clearly, more measurements over space and time, give us more information. Nevertheless, it is not reasonable to consider every accessible point of the body. Peripheral points located at the extremities should not be considered, since the initial condition is too uncertain. The initial temperature varies considerably, depending on e.g. the blood flow, the blood pressure, the activity and stress level of the person, the outside temperature, or the genetic predisposition. The useless points of measurement can either be neglected prior, or the great uncertainty can be displayed in the marginal variance of the respective point of measurement, the corresponding diagonal element of the noise covariance matrix.
Changes considering the temperature measurements at corpses occur by proving the relevance of the experimental design which can be deduced by the numerical results of Section 4. This thesis should lay the foundation for future changes.

Appendices

Algorithm 3 Optimal Design of Experiments: Gradient Descent

Input: Initial vector: $\boldsymbol{w_{new}} \in \mathbb{R}^q$, initial penalty multiplier: $\lambda \in \mathbb{R}$, precision to adjust λ : $\epsilon_2 \in \mathbb{R}$, restricted number of measurements: $restrict \in \mathbb{R}$, precision for gradient descent: $\epsilon_1 \in \mathbb{R}$

Output: Optimal design: $\boldsymbol{w}$ such that $\sum_i w_i \leq restrict$ and $0 \leq w_i \leq 1$

$weight \leftarrow restrict + 1$
while $weight \neq restrict$ **do**
  $\boldsymbol{w} \leftarrow \boldsymbol{w_{new}} + \vec{\mathbf{1}}$
  while $||\boldsymbol{w} - \boldsymbol{w_{new}}|| > \epsilon_1$ **do**
    $\boldsymbol{w} \leftarrow \boldsymbol{w_{new}}$
    $y \leftarrow \phi(\boldsymbol{w}) + \lambda p(\boldsymbol{w})$ ▷ Value of objective function
    $y_{new} \leftarrow y + 1$
    $\gamma \leftarrow 0.1$ ▷ Step size
    while $y_{new} > y$ **do**
      $\boldsymbol{w_{new}} \leftarrow \boldsymbol{w} - \gamma(\nabla(\phi(\boldsymbol{w}) + \lambda p(\boldsymbol{w})))$ ▷ Gradient descent
      for each component of $\boldsymbol{w_{new}}$ **do** ▷ Projection onto set of constraints
        if $w_{new}(j) < 0$ **then** $w_{new}(j) \leftarrow 0$
        else
          if $w_{new}(j) > 1$ **then** $w_{new}(j) \leftarrow 1$
          end if
        end if
      end for
      $y_{new} \leftarrow \phi(\boldsymbol{w_{new}}) + \lambda p(\boldsymbol{w_{new}})$ ▷ Potential new value of objective function
      $\gamma \leftarrow 0.5\gamma$ ▷ Adjust step size
    end while
  end while
  $\boldsymbol{w} \leftarrow \boldsymbol{w_{new}}$
  $weight \leftarrow \sum_i w_i$ ▷ Allowed measurements of optimal solution w
  $\lambda \leftarrow \lambda + \epsilon_2 * (weight - restrict)$ ▷ Adjust penalty term to meet constraint
end while

Table 2: Optimal Position of Points of Measurement

Time in seconds	$t_{e_1} = 100$	$t_{e_2} = 1300$	$t_{e_3} = 2500$	$t_{e_4} = 3700$	$t_{e_5} = 4900$	$t_{e_6} = 6100$
One Measurement	Top	Bottom	Bottom	Surface	Surface	Surface
Three Measurements	Top	Bottom	Bottom	Surface	Surface	Surface
	Bottom	Bottom	Bottom	Surface	Surface	Surface
	Bottom	Surface	Bottom	Inside	Inside	Inside
Seven Measurements	Top	Bottom	Bottom	Surface	Surface	Surface
	Top	Bottom	Bottom	Surface	Surface	Surface
	Bottom	Bottom	Bottom	Surface	Surface	Surface
	Bottom	Bottom	Bottom	Surface	Surface	Surface
	Bottom	Bottom	Bottom	Bottom	Surface	Surface
	Bottom	Bottom	Bottom	Inside	Inside	Inside
	Bottom	Surface	Surface	Inside	Inside	Inside

Acknowledgement. I thank my colleagues from the research group *Computational Medicine* for their collaboration during my work of the master thesis and the related research. You supported me greatly. I would particularly like to single out my supervisor Martin Weiser for his valuable guidance, and my two colleagues, Bodo Erdmann and Dennis Jentsch, for their continuous support and patience. Partial funding by the *University of Jena* is gratefully acknowledged.

Bibliography

[1] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A-Optimal Design of Experiments for Infinite-Dimensional Bayesian Linear Inverse Problems with Regularized L_0-Sparsification. *SIAM Journal Scientific Computing*, 36:2122–2148, 2014.

[2] A. Alexanderian, N. Petra, G. Stadler, and O. Ghattas. A Fast and Scalable Method for A-Optimal Design of Experiments for Infinite-dimensional Bayesian Nonlinear Inverse Problems. *SIAM Journal Scientific Computing*, 38:243–272, 2016.

[3] H. Arellano-Garcia, J. Schöneberger, and S. Körkel. Optimale Versuchsplanung in der chemischen Verfahrenstechnik. *Chemie Ingenieur Technik*, 79:1625–1638, 2007.

[4] A. C. Atkinson and A. N. Donev. *Optimum Experimental Designs*. Oxford University Press, 1992.

[5] G. Aumüller, G. Aust, and A. Doll. *Duale Reihe Anatomie*. Thieme, 2010.

[6] I. Bauer, S. Körkel, H. G. Bock, and J. P. Schlöder. Optimale Versuchsplanung für nichtlineare Prozesse. *Schriftliche Projektpräsentation zum BMBF-Verbundvorhaben. DECHEMA e.V.*, 1998.

[7] A. Björck. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, 1996.

[8] S. Boyd and L. Vandenberghe. Subgradients, 2008. Lecture Notes, Accessed: 12-20-2016, Available at `https://see.stanford.edu/materials/lsocoee364b/01-subgradients_notes.pdf`.

[9] B. Carlowitz. *Kunststoff Tabellen*. Carl Hanser Verlag, 1996.

[10] J. Clausen. Branch and Bound Algorithms – Principles and Examples. Technical report, University of Copenhagen, 1999.

[11] Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. Leitlinien der Deutschen Gesellschaft für Rechtsmedizin. 2007. Accessed: 12-01-2016, Available at `http://www.awmf.org/uploads/tx_szleitlinien/054-002l_S1_Leichenschau_2013-01.pdf`.

[12] P. Deuflhard and F. Bornemann. *Numerische Mathematik 1 - Eine algorithmisch orientierte Einführung*. de Gruyter, 2008.

[13] P. Deuflhard and A. Hohmann. *Numerische Mathematik 2 - Gewöhnliche Differentialgleichungen*. de Gruyter, 2008.

[14] P. Deuflhard and M. Weiser. *Numerische Mathematik 3 - Adaptive Lösung partieller Differentialgleichungen*. de Gruyter, 2011.

[15] R. B. Ellis. *Statistical Inference Basic Concepts*. Prentice-Hall, 1974.

[16] Weinberg F. *Branch and Bound: Eine Einführung*, volume 2. Springer Verlag, 1973. Lecture Notes in Economics and Mathematical Systems.

[17] J. B. J. Fourier. *Analytical Theory of Heat*. Cambridge University Press, 1878.

[18] A. Frommer. Numerische Methoden der nichtlinearen Optimierung. Technical report, Bergische Universität Wuppertal, 2005.

[19] R. G. Gallager. *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[20] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman & Co, 1979.

[21] C. Grossmann, H. Roos, and M. Stynes. *Numerical Treatment of Partial Differential Equations*. Springer, 2007.

[22] E. Haber, L. Horesh, and L. Tenorio. Numerical methods for Experimental Design of large-scale linear ill-posed Inverse Problems. *Inverse Problems*, 24:243–272, 2008.

[23] W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57:97–109, 1970. Accessed: 11-20-2016, Available at `http://links.jstor.org/sici?sici=0006-3444%28197004%2957%3A1%3C97%3AMCSMUM%3E2.0.CO%3B2-C`.

[24] D. Hömberg. Nichtlineare Optimierung. Technical report, Technische Universität Berlin, 2005. Accessed: 12-20-2016, Available at `http://www3.math.tu-berlin.de/Vorlesungen/WS05/NonLinOpt/skript/nonlin-opt.pdf`.

[25] KHP K. Werkstoffdatenblatt: PE 1000. Kunststofftechnik K, ed. Aurachtal, 2015. Accessed: 12-04-2016, Available at `http://www.khp-kunststoffe.de/Downloads/Polyolefine/Datenblatt_4_PE_1000.pdf`.

[26] J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*. Springer, 2010.

[27] R. M. Karp. *Reducibility among Combinatorial Problems*. Plenum Press, New York, 1972.

[28] K. Königsberger. *Analysis 2*. Springer, 1991.

[29] S. Körkel and E. Kostina. Numerical Methods for Nonlinear Experimental Design. In *Modelling, Simulation and Optimization of Complex Processes*, pages 255–272. Springer, 2004.

[30] J. D. C. Little, K. G. Murty, D. W. Sweeney, and K. Caroline. An Algorithm for the Traveling Salesman Problem. 1963.

[31] A. K. Louis. *Inverse und schlecht gestellte Probleme*. Stuttgart: Teubner, 1989.

[32] J.R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, 1988.

[33] G. Mall and W. Eisenmenger. Estimation of Time since Death by heat-flow Finite-Element model. Part I: Method, Model, Calibration and Validation. *Legal Medicine*, 7:69–80, 2005.

[34] W. Martienssen and H. Warlimont. *Springer Handbook of Condensed Matter and Materials Data*. Springer Science & Business Media, 2005.

[35] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *Journal of Chemical Physics*, 21:1087–1092, 1953.

[36] H. Muggenthaler, I. Sinicina, M. Hubig, and G. Mall. Database of post-mortem rectal cooling cases under strictly controlled conditions: a useful tool in death time estimation. *Legal Medicine*, 126:79–87, 2012.

[37] B. K. Natarajan. Sparse Approximate Solutions to Linear Systems. *SIAM Journal on Computing*, 24:227–234, 1995.

[38] A. Pázman. *Foundations of Optimum Experimental Design*. D. Reidel Publishing Co., 1986.

[39] N. Peras, Y. Liu, and P. Bunting. *3D Imaging, Analysis and Applications*. Springer, 2012.

[40] K. B. Petersen and M. S. Pedersen. The Matrix Cookbook, 2012.

[41] F. Pukelsheim. *Optimal Design of Experiments*. SIAM, 1993.

[42] E. A. Rady, M. M. E. Abd El-Monsef, and M. M. Seyam. Relationships among Several Optimality Criteria. Technical report, ISSR, Cairo University, Faculty of Science, Tanta University, 2009.

[43] T. Richter. Die Finite Elemente Methode für partielle Differentialgleichungen. Technical report.

[44] C. Robert and G. Casella. *Introducing Monte Carlo Methods with R*. Springer Verlag, 2010.

[45] C. W. Rowley, T. Colonius, and R. M. Murray. Model reduction for compressible flows using POD and Galerkin projection. *Physica D*, 189:115–129, 2004.

[46] S. Sawyer. The Metropolitan-Hastings Algorithm and Extensions. Technical report, Washington University, 2010.

[47] S. Schenkl, H. Muggenthaler, M. Hubig, and G. Mall. Automatic CT based Finite Element Model Generation for Temperature based Death Time Estimation: Feasibility Study and Sensitivity Analysis. under review, 2016. Scientific Exchange and Discussions.

[48] D. Stalling, M. Westerhoff, and H.-C. Hege. *ZIBAmira: A Highly Interactive System for Visual Data Analysis*. Elsevier, 2005.

[49] Princton University. Singular Value Decomposition SVD). Accessed: 11-20-2016, Available at `http://www.cs.princeton.edu/courses/archive/spring12/cos598C/svdchapter.pdf`.

[50] S. Volkwein. Model Reduction Using Proper Orthogonal Decomposition. Technical report, Karl-Franzens-Universität Graz, 2008.

[51] M. Weiser. *Inside Finite Elements*. de Gruyter, 2005.

[52] D. W. Widder. *The Heat Equation*. Academic Press, 1976.

[53] M. A. Woodbury. Inverting Modified Matrices. Technical report, Princeton University, 1950.

[54] ZIB. Amira and Avizo. Accessed: 11-20-2016, Available at `http://www.zib.de/software/amira`.