Optimum Experimental Design with PDE Constraints for Identification of Model Uncertainty in Load-bearing Systems

Vom Fachbereich Mathematik
der Technischen Universität Darmstadt
zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften
(Dr. rer. nat.)

genehmigte

Dissertation

von

Alexander Matei, M.Sc.

aus Cottbus, Deutschland

Optimum Experimental Design with PDE Constraints for Identification of Model Uncertainty in Load-bearing Systems

Autor:
Alexander Matei, geb. in Cottbus

Acknowledgments

I would like to express my special gratitude to Prof. Dr. Stefan Ulbrich, my supervisor, for his continual support and guidance during my time as a Ph.D. candidate from late 2017 to early 2022. Thank you, Stefan, for motivating me to choose this topic and for letting me be part of your research network. I am grateful for all the opportunities I had during various international conferences, like the GAMM, ISMP, ICCOPT, DMV, ECMI, SIAM OP and others, which you made possible for me to participate. Furthermore, I thank you for the different positions I was able to occupy: as scientific assistant in teaching or as member of a Collaborative Research Center (CRC). Working with students and engineers gave me a rich experience. I learned more self-reliance in problem solving, to be patient with the journal publication process and I deepened my ability to persist in chasing my goals. The training I have received in these four and a half years had a positive impact on my personality and significantly improved my presentation and communication skills by the many talks I could give which I gratefully acknowledge.

I also want to give my special thanks to Prof. Dr. Marc Pfetsch for acting as a referee for this thesis and for his numerous comments which were very helpful. Moreover, I enjoyed our collaboration during the closing phase of the CRC 805 "Mastering Uncertainty in Load-carrying Structures in Mechanical Engineering" where we worked on a book project.

I thank all my present colleagues and former colleagues in our optimization research group for the good time we had working together and for the various activities we enjoyed after work. I am especially thankful to Dr. Tristan Gally, Dr. Anja Kuttich-Meinlschmidt and to Dr. Florian Hoppe for being coauthors of my first journal paper. I am grateful to Dr. Philip Kolvenbach for his help with finite element code and for his insights in efficient mathematical programming. Special thanks goes to all who proofread parts of this dissertation: Dr. Johann Schmitt, Dr. Anne-Therese Rauls, Björn Polenz, Elisa Corbean, Isabel Jacob, Lea Rehlich, Marcel Steinhardt and Dr. Paloma Schäfer Aguilar. My sincere gratitude goes to Dr. Christopher Balzereit

who proofread the entire thesis and to Edwin Rosado for his helpful comments on the introduction. I also want to thank Matthias Deutschen for some fruitful discussions on Bayesian parameter estimation and Jianhang He for his insights in local search algorithms.

I thank my wife, Anna, for supporting me emotionally and for believing in me. Thank you for your love and patience.

Most importantly, I want to thank my Lord and Savior Jesus Christ for giving me health, strength and wisdom in these times. To Him be all honor and glory.

> *You are my strength, I sing praise to You;*
> *You, God, are my fortress, my God on whom I can rely.*
> – Psalm 59:17, NIV

Zusammenfassung

In den Natur- und Ingenieurswissenschaften ist der Gebrauch von mathematischen Modellen *das* Standardwerkzeug, um physikalische Phänomene zu beschreiben, technische Prozessketten zu steuern und Designprobleme zu lösen. Insbesondere dynamische Prozesse sind von großem Interesse. Eine einwandfreie Modellierung dynamischer Systeme stellt nicht selten eine Herausforderung dar. Unsicherheit wird zunehmend in allen Phasen des Produktlebenszyklus wahrgenommen: im Design, in der Produktion und in der Nutzung. Diese Unsicherheit hat ihren Ursprung in uns unbekannten, unvollständigen oder unpräzisen Modellierungsannahmen bzw. Gleichungen und sie äußert sich aufgrund von Unwissen bzw. Vernachlässigung von Wissen, schlechten numerischen Approximationsverfahren oder menschlichem Versagen. Deswegen ist es notwendig, Modellunsicherheit in einem möglichst frühen Stadium der Produktentwicklung zu beherrschen.

In dieser Dissertation wird ein neuer Algorithmus zur Identifizierung von Modellunsicherheit basierend auf Methoden des optimalen Designs von Experimenten mit partiellen Differentialgleichungen (PDE) als Nebenbedingungen und der Theorie der statistischen Testverfahren vorgestellt. In einem ersten Schritt stellen wir fünf verschiedene Ansätze zur Parameterschätzung aus verrauschten Daten vor, die entweder auf der frequentistischen oder der Bayes'schen Sicht von Wahrscheinlichkeit aufbauen, und bestimmen die a posteriori Wahrscheinlichkeitsverteilung der geschätzten Parameter. In einem zweiten Schritt stellen wir den klassischen Ansatz zu optimalem Design von Experimenten vor und zeigen, dass dieser ungeeignet ist wenn PDE Nebenbedingungen mit einbezogen werden. Folglich gehen wir auf moderne Methoden zur optimalen Sensorplatzierung ein und geben eine Erweiterung zur optimalen Wahl von Randdaten an. Dazu stellen wir ein Optimierungsproblem mit PDE Nebenbedingungen auf, welches die Güte der Parameterschätzung, die Kosten der eingesetzten Sensoren und einen Regularisierer für die Randdaten minimiert. Der Kostenterm ist ein Strafterm, der zu dünnbesetzten Sensorvariablen führt. Zur Lösung dieses Optimierungsproblems benutzen wir das Adjungiertenkal-

kül. Aus den Messdaten, die innerhalb dieses optimierten Experiments gesammelt wurden, werden dann die Werte der Modellparameter mit kleiner Varianz geschätzt. In einem dritten Schritt konstruieren wir einen Hypothesentest, um die Annahme zu falsifizieren, dass wiederholte Modellkalibrierung und Validierung dieselben Parameterwerte innerhalb eines kleinen Konfidenzgebietes liefern. Falls ein neuer Messdatensatz zu Werten außerhalb dieses Vertrauensbereichs führt, dann ist Modellunsicherheit mit einer gegebenen kleinen Wahrscheinlichkeit für den Fehler erster Art detektiert worden. Wir beweisen, dass im Fall von linearen Modellen, unser Algorithmus mit der bereits gegebenen kleinen Wahrscheinlichkeit fälschlicherweise Modellunsicherheit detektiert. Außerdem beweisen wir, dass je kleiner das Konfidenzgebiet ist, desto besser werden falsche lineare Modell auch als solche erkannt.

Im letzten Teil dieser Dissertation wenden wir unseren Algorithmus zur Identifizierung von Modellunsicherheit auf verschiedene Modelle zweier technischer Systeme an. Das Erste ist eine mechanische Presse zur Umformung von Bauteilen unter sehr großen Kräften. Das Zweite ist ein zweidimensionaler Prototyp der oberen Tragestruktur eines Flugzeugfahrwerks. Zum Schluss diskutieren wir die numerischen Ergebnisse und geben einen Ausblick, wie ein moderner Ansatz zur optimalen Wahl von Randdaten für ein optimales Experiment mit PDE Nebenbedingungen bei hochdimensionalen Parametern aussehen könnte.

Abstract

In science and technology, mathematical models are the standard tool to describe physical phenomena, to master engineering processes and to solve design problems. Especially dynamic processes are of profound interest and their accurate modeling is still challenging. One often encounters uncertainty in the design, production and usage phase of real products. Much of this uncertainty originates in unknown, incomplete or inadequate modeling assumptions and equations due to lack or disregard of knowledge, bad numerical approximation schemes and human shortcomings. Therefore, it is important to master model uncertainty in an early stage of product development.

In this thesis we introduce a novel algorithm to identify model uncertainty based on methods from optimum experimental design with partial differential equation (PDE) constraints and statistical hypothesis testing. We first introduce five different approaches based on a frequentist and a Bayesian probabilistic perspective to estimate the parameter's a posteriori probability distribution from noisy data. Second, we show that the classical approach to optimum experimental design is insufficient to handle PDE constraints. Furthermore, we examine modern approaches to optimal sensor placement and make an extension to optimal input configuration. In so doing, we introduce a PDE-constrained optimization problem, which adds a cost term to sparsify the number of used sensors and a smooth regularization for the inputs to the objective function, and solve it with an adjoint approach.

The data which are collected in an optimally designed experiment are used to infer parameter estimates that have a small variance. In the third step, we construct a hypothesis test to falsify the assumption that repeated calibration and validation procedures should yield parameter values in the same small confidence region. If a new set of data leads to estimates that lie outside of this confidence region, then model uncertainty is detected with a given small threshold to the Type I error probability. We prove that for linear models the probability that our algorithm falsely identifies model uncertainty is identical to the small test level. We also prove that

the smaller the confidence region the better the rejection of false linear models.

Finally, we apply our algorithm to detect model uncertainty in mathematical models of a forming machine and in the linear-elastic model of vibrations in a truss. In the first case, we have a real technical system and in the second case we have a two-dimensional representative of the upper truss structure of a landing gear. We conclude this thesis with an evaluation of the numerical results and we give an outlook on large-scale problems in optimal input configuration with PDE constraints.

Notation and Symbols

$\mathbb{N}$	natural numbers $1, 2, \ldots$
$\mathbb{R}$	real numbers
ODE(s), PDE(s)	ordinary/partial differential equation(s)
PE	parameter estimation problem(s)
OED	optimum experimental design
CDF	cumulative probability distribution function
FEM	finite element method
min, max	minimize/maximize
s.t.	subject to
const.	constant
log	natural logarithm
$\exp(\cdot)$	exponential function
$\mathrm{sign}(\cdot)$	signum function
$\emptyset$	empty set
I	identity matrix
$x^\top, A^\top$	transpose of a vector $x \in \mathbb{R}^n$/a matrix $A \in \mathbb{R}^{m \times n}$
$A \succ 0$	symmetric, positive definite matrix $A \in \mathbb{R}^{n \times n}$
$S_n^+(\mathbb{R})$	set of all real symmetric, positive definite $n \times n$ matrices
$\|x\|_2$	Euclidean norm of a vector $x \in \mathbb{R}^n$
$\|x\|_C^2$	$x^\top C x$ with a symmetric, positive definite matrix $C \in \mathbb{R}^{n \times n}$
$\limsup$	limes superior
$\mathrm{ess\,sup}$	essential supremum
$\lim_{x \searrow x_0} f(x)$	one-sided limit $\lim_{t \to 0} f(x_0 + |t|)$
$o(\|x - x_0\|)$	class of functions f such that $\lim_{x \to x_0} \|f(x)\| \cdot \|x - x_0\|^{-1} = 0$
$\mathbb{1}_A(\cdot)$	characteristic function for a subset $A \subset B$
$\mathrm{conv}\, M$	convex hull of a set M
$\mathrm{Lin}(w_1, \ldots, w_n)$	linear hull of the vectors/functions $w_1, \ldots, w_n$

range(A)	image of the linear mapping $A : \mathbb{R}^m \to \mathbb{R}^n, x \mapsto Ax$
Diag($\sigma_1, \ldots, \sigma_n$)	diagonal matrix $S \in \mathbb{R}^{n \times n}$ with $S_{kk} = \sigma_k$
rep($w_1, \ldots, w_n; m$)	vector $v \in \mathbb{R}^{n \cdot m}$ of m copies of $(w_1, \ldots, w_n)^\top \in \mathbb{R}^n$
$\partial_x f$	Jacobian $\mathscr{J}_x(f)$ of f with respect to x
$\nabla_x f, \nabla f$	gradient $\mathscr{J}_x(f)^\top$ of f with respect to x (x omitted if it is clear)
$\partial^2_{xx} f$	Hessian of f
$\left\langle \partial^2_{xx} f, (v_1, v_2) \right\rangle$	second directional derivative of f in the directions (v_1, v_2)
$\mathrm{d}/\mathrm{d}t$	total derivative with respect to t
$y'(t), \dot{y}$	derivative of a one-dimensional function $y : \mathbb{R} \to \mathbb{R}$
$P[Z < c]$	probability that the random variable Z is less than c
$P[A \mid B]$	conditional probability of two events A and B
$\mathbb{E}_z[f(Z)]$	expectation value of a function f of a random variable Z
$\mathrm{Var}[Z]$	variance matrix of a random variable Z
$\mathrm{Cov}[X, Y]$	covariance of two random variables X, Y
$\mathcal{N}(a, b^2)$	normal distribution with mean value a and variance b^2
$\mathcal{N}(\mu, \Sigma)$	multivariate Gaussian distribution with mean vector $\mu \in \mathbb{R}^n$ and variance matrix $\Sigma \in \mathbb{R}^{n \times n}$
∂G	boundary of a domain $G \subset \mathbb{R}^d$
$L^2(G)$	space of quadratically integrable functions over a domain G
$[\cdot, \cdot]_G$	L^2 scalar product over the domain G
$H^1(G)$	Sobolev space of all functions $f \in L^2(G)$ such that $\nabla f \in L^2(G)$ in a weak sense
$L^\infty(0, T)$	space of essentially bounded measurable functions over $(0, T)$
$L^2(G)^d$	d-dimensional vector space of $L^2(G)$ functions
$L^\infty(0, T; H^1(G)^2)$	Bochner space of vector valued functions $f(t, \cdot) \in H^1(G)^2$ such that $\|f(\cdot, x)\|_{H^1(G)^2} \in L^\infty(0, T)$

Contents

Introduction

What we observe is not nature itself, but nature exposed to our method of questioning.
– W. Heisenberg, Physics and Philosophy: The Revolution in Modern Science

Mathematics has been the most successful language to describe and predict natural phenomena, engineering processes and economic behavior. It is precisely because of our mental ability to transfer aspects of reality into the abstract world that enables us to make generalized statements. *Models* are, in a sense, an image of reality into mathematical abstraction. They contain a physical law, a working principle or just axiomatic assumptions encapsulated in the language of functional relations. Moreover, their scope and complexity is specified by the designer. These functional relations rule between state variables, like physical quantities such as displacements, velocities or accelerations, model parameters, like material constants or geometric properties, and inputs, like boundary forces or initial conditions. However, the modeling process is not free from uncertainty and errors. Various types and sources of uncertainty exist each of which have a different impact on the descriptive and predictive quality of models. It has been observed that the dominant source of model uncertainty is lack or disregard of knowledge, i.e., unknown, missing or simplified functional relations. Therefore, mathematical models are subject to a verification, calibration and validation process [100] in which the model output is *compared to new observations* and the model itself is evaluated, i.e., model parameters are tuned and functional terms are altered, removed or redesigned.

The key to improve our models of reality is experimental evidence. Quantities that are measured in an experiment or collected from individuals or repositories are called *data*. However, it is very likely that the value, interval or distribution of data is unknown, incomplete or insufficient. This mainly affects the model calibration

process in which model parameters are tuned such that the model's output correctly represents the quantity of interest. In general, the true values of the model parameters are not known, thus, the better the data the more precise the parameter estimates. In principle, it is not yet possible to conduct a perfect experiment yielding noiseless data. There is an intrinsic randomness in the measurement process itself which is often called aleatoric uncertainty [85], at least on the quantum scale [33, 62] but also beyond, and which cannot be removed. Thus, data uncertainty propagates to the model parameters and eventually to the model's output, impairing the prediction quality. In order to mitigate these effects the experimenter can optimize the design of the experiment from which measurements are collected such that the variance of the estimated model parameters becomes small. However, the impact of model uncertainty is considered to be more severe than the effects of data uncertainty.

The acquisition of scientific knowledge is based on philosophic principles which have been developed and debated over centuries. In opposition to inductive reasoning, Popper [106] introduced another philosophy of science known as the falsification principle. In this view a hypothesis in order to be considered scientific must be able to be tested and refuted by experimental evidence. Thus, all present scientific knowledge is provisional and subject to change if new data arise which contradict the model. Even though this view has been criticized for being too rigid it can be considered as safe to distinguish science from non-science. In fact, scientific knowledge progresses if a hypothesis is refuted by contrary evidence and a new theory is constructed which better explains the measurement data.

A historic example where the falsification principle was successfully employed is black body radiation. The Rayleigh-Jeans law states that the energy distribution upon the eigenfrequencies is continuous such that the energy density of a black body radiator is inverse proportional to the fourth power of the wave length. This law was validated by experiments with light in the infrared range. However, for ultraviolet radiation the Rayleigh-Jeans model predicted a very high energy density that goes to infinity if the wavelength approaches zero. In the sequel, the Rayleigh-Jeans law was falsified by experimental evidence from the sun emission spectrum. It was M. Planck who came up with the revolutionary idea that the energy distribution upon each eigenfrequency of a black body radiator is quantized, leading to a much better model which is today known as Planck's radiation law. For more details the reader may consult [33, Sec. 3.1].

We contribute to the Popperian view of science by introducing an algorithm to falsify a mathematical model in the presence of controlled data uncertainty.

Real technical systems, such as load-bearing structures, typically have a multitude of uncertain parameters and often exhibit unexpected physical effects. Thus, one frequently encounters uncertainty in all phases of their life-cycle. In this work we consider two systems with high safety requirements: a highly loaded forming machine and a landing gear's upper truss structure. The deflection of several components of a forming press during its motion is of high technical importance. We seek a model that accurately predicts the elastic behavior at low excitation frequencies. Because of the complexity of the forming machine and its components, a less accurate lumped parameter model together with a friction model is employed. Hence, it is necessary to estimate some model parameters, such as stiffnesses of bars and beams, *after* the assembly of the machine. Using these estimates, the model's correctness can be evaluated and it can be falsified by a hypothesis test. By similar reasoning, the material constants of the upper truss structure of an airplane's landing gear must be estimated from data in the usage phase to test the model's fitness. It is not uncommon due to manufacturing tolerances, lack of materials or human shortcomings in the production process that the original model may fail to predict the behavior of the *real* truss under load or excitation.

1.1 Our Approach to Model Uncertainty

In this thesis we introduce a novel algorithm to detect and quantify model uncertainty using parameter estimation, optimum experimental design and statistical hypothesis testing. We assume that the model parameters have a true value and that it is possible to approximate this value by repeated inference from data. By an optimally designed experiment we minimize the Gaussian confidence ellipsoid, determined by the covariance matrix of the a posteriori probability distribution, where the true values are assumed to be. If the model is correct, then repeated calibration and validation procedures should yield parameter values inside this confidence ellipsoid. However, if a set of measurements leads to parameters that lie outside this region, then the model is incorrect and we reject it in the context of a hypothesis test with a threshold to the Type I error probability. We use the p-value of the test not only to detect but also to quantify model uncertainty: the lower the p-value, the higher the model uncertainty.

The ideas of our approach originated at the Collaborative Research Center (CRC) 805 within the Subproject A3. We were able to apply our algorithm to a real-sized technical system thanks to a collaboration with the Institute for Production Engineering and Forming Machines (PtU) in the working group of Prof. Dr.-Ing. Peter Groche at TU Darmstadt.

1.2 Literature Review

Parameter estimation and inverse problems are a well-studied but still ongoing field of research [10, 22, 125]. Methods for parameter estimation and their a posteriori probability distribution range from (nonlinear) least-squares [11, 34, 127], (asymptotic) maximum likelihood [31, 112] and (asymptotic) Bayesian inference [125, 130]. These methods differ due to their probabilistic perspective and especially for nonlinear models. What is more, parameter estimation plays an important role in machine learning [96].

Optimum experimental design is a broad field of research [43, 47, 108, 133]. The pure sensor placement problem has been studied for various objective functions like measures of the Fisher Information matrix, Bayes factors, condition numbers or via a modal analysis [26, 46, 47, 63, 102]. Robust formulations like the min-max approach have been introduced by [9, 51, 132] to deal with uncertain priors and data uncertainty. We refer to these methods as the classical approach to optimal design of experiments. A modern perspective, however, considers PDE constraints and comes from an infinite-dimensional setting. The modern approach to optimum experimental design is adopted by Alexanderian et al. [3, 4] who minimize randomized trace estimators of the asymptotic covariance matrix of the parameters to be estimated governed by PDE constraints. Koval et al. [82] additionally consider the design problem in the presence of a given model uncertainty distribution. Neitzel et al. [98] introduce a sparse optimal control approach to the sensor placement problem using Dirac measures. On the other hand, the classical approach to optimal input configuration has been developed in the presence of ODE constraints in various engineering applications [27, 69, 95, 97, 124]. A sensitivity approach to derivative-based optimization was suggested by [12, 81] in order to find optimal input configurations. In our case, however, the dimension of the input space is large and thus we develop an adjoint method for PDE-constrained optimal input configuration.

Recently, the topic of model uncertainty has received more attention and has developed a considerable momentum [7, 91, 92, 100, 105, 111]. There is abundant literature on the detection, quantification and control of model uncertainty. We distinguish between *non-probabilistic methods*, such as residual analysis [120, 137], interval simulation [107] or polynomial fits [121], and *probabilistic Bayesian inference methods*, such as modeling the discrepancy function as a stochastic process [74], Bayesian model updating [140] or Kalman filtering [72], as well as *probabilistic frequentist methods*, such as validation metrics [89, 111] or quantile and density estimations [56, 77]. Bayesian methods rely on the Bayes' rule for comput-

ing posterior probability distributions whereas frequentist methods are based on the classical definition of probability which is the relative frequency of an event. Our novel algorithm to identify model uncertainty is based on a probabilistic frequentist perspective where it is assumed that the true values of the model parameters lie in a confidence region determined by the likelihood function.

Our approach is similar to Stegmaier et al. [122] who also use optimal design of experiments to minimize the parameter's confidence region but employ the Kullback-Leibler divergence as a model selection tool. Our work extends the idea presented in [10, Sec. 2.2] and in [47, Fig. 2] that optimum experimental design and parameter estimation can be used for model discrimination if the estimated parameters lie outside of an established confidence region.

1.3 Structure of the Thesis

This thesis is organized as follows. Chapter 2 introduces the parameter estimation problem and its a posteriori probability distribution. For models with nonlinear parameter dependence we present five ways to approximate the covariance matrix of the estimated parameters which are based on either a frequentist or a Bayesian view of probability. We further discuss the influence of regularization upon the a posteriori distribution.

In Chapter 3 we first analyze the classical optimal design of experiments problem and then introduce a modern approach to this topic which allows for PDE constraints to be present. Moreover, we extend the pure sensor placement problem to incorporate the choice of optimal inputs. At the end of this chapter we discuss advantages and disadvantages of the three most common design criteria used to measure the size of the covariance matrix.

Chapter 4 deals with the central topic of this thesis. We first specify the definition of a mathematical model and what we mean by model uncertainty. Second, we shortly introduce various approaches to model uncertainty from the literature and briefly mention the extent of data uncertainty upon the model's prediction quality. We finally present our algorithm to detect and quantify model uncertainty and prove that if a linear model is correct, then the probability that our algorithm falsely detects model uncertainty is identical to the small test level. Moreover, we prove that an optimally designed experiment correctly provokes the rejection of false linear models.

In Chapter 5 we apply our method to identify model uncertainty in three quasi-static models of a mechanical servo press and in the dynamic linear-elastic model of vibrations in a truss.

We summarize the main conclusions of this thesis in Chapter 6 and give a short outlook on possible extensions of our approach to optimal input configuration with PDE constraints for large-scale problems.

Parameter Estimation and its A Posteriori Probability Distribution

In science and engineering, one frequently relates physical parameters θ, like material properties, and inputs u, like initial conditions or acting forces, to the quantity of interest which can be observed by measurements $z(u)$, which we call data. All functional relations that are known and relevant to us are summarized in a mapping $(\theta, u) \mapsto \eta(\theta, u)$ called *model*. We further assume that the model η is correct, i.e., for all inputs u the model explains the true (but unknown) value of the quantity of interest $z^{\star}(u)$:

$$\eta(\theta^{\star}, u) = z^{\star}(u), \tag{2.1}$$

where $\theta^{\star}$ are the true (but unknown) parameter values.

In general, data are subject to statistically independent noise which we model by a (multivariate) Gaussian random variable $\mathcal{E} \sim \mathcal{N}(0, \Sigma)$, where Σ is a diagonal matrix with strictly positive diagonal entries. Thus, the data $z(u)$ are a realization of a Gaussian random variable $Z(u) \sim \mathcal{N}(z^{\star}(u), \Sigma)$. A more realistic version of (2.1) is

$$\begin{aligned} Z(u) &= z^{\star}(u) + \mathcal{E}, \\ &= \eta(\theta^{\star}, u) + \mathcal{E}. \end{aligned} \tag{2.2}$$

We call the task to find for given data $z(u)$ the best approximation $\bar{\theta}$ of the true parameters $\theta^{\star}$ in the sense that

$$z(u) = \eta(\bar{\theta}, u) + \varepsilon,$$

where ε is a realization of the noise variable $\mathscr{E}$, a *parameter estimation problem*. Since this is also dependent upon the specific model η it is sometimes called model-based parameter estimation.

However, if the data are subject to uncertainty, modeled by Gaussian noise, so are the estimated parameters $\bar{\theta}$ by the laws of uncertainty propagation. In this chapter we introduce well-known methods for parameter estimation problems (PE) and compute their statistical properties like mean value and covariance matrix, based on [10, 23, 36] and [125]. We present two probabilistic perspectives here: a frequentist and a Bayesian view.

This chapter is structured as follows. Section 2.1 deals with linear models and the linear least-squares problem where we also consider Tikhonov regularization. In Section 2.2 we introduce nonlinear inverse problems and Newton's method. We also discuss simplifications of the Newton step like the Gauss-Newton and Levenberg-Marquardt approach. Section 2.3 introduces the Bayesian framework to estimate model parameters and their posterior probability distribution. In Section 2.4 we bring in asymptotic maximum likelihood theory which leads to a different covariance approximation. Section 2.5 discusses confidence regions derived from these approaches. In Section 2.6 we give a short summary.

2.1 Linear Regression

Let $\Theta \subset \mathbb{R}^{n_\theta}$ be the set of admissible model parameters, $U \subset \mathbb{R}^{d_u}$ be the set of inputs and $Q_I \subset \mathbb{R}^{d_z}$ be the domain of the quantity of interest. We consider the linear model

$$\eta : \Theta \times U \to Q_I, \ (\theta, u) \mapsto \eta(\theta, u) = B(u)\,\theta$$

which maps parameters θ and inputs u to the quantity of interest by employing the elemental system matrix $B(u) \in \mathbb{R}^{d_z \times n_\theta}$.

Let $u := \{u_1, \dots, u_m\} \subset U$ be a collection of $m \in \mathbb{N}$ inputs for which corresponding data $z(u_1), \dots, z(u_m) \in Q_I$ are acquired. In order to improve the information gain and accuracy of the measurement process, we repeatedly measure the quantity of interest n_z times and acquire $z_1(u_i), \dots, z_{n_z}(u_i) \in Q_I$ *measurement series* for each input $u_i \in U$. Thus, in view of (2.2) we have for all $j = 1, \dots, n_z$ and $i = 1, \dots, m$ the following equation:

$$z_j(u_i) = B(u_i)\,\theta^\star + \varepsilon_{ij}, \tag{2.3}$$

where ε_{ij} are independent realizations of $\mathscr{E}_d \sim \mathscr{N}(0, \Sigma_d)$ each with the same variance

matrix $\Sigma_d = \text{Diag}(\sigma_1^2, \ldots, \sigma_{d_z}^2) \in \mathbb{R}^{d_z \times d_z}$. For convenience, let

$$\tilde{z}(u_i) := \begin{pmatrix} z_1(u_i) \\ \vdots \\ z_{n_z}(u_i) \end{pmatrix}, \quad z(u) := \begin{pmatrix} \tilde{z}(u_1) \\ \vdots \\ \tilde{z}(u_m) \end{pmatrix}, \quad \tilde{A}(u_i) := \begin{pmatrix} B(u_i) \\ \vdots \\ _{n_z \text{ times}} \\ \vdots \\ B(u_i) \end{pmatrix} \text{ and } A(u) := \begin{pmatrix} \tilde{A}(u_1) \\ \vdots \\ \tilde{A}(u_m) \end{pmatrix},$$

such that $z(u)$ and $\eta(\theta, u) := A(u)\theta$ have dimension $n := n_z d_z m$. Here, $A(u) \in \mathbb{R}^{n \times n_\theta}$ is the overall system matrix which we use in the sequel. Similarly, we define the diagonal matrix $\Sigma \in \mathbb{R}^{n \times n}$ to be made of $n_z m$ blocks where each block is identical to the noise covariance matrix Σ_d, i.e., $\Sigma := \text{Diag}(\text{rep}(\sigma_1^2, \ldots, \sigma_{d_z}^2 ; n_z m))$, where $\text{rep}(\cdot\, ; n_z m)$ produces $n_z m$ copies of its input vector. Furthermore, set $Q := Q_I^{n_z m} \subset \mathbb{R}^n$. In the sequel, we want to find an estimate $\bar{\theta}$ that best fits the model to the data:

$$z(u) = A(u)\,\bar{\theta} + \varepsilon,$$

where ε is a realization of the noise variable $\mathcal{E} \sim \mathcal{N}(0, \Sigma)$. We start by introducing the least-squares method based on [10].

2.1.1 Least-squares Estimators

From a probabilistic frequentist view, if we know the probability distribution of the data, then we aim to find parameters such that these observations are most likely noisy realizations of the model output. This framework is also called maximum likelihood approach. For some given input u_i, let $f_i(\cdot)$ be the probability density function of the random variable $\mathcal{E}_d$. Now, let $u \subset U$ be a set of inputs. Then the joint probability density function $f(\cdot)$ can be written as

$$z(u) \mapsto f(z(u)) = \prod_{i=1}^{d_u} \prod_{j=1}^{n_z} f_i(z_j(u_i)).$$

For some fixed $z(u)$, we call $\theta \mapsto L(\theta) := f(z(u) - A(u)\,\theta)$ the *likelihood function*. Evidently, $L(\theta)$ is the joint probability that $z(u) - A(u)\,\theta$ is a realization of the random variable $\mathcal{E}$. We consider measurement series $z_j(u_i)$ with known variance Σ_d such that

$$L(\theta) = C \prod_{i=1}^{d_u} \prod_{j=1}^{n_z} \exp\left(-\frac{1}{2} [B(u_i)\,\theta - z_j(u_i)]^\top \Sigma_d^{-1} [B(u_i)\,\theta - z_j(u_i)] \right),$$

where $C > 0$ is a normalization constant. Since the logarithm is monotonically increasing, maximizing $L(\theta)$ over θ is equivalent to maximizing $\log L(\theta)$. Ignoring the factor C in front of the exponential function yields

$$\max_{\theta} \ \log L(\theta) = \max_{\theta} \ -\frac{1}{2} \sum_{i=1}^{d_u} \sum_{j=1}^{n_z} [B(u_i)\,\theta - z_j(u_i)]^\top \Sigma_d^{-1} [B(u_i)\,\theta - z_j(u_i)]$$

$$= \max_{\theta} \ -\frac{1}{2} [A(u)\,\theta - z(u)]^\top \Sigma^{-1} [A(u)\,\theta - z(u)],$$

which is equivalent to the following problem:

Problem 2.1 (PE-LS). Let $u \subset U$ be a set of inputs and $z(u) \in Q$ be the corresponding data as introduced above. Furthermore, let $\Sigma \in \mathbb{R}^{n \times n}$ be a symmetric, positive definite noise matrix. Then a *linear least-squares estimate* $\bar{\theta}_{LS} \in \mathbb{R}^{n_\theta}$ is a solution of the unconstrained optimization problem

$$\min_{\theta} \ \frac{1}{2} \|A(u)\,\theta - z(u)\|_{\Sigma^{-1}}^2 . \tag{2.4}$$

The solution of Problem 2.1 can be obtained directly:

Proposition 2.2. Let $u \subset U$ be a set of inputs, let the system matrix $A(u) \in \mathbb{R}^{n \times n_\theta}$ have full column-rank and let $\Sigma \in \mathbb{R}^{n \times n}$ be a symmetric, positive definite noise matrix. Furthermore, let some data $z(u) \in Q$ corresponding to the inputs be given. Then the unique solution $\bar{\theta}_{LS} = \bar{\theta}_{LS}(u; z)$ to Problem 2.1 is given by the *normal equations*

$$A(u)^\top \Sigma^{-1} A(u)\,\bar{\theta}_{LS}(u; z) = A(u)^\top \Sigma^{-1} z(u). \tag{2.5}$$

Proof. We denote by $J(\theta)$ the objective function in (2.4). Differentiation yields

$$J'(\bar{\theta}_{LS}) = A(u)^\top \Sigma^{-1} A(u)\,\bar{\theta}_{LS} - A(u)^\top \Sigma^{-1} z(u) \overset{!}{=} 0$$

and thus

$$A(u)^\top \Sigma^{-1} A(u)\,\bar{\theta}_{LS} = A(u)^\top \Sigma^{-1} z(u).$$

Since Σ is symmetric, positive definite and the system matrix $A(u)$ has full column-rank, the matrix $A(u)^\top \Sigma^{-1} A(u)$ is invertible. We have indeed a minimum since $J''(\bar{\theta}_{LS}) = 2A(u)^\top \Sigma^{-1} A(u)$ is positive definite as well. ∎

We now consider Problem 2.1 from a probabilistic point of view. The following is a natural consequence of Proposition 2.2:

Proposition 2.3. Let $u \subset U$ be a collection of inputs and $z(u) \in Q$ be the corresponding data vector as a realization of the random variable $Z(u) \sim \mathcal{N}(z^\star(u), \Sigma)$, where $z^\star = z^\star(u)$ is its mean value and Σ is its symmetric, positive definite covariance matrix. Moreover, let $A(u)$ have full column-rank. Then the optimal parameters $\overline{\theta}_{LS} = \overline{\theta}_{LS}(u; z)$ obtained by the solution of Problem 2.1 are realizations of Gaussian random variables with mean $\theta_{LS}(u, z^\star)$ and covariance matrix $C_{LS}(u)$:

$$\theta_{LS}(u, z^\star) = (A(u)^\top \Sigma^{-1} A(u))^{-1} A(u)^\top \Sigma^{-1} z^\star(u), \qquad C_{LS}(u) = \left(A(u)^\top \Sigma^{-1} A(u)\right)^{-1}.$$

Proof. By Proposition 2.2 we have

$$\begin{aligned}
\theta_{LS}(u; z^\star) = \mathbb{E}_z[\overline{\theta}_{LS}] &= (A(u)^\top \Sigma^{-1} A(u))^{-1} A(u)^\top \Sigma^{-1} \mathbb{E}_z[Z(u)] \\
&= (A(u)^\top \Sigma^{-1} A(u))^{-1} A(u)^\top \Sigma^{-1} z^\star(u).
\end{aligned}$$

Let $F(u) := A(u)^\top \Sigma^{-1} A(u)$ for convenience. The definition of the covariance of a random variable then yields

$$\begin{aligned}
C_{LS}(u) &= \mathbb{E}_z\left[(\overline{\theta}_{LS} - \theta_{LS}(u; z^\star)) \cdot (\overline{\theta}_{LS} - \theta_{LS}(u; z^\star))^\top\right] \\
&= F(u)^{-1} A(u)^\top \Sigma^{-1} \mathbb{E}_z\left[(Z(u) - z^\star(u)) \cdot (Z(u) - z^\star(u))^\top\right] \Sigma^{-1} A(u) F(u)^{-\top} \\
&= F(u)^{-1} A(u)^\top \Sigma^{-1} \Sigma \Sigma^{-1} A(u) F(u)^{-\top} \\
&= F(u)^{-\top}. \qquad \blacksquare
\end{aligned}$$

Remark 2.4. The inverse of the covariance matrix $M(u) := C_{LS}(u)^{-1} = A(u)^\top \Sigma^{-1} A(u)$ is also called *information matrix*. If Σ is a diagonal matrix with strictly positive entries $\sigma_i^2 = \Sigma_{ii}$, then

$$M(u) = \sum_{i=1}^n \sigma_i^{-2} A_i(u)^\top A_i(u),$$

where $A_i(u)^\top A_i(u) \in \mathbb{R}^{n_\theta \times n_\theta}$ is a rank-1 update with the row vectors $A_i(u) = A_{i\cdot}(u)$ for each $i = 1, \ldots, n$.

Moreover, we introduce the notion of bias in statistical estimators:

Definition 2.5. Let $\overline{\theta}_E(z)$ be a statistical estimator of the true value $\theta^\star$. Then we call the estimator *unbiased* if $\mathbb{E}_z[\overline{\theta}_E(Z)] = \theta^\star$.

Corollary 2.6. If the linear model is correct, then $\overline{\theta}_{\mathrm{LS}}$ is unbiased.

Proof. Let $\theta^\star$ be the true value of the model parameters. We have $A(u)\,\theta^\star = z^\star$ and by Proposition 2.3

$$
\begin{aligned}
\mathbb{E}_z[\overline{\theta}_{\mathrm{LS}}(u;z)] &= (A(u)^\top \Sigma^{-1} A(u))^{-1} A(u)^\top \Sigma^{-1} z^\star(u) \\
&= (A(u)^\top \Sigma^{-1} A(u))^{-1} A(u)^\top \Sigma^{-1} A(u)\,\theta^\star = \theta^\star.
\end{aligned}
$$
∎

Remark 2.7. Note, that we have treated the unconstrained case so far. If $\theta^\star$ is an interior point of $\Theta \subset \mathbb{R}^{n_\theta}$, the data $z(u)$ are reasonable for all admissible $u \subset U$ and the model is accurate enough, then this may be sufficient. However, for the constrained problem

$$
\begin{aligned}
&\min_\theta \;\; \frac{1}{2}\,\|A(u)\,\theta - z(u)\|^2_{\Sigma^{-1}}, \\
&\text{s.t. }\; \theta \in \Theta,
\end{aligned}
$$

a projection method may be suited, see [127] if $\theta \in \Theta$ describes box constraints.

2.1.2 Tikhonov Regularization

We consider the case where Problem 2.1 is ill-posed which happens if $\mathrm{rank}(A) < n_\theta$. A widely used method to deal with this ill-posedness is *Tikhonov regularization* around an arbitrary value $\theta_0 \in \mathbb{R}^{n_\theta}$. Adding a damping term to the objective function yields

Problem 2.8 (PE-DLS). Let $u \subset U$ be a set of inputs and $z(u) \in Q$ be the corresponding data. Moreover, let $\Sigma \in \mathbb{R}^{n \times n}$ be a symmetric, positive definite matrix and $\theta_0 \in \mathbb{R}^{n_\theta}$ be given. Then a *damped linear least-squares estimate* $\overline{\theta}_{\mathrm{DLS}} \in \mathbb{R}^{n_\theta}$ is a solution of the unconstrained optimization problem

$$
\min_\theta \;\; \frac{1}{2}\,\|A(u)\,\theta - z(u)\|^2_{\Sigma^{-1}} + \frac{\alpha}{2}\,\|\theta - \theta_0\|^2_2, \tag{2.6}
$$

where $\alpha > 0$ is the *damping parameter*.

We rewrite (2.6) to

$$
\min_\theta \;\; \frac{1}{2}\,\left\| \begin{pmatrix} \Sigma^{-\frac{1}{2}} A(u) \\ \sqrt{\alpha} I \end{pmatrix} \theta - \begin{pmatrix} \Sigma^{-\frac{1}{2}} z \\ \sqrt{\alpha}\,\theta_0 \end{pmatrix} \right\|^2_2 \tag{2.7}
$$

and observe that the last n_θ rows of the augmented matrix in (2.7) are linearly independent such that we have a full-rank linear least-squares problem.

Corollary 2.9. Let $u \subset U$ be a set of inputs and $z(u) \in Q$ be the corresponding data. Moreover, let $\Sigma \in \mathbb{R}^{n \times n}$ be a symmetric, positive definite matrix. Let $\theta_0 \in \mathbb{R}^{n_\theta}$ and $\alpha < 0$ be given. Then the unique solution $\overline{\theta}_{\mathrm{DLS}}(u; z)$ of Problem 2.8 is given by

$$\left(A(u)^\top \Sigma^{-1} A(u) + \alpha I\right) \overline{\theta}_{\mathrm{DLS}}(u; z) = A(u)^\top \Sigma^{-1} z(u) + \alpha \theta_0.$$

Proof. As mentioned above, (2.7) is a full-rank linear least-squares problem. Applying the normal equations, cf. Proposition 2.2, yields

$$\left(A(u)^\top \Sigma^{-\frac{1}{2}}, \ \sqrt{\alpha} I\right) \begin{pmatrix} \Sigma^{-\frac{1}{2}} A(u) \\ \sqrt{\alpha} I \end{pmatrix} \overline{\theta}_{\mathrm{DLS}}(u; z) = \left(A(u)^\top \Sigma^{-\frac{1}{2}}, \ \sqrt{\alpha} I\right) \begin{pmatrix} \Sigma^{-\frac{1}{2}} z \\ \sqrt{\alpha} \theta_0 \end{pmatrix},$$

which simplifies to the assertion. ∎

We also mention the statistical properties of the damped linear least-squares estimator:

Corollary 2.10. Let $u \subset U$ be a collection of inputs and $Z(u) \sim \mathcal{N}(z^\star(u), \Sigma)$ be the corresponding data random variable where Σ is symmetric and positive definite. Moreover, let $\theta_0 \in \mathbb{R}^{n_\theta}$, $\alpha > 0$ be given. Then the parameters $\overline{\theta}_{\mathrm{DLS}} = \overline{\theta}_{\mathrm{DLS}}(u; z)$ obtained by the solution of Problem 2.8 are realizations of Gaussian random variables with mean $\theta_{\mathrm{DLS}}(u; z^\star)$ and covariance matrix $C_{\mathrm{DLS}}(u)$ given by

$$\begin{aligned} \theta_{\mathrm{DLS}}(u; z^\star) &= \left(A(u)^\top \Sigma^{-1} A(u) + \alpha I\right)^{-1} \left(A(u)^\top \Sigma^{-1} z^\star(u) + \alpha \theta_0\right) \\ C_{\mathrm{DLS}}(u) &= \left(A(u)^\top \Sigma^{-1} A(u) + \alpha I\right)^{-1} A(u)^\top \Sigma^{-1} A(u) \left(A(u)^\top \Sigma^{-1} A(u) + \alpha I\right)^{-\top}. \end{aligned} \tag{2.8}$$

Proof. By Corollary 2.9 we have

$$\theta_{\mathrm{DLS}}(u; z^\star) = \mathbb{E}_z[\overline{\theta}_{\mathrm{DLS}}] = \left(A(u)^\top \Sigma^{-1} A(u) + \alpha I\right)^{-1} \left(A(u)^\top \Sigma^{-1} z^\star(u) + \alpha \theta_0\right).$$

Let $F(u) := A(u)^\top \Sigma^{-1} A(u) + \alpha I$ for convenience. Then

$$\begin{aligned} C_{\mathrm{DLS}}(u) &= \mathbb{E}_z\left[(\overline{\theta}_{\mathrm{DLS}} - \theta_{\mathrm{DLS}}(u; z^\star)) \cdot (\overline{\theta}_{\mathrm{DLS}} - \theta_{\mathrm{DLS}}(u; z^\star))^\top\right] \\ &= F(u)^{-1} A(u)^\top \Sigma^{-1} \mathbb{E}_z[(Z(u) - z^\star(u)) \cdot (Z(u) - z^\star(u))^\top] \Sigma^{-1} A(u) F(u)^{-\top} \\ &= F(u)^{-1} A(u)^\top \Sigma^{-1} \Sigma \Sigma^{-1} A(u) F(u)^{-\top} \\ &= F(u)^{-1} A(u)^\top \Sigma^{-1} A(u) F(u)^{-\top}. \end{aligned}$$
∎

Remark 2.11. The damped linear least-squares estimator is biased even if the linear model is correct since

$$\mathbb{E}_z[\overline{\theta}_{\text{DLS}}(u;z)] = \left(A(u)^\top \Sigma^{-1} A(u) + \alpha I\right)^{-1} \left(A(u)^\top \Sigma^{-1} A(u)\,\theta^\star + \alpha \theta_0\right) \neq \theta^\star$$

if $\theta_0 \neq \theta^\star$, where $\theta^\star$ is the true value of the model parameters.

So far, the damping term introduced by α and θ_0 has no meaningful statistical interpretation. Moreover, the choice of θ_0 is very subjective which may lead to a significantly large bias in the estimator. The disadvantages of using $C_{\text{DLS}}(u)$ as a covariance matrix of the estimated model parameters are discussed in Chapter 3.

Finally, we want to mention that for large-scale problems, i.e., the dimensions of the system matrix $A(u)$ being large, Conjugate Gradient or Krylov methods to solve the normal equations are more efficient, see [10, 113]. Furthermore, if $A(u)$ contains many zeros a sparse approach to the storage of its non-zero elements is advised.

2.2 Nonlinear Inverse Problems

In this section we do not assume that the model $\eta(\theta, u)$ is linear in θ but consider the general nonlinear case. We are mainly interested in inverse problems that involve partial differential equations (PDEs) which describe the underlying physical laws. Therefore, we introduce state variables $y \in Y \subset \mathbb{R}^{d_y}$ which obey the modeled physics summarized by a discretized operator $e : \Theta \times U \times Y \to Y$ in a *state equation*

$$e(\theta, u, y) = 0. \tag{2.9}$$

We denote the partial derivatives shortly by $\partial_\theta e := \partial_\theta e(\theta, u, y)$, $\partial_u e := \partial_u e(\theta, u, y)$ and $\partial_y e := \partial_y e(\theta, u, y)$ at some $(\theta, u, y) \in \Theta \times U \times Y$ if this is clear and in case they exist. By $\partial_y(\cdot)$ we mean the Jacobian $\mathscr{J}_y(\cdot)$. Since e is the discretization in space (and time) of a PDE we require that this discretized PDE always has a unique solution y. In this section we also require

Assumption 2.12. For every $\theta \in \Theta$, $u \in U$ and $y \in Y$ the following holds:
  (a) $\partial_y e$ and $\partial_\theta e$ exist and are continuous,
  (b) $\partial^2_{yy} e$, $\partial^2_{y\theta} e$, $\partial^2_{\theta y} e$ and $\partial^2_{\theta\theta} e$ exist and are continuous,
  (c) $\partial_y e$ is invertible.

By the Theorem of Schwarz, Assumption 2.12 (b) implies that $\partial^2_{y\theta} e = \partial^2_{\theta y} e$. Moreover, by the Implicit Function Theorem, Assumption 2.12 (a) and (c) imply that

there exists a unique solution $y(\theta; u)$ to Equation (2.9) which is continuously differentiable with respect to θ.

In general, not all state variables are of equal interest. Therefore, we define an *observation operator* as $(y, \theta) \mapsto h(y, \theta) \in Q_I$ which maps states and parameters to the quantity of interest that can be observed through measurements. The overall reduced model $\eta(\theta, u) := h(y(\theta; u), \theta)$ is composed of the solution operator of the state equation and the observation operator. We also require in this section

Assumption 2.13. The observation operator $(y, \theta) \mapsto h(y, \theta)$ is twice continuously differentiable.

Let $u := \{u_1, \ldots, u_m\} \subset U$ be a collection of $m \in \mathbb{N}$ inputs for which corresponding data $z(u_1), \ldots, z(u_m) \in Q_I$ are acquired. We again collect $n_z \in \mathbb{N}$ measurement series for each input to improve the information gain and adjust the dimensions of the data and model output vector to $n := n_z \cdot d_z \cdot m$ as in the linear case:

$$\tilde{z}(u_i) := \begin{pmatrix} z_1(u_i) \\ \vdots \\ z_{n_z}(u_i) \end{pmatrix}, \quad z(u) := \begin{pmatrix} \tilde{z}(u_1) \\ \vdots \\ \tilde{z}(u_m) \end{pmatrix} \in \mathbb{R}^n, \quad \tilde{h}(y_i, \theta) := \begin{pmatrix} h(y_i, \theta) \\ \vdots \\ \phantom{n_z \text{ times}} \\ \vdots \\ h(y_i, \theta) \end{pmatrix} \begin{matrix} \\ n_z \text{ times} \\ \\ \end{matrix} \tag{2.10}$$

and $\eta(\theta, u) := \big(\tilde{h}(y_1(\theta; u_1), \theta)^\top, \ldots, \tilde{h}(y_m(\theta; u_m), \theta)^\top\big)^\top \in \mathbb{R}^n$. We now naturally have

Problem 2.14 (PE-NLS). Let Assumption 2.12 be satisfied, $u \subset U$ be a set of inputs and $z(u) \in Q$ be the corresponding data. Let $\sigma_1, \ldots, \sigma_{d_z} > 0$, $\theta_0 \in \mathbb{R}^{n_\theta}$ and $\alpha \geq 0$ be given. Then a *(damped) nonlinear least-squares estimate* $\overline{\theta}_{\text{NLS}} \in \mathbb{R}^{n_\theta}$ is a (local) solution of the constrained optimization problem

$$\min_{\theta, y_1, \ldots, y_m} \sum_{j=1}^{m} \sum_{k=1}^{d_z} \sum_{i=1}^{n_z} \frac{1}{2\sigma_k^2} \big(h_k(y_j, \theta) - z_{ik}(u_j)\big)^2 + \frac{\alpha}{2} \|\theta - \theta_0\|_2^2, \tag{2.11}$$

$$\text{s.t.} \quad e(\theta, u_j, y_j) = 0, \quad \text{for all } j = 1, \ldots, m.$$

We can write a reduced formulation of (2.11) by inserting the reduced model into the objective function and introducing the diagonal matrix $\Sigma \in \mathbb{R}^{n \times n}$ as before:

$$\min_{\theta} \frac{1}{2} \|\eta(\theta, u) - z(u)\|_{\Sigma^{-1}}^2 + \frac{\alpha}{2} \|\theta - \theta_0\|_2^2. \tag{2.12}$$

2.2.1 Newton's Approach

We consider the reduced problem first. Let $f = f(\theta; z, u)$ be the objective function in (2.12) and in addition to this setting let Assumption 2.13 be satisfied. Newton's method requires the gradient and the Hessian of f with respect to θ. We have

$$\nabla_\theta f(\theta; z, u) = \partial_\theta \eta(\theta, u)^\top \Sigma^{-1}(\eta(\theta, u) - z(u)) + \alpha(\theta - \theta_0) \qquad (2.13)$$

and

$$\partial^2_{\theta\theta} f(\theta; z, u) = \partial_\theta \eta(\theta, u)^\top \Sigma^{-1} \partial_\theta \eta(\theta, u) + \alpha I + \sum_{i=1}^n \sigma_i^{-2}(\eta_i(\theta, u) - z_i(u)) \partial^2_{\theta\theta} \eta_i(\theta, u).$$

Writing $J(\theta; u) := \partial_\theta \eta(\theta, u)$ and $S(\theta; z, u) := \sum_{i=1}^n \sigma_i^{-2}(\eta_i(\theta, u) - z_i(u)) \partial^2_{\theta\theta} \eta_i(\theta, u)$ we reformulate the Hessian as

$$\partial^2_{\theta\theta} f(\theta; z, u) = H(\theta; z, u) = J(\theta; u)^\top \Sigma^{-1} J(\theta; u) + \alpha I + S(\theta; z, u). \qquad (2.14)$$

The exact expressions for $J(\theta; u)$ and $S(\theta; z, u)$ are computed in Appendix A.4. Provided that $H(\theta; z, u)$ is invertible in a neighborhood of a (local) solution $\overline{\theta}_{\mathrm{NLS}}$ of Problem 2.14 we can apply Newton steps from a suitable starting point θ^0 as depicted in Algorithm 2.1. The following result is standard:

Proposition 2.15. Let $B_r(\overline{\theta}_{\mathrm{NLS}}) \subset \mathbb{R}^{n_\theta}$ be a sufficiently small neighborhood of the local minimizer $\overline{\theta}_{\mathrm{NLS}}$ such that $f \in C^2(B_r(\overline{\theta}_{\mathrm{NLS}}), \mathbb{R})$, $\nabla_\theta f(\overline{\theta}_{\mathrm{NLS}}) = 0$ and the Hessian be Lipschitz and positive definite in $B_r(\overline{\theta}_{\mathrm{NLS}})$. Moreover, let $\theta^0 \in B_r(\overline{\theta}_{\mathrm{NLS}})$ be given. Then Algorithm 2.1 either terminates with $\theta^k = \overline{\theta}_{\mathrm{NLS}}$ for some $k \in \mathbb{N}$ or the sequence of iterations $(\theta^k)_{k \in \mathbb{N}}$ converges q-quadratically to $\overline{\theta}_{\mathrm{NLS}}$, i.e.,

$$\|\theta^{k+1} - \overline{\theta}_{\mathrm{NLS}}\| \leq C \|\theta^k - \overline{\theta}_{\mathrm{NLS}}\|^2, \quad \text{for all } k \geq 0, \text{ where } C > 0.$$

Proof. See [129, Thm. 10.8], for example. ∎

Since the relationship between data and estimated parameters is nonlinear, $\overline{\theta}_{\mathrm{NLS}}$ is no longer a realization of a multivariate normal distribution. However, if small perturbations in the data cause small changes in the parameter estimate a linearization of the mapping $Z \mapsto \overline{\theta}_{\mathrm{NLS}}(Z; u)$ at the estimate $\overline{\theta}_{\mathrm{NLS}}(z; u)$ may be justified:

$$\overline{\theta}_{\mathrm{NLS}}(Z; u) = \overline{\theta}_{\mathrm{NLS}}(z; u) + \partial_z \overline{\theta}_{\mathrm{NLS}}(z; u)(Z - z) + o(\|Z - z\|).$$

Algorithm 2.1 (Newton for PE [10, 129]).

Input: $f(\theta) \in C^2(\mathbb{R}^{n_\theta}, \mathbb{R})$, initial guess θ^0, small tolerance $\delta > 0$.
Output: sequence $\theta^1, \theta^2, \ldots$ that converges to the solution $\overline{\theta}_{\mathrm{NLS}}$ of (2.12).

  1: Set $k := 0$.
  2: **if** $\|\nabla_\theta f(\theta^k)\|_2 \leq \delta$ **then stop**.
  3: Calculate $\nabla_\theta f(\theta^k)$ and $\partial_{\theta\theta}^2 f(\theta)$ from (2.13) and (2.14).
  4: Solve $\partial_{\theta\theta}^2 f(\theta) \Delta \theta^k = -\nabla_\theta f(\theta^k)$.
  5: Determine step size $\alpha^k > 0$ via line search.
  6: Set $\theta^{k+1} := \theta^k + \alpha^k \Delta \theta^k$.
  7: Update $k := k + 1$ and go to line 2.

Following [11], we compute $\partial_z \overline{\theta}_{\mathrm{NLS}}(z; u)$ by applying the Implicit Function Theorem. Let the assumptions of Proposition 2.15 be satisfied. Then the Hessian $H(\theta; z, u)$ is (Lipschitz) continuous and invertible for all θ in a small neighborhood of $\overline{\theta}_{\mathrm{NLS}}(z; u)$. Since the mixed partial derivatives $\partial_{\theta z}^2 f(\overline{\theta}_{\mathrm{NLS}}(z; u); z, u)$ exist and are continuous, the Implicit Function Theorem applied to the first order optimality condition

$$\partial_\theta f(\overline{\theta}_{\mathrm{NLS}}(z; u); z, u) = 0$$

yields a mapping $Z \mapsto \overline{\theta}_{\mathrm{NLS}}(Z; u)$ which is continuously differentiable and its derivative $\partial_z \overline{\theta}_{\mathrm{NLS}}(z; u)$ satisfies

$$\partial_{\theta\theta}^2 f(\overline{\theta}_{\mathrm{NLS}}(z; u); z, u) \partial_z \overline{\theta}_{\mathrm{NLS}}(z; u) \tilde{Z} = -\partial_{\theta z}^2 f(\overline{\theta}_{\mathrm{NLS}}(z; u); z, u) \tilde{Z},$$

for any direction $\tilde{Z} \in Q$, and thus with (2.14)

$$\partial_z \overline{\theta}_{\mathrm{NLS}}(z; u) = H(\overline{\theta}_{\mathrm{NLS}}(z; u); z, u)^{-1} J(\overline{\theta}_{\mathrm{NLS}}(z; u); u)^\top \Sigma^{-1}.$$

By definition, we get a local, data-dependent covariance matrix at the (damped) nonlinear least-squares estimate:

$$\begin{aligned}
C_{\mathrm{NLS}}(\overline{\theta}_{\mathrm{NLS}}, z, u) &= \mathbb{E}_z \left[\partial_z \overline{\theta}_{\mathrm{NLS}}(z; u)(Z - z)(Z - z)^\top \partial_z \overline{\theta}_{\mathrm{NLS}}(z; u)^\top \right] \\
&= H(\overline{\theta}_{\mathrm{NLS}}; z, u)^{-1} J(\overline{\theta}_{\mathrm{NLS}}; u)^\top \Sigma^{-1} \mathbb{E}_z \left[(Z - z)(Z - z)^\top \right] \\
&\quad \Sigma^{-1} J(\overline{\theta}_{\mathrm{NLS}}; u) H(\overline{\theta}_{\mathrm{NLS}}; z, u)^{-\top} \\
&= H(\overline{\theta}_{\mathrm{NLS}}; z, u)^{-1} J(\overline{\theta}_{\mathrm{NLS}}; u)^\top \Sigma^{-1} J(\overline{\theta}_{\mathrm{NLS}}; u) H(\overline{\theta}_{\mathrm{NLS}}; z, u)^{-\top}.
\end{aligned}$$

We have proved the following:

Proposition 2.16. Let $u \subset U$ be given and the data $z = z(u)$ be a realization of $Z(u) \sim \mathcal{N}(z^{\star}(u), \Sigma)$. Let the assumptions of Proposition 2.15 be satisfied and let $\overline{\theta}_{\mathrm{NLS}} = \overline{\theta}_{\mathrm{NLS}}(z;u)$ be the output of Algorithm 2.1. Then, in a first order approximation, the estimated parameters are also Gaussian and the covariance matrix is given by

$$C_{\mathrm{NLS}}(\overline{\theta}_{\mathrm{NLS}}, z, u) = H(\overline{\theta}_{\mathrm{NLS}}; z, u)^{-1} J(\overline{\theta}_{\mathrm{NLS}}; u)^{\top} \Sigma^{-1} J(\overline{\theta}_{\mathrm{NLS}}; u) H(\overline{\theta}_{\mathrm{NLS}}; z, u)^{-\top}. \quad (2.15)$$

The disadvantages of using $C_{\mathrm{NLS}}(\overline{\theta}_{\mathrm{NLS}}, z, u)$ for the damped case, i.e., when $\alpha > 0$, are discussed in Chapter 3.

For large-scale problems, it is impracticable to compute the full Hessian. Therefore, consider the optimization problem (2.11) and apply the method of Lagrange multipliers. Let $g(\theta, y_1, \dots, y_m; z)$ be the objective function in (2.11). Then the Lagrangian is defined as

$$\mathfrak{L}_{\mathrm{PE}}(\theta, y_1, \dots, y_m, \lambda_1, \dots, \lambda_m) := g(\theta, y_1, \dots, y_m; z) + \sum_{i=1}^{m} \lambda_i^{\top} e(\theta, u_i, y_i),$$

where $\lambda_i \in Y$ are Lagrange multipliers, and the first order optimality conditions are given by

$$0 = \nabla_{\lambda_i} \mathfrak{L}_{\mathrm{PE}} = e(\theta, u_i, y_i), \quad (2.16)$$

$$0 = \nabla_{y_i} \mathfrak{L}_{\mathrm{PE}} = \nabla_{y_i} g(\theta, y_1, \dots, y_m; z) + \partial_{y_i} e(\theta, u_i, y_i)^{\top} \lambda_i, \quad (2.17)$$

$$0 = \nabla_{\theta} \mathfrak{L}_{\mathrm{PE}} = \nabla_{\theta} g(\theta, y_1, \dots, y_m; z) + \sum_{j=1}^{m} \partial_{\theta} e(\theta, u_j, y_j)^{\top} \lambda_j, \quad (2.18)$$

for $i = 1, \dots, m$. The right-hand side of (2.18) is the gradient of the objective function provided that y_i and λ_i are solutions of the state and adjoint equations (2.16)–(2.17), respectively. Solving systems of the form $H(\theta; z, u)\xi = r$ requires to find $(v_1, \dots, v_m, q_1, \dots, q_m, \xi)$ which satisfy

$$\partial_{y_i} e(\theta, u_i, y_i) q_i + \partial_{\theta} e(\theta, u_i, y_i)\xi = 0, \quad (2.19)$$

as well as

$$\begin{aligned}
\partial_{y_i} e(\theta, u_i, y_i)^{\top} v_i &+ \left\langle \partial^2_{y_i y_i} e(\theta, u_i, y_i)^{\top}, (\lambda_i, q_i) \right\rangle + \partial^2_{y_i y_i} g(\theta, y_1, \dots, y_m; z) q_i \\
&+ \left\langle \partial^2_{y_i \theta} e(\theta, u_i, y_i)^{\top}, (\lambda_i, \xi) \right\rangle + \partial^2_{y_i \theta} g(\theta, y_1, \dots, y_m; z)\xi = 0,
\end{aligned} \quad (2.20)$$

for $i = 1, \ldots, m$ and

$$
\partial_{\theta\theta}^2 g(\theta, y_1, \ldots, y_m; z)\xi + \sum_{i=1}^{m} \Big[\partial_\theta e(\theta, u_i, y_i)^\top v_i + \partial_{\theta y_i}^2 g(\theta, y_1, \ldots, y_m; z)q_i
$$
$$
+ \big\langle \partial_{\theta y_i}^2 e(\theta, u_i, y_i)^\top, (\lambda_i, q_i) \big\rangle + \big\langle \partial_{\theta\theta}^2 e(\theta, u_i, y_i)^\top, (\lambda_i, \xi) \big\rangle \Big] = r, \tag{2.21}
$$

compare [4]. This system is linear in $(v_1, \ldots, v_m, q_1, \ldots, q_m, \xi)$ and can be solved by Krylov methods requiring only matrix-vector products, see [113]. Thus, for large-scale problems the Newton step $\Delta\theta^k$ in line 4 of Algorithm 2.1 can be efficiently computed without implementing the full Hessian.

2.2.2 Gauss-Newton and Levenberg-Marquardt Methods

If the residuals $r_i(\theta, u) := \eta_i(\theta, u) - z_i(u)$ in (2.14) are small and the model is mildly nonlinar, i.e., $\| \partial_{\theta\theta}^2 \eta_i(\theta, u) \|_2$ is not too large, then the term $S(\theta; z, u)$ can be neglected and we have the *Gauss-Newton (GN) approximation* of the Hessian:

$$
H(\theta; z, u) \approx H_{\mathrm{GN}}(\theta; u) := J(\theta; u)^\top \Sigma^{-1} J(\theta; u) + \alpha I. \tag{2.22}
$$

Note, that H_{GN} is independent of data and free from second derivatives of the model. In a Gauss-Newton solver scheme, the Newton step in line 4 of Algorithm 2.1 is thus replaced by

$$
\big(J(\theta^k; u)^\top \Sigma^{-1} J(\theta^k; u) + \alpha I \big) \Delta\theta^k = J(\theta^k; u)^\top \Sigma^{-1} r(\theta^k, u) + \alpha(\theta^k - \theta_0). \tag{2.23}
$$

If Tikhonov regularization is active, then the matrix on the left-hand side is invertible since $J(\theta^k; u)^\top \Sigma^{-1} J(\theta^k; u)$ is positive semidefinite and αI is obviously positive definite. On the other hand, if $\alpha = 0$, we have to ensure that $J(\theta; u)$ has full column-rank for all θ in a small neighborhood of the solution. Since this may not always be the case, the *Levenberg-Marquardt method* modifies (2.23) to

$$
\big(J(\theta^k; u)^\top \Sigma^{-1} J(\theta^k; u) + \gamma^k I \big) \Delta\theta^k = J(\theta^k; u)^\top \Sigma^{-1} r(\theta^k, u), \tag{2.24}
$$

where $\gamma^k > 0$ is adjusted after every iteration. For large values of γ^k we achieve a steepest descent step, for small ones we get close to a Gauss-Newton step.

As in the linear case, if Tikhonov regularization is active, then the Gauss-Newton method leads to a biased solution of the nonlinear least-squares problem. However, the Levenberg-Marquardt approach (2.24) does not alter the solution at convergence. The additional term $\gamma^k I$ only stabilizes the solver that is used in (2.24), but

does not enter the objective function in (2.12). See also [10].

For a quasi-Newton modification of the Gauss-Newton step we refer to [34]. The convergence analysis is very similar as for Newton-type methods. In general, the convergence rate is linear and in some cases even superlinear. However, for large residuals $r_i(\theta, u)$ or for strongly nonlinear models $\eta(\theta, u)$ the Gauss-Newton method does not converge [22, 23].

Another view of the Gauss-Newton method is based on a sequence of linear approximations of the model $\eta(\theta, u)$. Let θ^k be the current approximation. Then the correction $\Delta\theta^k$ is the solution of the (damped) linear least-squares problem

$$\min_{\Delta\theta^k} \; \frac{1}{2} \left\| r(\theta^k, u) + J(\theta^k; u)\Delta\theta^k \right\|_{\Sigma^{-1}}^2 + \frac{\alpha}{2} \left\| \theta^k + \Delta\theta^k - \theta_0 \right\|_2^2. \tag{2.25}$$

By applying Corollary 2.9, we obtain

$$\Delta\theta^k = - \left(J(\theta^k; u)^\top \Sigma^{-1} J(\theta^k; u) + \alpha I \right)^{-1} \left(J(\theta^k; u)^\top \Sigma^{-1} r(\theta^k, u) + \alpha(\theta^k - \theta_0) \right)$$

which is identical to (2.23).

The statistical interpretation of the (damped) nonlinear least-squares estimate in Proposition 2.16 stays the same if one uses the Gauss-Newton method except that one replaces the full Hessian $H(\bar{\theta}_{\mathrm{NLS}}; z, u)$ by the Gauss-Newton approximation $H_{\mathrm{GN}}(\bar{\theta}_{\mathrm{NLS}}; u)$ in (2.15) which yields

$$C_{\mathrm{GN}}(\bar{\theta}_{\mathrm{NLS}}, u) = H_{\mathrm{GN}}(\bar{\theta}_{\mathrm{NLS}}; u)^{-1} J(\bar{\theta}_{\mathrm{NLS}}; u)^\top \Sigma^{-1} J(\bar{\theta}_{\mathrm{NLS}}; u) H_{\mathrm{GN}}(\bar{\theta}_{\mathrm{NLS}}; u)^{-\top}. \tag{2.26}$$

Again, the disadvantages of using $C_{\mathrm{GN}}(\bar{\theta}_{\mathrm{NLS}}, u)$ for the damped case are discussed in Chapter 3.

Remark 2.17. We again neglected the constraints $\theta \in \Theta$. As before, this may not be a problem if $\theta^\star$ is an interior point of $\Theta \subset \mathbb{R}^{n_\theta}$, the data $z(u)$ are reasonable for all admissible $u \subset U$ and the model is accurate enough. Otherwise, one can reformulate the inequality constraints as a penalty term in the objective function of Problem 2.14 and use penalty methods [129].

2.3 The Bayesian Perspective

This section is based on [10] and [125]. In the Bayesian approach to PE we encounter philosophically different ideas. In the perspective introduced before, we first computed a (deterministic) parameter estimate which is the solution to an optimization problem and *then* projected statistical properties upon it. We have always

assumed that there is a true (but unknown) value $\theta^\star$ of the model parameters which we want to approximate very well by an optimization procedure using data. This view of probability is also known as the *frequentist's perspective*. On the other hand, in the Bayesian approach the model parameters are random variables themselves and their estimation has a probability distribution known as *posterior*. One obtains the posterior by applying Bayes' formula to density functions. In the classical setting of conditional probability the Bayes' rule is given by

$$P[A \,|\, B] = \frac{P[B \,|\, A]\, P[A]}{P[B]},$$

for two events A and B.

A second difference is that the Bayesian view naturally incorporates prior knowledge about the parameters into the modeling process. We call all such information the *prior*, and in terms of Gaussian probability distributions it suffices to consider a prior mean θ_0 and a prior covariance Γ. As we will see, the Tikhonov regularization term introduced in the last two sections now has a statistical meaning and its bias upon the solution can be interpreted as incorporating prior knowledge into the modeling process.

Let $\pi_0 \sim \mathcal{N}(\theta_0, \Gamma)$ be a multivariate Gaussian prior and $u \subset U$ be a collection of inputs. We consider observational noise ε in the data $z = z(u)$ as before:

$$z = \eta(\theta, u) + \varepsilon,$$

where ε is a realization of the random variable $\mathscr{E} \sim \mathcal{N}(0, \Sigma)$ with Σ being a symmetric, positive definite covariance matrix Σ. We call $\rho(\cdot)$ the probability density function of the random variable $\mathscr{E}$. Then the data likelihood function $\pi(\cdot \,|\, \theta)$ has the density $\rho(\cdot - \eta(\theta, u))$. According to Bayes' formula the parameter's posterior probability distribution $\pi(\theta \,|\, z)$ is proportional to $\pi(z \,|\, \theta)\pi_0(\theta)$, more precisely

$$\begin{aligned}
\pi(\theta \,|\, z) &= c\, \pi(z \,|\, \theta)\pi_0(\theta) \\
&= c \exp\left(-\frac{1}{2} \|\eta(\theta, u) - z\|^2_{\Sigma^{-1}} - \frac{1}{2} \|\theta - \theta_0\|^2_{\Gamma^{-1}} \right),
\end{aligned} \tag{2.27}$$

where the proportionality constant c is given such that $\pi(\theta \,|\, z)$ is normalized:

$$\frac{1}{c} = \int_{\mathbb{R}^{n_\theta}} \rho(z - \eta(\theta, u))\pi_0(\theta)\, \mathrm{d}\theta.$$

In general, it is difficult to work with the entire distribution. Therefore, one consid-

ers a point that maximizes the posterior $\pi(\theta\,|\,z)$ which is equivalent to minimizing the negative exponent in the exponential function in (2.27).

Definition 2.18. The *maximum a posteriori estimator (MAP)* is a point $\overline{\theta}_{\mathrm{MAP}} \in \mathbb{R}^{n_\theta}$ such that the posterior $\pi(\theta\,|\,z)$ given in (2.27) becomes maximal:

$$\overline{\theta}_{\mathrm{MAP}}(z;u) := \underset{\theta}{\mathrm{argmin}}\ \frac{1}{2}\,\|\eta(\theta,u)-z\|_{\Sigma^{-1}}^2 + \frac{1}{2}\,\|\theta-\theta_0\|_{\Gamma^{-1}}^2. \qquad (2.28)$$

We observe that Tikhonov regularization is a particular instance of (2.28), namely, for $\Gamma = \alpha^{-1}I$. Thus, in the Bayesian context it has a statistical interpretation.

In general, the posterior is not Gaussian if the model η is nonlinear. However, for linear models $\eta(\theta,u) = A(u)\,\theta$ the posterior is a multivariate normal distribution with the following characteristics:

Proposition 2.19. Let $\Gamma \in \mathbb{R}^{n_\theta \times n_\theta}$, $\Sigma \in \mathbb{R}^{n \times n}$ be symmetric and positive definite and $\eta(\theta,u) = A(u)\,\theta$ be a linear model. Moreover, let $z = z(u) \in Q$ be given. Then the density function of the posterior $\pi(\theta\,|\,z)$ is proportional to

$$\pi(\theta\,|\,z) \propto \exp\left(-\frac{1}{2}\left\|\theta-\overline{\theta}_{\mathrm{LMAP}}\right\|_{C_{\mathrm{LB}}(u)^{-1}}^2\right),$$

with mean $\overline{\theta}_{\mathrm{LMAP}}$ and covariance $C_{\mathrm{LB}}(u)$ given by

$$\begin{aligned}
\overline{\theta}_{\mathrm{LMAP}}(z;u) &= \left(A(u)^\top \Sigma^{-1}A(u) + \Gamma^{-1}\right)^{-1}\left(A(u)^\top \Sigma^{-1}z(u) + \Gamma^{-1}\theta_0\right),\\
C_{\mathrm{LB}}(u) &= \left(A(u)^\top \Sigma^{-1}A(u) + \Gamma^{-1}\right)^{-1}.
\end{aligned} \qquad (2.29)$$

Proof. See [125, Thm. 2.4]. ∎

Remark 2.20. There is a striking difference between (2.8) and (2.29) in the covariance matrix even for the specific choice $\Gamma = \alpha^{-1}I$. The reason is that a Bayesian PE is fundamentally different from the classical approach where one solves a damped linear least-squares problem and *then* infers statistical properties of the solution. These differences have a significant impact on optimum experimental design which we present in Chapter 3.

The Bayesian approach can be extended to nonlinear models. After finding the MAP solution of (2.28) by a Newton-type method, see Section 2.2, one linearizes

the model around $\overline{\theta}_{\mathrm{MAP}}$:

$$\eta(\theta, u) = \eta(\overline{\theta}_{\mathrm{MAP}}, u) + J(\overline{\theta}_{\mathrm{MAP}}; u)(\theta - \overline{\theta}_{\mathrm{MAP}}) + o(\|\theta - \overline{\theta}_{\mathrm{MAP}}\|),$$

where $J(\overline{\theta}_{\mathrm{MAP}}; u) = \partial_\theta \eta(\overline{\theta}_{\mathrm{MAP}}, u)$. Then the expression for the approximated covariance matrix is a direct consequence of Proposition 2.19:

$$C_{\mathrm{B}} = \left(J(\overline{\theta}_{\mathrm{MAP}}; u)^\top \Sigma^{-1} J(\overline{\theta}_{\mathrm{MAP}}; u) + \Gamma^{-1} \right)^{-1}. \tag{2.30}$$

What is more, the Bayesian approach is naturally suited for ill-conditioned problems where the system matrix $A(u)$ or the Jacobian $J(\overline{\theta}_{\mathrm{MAP}}; u)$ become rank deficient. This is due to incorporating prior knowledge (θ_0, Γ) about the parameters directly into the model.

In most cases, adding inequality constraints $\theta \in \Theta$ is not needed since we expect the MAP point to be close to the prior θ_0. Moreover, it is unreasonable to choose a prior (θ_0, Γ) whose, e.g., 95% confidence region does not lie entirely in Θ.

2.4 Asymptotic Maximum Likelihood Approach

In this section we follow [31] and [112] without introducing all concepts rigorously. As before, let $\Theta \subset \mathbb{R}^{n_\theta}$ be the set of model parameters and let $m = 1$, i.e., let one input $u \in U$ be given. We consider the nonlinear model $\eta(\theta, u) = h(y(\theta; u), \theta)$ without repeating the output vector n_z times. Recall, that n_z is the number of measurement series where we repeatedly measure the quantity of interest yielding independent, *identically* distributed multivariate Gaussian data $z_1, \ldots, z_{n_z}$ each with known variance matrix $\Sigma = \mathrm{Diag}(\sigma_1^2, \ldots, \sigma_{d_z}^2) \in \mathbb{R}^{d_z \times d_z}$ and unknown mean $z^\star$.

As in Subsection 2.1.1 the likelihood function takes the form

$$L(\theta; n_z) := \prod_{i=1}^{n_z} L_i(\theta) = \prod_{i=1}^{n_z} \exp\left[-\frac{1}{2}(\eta(\theta, u) - z_i)^\top \Sigma^{-1}(\eta(\theta, u) - z_i) \right],$$

where we omitted the normalization factors. Moreover, let $\ell_i(\theta) := \log L_i(\theta)$ and the overall log-likelihood function be defined as

$$\ell(\theta; n_z) := \log L(\theta; n_z) = \sum_{i=1}^{n_z} \ell_i(\theta) = -\frac{1}{2} \sum_{i=1}^{n_z} (\eta(\theta, u) - z_i)^\top \Sigma^{-1}(\eta(\theta, u) - z_i). \tag{2.31}$$

Definition 2.21. The *maximum likelihood estimator* $\bar{\theta}_{\mathrm{MLE}}(n_z) \in \Theta$ is a value that maximizes the log-likelihood function:

$$\bar{\theta}_{\mathrm{MLE}}(n_z) := \underset{\theta \in \Theta}{\mathrm{argmax}}\ \ell(\theta; n_z).$$

We call $\bar{\theta}_{\mathrm{LE}}(n_z) \in \mathbb{R}^{n_\theta}$ just a *likelihood estimator* if it is a solution of

$$\partial_\theta \ell(\theta; n_z) = 0. \tag{2.32}$$

Asymptotic distribution theory determines conditions under which a (maximum) likelihood estimator converges to the true (but unknown) value $\theta^\star$ if $n_z \to \infty$, and characterizes their distribution in the limit case. We present a few concepts of this theory in Appendix A.3 in order to make the following statements meaningful.

For the rest of this section, let Assumption 2.12 and Assumption 2.13 be satisfied, Θ be compact and $\theta^\star$ be an interior point of Θ. First, we observe

$$\mathbb{E}_z[\nabla_\theta \ell_i(\theta)] = -\mathbb{E}_z[J(\theta; u)^\top \Sigma^{-1}(\eta(\theta, u) - Z)] = -J(\theta; u)^\top \Sigma^{-1}(\eta(\theta, u) - z^\star),$$

where $J(\theta; u) = \partial_\theta \eta(\theta, u)$. Particularly,

$$\mathbb{E}_z[\nabla_\theta \ell_i(\theta^\star)] = 0,$$

if the model is correct, i.e., if $\eta(\theta^\star, u) = z^\star$ which we also assume from now on.

Definition 2.22. The *Fisher information matrix* $F(\theta^\star, u) \in \mathbb{R}^{n_\theta \times n_\theta}$ is defined per observed sample as

$$F(\theta^\star, u) := \mathbb{E}_z[\nabla_\theta \ell_i(\theta^\star) \nabla_\theta \ell_i(\theta^\star)^\top],$$

if this expectation exists.

In our case

$$\begin{aligned}
F(\theta^\star, u) &= J(\theta^\star; u)^\top \Sigma^{-1} \mathbb{E}_z[(\eta(\theta^\star, u) - Z)(\eta(\theta^\star, u) - Z)^\top] \Sigma^{-1} J(\theta^\star; u) \\
&= J(\theta^\star; u)^\top \Sigma^{-1} J(\theta^\star; u),
\end{aligned}$$

which has a similarity to the Gauss-Newton covariance matrix (2.26) in absence of damping. We further have a useful identity:

Lemma 2.23. Let $\partial_{\theta\theta}^2 \ell_i(\theta^\star)$ have finite expectation and $F(\theta^\star, u)$ exist. Then

$$-\mathbb{E}_z[\partial_{\theta\theta}^2 \ell_i(\theta^\star)] = F(\theta^\star, u).$$

Proof. See [112, Lem. 14.4] or [44]. ∎

Our main result of this section is a statement about asymptotic normality of the likelihood estimator:

Proposition 2.24. Let the following additional assumptions hold:
   (a) $\partial_{\theta\theta}^2 \ell_i(\theta^\star)$ has finite expectation,
   (b) if $\eta(\theta, u) = \eta(\theta^\star, u)$, then $\theta = \theta^\star$ and
   (c) $F(\theta^\star, u)$ exists and is positive definite.
Let $\overline{\theta}_{\mathrm{LE}}(n_z)$ be a likelihood estimator. Then $\overline{\theta}_{\mathrm{LE}}(n_z)$ converges in probability to $\theta^\star$ and $\sqrt{n_z}(\overline{\theta}_{\mathrm{LE}}(n_z) - \theta^\star)$ converges in distribution to $\mathcal{N}(0, F(\theta^\star, u)^{-1})$, for $n_z \to \infty$.

Proof. The proof is an application of the central limit theorem (Proposition A.13) in the multivariate case, see [30, p. 500 et seq.] or [44, Thm. 18]. The other technical assumptions stated there are fulfilled in our setting since we consider only Gaussian probability density functions. ∎

Remark 2.25. A few comments are at hand:
   (a) The statement of this proposition does *not* say that $\overline{\theta}_{\mathrm{MLE}}(n_z)$ is asymptotically normal. There is only a sequence of likelihood estimators $\overline{\theta}_{\mathrm{LE}}(n_z)$ that solve (2.32) and are asymptotically normal. For nonlinear models, multiple solutions of (2.32) may exist and it may be difficult to obtain $\overline{\theta}_{\mathrm{MLE}}(n_z)$. Further details can be found in [44].
   (b) As a practical application and in view of Lemma 2.23, one can approximate $\overline{\theta}_{\mathrm{LE}}(n_z)$ for large n_z by a Gaussian distribution with covariance matrix

$$C_{\mathrm{AL}} = \left(-n_z \mathbb{E}_z[\partial_{\theta\theta}^2 \ell_i(\theta^\star)]\right)^{-1}.$$

Since $\theta^\star$ is unknown we choose to insert $\overline{\theta}_{\mathrm{LE}}(n_z)$ and arrive at

$$C_{\mathrm{AL}} \approx \left[-n_z \frac{1}{n_z} \sum_{i=1}^{n_z} \partial_{\theta\theta}^2 \ell_i(\overline{\theta}_{\mathrm{LE}}(n_z))\right]^{-1} = H(\overline{\theta}_{\mathrm{LE}}(n_z); z, u)^{-1}, \qquad (2.33)$$

where $H(\,\cdot\,; z, u)$ is exactly the Hessian from (2.14) for $m = 1$ and $\alpha = 0$.

This approach does not take into account prior information (θ_0, Γ) about the

model parameters yet. To do this, the asymptotic theory needs to be extended to the Bayesian setting which the reader can find in [130], for example. Following [11], we only motivate the incorporation of the term Γ^{-1} into the Hessian approximation of the covariance matrix of a likelihood estimate in Equation (2.33).

Let $\bar{\theta}_{\mathrm{MAP}}$ be the maximum a posteriori estimator (MAP) which is the solution of the optimization problem (2.28). This estimate maximizes the logarithm of the posterior distribution $\pi(\theta \,|\, z)$. Then a quadratic Taylor model is given by

$$\log \pi(\theta \,|\, z) = \log \pi(\bar{\theta}_{\mathrm{MAP}} \,|\, z) - \frac{1}{2}(\theta - \bar{\theta}_{\mathrm{MAP}})^{\top} H_{\mathrm{B}}(\bar{\theta}_{\mathrm{MAP}}; z, u)(\theta - \bar{\theta}_{\mathrm{MAP}})$$
$$+ o(\|\theta - \bar{\theta}_{\mathrm{MAP}}\|^2),$$

where in this case

$$H_{\mathrm{B}}(\bar{\theta}_{\mathrm{MAP}}; z, u) = J(\bar{\theta}_{\mathrm{MAP}}; u)^{\top} \Sigma^{-1} J(\bar{\theta}_{\mathrm{MAP}}; u) + \Gamma^{-1} + S(\bar{\theta}_{\mathrm{MAP}}; z, u), \qquad (2.34)$$

compare also (2.14). It follows that

$$\pi(\theta \,|\, z) \approx \exp\left(-\frac{1}{2}\left\|\theta - \bar{\theta}_{\mathrm{MAP}}\right\|^2_{H_{\mathrm{B}}(\bar{\theta}_{\mathrm{MAP}}; z, u)}\right),$$

and thus the posterior is approximately Gaussian with mean $\bar{\theta}_{\mathrm{MAP}}$ and covariance matrix

$$C_{\mathrm{AB}} = H_{\mathrm{B}}(\bar{\theta}_{\mathrm{MAP}}; z, u)^{-1}. \qquad (2.35)$$

We call this expression the *asymptotic Bayesian covariance approximation (AB)*.

2.5 Confidence Regions

In this section we follow [36] and [10]. Given an estimate $\bar{\theta}$ and its covariance matrix C one is interested in a confidence region $K \subset \mathbb{R}^{n_\theta}$ in which it is very likely that the true values $\theta^\star$ of the model parameters lie. Confidence regions are the multidimensional analogue of confidence intervals.

For linear models and a linear least-squares estimator the expression

$$(\theta^\star - \bar{\theta}_{\mathrm{LS}})^{\top} C_{\mathrm{LS}}^{-1}(\theta^\star - \bar{\theta}_{\mathrm{LS}})$$

has a chi-squared distribution, see Proposition 2.3 and Appendix A.2, and the confi-

dence region for $\theta^\star$ is exactly an n_θ-dimensional ellipsoid

$$K(\alpha, \overline{\theta}_{\mathrm{LS}}, C_{\mathrm{LS}}) := \left\{ \theta \in \mathbb{R}^{n_\theta} : (\theta - \overline{\theta}_{\mathrm{LS}})^\top C_{\mathrm{LS}}^{-1} (\theta - \overline{\theta}_{\mathrm{LS}}) \leq \gamma_{n_\theta}(1 - \alpha) \right\} \qquad (2.36)$$

for a given confidence level $1 - \alpha$, where $\alpha \in (0, 1)$ and $\gamma_{n_\theta}(\cdot)$ is the quantile function of the chi-squared distribution $\chi^2_{n_\theta}$ with n_θ degrees of freedom. In other words, the probability that $\theta^\star \in K(\alpha, \overline{\theta}_{\mathrm{LS}}, C_{\mathrm{LS}})$ is $1 - \alpha$.

For damped linear least-squares estimators, which are usually biased, confidence regions of this kind are misleading.

In case of nonlinear models the specific form of the confidence region is nonlinear as well and it depends on the probabilistic perspective. However, if *linearization* methods are applied, then the confidence region is in a first order approximation an n_θ-dimensional ellipsoid

$$K(\alpha, \overline{\theta}, C) := \left\{ \theta \in \mathbb{R}^{n_\theta} : (\theta - \overline{\theta})^\top C^{-1} (\theta - \overline{\theta}) \leq \gamma_{n_\theta}(1 - \alpha) \right\}, \qquad (2.37)$$

around the estimate $\overline{\theta}$ given by the Newton, Gauss-Newton, (asymptotic) Bayesian or asymptotic maximum likelihood method, and with covariance matrix C given by (2.15), (2.26), (2.30), (2.35) or (2.33), respectively. This expression of the confidence region is convenient for us for several reasons and we use it in our optimum experimental design approach, see Chapter 3.

On the other hand, the likelihood method yields a *nonlinear approximation* of the confidence region for nonlinear models. If $\ell(\theta; n_z)$ is the log-likelihood function from (2.31) and $\overline{\theta}_{\mathrm{LE}}$ is a likelihood estimate, then

$$\left\{ \theta \in \mathbb{R}^{n_\theta} : \ell(\overline{\theta}_{\mathrm{LE}}; n_z) - \ell(\theta; n_z) \leq \gamma_{n_\theta}(1 - \alpha) \right\} \qquad (2.38)$$

is no longer ellipsoidal. For linear models, (2.38) and (2.36) coincide. In general, the likelihood approximation of the confidence region has several disadvantages. The set in (2.38) can be disconnected and unbounded. Furthermore, it is very expensive to determine its shape. We refer to [14] for more details.

The lack of fit method is an exact way to determine nonlinear confidence regions but it is still more expensive than the likelihood method [36]. Thus, we refrain from further details.

2.6 Summary

In this chapter we introduced several approaches and probabilistic perspectives to PE. Given a data vector $z(u) \in Q$ for some inputs $u \subset U$ and a model η we developed estimators $\bar{\theta}$ of the true (but unknown) model parameters $\theta^\star$ for both linear and nonlinear models. Moreover, we analyzed their statistical properties. We conclude that in absence of damping or priors there are three reasonable ways to approximate the covariance matrix of the Gaussian distribution of the estimated parameters:

$$C_{\mathrm{NLS}}(\bar{\theta}_{\mathrm{NLS}}, z, u) = H(\bar{\theta}_{\mathrm{NLS}}; z, u)^{-1} J(\bar{\theta}_{\mathrm{NLS}}; u)^\top \Sigma^{-1} J(\bar{\theta}_{\mathrm{NLS}}; u) H(\bar{\theta}_{\mathrm{NLS}}; z, u)^{-\top}, \quad (2.39)$$

$$C_{\mathrm{GN}}(\bar{\theta}_{\mathrm{NLS}}, u) = \left[J(\bar{\theta}_{\mathrm{NLS}}; u)^\top \Sigma^{-1} J(\bar{\theta}_{\mathrm{NLS}}; u) \right]^{-1}, \quad (2.40)$$

$$C_{\mathrm{AL}}(\bar{\theta}_{\mathrm{LE}}, z, u) = H(\bar{\theta}_{\mathrm{LE}}; z, u)^{-1}, \quad (2.41)$$

and two ways if priors are used:

$$C_{\mathrm{B}}(\bar{\theta}_{\mathrm{MAP}}, u) = \left(J(\bar{\theta}_{\mathrm{MAP}}; u)^\top \Sigma^{-1} J(\bar{\theta}_{\mathrm{MAP}}; u) + \Gamma^{-1} \right)^{-1}, \quad (2.42)$$

$$C_{\mathrm{AB}}(\bar{\theta}_{\mathrm{MAP}}, z, u) = H_{\mathrm{B}}(\bar{\theta}_{\mathrm{MAP}}; z, u)^{-1}, \quad (2.43)$$

where the subscripts stand for NLS = Nonlinear Least Squares, GN = Gauss-Newton, B = Bayes and AB/AL = asymptotic Bayes/asymptotic likelihood.

In [34] the authors suggest that (2.39) is more useful if the sample size is small and in cases where the conditions of the asymptotic maximum likelihood theory are violated. Moreover, they find (2.40) not justified if the residuals are large. However, (2.40) is favored in [36] among the three variants. In the presence of priors, (2.42) is quite standard [47] but one frequently finds (2.43) as well in the literature, see, e.g., [4].

Note that for linear models the expressions (2.39)–(2.41) coincide. The same holds true for (2.42)–(2.43). The extent and severity of the model's nonlinearity decides how much these approximations of the parameter's covariance differ.

Optimum Experimental Design with PDE Constraints

It is natural to ask the question how good the solution of the parameter estimation problem (PE) actually is in terms of statistical measures. In optimum experimental design (OED) we address the problem of finding a perfect experiment determined by sensor locations and acting inputs such that the parameter estimate, which is inferred from data collected in such an experiment, has minimum variance. This approach leads to another optimization problem and we add to its complexity by introducing constraints that consist of discretized partial differential equations (PDEs).

In order to illustrate the problem more rigorously, let $G \subset \mathbb{R}^d$ be a bounded domain with Lipschitz boundary $\partial G \subset \mathbb{R}^d$, see Figure 3.1. We consider G to be fixed at $\Gamma_D \subset \partial G$ indicated by red straight lines, free at $\Gamma_F \subset \partial G$ and acted at $\Gamma_N \subset \partial G$ with time-dependent inputs $u(t)$ indicated by blue arrows. Here, the state variables $y(t) \in Y$, parameters $\theta \in \Theta$ and inputs $u \in U$ satisfy a time-variant PDE which we consider in a discretized form. The bullets in Figure 3.1 represent sensors that are placed on Γ_F to collect measurements $z \in Q_I$. These data are used to compute a parameter estimate $\overline{\theta} = \overline{\theta}(z)$. The main

Figure 3.1 OED Motivation.

question in this chapter is where to locate the *sensors* and how to choose the *inputs* $u(t) \in U$ such that a version of the covariance matrix $C(\overline{\theta}, u(t))$ of the parameter estimate, cf. Section 2.6, becomes "minimal" in a sense that will be explained below. We interpret small variances as a big information gain and this interpretation
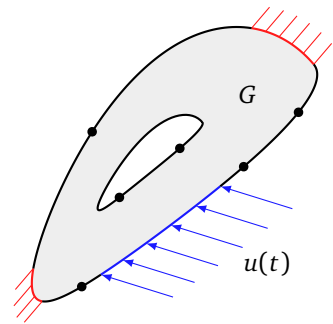
is supported by the notion of the information matrix, see also Remark 2.4.

This chapter is structured as follows. Section 3.1 introduces the classic approach to optimal design of experiments as it was developed by Fedorov and Leonov [43], Pukelsheim [108] and others. In Section 3.2 we make the transition to more modern approaches to optimal sensor placement, specifically when PDE constraints are present. We extend this method to optimal input configuration in Section 3.3 and finally discuss advantages and disadvantages of common design criteria in Section 3.4.

3.1 The Classical Method for Optimal Design of Experiments

In this section we follow [43, Chap. 2]. Let us adopt the notation of Section 2.1 and consider linear models $\eta(\theta, u) = A(u)\theta$ where the system matrix $A(u) \in \mathbb{R}^{n \times n_\theta}$ has full column-rank. Moreover, let $X \subset \mathbb{R}^{d_u}$ be closed and bounded. In classical optimal design of experiments the time-invariant input u is understood to be a collection of m design points $x_1, \ldots, x_m \in X$ and the number of repeated observations $r_i \in \mathbb{N}$ taken at x_i, $i = 1, \ldots, m$. An experiment is performed with m sensors positioned at $x_1, \ldots, x_m$ which collect a fixed number N of observations. Thus, we always have the constraint $r_1 + \ldots + r_m = N$. We refer to the collection

$$u = \{(x_1, r_1), \ldots, (x_m, r_m)\} \subset X \times \mathbb{N} =: U, \ \text{where} \ \sum_{i=1}^{m} r_i = N,$$

and to the number m as the *design of the experiment* from now on. The design enters the system matrix $A(u)$ in the following way:

$$\tilde{A}(u_i) := \begin{pmatrix} B(x_i) \\ \vdots \\ r_i \text{ times} \\ \vdots \\ B(x_i) \end{pmatrix}, \ \text{for all } i = 1, \ldots, m \quad \text{and} \quad A(u) := \begin{pmatrix} \tilde{A}(u_1) \\ \vdots \\ \tilde{A}(u_m) \end{pmatrix},$$

where $B(x_i)$ is the elemental system matrix at the design point x_i. Thus, $A(u)$ has $n := d_z \cdot (r_1 + \ldots + r_m)$ rows.

Using Proposition 2.3, let the linear least-squares estimate $\bar{\theta}_{\text{LS}}$ and its information matrix $M(u) = C_{\text{LS}}(u)^{-1}$ be given. It is our aim to find an optimal design

$(\bar{u}, \bar{m}) \in U \times \mathbb{N}$ such that $C_{\mathrm{LS}}(u)$ is minimal (or equivalently $M(u)$ is maximal) in the sense of the Loewner order of symmetric, positive definite matrices. This order is defined in the following way.

Definition 3.1. Let $A, B \succ 0$, i.e., A and B are symmetric, positive definite matrices with compatible dimensions. Then we say $A \succ B$ if, and only if, $A - B \succ 0$.

Thus, one may suggest the following optimization problem:

$$
\min_{u,m} \; C_{\mathrm{LS}}(u)
$$

$$
\text{s.t.} \;\; u = \{(x_1, r_1), \ldots, (x_m, r_m)\} \subset U, \; \sum_{i=1}^{m} r_i = N. \tag{3.1}
$$

The solution of this problem yields the optimal number of used sensors $\bar{m}$, their optimal positions $\bar{x}_i$ and the optimal number of repeated observations $\bar{r}_i$ at $\bar{x}_i$, for $i = 1, \ldots, \bar{m}$. However, there are examples with very simple models η where the optimization problem (3.1) has no solution, see [43, p. 50]. It is therefore inevitable to consider less ambitious objectives instead.

Definition 3.2. Let $S_{n_\theta}^+(\mathbb{R})$ be the set of all real $n_\theta \times n_\theta$ matrices which are symmetric, positive definite. We call a continuous function $\phi : S_{n_\theta}^+(\mathbb{R}) \to (0, \infty)$ a *criterion of optimality*, which maps the information matrix $M(\cdot) \in S_{n_\theta}^+(\mathbb{R})$ to a scalar value, if it has the following structure: $\phi[M(\cdot)] = \Psi[M(\cdot)^{-1}] = \Psi[C(\cdot)]$, where the function $\Psi(\cdot)$ is either $\Psi_{\mathrm{A}}(C) = \mathrm{tr}(C)$, $\Psi_{\mathrm{D}}(C) = \det(C)$ or $\Psi_{\mathrm{E}}(C) = \lambda_{\max}(C)$.

The functions $\Psi_{\mathrm{A/D/E}}(\cdot)$ are also called *design criteria* and we discuss their properties, advantages and disadvantages in Section 3.4. For theoretical reasons, we consider the mapping $M \mapsto \phi[M]$ in this chapter. The objective function in (3.1) is now replaced by a criterion of optimality:

Problem 3.3. Let $\bar{\theta}_{\mathrm{LS}}$ be the linear least-squares estimate and $M(\cdot) = C_{\mathrm{LS}}(\cdot)^{-1}$ be its information matrix. Moreover, let $\phi : S_{n_\theta}^+(\mathbb{R}) \to (0, \infty)$ be a criterion of optimality. Then an *optimal discrete design* $(\bar{u}, \bar{m}) \in U \times \mathbb{N}$ of an experiment with N observations is the solution of the constrained optimization problem

$$
\min_{u,m} \; \phi[M(u)]
$$

$$
\text{s.t.} \;\; u = \{(x_1, r_1), \ldots, (x_m, r_m)\} \subset U, \; \sum_{i=1}^{m} r_i = N. \tag{3.2}
$$

Notice that (3.2) is discrete since m and $r_1, \ldots, r_m$ are natural numbers. This makes it extremely difficult to solve.

Remark 3.4. The usage of the damped linear least-squares estimate and its covariance matrix $C_{\mathrm{DLS}}(\cdot)$, see Corollary 2.10, leads to unreasonable optimal discrete designs. If there is a design (u, m) such that $A(u)^\top \Sigma^{-1} A(u) = 0$, then $C_{\mathrm{DLS}}(u)$ is the null matrix and $\Psi[C_{\mathrm{DLS}}(u)] = 0$. Thus, $\bar{u} = u$ and $\bar{m} = m$ but such designs are undesirable. Even a restriction of the set of design points X to avoid such cases is not a solution to this ill-posedness since any design criterion $\Psi(\cdot)$ that minimizes $C_{\mathrm{DLS}}(\cdot)$ actually minimizes $A(u)^\top \Sigma^{-1} A(u)$. But $A(u)^\top \Sigma^{-1} A(u)$ is the information gain which we want to *maximize*. Therefore, in OED we discard any covariance matrix that includes Tikhonov regularization. However, regularization can be realized within a Bayesian framework.

An equivalent formulation of Problem 3.3 introduces weights $p_i = r_i/N$ and uses them in the design variable u instead of r_i. Then, clearly, $p_1 + \ldots + p_m = 1$ and the normalized information matrix can be written as

$$\widetilde{M}(u) := \frac{1}{N} M(u) = \sum_{i=1}^{m} p_i F(x_i), \tag{3.3}$$

where $F(x_i) = B(x_i)^\top \Sigma^{-1} B(x_i)$ is the Fisher information matrix which is independent of the parameter estimate in our linear setting, cf. Definition 2.22. We denote Equation (3.3) as the *additive property* of the information matrix. The difficulties which arise when solving (3.2) can be circumvented if one allows *continuous* weights $p_i \in [0, 1]$. Thus, a continuous design u_c of an experiment consists of the collection

$$u_c = \{(x_1, p_1), \ldots, (x_m, p_m)\} \subset X \times [0, 1] =: U_c, \quad \text{where} \quad \sum_{i=1}^{m} p_i = 1,$$

and the number m of design points. There is no restriction on m in this case. We define a convex combination of two continuous designs

$$u_c = \alpha u_{c1} + (1 - \alpha) u_{c2}, \tag{3.4}$$

where $\alpha \in (0, 1)$ in the following way. If a design point x_{j1} belongs to u_{c1} only, then its weight in the new design is $p_j := \alpha p_{j1}$. Similarly, if the design point x_{j2} belongs to u_{c2} only, then its weight in the new design is $p_j := (1 - \alpha) p_{j2}$. Design points that are common to both u_{c1} and u_{c2} have the new weight $p_j := \alpha p_{j1} + (1 - \alpha) p_{j2}$. Thus, $p_1 + \ldots + p_m = 1$ is always guaranteed. The number m of new design points is the

cardinality of the set $\{x_{j1} : j = 1, \ldots, m_1\} \cup \{x_{j2} : j = 1, \ldots, m_2\}$.

Having these notions settled, we observe that the definition of the continuous weights implies that any (discrete) probability measure on X can be a design whose density function is given by the Fisher information matrix. Thus, in view of (3.3) and (3.4) a convex combination of normalized information matrices is well defined:

$$\widetilde{M}(\alpha u_{c1} + (1 - \alpha) u_{c2}) = \alpha \widetilde{M}(u_{c1}) + (1 - \alpha) \widetilde{M}(u_{c2}), \tag{3.5}$$

where $\alpha \in (0, 1)$.

We are ready to restate Problem 3.3 in a continuous version:

Problem 3.5. Let $\bar{\theta}_{\mathrm{LS}}$ be the linear least-squares estimate and $\widetilde{M}(\cdot) = \frac{1}{N} C_{\mathrm{LS}}(\cdot)^{-1}$ be its normalized information matrix. Moreover, let $\phi : S_{n_\theta}^+(\mathbb{R}) \to (0, \infty)$ be a criterion of optimality. Then an *optimal continuous design* $(\bar{u}_c, \overline{m}) \in U_c \times \mathbb{N}$ of an experiment is the solution of the constrained optimization problem

$$\min_{u_c, m} \ \phi[\widetilde{M}(u_c)]$$

$$\text{s.t.} \ \ u_c = \{(x_1, p_1), \ldots, (x_m, p_m)\} \subset U_c, \ \sum_{i=1}^{m} p_i = 1. \tag{3.6}$$

In order to guarantee that a solution to (3.6) exists and that it has desirable properties we make the following assumption:

Assumption 3.6. Let the following hold:

(a) X is compact,

(b) the Fisher information matrix $F(x)$ is continuous with respect to $x \in X$,

(c) $\phi(\cdot)$ is convex, i.e., for all $M_1, M_2 \in S_{n_\theta}^+(\mathbb{R})$ and $\alpha \in (0, 1)$

$$\phi(\alpha M_1 + (1 - \alpha) M_2) \leq \alpha \phi(M_1) + (1 - \alpha) \phi(M_2),$$

(d) $\phi(\cdot)$ is a monotonically decreasing function, i.e., for all $M_1, M_2 \in S_{n_\theta}^+(\mathbb{R})$

$$M_1 \succ M_2 \ \Rightarrow \ \phi(M_1) \leq \phi(M_2),$$

(e) there exists $q \in \mathbb{R}$ such that $U_c(q) := \{u_c \in U_c : \phi[\widetilde{M}(u_c)] \leq q\}$ is non-empty.

We can state the main result of this section:

Proposition 3.7. Let Assumption 3.6 and Equation (3.5) be satisfied. Then there exists an optimal continuous design $(\bar{u}_c, \overline{m}) \in U_c \times \mathbb{N}$ which solves Problem 3.5 in such a way that $\overline{m} \leq n_\theta(n_\theta + 1)/2$. Moreover, the set of optimal designs is convex.

Proof. See [43, Thm. 2.2]. ∎

Remark 3.8. A few comments are at hand:
  (a) In general, the solution to Problem 3.5 is not unique. The experimenter still has the difficulty to choose a suitable optimal continuous design. However, this may not necessarily be a disadvantage. If $\phi(\cdot)$ is even strictly convex, then the optimal normalized information matrix is unique. This follows by a direct contradiction argument.
  (b) A solution to (3.6) may not be a solution to (3.2). In fact, $r_i = Np_i$ may not be an integer. In this case rounding strategies are often acceptable where r_i is rounded to the nearest integer such that $r_1 + \ldots + r_m = N$ is satisfied. However, a rounded solution may not be an optimal discrete design.
  (c) Efficient numerical techniques to solve Problem 3.5 rely on a particular representation of the derivate of $\phi(\cdot)$ and on iteratively adding design points. An algorithm based on first-order derivatives can be found in [43, Chap. 3], for example.

   If prior information about the model parameters is available as it is the case in a Bayesian approach, then the covariance matrix is given by $C_{\mathrm{LB}}(u)$, see (2.29), and the normalized information matrix is defined as

$$\widetilde{M}_{\mathrm{LB}}(u) = \frac{1}{N} C_{\mathrm{LB}}(u)^{-1} = \widetilde{M}(u) + \frac{1}{N} \Gamma^{-1}, \qquad (3.7)$$

where $\widetilde{M}(u)$ is given by (3.3). It can easily be seen that if the criterion of optimality $\phi(\cdot)$ satisfies Assumption 3.6 (c) and (d), then so does the function $\phi(\cdot + \Gamma^{-1}/N)$. Thus, in this setting Proposition 3.7 holds as well. Note, however, that Bayesian optimal continuous designs depend on the number of observations N.

   The classical approach to optimal design of experiments can be extended to the case where nonlinear models $\eta(\theta, u)$ which are *not* derived from PDEs enter the stage. As in the linear case we require the model to have the following structure: $\eta(\theta, u) := (\xi(\theta, u_1), \ldots, \xi(\theta, u_m))^\top$, i.e., in each row of η enters only one of the design variables u_i via an elemental nonlinear function $\xi(\theta, u_i)$. In a nonlinear setting the normalized information matrix can only be chosen from the Gauss-Newton (GN) or Bayesian (B) approach since any other choice violates the additive property (3.3)

and thus also (3.5) does not hold. Thus, we have

$$\widetilde{M}_{\text{GN}}(\overline{\theta}_{\text{NLS}}, u) = \frac{1}{N} C_{\text{GN}}(\overline{\theta}_{\text{NLS}}, u)^{-1} \quad \text{or} \quad \widetilde{M}_{\text{B}}(\overline{\theta}_{\text{MAP}}, u) = \widetilde{M}_{\text{GN}}(\overline{\theta}_{\text{MAP}}, u) + \frac{1}{N} \Gamma^{-1},$$

depending on the probabilistic perspective. These expressions are reliant on the estimated parameters $\overline{\theta}_{\text{NLS}}$ or $\overline{\theta}_{\text{MAP}}$, respectively, see (2.40) and (2.42). As a consequence, we speak of a *locally* optimal continuous design $\left(\overline{u}_{\text{c}}(\overline{\theta}), \overline{m}(\overline{\theta})\right) \in U_{\text{c}} \times \mathbb{N}$ of an experiment if it is the solution of the following optimization problem in which the estimate $\overline{\theta}$ is fixed to $\overline{\theta}_{\text{NLS}}$ or $\overline{\theta}_{\text{MAP}}$:

$$
\begin{aligned}
\min_{u_{\text{c}}, m} \quad & \phi[\widetilde{M}_{\text{GN/B}}(\overline{\theta}_{\text{NLS/MAP}}, u_{\text{c}})] \\
\text{s.t.} \quad & u_{\text{c}} = \{(x_1, p_1), \ldots, (x_m, p_m)\} \subset U_{\text{c}}, \ \sum_{i=1}^{m} p_i = 1.
\end{aligned}
\tag{3.8}
$$

Existence of a locally optimal continuous design follows from an analog version of Proposition 3.7. Usually, an adaptive procedure is pursued in which (3.8) is solved with the current estimate $\overline{\theta} = \overline{\theta}^j$ and then a new $\overline{\theta}^{j+1}$ is computed with data obtained from the locally optimal continuous design $\left(\overline{u}_{\text{c}}(\overline{\theta}^j), \overline{m}(\overline{\theta}^j)\right)$. In the next step the normalized information matrix is re-evaluated at $\overline{\theta} = \overline{\theta}^{j+1}$ and (3.8) is solved again. The new design is accepted with a predefined probability or on other conditions, see [12] or [43, Sec. 5.3], and the iteration continues.

For (nonlinear) PDE models, however, the classical approach fails since the connection between the design points x_i and the variables which are involved in a PDE, see Section 2.2, is unclear. Therefore, modern approaches to OED with PDE constraints become a necessity.

3.2 A Modern Approach to Optimal Sensor Placement with PDE Constraints

The main idea in OED with PDE constraints is to link the concept of the observation operator, see Section 2.2, with the sensor positions. We start with the pure sensor placement problem. Our exposition in this section is based on [3, 4] and [12, 81].

Let us adopt the notation from Section 2.2 and let $\eta(\theta, u) := h(y(\theta; u), \theta)$ be a nonlinear PDE model where h is the observation operator and $y(\theta; u)$ is the unique solution to the state equation (2.9) for given parameters θ and inputs u. Furthermore, let Assumption 2.12 and Assumption 2.13 be satisfied. The state equation comes from a PDE defined on a bounded domain $G \subset \mathbb{R}^d$ with Lipschitz boundary

$\partial G = \Gamma_D \cup \Gamma_N \cup \Gamma_F$ as stated in the introduction of this chapter. In a first-discretize-then-optimize approach the state variable y is a large-scale vector and its components are defined on the mesh element nodes, e.g., finite element nodes, of the whole domain and its boundary. Our goal is to find optimal sensor positions $\bar{x}_k \in \Gamma_F$ such that a criterion of optimality $\phi(\cdot)$ of the estimated parameter's information matrix becomes minimal.

It is crucial to emphasize that *all* candidate sensor locations x_k are restricted to the free boundary part Γ_F of the domain. The observation operator maps precisely all components of the state vector that belong to Γ_F to the quantity of interest that can be measured by a sensor. Let n_s be the *fixed* number of all candidate sensor locations. We introduce a new variable $\omega \in \{0,1\}^{n_s}$ which assigns to each candidate location x_k a weight ω_k for $k = 1, \ldots, n_s$. The decision $\omega_k = 1$ means that experimental data are collected at x_k and $\omega_k = 0$ decides not to do so. Thus, the design of an experiment is fully specified by the weight variable ω.

Since an optimal design of an experiment determines optimal sensor locations at which data samples are collected, the weights also enter the PE. Let $\theta^\star$ be the true parameter value. Recall that the measurement process is subject to noise and we modeled this noise in the following equation:

$$z_j(u_i) = h(y(\theta^\star; u_i), \theta^\star) + \varepsilon_{ij},$$

for $j = 1, \ldots, n_z$ and $i = 1, \ldots, m$, where ε_{ij} is a realization of the Gaussian random variable $\mathcal{E}_d \sim \mathcal{N}(0, \Sigma_d)$ with $\Sigma_d = \text{Diag}(\sigma_1^2, \ldots, \sigma_{n_s}^2) \in \mathbb{R}^{n_s \times n_s}$ being a diagonal matrix with strictly positive diagonal entries, cf. Equation (2.3). In the present setting, the dimension d_z of the data vector equals the number n_s of candidate sensor locations. Our knowledge about the quantity of interest is controlled by the weights. If $\omega_k = 1$, then our measurement takes place and it has the variance σ_k^2. However, if $\omega_k = 0$, we do not measure at all, i.e., we know nothing about the quantity of interest. The last case can be interpreted in statistical terms as an infinite variance. Thus, we re-model the noise variable as $\mathcal{E}_d \sim \mathcal{N}(0, \Omega_d^{-1} \Sigma_d)$ with $\Omega_d := \text{Diag}(\omega_1, \ldots, \omega_{n_s})$ where a division by zero is set to infinity in accordance with the statistical interpretation above, see also [23, 80, 81].

In Section 2.2 we introduced the new dimension $n := n_z n_s m$ and adjusted the data vector z, the output of the observation operator h and the noise covariance matrix Σ_d by repeating its entries or blocks if necessary. We adopt this strategy for the noise matrix and obtain $\Sigma \in \mathbb{R}^{n \times n}$ as before. Allowing the experimenter to control the number of repeated observations at a candidate location x_k, we extend the number of weights by the number of measurement series n_z such that each

sensor location x_k has n_z weights that can be switched on or off. Thus, $\omega \in \{0,1\}^{n_\omega}$ where $n_\omega = n_z n_s$. Since the position of the sensors stays the same throughout the experiment regardless of the input u_i, we define $\Omega := \text{Diag}(\text{rep}(\omega_1, \ldots, \omega_{n_\omega}; m))$ such that $\Omega \in \mathbb{R}^{n \times n}$ and introduce the stochastic noise variable $\mathscr{E} \sim \mathscr{N}(0, \Omega^{-1}\Sigma)$. Thus, Σ and Ω are dimension compatible diagonal matrices. The weights enter the reduced PE in the following way:

$$\min_\theta \; \frac{1}{2} \left[\eta(\theta, u) - z(u) \right]^\top \Omega \Sigma^{-1} \left[\eta(\theta, u) - z(u) \right]. \tag{3.9}$$

Similarly, for the Bayesian case we have

$$\min_\theta \; \frac{1}{2} \| \eta(\theta, u) - z(u) \|^2_{\Omega \Sigma^{-1}} + \frac{1}{2} \| \theta - \theta_0 \|^2_{\Gamma^{-1}}. \tag{3.10}$$

where (θ_0, Γ) is the prior, see Section 2.3. As a consequence, the estimated parameters $\overline{\theta}_{\text{NLS/LE/MAP}} = \overline{\theta}_{\text{NLS/LE/MAP}}(z; \omega, u)$ depend on the weight variables. By a close inspection of the proofs of the expressions (2.39)–(2.43) we see that it suffices to replace the inverse of the noise matrix Σ^{-1} by $\Omega \Sigma^{-1}$ such that we can summarize the expressions for the covariance matrices of the estimated parameters as follows:

$$C_{\text{NLS}}(\overline{\theta}_{\text{NLS}}, z, \omega, u) = H(\overline{\theta}_{\text{NLS}}; z, \omega, u)^{-1} F(\overline{\theta}_{\text{NLS}}; \omega, u) H(\overline{\theta}_{\text{NLS}}; z, \omega, u)^{-\top}, \tag{3.11}$$

$$C_{\text{GN}}(\overline{\theta}_{\text{NLS}}, \omega, u) = F(\overline{\theta}_{\text{NLS}}; \omega, u)^{-1}, \tag{3.12}$$

$$C_{\text{AL}}(\overline{\theta}_{\text{LE}}, z, \omega, u) = H(\overline{\theta}_{\text{LE}}; z, \omega, u)^{-1}, \tag{3.13}$$

where $F(\overline{\theta}_{\text{NLS}}; \omega, u) := J(\overline{\theta}_{\text{NLS}}; u)^\top \Omega \Sigma^{-1} J(\overline{\theta}_{\text{NLS}}; u)$ and the Hessian is given by

$$H(\cdot \, ; z, \omega, u) = F(\cdot \, ; \omega, u) + S(\cdot \, ; z, \omega, u),$$

$$S(\cdot \, ; z, \omega, u) = \sum_{k=1}^n \Omega_{kk} \Sigma_{kk}^{-1} [\eta_k(\cdot \, , u) - z_k(u)] \partial^2_{\theta\theta} \eta_k(\cdot \, , u),$$

compare Equation (2.14). If priors are used, then

$$C_{\text{B}}(\overline{\theta}_{\text{MAP}}, \omega, u) = \left[J(\overline{\theta}_{\text{MAP}}; u)^\top \Omega \Sigma^{-1} J(\overline{\theta}_{\text{MAP}}; u) + \Gamma^{-1} \right]^{-1}, \tag{3.14}$$

$$C_{\text{AB}}(\overline{\theta}_{\text{MAP}}, z, \omega, u) = H_{\text{B}}(\overline{\theta}_{\text{MAP}}; z, \omega, u)^{-1}, \tag{3.15}$$

where in this case

$$H_{\text{B}}(\overline{\theta}_{\text{MAP}}; z, \omega, u) = J(\overline{\theta}_{\text{MAP}}; u)^\top \Omega \Sigma^{-1} J(\overline{\theta}_{\text{MAP}}; u) + \Gamma^{-1} + S(\overline{\theta}_{\text{MAP}}; z, \omega, u),$$

compare Equation (2.34).

In a first attempt we formulate the OED problem using the modern approach with binary weights $\omega \in \{0,1\}^{n_\omega}$ where the parameter estimate is kept constant:

Problem 3.9. Let $u \subset U$, $\bar{\theta}$ be a *fixed* parameter estimate and let the information matrix $M(\bar{\theta}, \omega, u) = C(\bar{\theta}, \omega, u)^{-1}$ be given where $C(\bar{\theta}, \omega, u)$ comes from one of the expressions in (3.11)–(3.15) depending on the probabilistic perspective. Moreover, let $\phi : S_{n_\theta}^+ \to (0,\infty)$ be a criterion of optimality. Then a *locally optimal discrete design* $\bar{\omega} \in \{0,1\}^{n_\omega}$ of an experiment with at most N observations is the solution of the constrained optimization problem

$$\min_{\omega} \ \phi[M(\bar{\theta}, \omega, u)]$$

$$\text{s.t.} \ \ \omega \in \{0,1\}^{n_\omega}, \ \sum_{k=1}^{n_\omega} \omega_k \leq N. \tag{3.16}$$

Although no PDE appears in the constraints of (3.16) this problem is still very challenging due to its discrete nature. In fact, for some important design criteria Problem 3.9 was proven to be NP-hard, see [6] and the references therein. In order to make (3.16) tractable we relax the binary constraint of the weight variables and allow them to take values in the interval $[0,1]^{n_\omega}$. Moreover, we remove the inequality constraint and add a penalty function to sparsify the solution. The final number of observations is thus controlled by a penalty parameter $\kappa > 0$. The larger κ is chosen, the more sparse the weight variable ω becomes. We also introduce a continuation strategy with a family of penalty functions to enforce binary weights. This approach was first proposed by Alexanderian et al. [3]. Recall, that the ℓ_0-"norm" $|x|_0$ of a real number $x \in \mathbb{R}$ is defined as

$$|x|_0 = \begin{cases} 0, & \text{if } x = 0, \\ 1, & \text{otherwise.} \end{cases}$$

For vectors $x \in \mathbb{R}^n$ we have $|x|_0 = |x_1|_0 + \ldots + |x_n|_0$.

Definition 3.10. We call a family of penalty functions $\mathscr{P}_\varepsilon : [0,1]^{n_\omega} \to [0,\infty)$ where $0 < \varepsilon \leq 1$ a smooth approximation of the ℓ_0-"norm" if
 (a) $\mathscr{P}_1(\omega) = \omega_1 + \ldots + \omega_{n_\omega}$,
 (b) $\mathscr{P}_\varepsilon(\omega) = \sum_{k=1}^{n_\omega} f_\varepsilon(\omega_k)$ with $f_\varepsilon : [0,1] \to [0,1]$ being continuously differentiable and $f_\varepsilon(1) = 1$, $f_\varepsilon(0) = 0$,
 (c) $\mathscr{P}_\varepsilon$ converges pointwise to the ℓ_0-"norm" for $\varepsilon \to 0$.

The functions $f_\varepsilon : [0,1] \to [0,1]$ have the following specific form, cf. [3]:

$$f_\varepsilon(x) := \begin{cases} x/\varepsilon, & \text{for } 0 \le x \le \frac{1}{2}\varepsilon, \\ p_\varepsilon(x), & \text{for } \frac{1}{2}\varepsilon < x \le 2\varepsilon, \\ 1, & \text{for } 2\varepsilon < x \le 1, \end{cases}$$

where p_ε is the unique third-order polynomial such that f_ε becomes continuously differentiable. Using these penalty functions $\mathscr{P}_\varepsilon$ we formulate a continuous relaxation of Problem 3.9:

Problem 3.11. Let $u \subset U$, $\bar\theta$ be a *fixed* parameter estimate and let the information matrix $M(\bar\theta, \omega, u) = C(\bar\theta, \omega, u)^{-1}$ be given where $C(\bar\theta, \omega, u)$ comes from one of the expressions in (3.11)–(3.15) depending on the probabilistic perspective. Moreover, let $\phi : S_{n_\theta}^+ \to (0, \infty)$ be a criterion of optimality and $\mathscr{P}_\varepsilon(\omega)$ be a smooth approximation of the ℓ_0-norm and $\kappa > 0$. Then a *locally optimal continuous design* $\bar\omega_\varepsilon \in [0,1]^{n_\omega}$ of an experiment is the solution of the constrained optimization problem

$$\begin{aligned} \min_\omega \quad & \phi[M(\bar\theta, \omega, u)] + \kappa \mathscr{P}_\varepsilon(\omega) \\ \text{s.t.} \quad & \omega \in [0,1]^{n_\omega} . \end{aligned} \tag{3.17}$$

The continuation strategy works in the following way. First, Problem 3.11 is solved for $\varepsilon = \varepsilon_0 := 1$ and we obtain the optimum $\bar\omega_0$. Next, we decrease ε, e.g., $\varepsilon_1 = \varepsilon_0/2$, and solve (3.17) with $\varepsilon = \varepsilon_1$ and $\bar\omega_0$ as initial guess (starting point for the solver). Subsequently, Problem 3.11 is solved for ε_i, i.e., with $\mathscr{P}_{\varepsilon_i}$ as penalty function, and the initial guess is given by $\bar\omega_{i-1}$, i.e., the solution of the preceding optimization problem. After a few reiterations $i = 1, 2, 3, \ldots$ we observe in practice that the optimal continuous design has the desired binary structure that remains unchanged as ε is further decreased. Such a continuation strategy is used in a similar way in topology optimization [17, 18]. Problem 3.11 can be solved by standard SQP-methods with BFGS updates [129]. Note that there is still no PDE involved in the constraints of (3.17).

Concerning the existence of a locally optimal continuous design we proceed in a similar way as in the classical approach:

Assumption 3.12. Let the following hold:
  (a) $\phi : S_{n_\theta}^+(\mathbb{R}) \to (0, \infty)$ is convex,
  (b) the mapping $\omega \mapsto M(\bar\theta, \omega, u)$ is continuous,

(c) for all $\omega, \widetilde{\omega} \in [0,1]^{n_\omega}$ and $\alpha \in (0,1)$ we have the following property:

$$M(\overline{\theta}, \alpha\omega + (1-\alpha)\widetilde{\omega}, u) = \alpha M(\overline{\theta}, \omega, u) + (1-\alpha)M(\overline{\theta}, \widetilde{\omega}, u),$$

(d) there exists $\omega_0 \in [0,1]^{n_\omega}$ such that $\phi[M(\overline{\theta}, \omega_0, u)] < \infty$.

In analogy to Proposition 3.7 we state

Proposition 3.13. Let Assumption 3.12 be satisfied. Then there exists a locally optimal continuous design $\overline{\omega}_\varepsilon \in [0,1]^{n_\omega}$ which is the solution of Problem 3.11. Moreover, for $\varepsilon = 1$ the set of optimal designs $W_{\mathrm{opt}}(c)$ which have the same cost $c > 0$

$$W_{\mathrm{opt}}(c) := \big\{ \overline{\omega}_1 \in [0,1]^{n_\omega} : \overline{\omega}_1 \text{ solves (3.17) and } \kappa\mathscr{P}_1(\overline{\omega}_1) = c \big\},$$

is convex.

Proof. Existence of locally optimal continuous design follows from Assumption 3.12 (b) and (d) and the fact that $[0,1]^{n_\omega}$ is compact.

For the second statement let $c > 0$ be fixed, $\omega, \widetilde{\omega} \in W_{\mathrm{opt}}(c)$ and $\alpha \in (0,1)$ be given. By definition of $W_{\mathrm{opt}}(c)$ we have

$$\phi[M(\overline{\theta}, \omega, u)] = \phi[M(\overline{\theta}, \widetilde{\omega}, u)] =: q.$$

Let $\omega^\star := \alpha\omega + (1-\alpha)\widetilde{\omega}$. Then, clearly, $\kappa\mathscr{P}_1(\omega^\star) = c$ and by Assumption 3.12 (a) and (c) it follows that

$$\phi[M(\overline{\theta}, \omega^\star, u)] \leq \alpha\phi[M(\overline{\theta}, \omega, u)] + (1-\alpha)\phi[M(\overline{\theta}, \widetilde{\omega}, u)] = q.$$

Thus, $\omega^\star$ solves (3.17). ∎

Remark 3.14. A few comments are at hand:
(a) If $M = C_{\mathrm{NLS}}^{-1}$, then Assumption 3.12 (c) is not satisfied. Consequently, the second statement in Proposition 3.13 does not hold for this type of information matrix.
(b) For $M = C_{\mathrm{GN/AL}}^{-1}$ we have to ensure that the principal part of these covariance matrices, i.e., $J(\overline{\theta}; u)^\top \Omega \Sigma^{-1} J(\overline{\theta}; u)$ stays invertible in all iterations. Therefore, we generally require that $J(\overline{\theta}; u)$ has full column-rank and that

$$\sum_{k=1}^{n_\omega} \omega_k \geq n_\theta. \tag{3.18}$$

However, we do not have to impose (3.18) as a constraint in (3.17) if we choose an appropriate penalty constant $\kappa > 0$.

(c) We again speak of a *locally* optimal design since the covariance matrix depends on the current parameter estimate $\bar{\theta}$. As explained at the end of the previous section, a sequential procedure can be applied to successively solve Problem 3.11 with a continuation strategy and then to compute a new parameter estimate (this requires repeated PDE solutions).

(d) In general, one cannot guarantee that a global solution of (3.16) is obtained from the limit of the sequence $\bar{\omega}_\varepsilon$ as $\varepsilon \to 0$. In practice, however, the continuation strategy provides good results.

In order to deal with the parameter-dependence of the optimal design we introduce an alternative to sequential design where an *inner* optimization problem is added to the constraints of (3.17). The estimated parameters become a function of the weights, i.e., we consider the mapping $\omega \mapsto \bar{\theta}(z; \omega, u)$ where $\bar{\theta}(z; \omega, u)$ is the solution of (3.9) or (3.10) depending on the probabilistic perspective. As in Subsection 2.2.1 we can argue by the Implicit Function Theorem that $\omega \mapsto \bar{\theta}(z; \omega, u)$ is well-defined and that this mapping is continuously differentiable. If an iterative procedure is applied to obtain optimal designs, then in each step $\omega^{k+1} = \omega^k + \Delta\omega^k$ a new estimate $\bar{\theta}(z; \omega^{k+1}, u)$ is computed. This involves computing solutions of the state equation. Since the information matrix $M(\theta, \omega, u)$ contains the derivatives $\partial_\theta \eta(\theta, u)$ and, if applicable, $\partial^2_{\theta\theta} \eta(\theta, u)$, they need to be re-evaluated at $\theta = \bar{\theta}(z; \omega^{k+1}, u)$ which requires additional PDEs to be solved, see Appendix A.4.

Problem 3.15. Let $u \subset U$ and let the information matrix $M(\theta, \omega, u) = C(\theta, \omega, u)^{-1}$ be given where $C(\theta, \omega, u)$ comes from one of the expressions in (3.11)–(3.15) depending on the probabilistic perspective. Moreover, let $\phi : S^+_{n_\theta} \to (0, \infty)$ be a criterion of optimality, $\mathscr{P}_\varepsilon(\omega)$ be a smooth approximation of the ℓ_0-norm and $\kappa > 0$. Then an *optimal continuous design* $\bar{\omega}_\varepsilon \in [0,1]^{n_\omega}$ of an experiment is the solution of the bilevel PDE-constrained optimization problem

$$\min_\omega \ \phi[M(\bar{\theta}(z; \omega, u), \omega, u)] + \kappa \mathscr{P}_\varepsilon(\omega)$$
$$\text{s.t. } \bar{\theta}(z; \omega, u) \text{ solves (3.9) or (3.10),} \qquad (3.19)$$
$$\omega \in [0,1]^{n_\omega}.$$

If there exists some $\omega_0 \in [0,1]^{n_\omega}$ such that $\phi[M(\bar{\theta}(z; \omega_0, u), \omega_0, u)] < \infty$ and if the mapping $\omega \mapsto M(\bar{\theta}(z; \omega, u), \omega, u)$ is continuous, then the existence of optimal continuous designs $\bar{\omega}_\varepsilon$ which solve (3.19) follows from standard arguments. These

assumptions are clearly fulfilled for all covariance estimates that are summarized in (3.11)–(3.15). However, Assumption 3.12 (c) is not satisfied.

For large-scale problems a direct computation of the full Hessian and $C_{\text{NLS/AL/AB}}$ is no longer tractable. In [4] randomized trace estimators are used for Bayesian optimal designs where the criterion of optimality is $\phi(M_{\text{AB}}) = \text{tr}(M_{\text{AB}}^{-1}) = \text{tr}(C_{\text{AB}})$. Within this approach, the matrix C_{AB} is never fully computed. Since

$$\text{tr}(C_{\text{AB}}) = \sum_{i=1}^{n_\theta} \mathbf{e}_i^\top H_{\text{B}}(\bar{\theta}_{\text{MAP}}; z, \omega, u)^{-1} \mathbf{e}_i, \tag{3.20}$$

where $\mathbf{e}_i$ is the i-th unit vector, only solutions of the systems $H_{\text{B}}(\bar{\theta}_{\text{MAP}}; z, \omega, u)\xi_i = \mathbf{e}_i$ are needed which we already examined in (2.19)–(2.21). Because of the large dimension n_θ, the computation of the trace in (3.20) is approximated by a randomized trace estimator:

$$\text{tr}(C_{\text{AB}}) \approx \sum_{i=1}^{n_{\text{tr}}} \tau_i^\top \xi_i, \tag{3.21}$$

$$\tau_i = H_{\text{B}}(\bar{\theta}_{\text{MAP}}; z, \omega, u)\xi_i$$

where $\tau_i \sim \mathcal{N}(0, I)$ are realizations of multivariate standard Gaussian random variables, see [4, Sec. 2.5]. The system of equations (2.19)–(2.21) and (2.16)–(2.18) can then be imposed as constraints to the minimization of the function in (3.21). For an efficient adjoint calculus of this optimization problem we refer to [4, Sec. 5.4].

Another modern approach to optimal sensor placement with PDE constraints is introduced in [98] where the design variable is treated as a positive Borel measure. Let $X \subset \mathbb{R}^{d_x}$ be a closed subset covering all possible sensor locations in space and $y(\theta, u; x)$ be the solution of the state equation in an infinite dimensional setting, i.e., the state is still a function of the space variable x. The measurement setup is defined as a weighted sum of Dirac delta functions

$$\omega(x, \lambda) := \sum_{j=1}^{n_s} \lambda_j \delta_{x_j}$$

where $x_j \in X$ and $\lambda_j \geq 0$ are non-negative weights. The information matrix can then be written as

$$M(\theta, \omega, u)_{k\ell} = \int_X \partial_{\theta_k} y(\theta, u; x) \partial_{\theta_\ell} y(\theta, u; x) \, d\omega(x, \lambda),$$

for $k, \ell = 1, \dots, n_\theta$. The corresponding optimization problem is solved with a conditional gradient method in measure space. In this thesis, however, we do not pursue this approach any further.

In some cases, a heuristic approach to optimal sensor placement provides sufficiently good results, see [2, Sec. 4.8] for a description of the greedy algorithm. However, solving Problem 3.9 with such a heuristic does not provide provably optimal results.

3.3 An Extension of the Modern Approach to Optimal Input Configuration

A relevant question is whether the modern approach to optimal sensor placement with PDE constraints can be extended such as to incorporate the input variable u into the design of the experiment. In this section, we add an optimal configuration framework to the previous context in order to minimize a criterion of optimality of the parameter's covariance matrix. We are specifically interested in time-variant PDE problems where the state and the input are time-dependent. Moreover, we build on our original problem formulation as stated in the introduction to this chapter. This section is based on the author's work in [93].

We use the notation of Section 2.2 with $m = 1$ again. Let Assumption 2.12 and Assumption 2.13 be satisfied. In this section the state equation is now a discretized PDE on the space *and* time domain. Let $T > 0$ and n_t denote the number of points in the discretized time domain $\{t_i \in [0, T] : t_i = i \cdot T / n_t, \ i = 1, \dots, n_t\}$. Since the positions of the sensors do not change in time throughout the experiment we adjust the dimension $n := n_z \cdot n_s \cdot n_t$ of the matrices $\Omega \in \mathbb{R}^{n \times n}$ and $\Sigma \in \mathbb{R}^{n \times n}$ in a similar way as done in Section 3.2 by copying its diagonal entries n_t times. The model output $\eta(\theta, u) \in \mathbb{R}^n$ is then a vector that covers the predicted quantity of interest at all time points and the data vector $z \in \mathbb{R}^n$ is the corresponding measured quantity of interest at all time points.

Let $\bar{\theta} \in \Theta \subset \mathbb{R}^{n_\theta}$ be a given parameter estimate and let $y(\bar{\theta}, u) \in Y \subset \mathbb{R}^{d_y n_t}$ be the unique solution to the state equation

$$e(\bar{\theta}, u, y) = 0, \tag{3.22}$$

for some input $u \in U \subset \mathbb{R}^{d_u n_t}$. We introduce new variables $s_i := \partial_{\theta_i} y(\bar{\theta}, u)$ for all $i = 1, \dots, n_\theta$ where each s_i is the solution to its sensitivity equation

$$\partial_y e(\bar{\theta}, u, y(\bar{\theta}, u)) s_i + \partial_{\theta_i} e(\bar{\theta}, u, y(\bar{\theta}, u)) = 0. \tag{3.23}$$

Existence and continuity of each s_i follows from Assumption 2.12.

The information matrix of our choice cannot depend on the data z, since a change in inputs $u + \Delta u$ would produce a model output that mismatches the data. In this case, the residuals $r_k = \eta_k(\bar{\theta}, u + \Delta u) - z_k$ would be unreasonably large unless one collects new data $z(u + \Delta u)$ to the corresponding input which usually is far too costly. Therefore, only the Gauss-Newton or Bayesian framework is applicable to our setting, cf. (3.12) and (3.14):

$$M_{\mathrm{GN}}(\bar{\theta}, \omega, u) = C_{\mathrm{GN}}(\bar{\theta}, \omega, u)^{-1} = J(\bar{\theta}; u)^{\top} \Omega \Sigma^{-1} J(\bar{\theta}; u),$$
$$M_{\mathrm{B}}(\bar{\theta}, \omega, u) = C_{\mathrm{B}}(\bar{\theta}, \omega, u)^{-1} = J(\bar{\theta}; u)^{\top} \Omega \Sigma^{-1} J(\bar{\theta}; u) + \Gamma^{-1},$$

where $\bar{\theta} = \bar{\theta}_{\mathrm{NLS/MAP}}$, depending on the probabilistic perspective, and

$$J(\bar{\theta}; u) := \partial_y h(y(\bar{\theta}, u), \bar{\theta}) \cdot s(y(\bar{\theta}, u), u) + \partial_\theta h(y(\bar{\theta}, u), \bar{\theta}).$$

Here, $y(\bar{\theta}, u)$ is the solution to (3.22) and

$$s(y(\bar{\theta}, u), u) = [s_1(y(\bar{\theta}, u), u), \dots, s_{n_\theta}(y(\bar{\theta}, u), u)] \in \mathbb{R}^{d_y n_t \times n_\theta}$$

is the matrix of the solutions $s_i(y(\bar{\theta}, u), u)$ to (3.23). We also require the following assumption:

Assumption 3.16. For all $\theta \in \Theta$, $u \in U$ and $y \in Y$ the derivatives $\partial_u e(\theta, u, y)$, $\partial_{yu}^2 e(\theta, u, y)$, $\partial_{uy}^2 e(\theta, u, y)$, $\partial_{u\theta}^2 e(\theta, u, y)$ and $\partial_{\theta u}^2 e(\theta, u, y)$ exist and are continuous.

For the OED problem, we fix an estimate $\bar{\theta}$ and minimize a criterion of optimality $\phi[M_{\mathrm{GN/B}}(\bar{\theta}, \omega, u)]$ where ω and u are the design variables. Note that $M_{\mathrm{GN/B}}$ does only implicitly depend on u via the state y and the sensitivity variable s. Therefore, we write $M_{\mathrm{GN/B}} = M_{\mathrm{GN/B}}(\bar{\theta}, \omega, y, s)$ for the formality of the Lagrange calculus in the following optimization problem. We again add the family of penalty functions $\mathscr{P}_\varepsilon(\omega)$ which we introduced in the previous section and a smooth regularizer $R(u)$ to the objective function. The regularization is chosen to smooth the optimal inputs in the time domain, e.g., one can choose a discretized H^1-norm in time if U is a discretization of the space $H^1(0, T; L^2(\Gamma_{\mathrm{N}}))$.

Problem 3.17. Let $\kappa, \beta > 0$ be constants, $\bar{\theta}$ be a *fixed* parameter estimate and let the information matrix $M_{\mathrm{GN/B}} = C_{\mathrm{GN/B}}^{-1}$ be given depending on the probabilistic perspective. Moreover, let $\phi : S_{n_\theta}^+ \to (0, \infty)$ be a criterion of optimality, $\mathscr{P}_\varepsilon(\cdot)$ be a smooth approximation of the ℓ_0-norm and $R(\cdot)$ be a smooth regularizer. Then

locally optimal sensor positions and input configurations $(\overline{\omega}_\varepsilon, \overline{u}) \in [0,1]^{n_\omega} \times U$ *for an experiment are a solution of the PDE-constrained optimization problem*

$$
\begin{aligned}
\min_{\omega, u, y, s} \quad & \phi[M_{\text{GN/B}}(\overline{\theta}, \omega, y, s)] + \kappa \mathscr{P}_\varepsilon(\omega) + \beta R(u) \\
\text{s.t.} \quad & 0 = e(\overline{\theta}, u, y), \\
& 0 = \partial_y e(\overline{\theta}, u, y) s_i + \partial_{\theta_i} e(\overline{\theta}, u, y), \quad i = 1, \dots, n_\theta \\
& \omega \in [0,1]^{n_\omega}, \ u \in U.
\end{aligned}
\tag{3.24}
$$

In the following we explain the continuation strategy with the family of penalty functions which needs to be adapted to this setting. First, Problem 3.17 is solved for $\varepsilon = \varepsilon_0 = 1$ and we obtain $(\overline{\omega}_0, \overline{u})$. In the next step we decrease ε and solve the following derived optimization problem with fixed $\overline{u}$ and $\overline{\omega}_0$ as initial guess:

$$
\begin{aligned}
\min_\omega \quad & \phi[M_{\text{GN/B}}(\overline{\theta}, \omega, y(\overline{\theta}, \overline{u}), s(y(\overline{\theta}, \overline{u}), \overline{u}))] + \kappa \mathscr{P}_\varepsilon(\omega) + \beta R(\overline{u}) \\
\text{s.t.} \quad & \omega \in [0,1]^{n_\omega}.
\end{aligned}
\tag{3.25}
$$

Subsequently, the continuation strategy is performed as usual with (3.25) such that after a few reiterations the optimal sensor weights have the desired binary structure.

We continue with a statement about the existence of locally optimal sensor positions and input configurations that solve (3.24), in analogy to our statement in Proposition 3.13. Therefore, we make the following assumption:

Assumption 3.18. Let the following hold:
(a) $\phi : S_{n_\theta}^+(\mathbb{R}) \to (0, \infty)$ is continuous,
(b) $U \subset \mathbb{R}^{d_u n_t}$ is compact,
(c) the mapping $(\omega, u) \mapsto M_{\text{GN/B}}(\overline{\theta}, \omega, y(\overline{\theta}, u), s(y(\overline{\theta}, u), u))$ is continuous,
(d) there exist $\omega_0 \in [0,1]^{n_\omega}$ and $u_0 \in U$ such that

$$
\phi[M_{\text{GN/B}}(\overline{\theta}, \omega_0, y(\overline{\theta}, u_0), s(y(\overline{\theta}, u_0), u_0))] < \infty.
$$

We now have the following result:

Proposition 3.19. Let Assumption 3.18 be satisfied. Then there exist locally optimal sensor positions and input configurations $(\overline{\omega}_\varepsilon, \overline{u}) \in [0,1]^{n_\omega} \times U$ which are the solution of Problem 3.17.

Proof. The proof follows from standard arguments. ∎

The comments in Remark 3.14 (b) and (c) are also valid in the present setting.

In order to use derivative-based optimization techniques for the solution of Problem 3.17 let $J(\omega, u, y, s)$ be the objective function in (3.24) and denote the reduced objective function as

$$\widehat{J}(\omega, u) := \phi[M_{\mathrm{GN/B}}(\overline{\theta}, \omega, y(\overline{\theta}, u), s(y(\overline{\theta}, u), u))] + \kappa \mathscr{P}_{\varepsilon}(\omega) + \beta R(u).$$

For simplicity, we ignore the inequality constraints and define the Lagrangian as

$$\mathfrak{L}_{\mathrm{OED}}(\omega, u, y, s, \mu, \lambda) := J(\omega, u, y, s) + \mu^{\top} e(\overline{\theta}, u, y)$$
$$+ \sum_{i=1}^{n_{\theta}} \lambda_i^{\top} \big[\partial_y e(\overline{\theta}, u, y) s_i + \partial_{\theta_i} e(\overline{\theta}, u, y) \big],$$

where $\mu, \lambda_1, \dots, \lambda_{n_{\theta}} \in Y$ are the adjoint variables that are obtained by solving the adjoint equations

$$0 = \nabla_y \mathfrak{L}_{\mathrm{OED}} = \nabla_y J(\omega, u, y, s) + \partial_y e(\overline{\theta}, u, y)^{\top} \mu$$
$$+ \sum_{i=1}^{n_{\theta}} \Big(\big\langle \partial_{yy}^2 e(\overline{\theta}, u, y), (s_i, \lambda_i) \big\rangle + \partial_{\theta_i y}^2 e(\overline{\theta}, u, y)^{\top} \lambda_i \Big), \qquad (3.26)$$
$$0 = \nabla_{s_i} \mathfrak{L}_{\mathrm{OED}} = \nabla_{s_i} J(\omega, u, y, s) + \partial_y e(\overline{\theta}, u, y)^{\top} \lambda_i.$$

The full derivative of the reduced objective function is then given by

$$\nabla_{\omega} \widehat{J}(\omega, u) = \nabla_{\omega} \mathfrak{L}_{\mathrm{OED}} = \nabla_{\omega} J(\omega, u, y, s),$$
$$\nabla_u \widehat{J}(\omega, u) = \nabla_u \mathfrak{L}_{\mathrm{OED}} = \nabla_u J(\omega, u, y, s) + \partial_u e(\overline{\theta}, u, y)^{\top} \mu$$
$$+ \sum_{i=1}^{n_{\theta}} \Big[\big\langle \partial_{yu}^2 e(\overline{\theta}, u, y), (s_i, \lambda_i) \big\rangle + \partial_{\theta_i u}^2 e(\overline{\theta}, u, y)^{\top} \lambda_i \Big].$$

We show in Appendix A.5 how to efficiently compute $\nabla_y J(\omega, u, y, s)$, $\nabla_{\omega} J(\omega, u, y, s)$ and $\nabla_{s_i} J(\omega, u, y, s)$.

For large-scale problems, solving $n_{\theta} + 1$ PDEs in Problem 3.17 is too costly. In a first attempt, one could apply the procedure introduced in [4] using randomized trace estimators for the inverse Hessian C_{AB}. However, this Hessian depends on the data and the parameters are obtained by solving (2.16)–(2.18) as an inner optimization problem. This becomes precarious in our setting since a change in inputs would produce a model output that mismatches the data resulting in unreasonable parameter values. In Chapter 6 we briefly present a possible solution to this drawback as an outlook.

3.4 Discussion of Design Criteria

In this section we discuss the three most popular criteria of optimality $\phi(\cdot)$ and some of their extensions. We mainly follow [43] and [47].

Recall that the linearized confidence region around the estimated parameters at a fixed confidence level $1 - \alpha$, where $\alpha \in (0,1)$, is an n_θ-dimensional ellipsoid

$$K(\alpha, \overline{\theta}, C) := \left\{ \theta \in \mathbb{R}^{n_\theta} : (\theta - \overline{\theta})^\top C(\overline{\theta}, \omega, u)^{-1} (\theta - \overline{\theta}) \le \gamma_{n_\theta}(1 - \alpha) \right\},$$

compare Equation (2.37). The matrix $C(\overline{\theta}, \omega, u)$ is given by one of the expressions in (3.11)–(3.15). Size and volume of $K(\alpha, \overline{\theta}, C)$ are determined by the covariance matrix C and the quantile function $\gamma_{n_\theta}(\cdot)$. A minimization of the following design criteria $\Psi(C)$ has distinctive geometrical properties which we will point out in the sequel. In view of Lemma 3.20, we again consider the information matrix $M(\overline{\theta}, \omega, u) = C(\overline{\theta}, \omega, u)^{-1}$ and the function $\phi : S^+_{n_\theta}(\mathbb{R}) \to (0, \infty)$.

- **A-criterion:** let $\phi_A(M) = \mathrm{tr}\left(M^{-1}\right) = \mathrm{tr}(C) = \Psi_A(C)$ be the sum of eigenvalues of the covariance matrix. It is proportional to the average variance $n_\theta^{-1} \mathrm{tr}(C)$ of the parameter estimates. Additionally, $\Psi_A(C)$ is equal to the square of half the length of the diagonal of an n_θ-dimensional cuboid which circumscribes the ellipsoid $K(\alpha, \overline{\theta}, C)$.

- **D-criterion:** let $\phi_D(M) = \det\left(M^{-1}\right) = \det(C) = \Psi_D(C)$ be the product of eigenvalues of the covariance matrix. The volume of $K(\alpha, \overline{\theta}, C)$ is given by

$$\mathrm{Vol}(K) = \frac{\pi^{n_\theta/2}}{\Gamma(n_\theta/2 + 1)} \gamma_{n_\theta}(1 - \alpha)^{n_\theta} \sqrt{\det(C)},$$

  where $\Gamma(\cdot)$ is the Gamma function, see [15]. Since the square root function is monotonic, a minimization of $\Psi_D(C)$ corresponds to minimizing $\mathrm{Vol}(K)$.

- **E-criterion:** let $\phi_E(M) = \lambda_{\max}\left(M^{-1}\right) = \lambda_{\max}(C) = \Psi_E(C)$ be the maximal eigenvalue of the covariance matrix. The length of the largest principal axis of $K(\alpha, \overline{\theta}, C)$ is given by $2\sqrt{\lambda_{\max}(C)}$. Thus, minimizing $\Psi_E(C)$ leads to minimizing the "linear expansion" of the confidence ellipsoid.

These criteria of optimality have the following desirable properties:

Lemma 3.20. The *A*-, *D*- and *E*-criterion are continuous and convex matrix functions $\phi_{A/D/E} : S^+_{n_\theta}(\mathbb{R}) \to (0, \infty)$. Moreover, $\phi_A(\cdot)$ is even strictly convex.

Proof. It is well known that $\mathrm{tr}(\cdot)$, $\det(\cdot)$ and $\lambda_{\max}(\cdot)$ are continuous functions. The inverse mapping $M \mapsto M^{-1}$ is continuous too, compare Lemma A.1. We refer to

[25, Thm. 2.10] in order to prove that $\phi_A(\cdot)$ is strictly convex. The convexity of $\phi_D(\cdot)$ is treated in [84]. Concerning $\phi_E(\cdot)$, first note that for all $M_1, M_2 \in S_{n_\theta}^+(\mathbb{R})$ and $\alpha \in (0, 1)$ we have

$$x^\top [\alpha M_1 + (1-\alpha)M_2]^{-1} x \leq x^\top \left[\alpha M_1^{-1} + (1-\alpha)M_2^{-1}\right] x, \qquad (3.27)$$

for all $x \in \mathbb{R}^{n_\theta}$, see [117, Eq. 10.61]. Consequently, for $M := \alpha M_1 + (1-\alpha)M_2$ we have

$$\begin{aligned}
\lambda_{\max}(M^{-1}) &:= \max_x \left\{x^\top [\alpha M_1 + (1-\alpha)M_2]^{-1} x : \|x\|_2 = 1\right\}, \\
&\leq \max_x \left\{x^\top \left[\alpha M_1^{-1} + (1-\alpha)M_2^{-1}\right] x : \|x\|_2 = 1\right\}, \\
&\leq \alpha \max_x \left\{x^\top M_1^{-1} x : \|x\|_2 = 1\right\} + (1-\alpha) \max_x \left\{x^\top M_2^{-1} x : \|x\|_2 = 1\right\}, \\
&= \alpha \lambda_{\max}(M_1^{-1}) + (1-\alpha)\lambda_{\max}(M_2^{-1}),
\end{aligned}$$

where the first inequality holds due to (3.27) and the second follows from the convexity of the maximum function. ∎

Therefore, it is evident that the A-, D- and E-criterion satisfy Assumption 3.6 (c), Assumption 3.12 (a) and Assumption 3.18 (a).

We discuss (directional) differentiability of these criteria in Appendix A.1, especially for the nonsmooth E-criterion.

We finally present advantages and disadvantages of each of these criteria. First, the D-criterion is invariant with respect to non-degenerate transformations of the model parameters. This property is particularly helpful when bijective coordinate transformations have been applied in the modeling. Moreover, D-optimal designs usually have a small number of different design points but more replications of them. A drawback of this criterion is that it tends to emphasize the parameter component that is most sensitive to changes in the design [47]. This may lead to degenerated confidence ellipsoids which have a small volume but still have a large principal axis. An advantage of the A-criterion is its adaptability to large-scale problems where the dimension n_θ of the parameter space becomes large [4]. A major drawback, however, lies in its disregard of the off-diagonal elements of the covariance matrix and thus any information about cross-correlations among parameters is lost. On the other hand, the E-criterion effectively minimizes the largest principal axis of the confidence ellipsoid. However, its usage requires a cautious scaling of the parameters and non-smooth methods for derivative-based optimization, see Equation (A.3).

An alternative to the standard criteria which where introduced above would be

to use the condition number of the covariance matrix, i.e.,

$$\mathrm{cond}(C) = \frac{\lambda_{\max}(C)}{\lambda_{\min}(C)},$$

in order to reduce cross-correlations of the parameters. However, this criterion operates on the shape of the confidence ellipsoid only, making it more spherical with a very large volume. Another approach to deal with parameter cross-correlations is to introduce constraints on the non-diagonal entries of the cross-correlation matrix [48]. This leads to optimization problems with state constraints which are difficult to handle especially in PDE-constrained optimization. For other criteria, which are less used, the reader may consult [108, 133].

In the next chapters, we use the design criterion $\Psi(C)$ instead of the function $\phi(M)$, and the covariance matrix C instead of the information matrix M.

Uncertainty Detection in Mathematical Models

In science and technology, mathematics is the most successful language used to describe and predict natural or engineering processes that are observed in nature or which are of common interest. It is precisely because mathematical models are employed in our way of thinking that we are able to quantify and control the behavior and outcome of these processes. However, our knowledge is finite and thus limited to accumulations from the past and from our present experience. Therefore, any image of reality in the domain of mathematical abstraction is subject to change if new data arise that contradict or refine parts of our knowledge. We denote this tentativeness or insecurity that occurs in the mathematical modeling by model uncertainty.

In general, models are constructed by evaluating data, applying conservation principles, incorporating prior knowledge and using experience or intuition. In order to make models successful it is vital to subject the modeling process to a calibration and validation cycle, see Figure 4.1. In this loop, natural phenomena are exposed to our method of questioning and observational data, which are part of our mathematical abstraction, are acquired. These observations are used to calibrate the model by solving inverse problems. A validation procedure follows after the model is em-
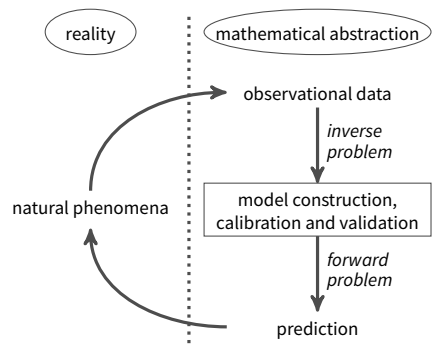
Figure 4.1 Model building cycle [119].

ployed to generate a prediction. By comparing the model-output with new data, the cycle continues and the model is either rejected or updated. It is evident that uncertainty in the model as well as in the data significantly affects the success or outcome of this calibration and validation procedure. This motivates a close study of the effects uncertainty has upon both models and data.

In this chapter we first specify what we mean by the terms mathematical model and model uncertainty. In Section 4.2 we elaborate on the different sources and types of model uncertainty. Furthermore, we highlight the influence of uncertain data on the model's prediction quality in Section 4.3. We give a literature overview of common methods to detect, quantify and control model uncertainty in Section 4.4. Our main contribution to the topic is a novel algorithm that detects and quantifies model uncertainty using parameter estimation, optimum experimental design and hypothesis testing which we present in Section 4.5. This algorithm originated at the Collaborative Research Center (CRC) 805 within Subproject A3 and appears in several works that emerged from the CRC.

4.1 Fundamental Definitions

This section is based on ideas which the author developed in [105]. We start by giving a clearer notion of our terms. A model is an image of reality in the domain of mathematical abstraction. It mainly consists of functional relations between model parameters θ, like material constants, inputs u, like boundary values or initial conditions, (passive) states y and outputs q, which usually are the quantities of interest. Let $\Theta \subset \mathbb{R}^{n_\theta}$, $U \subset \mathbb{R}^{d_u}$, $Y \subset \mathbb{R}^{d_y}$ and $Q_\mathrm{I} \subset \mathbb{R}^{d_z}$ be bounded and closed, respectively. Then we state:

> **Definition 4.1 (Mathematical Model).** We call the set of functions and equations
>
> $$\mathcal{M} := \{e : \Theta \times U \times Y \to Y, h : Y \times \Theta \to Q_\mathrm{I}, e(\theta, u, y) = 0, h(y, \theta) = q\}$$
>
> a *mathematical model* consisting of the functional relations between the model variables θ, u, y and q. Here, $e(\theta, u, y) = 0$ is the *state equation* and $h : Y \times \theta \to Q_\mathrm{I}$ is the *observation operator*.

As in Chapter 2, we assume that the state equation yields for any given $\theta \in \Theta$ and $u \in U$ a unique solution $y(\theta, u) \in Y$ which is then mapped by the observation operator to the output $h(y(\theta, u), \theta) = q \in Q_\mathrm{I}$, i.e., the quantity of interest. The latter can be observed in experiments through the measurement process yielding (noisy) data z. We denote by $z^\star$ the *true (but unknown) value* of the quantity of interest and
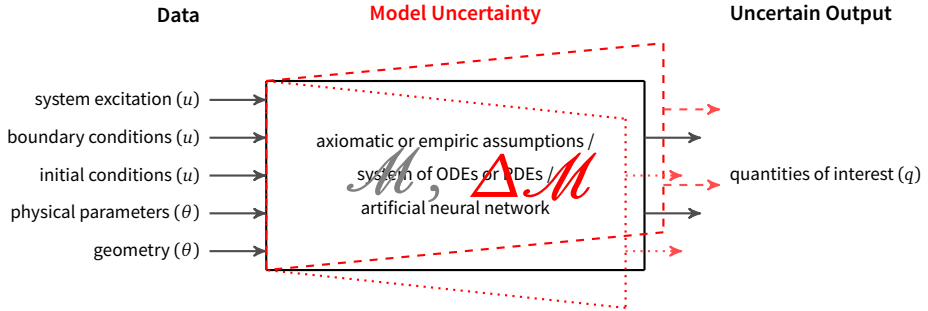
Figure 4.2 The model-output q is affected by model uncertainty $\Delta\mathcal{M}$ even if the data (parameters θ and inputs u) are exact; based on [111].

we say that our model $\mathcal{M}$ is *correct* if its output q is identical to $z^\star$ for all given inputs $u \in U$ and the *true (but unknown)* parameter value $\theta^\star \in \Theta$. A model is said to be *consistent* if it is free from logical errors. In Section 4.5, we develop an algorithm to detect and quantify any substantial deviation of the output q from $z^\star$ by evaluating the model's consistency.

Model uncertainty is present if the functional relations are (partly) unknown, incomplete, inadequate or reduced to approximations of more complex (but expensive) relations. Thus, if model uncertainty exists for the relevant process, then the set of equations

$$e(\theta, u, y) = 0,$$
$$h(y, \theta) - q = 0$$

does no longer represent reality. The "real" functional relations are often represented by introducing additional terms:

$$e(\theta, u, y) + \delta e(\theta, u, y, \ldots) = 0,$$
$$h(y, \theta) + \delta h(y, \theta, \ldots) - q = 0,$$

where $\delta e(\theta, u, y, \ldots)$ is the *discrepancy function* of the state equation and $\delta h(y, \theta, \ldots)$ is the modeling error of the observation operator. We denote by $\Delta\mathcal{M} := \{\delta e, \delta h\}$ the uncertainty of the model $\mathcal{M}$. In general, the discrepancy function δe is unknown, whereas δh is quite often known from experimental practice and can be quantified, see [75] in case of strain gauges, for example. Therefore, we focus on the uncertainty in the functional relations of the state equation. Evidently, the model-output is significantly affected by model uncertainty even if the data have zero noise, see

Figure 4.2.

Since computer models are very dominant in the literature we reformulate our equations using explicit functions. This is possible since we assumed that the state equation always has a unique solution $y(\theta, u)$. Thus, a model given in explicit form represents an input-output-relation that can be implemented in a computer program. Let $\eta(\theta, u) := h(y(\theta, u), \theta)$ be this reduced model and $\delta\eta(\theta, u, \ldots)$ be its uncertainty function. Then the "real" functional relations are given by

$$\eta(\theta, u) + \delta\eta(\theta, u, \ldots) = q.$$

In the following, we give some sources for $\delta\eta$ and classify its different types.

4.2 Sources and Types of Model Uncertainty

Following [111] and [105, Sec. 2.2], the main source of model uncertainty is *lack of knowledge* about the functional relations between the involved variables. Any mechanism that is unnoticed, unknown and nameless to us falls under this category. An example could be given from materials science and engineering. When a novel composite material is developed, then its reaction to the environment, its behavior under pressure or stress and its wear needs to observed, modeled and validated. As long as there is only a few data available for verification, calibration and validation, lack of knowledge is dominant.

A second major source of model uncertainty is *disregard of knowledge*. Every effect that is known to us but which is neglected, ignored or kept out of consideration in the modeling we categorize as disregarded. If a linear elastic spring, for example, is placed under load, then the deformation of the spring is proportional to the loading force provided that the latter stays below a material-dependent threshold. If the magnitude of the load is above this limit the spring shows nonlinear, hyper-elastic or even plastic behavior. Since these effects are difficult to model they are often neglected.

Another source of model uncertainty is the *numerical scheme* used to approximately solve equations that do not admit analytical solutions or whose analytical solutions are impracticable to compute. Numerical errors include discretization errors, iterative convergence errors, rounding errors, and also implementation errors in the computer code. If there are estimators for all these errors, then their impact can, in principle, be removed. In engineering applications, a classic approach to the discretization of a PDE is the finite element method (FEM) for which a large class of error estimates is available [109].

Besides numerical issues, *erroneous solution techniques* to the mathematical equations could yield non-physical solutions which are useless. The weak solutions of a scalar hyperbolic PDE, for example, are non-unique and one needs to select an entropy-solution in order to catch the appropriate physics [99]. However, this source of uncertainty is easily recognized and often taken care of by mathematicians.

Additionally, *human shortcomings* also contribute to model uncertainty.

In view of the different sources of model uncertainty one can distinguish the following types. First, model uncertainty addresses the functional structure itself. Incomplete, inadequate or missing relations between all involved variables are summarized in this type. Second, numerical uncertainty comprises all rounding, discretization and convergence problems and implementation errors in the computer code. Lastly, the human factor also induces model uncertainty. In this dissertation we focus on the detection and quantification of the first type, namely, uncertainty in the model.

The model-output is significantly influenced by the assumptions made during the modeling process: shall the functional relations be of axiomatic or empiric type, linear or non-linear, time-invariant or transient [92]? Moreover, the scope and complexity of the model determine its applicability and computational tractability. It is evident that any computer simulation that is based upon uncertain models inevitably leads to misleading predictions of the quantity of interest. As a consequence the model's usefulness, correctness and consistency need to be verified and revalidated. To do this, one needs to carefully select the adequate verification and validation processes, see Section 4.4. In the following, we investigate the role of data and data uncertainty upon the model's prediction quality and calibration parameters.

4.3 The Influence of Data Uncertainty upon Models

This section is based on [105, Sec. 2.1]. In general, every measured or collected quantities, which we call *data*, are subject to uncertainty. We speak of *data uncertainty* if the value, interval or distribution of data is unknown, incomplete or insufficient. The functional relations as well as the scope and complexity of a model are clearly independent from how data are at hand: as a single value, an interval or randomly distributed. However, data often function as inputs to the model and thus, uncertainty propagates to the model-output. In this way uncertain data affect the prediction quality of a model but does not contribute to model uncertainty in a strict sense. Figure 4.3 shows how uncertainty in the data is propagated through the model whereas model uncertainty is an intrinsic tentativeness in the functional
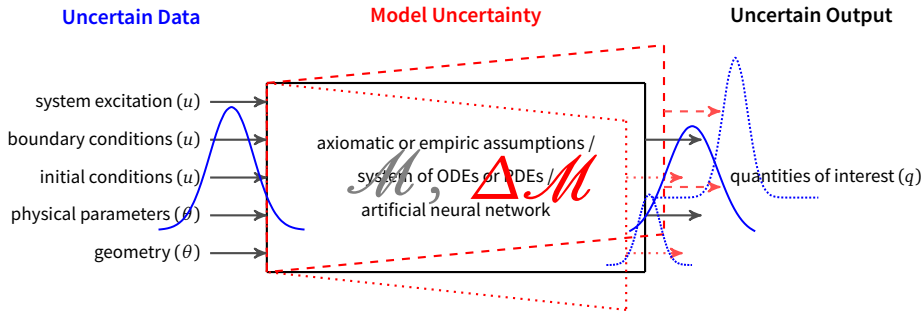
Figure 4.3 The model-output q is affected by both, model uncertainty $\Delta\mathscr{M}$ and uncertain parameters θ and inputs u. Data uncertainty is propagated through the model; based on [111].

relations which has a far greater impact on the model's prediction quality.

The output of the model $q = \eta(\theta, u)$ is usually compared to data z that are collected from real experiments, simulations or archives. This data are used to estimate the model parameters θ (geometrical or material properties) in a calibration procedure, for example. In addition, the model's inputs u (initial or boundary conditions) are also uncertain since they are measured quantities. When we speak of data uncertainty and its relation to models we thus focus on parametric uncertainty and uncertain inputs.

In general, data uncertainty is categorized into aleatoric or epistemic [35]. *Aleatoric uncertainty* is an intrinsic randomness of a phenomenon, it is irreducible, objective and often characterized by a probability distribution [85, 131, 142]. In experimental practice this type is often referred to as noise. On the other hand, *epistemic uncertainty* arises due to lack of knowledge about the origins of data or simply because of insufficient or incomplete data [35, 131]. It is subjective and, in principle, reducible. An example could be the unwanted systematic bias in the measurement of seismic waves nearby a construction area. Both types of data uncertainty can be quantified with non-probabilistic and probabilistic methods [87]. In the following we give a short literature overview of these methods.

4.3.1 A Probabilistic Framework for Data Uncertainty

The first type of methods to deal with data uncertainty is probabilistic in its nature. If data can be modeled as a realization of a random variable Ξ with a distribution $P[\Xi]$, then we speak of *stochastic uncertainty*. The distribution can be approximated or estimated via a parametric or non-parametric approach and within a frequentist

or Bayesian inference perspective. All these approaches operate under the assumption that an event is possible with a known probability.

In *parametric* stochastic uncertainty, one assumes that the probability distribution $P_\theta[\Xi]$ of the random variable Ξ depends on finite dimensional parameters θ which uniquely characterize the distribution. Measurement errors or production tolerances are typically modeled by Gaussian distributions which are fully described by two parameters (θ_1, θ_2): mean $\theta_1 = \bar{\Xi}$ and standard deviation $\theta_2 = \sigma(\Xi)$. If sufficient data are available, a maximum likelihood approach gives an estimate of these parameters [71]. Otherwise, one can use prior knowledge from physics or engineering to approximate their value.

Non-parametric stochastic uncertainty, however, does not use parameters as a description but is only based upon observations. Kernel density estimators can be applied, for example, to estimate the density function $\rho(\cdot)$ of the distribution $P[\Xi]$, see [103]. In most cases, though, if the sample size is small and the assumption on the distribution type is correct, then a maximum likelihood estimation yields better results than a kernel density estimator.

If the distribution of the probabilistic model of data uncertainty is thus estimated, a common class of methods used in computer simulations to predict a probabilistic output are Monte Carlo methods [57, 116]. These strongly depend on the perspective from which statistical inference is made: frequentist or Bayesian. From a *frequentist* perspective, the term probability is defined as the relative frequency of an occurring event if a statistical experiment is repeated a large number of times. On the other hand, a *Bayesian* view of probability is much more subjective since it incorporates prior knowledge and assumptions about the distribution into the process. The prior is usually updated using Bayes' rule to obtain the posterior whenever new data are available. On an algorithmic level we refer to Markov Chain Monte Carlo techniques [121]. Bayesian inference is commonly used for parameter calibration.

4.3.2 Non-Probabilistic Approaches to Data Uncertainty

We say that data are subject to non-probabilistic uncertainty if they *cannot* be modeled as a realization of a random variable with a given probability distribution. In order to quantify the uncertainty, fuzzy set theory or a direct interval analysis may be suitable. These approaches mainly determine whether an event is possible or impossible [87].

Within the *fuzzy set theory* framework introduced by Zadeh [141], data elements are expressed as being members of a subset $A \subseteq F$ relative to a superset F. The characteristic function $\mathbb{1}_A : F \to \{0, 1\}$ describes membership as $\mathbb{1}_A(x) = 1$ if $x \in A$ and

non-membership as $\mathbb{1}_A(x) = 0$ otherwise. Other (smooth) membership functions are also possible. Often, it is better to divide the domain of the membership function into intervals to determine the possibility of a data element being inside a certain interval. An overview of applied fuzzy arithmetic is given in [61]. However, only a few practical applications of this theory exist.

Another approach used to quantify non-probabilistic data uncertainty is *interval analysis*. Here it is assumed that the data values are between fixed minima and maxima. A pessimistic perspective then searches for the worst-case scenario. Assuming that each parameter lies in an interval given by min/max values, the basic rules of arithmetic apply: addition, subtraction, multiplication and division. Depending on how often and in what form the parameters occur in the model equations, the intervals become larger when propagated. This process may be time consuming when implemented. Another drawback is the fact that interval analysis tends to overestimate uncertainty since only extreme values are considered which occur rather seldom in practice [1]. In higher dimensions, extreme values are replaced by compact uncertainty sets, like ellipsoids. For these the worst-case analysis becomes more complicated. As shown in [16, 49, 79] robust optimization techniques are suitable to quantify and master data uncertainty in this pessimistic setting.

4.4 Common Methods for Analysis, Quantification and Evaluation of Model Uncertainty

In this section we give an overview of the different approaches to handle model uncertainty following the author's work in [105, Sec. 2.2]. The literature can be classified into two main perspectives on mastering model uncertainty: a *deterministic analysis* and a *probabilistic frequentist* or *Bayesian approach*. The latter is much more subjective since it needs prior knowledge as it was the case for data uncertainty, see Section 4.3. However, a frequentist view does not take into account prior information even if it is available and reliable. The deterministic analysis, on the other hand, has the difficulty of finding the adequate functional correction terms. In spite of these disadvantages, each approach has its own use cases of success. In the following, we present the most common methods to identify, quantify and master model uncertainty.

4.4.1 Deterministic Approaches

Simani et al. [120] develop a strategy for fault diagnosis where bounded error terms are incorporated into the model equations. To give a short example, let η be the

solution operator to the differential equations

$$\begin{aligned} \partial_t y(t) &= A(\theta)y(t) + Bu(t), \qquad \text{for } t > 0, \\ y(0) &= Bu(0), \end{aligned}$$

(4.1)

where $A(\theta)$ and B are suitable matrices. Then model uncertainty can be described by including the term $\Delta(t) \in [c_{\min}(t), c_{\max}(t)]$, where $c_{\min}(t) < c_{\max}(t)$, to the first equation in (4.1):

$$\partial_t y(t) = A(\theta)y(t) + Bu(t) + \Delta(t), \qquad \text{for } t > 0. \tag{4.2}$$

Using interval analysis [1] or a worst-case-scenario treatment [16, 120] for (4.2), the uncertainty $\delta\eta$ of the solution operator can be bounded. Thus, the model-output is uncertain but stays in a bounded uncertainty set $\mathcal{U}_\eta$. By employing a residual analysis with thresholding, the engineer can recognize component failures while considering the effects of model uncertainty. However, the approach relies on the assumption that the uncertainty stays within a bounded set $\mathcal{U}_\eta$ which may not be the case in practice.

In [121] and [42] it is mentioned to approximate the discrepancy function $\delta\eta$ by a polynomial p due to its computational tractability:

$$\eta(\theta, u) + p(\theta, u, \vartheta) = q.$$

In this setting, the polynomial also depends on hyper-parameters ϑ. We say that the original set of physical or empiric parameters is augmented by these non-physical entities. However, it is nearly impossible to gain a physics-based understanding of missing functional relations from these polynomial parameters. In practice, the augmented parameter set needs to be calibrated which is done by an optimization approach, see Chapter 2.

4.4.2 Probabilistic Framework

There is extensive literature on *Bayesian inference-based methods* for model calibration and validation [58, 88, 114, 134]. In [91] an interval-hypothesis test is developed to compare various models of a suspension strut concerning their basic functional assumptions of axiomatic or empiric type.

Another important approach was introduced by Kennedy and O'Hagan [74] who

model the discrepancy function $\delta\eta$ as a stochastic process:

$$\eta(\theta, u) + \mathcal{V}(\theta, u, \vartheta) + \mathcal{E} = Z, \qquad (4.3)$$

where $\mathcal{V} \sim \mathcal{N}(m(u, \theta, \vartheta), C(u, \theta, \vartheta))$ is a multivariate Gaussian random variable with mean $m(u, \theta, \vartheta)$ and covariance function $C(u, \theta, \vartheta)$. In this setting, $\mathcal{E} \sim \mathcal{N}(0, \Sigma)$ is the random variable describing the noise that is commonly observed in the measurement process. The discrepancy function in (4.3) depends on hyper-parameters ϑ which need to be calibrated within a Bayesian optimization scheme. An adaptation for time-dependent problems is developed in [13]. Arendt et al. [8] use this approach to update models. However, in [24] it was shown that the success of modeling $\delta\eta$ as in (4.3) depends on accurate prior knowledge about $\mathcal{V}$ which may be difficult to obtain. Gu and Wang [60] use a scaled Gaussian process for model calibration and output prediction by overcoming the conceptual differences to least-squares calibration and thus, improve the method introduced in [74].

A very efficient approach to master uncertainty in statistical state estimation of linear and nonlinear problems was developed by Kalman [72] and is today known as Kalman filtering. The basic idea is to employ multivariate Gaussian distributions in the state equation (4.1) as follows:

$$\partial_t y(t) = A(\theta) y(t) + Bu(t) + \Upsilon(t), \qquad \text{for } t > 0, \qquad (4.4)$$

where $\Upsilon(t) \sim \mathcal{N}(0, M(t))$ is dynamical noise with zero mean and time-dependent covariance matrix $M(t)$. The model-output is given by the solution of (4.4) inserted into the observation operator $h(y(t; \theta, u), \theta)$ which is compared to measurement data at each time point $t > 0$:

$$Z(t) = h(y(t; \theta, u), \theta) + \mathcal{E}(t),$$

where $\mathcal{E}(t) \sim \mathcal{N}(0, \Sigma(t))$ is observational noise with zero mean and covariance matrix $\Sigma(t)$. In this model, the set of all state variables is a stochastic process, a Markov Chain. Because the measurements are also dynamic random variables one often speaks of a Hidden Markov Model, see [54]. Equation (4.4) is then discretized and for each time step the evolving state variable is estimated using Bayes' rule. The prediction is corrected by actual measurement points if available. A drawback of this approach is the numerical instability that could arise when solving (4.4). What is more, the Kalman estimation needs a prior initialization of the state which may be uncertain as well.

Tuomi et al. [128] address the question whether a set of measurements is ad-

equately described by a single model with the same parameter values. Let N independent measurements $z_1, \dots, z_N$ be given and let $P[Z_1, \dots, Z_N | \mathcal{M}]$ be their joint probability distribution. Then the model $\mathcal{M}$ is said to be an inadequate description of the independent measurements if for some small $0 < r < 1$ the inequality

$$\frac{P[Z_1 = z_1, \dots, Z_N = z_N | \mathcal{M}]}{\prod_{k=1}^{N} P[Z_k = z_k | \mathcal{M}]} < r \tag{4.5}$$

holds. This definition is only based on the statistical independence of the measurements and them being modeled by a single model $\mathcal{M}$ with the same parameters. In (4.5) the term on the left side is sometimes called Bayes' factor. Other ways to derive similar factors that evaluate the fitness of different competing models can be found in [28, 73]. However, this approach depends on prior information about model parameters.

From a *frequentist's* perspective, a natural approach to quantify model uncertainty or the discrepancy term is to use validation metrics. According to [111], a simple way to do this is to compare the cumulative density function of the simulation model $F^{\mathcal{M}}(\cdot)$ with the cumulative density function of the experimental findings $F^{\mathcal{E}}(\cdot)$. These functions are empirical (one-dimensional) probability distributions:

$$F^{\mathcal{M}}(t) := \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{(-\infty, t]}(\eta(\theta, u_k)), \qquad F^{\mathcal{E}}(t) := \frac{1}{N} \sum_{k=1}^{N} \mathbb{1}_{(-\infty, t]}(z_k),$$

for N observations. Then an area validation metric can be defined as

$$d\big(F^{\mathcal{M}}, F^{\mathcal{E}}\big) := \sum_{k=1}^{N} \big| F^{\mathcal{M}}(t) - F^{\mathcal{E}}(t) \big| \Delta t,$$

where $\Delta t > 0$ is a predefined mesh size. Other validation metrics can be found in [89, 143]. Within a model selection process, an arbitrary threshold is imposed on the metric or a hypothesis test is performed.

For more complex technical systems, one often uses surrogate models to quantify their uncertainty. These methods are based on computer simulations and a small sample of measurements to estimate quantiles or densities of probability functions [56, 76, 77].

Another approach to quantify the model discrepancy term is developed in [138] which relies on bootstrap samples in a regression estimation. The authors also use smoothing splines and artificial neural networks.

4.5 A Novel Algorithm to Detect and Quantify Model Uncertainty

In this section, we come to the main contribution of this thesis. We develop a novel algorithm to detect and quantify model uncertainty using parameter estimation, optimum experimental design and statistical hypothesis testing. All these topics have been introduced in the previous chapters; hypothesis testing is summarized in Appendix A.6. This exposition is based on the author's work in [50, 93], see also [115, Sec. 4.3.1], which originated at the CRC 805 in the Subproject A3. We only consider models for which the following assumption holds:

Assumption 4.2. All model parameters $\theta \in \Theta$ are *physical* in the sense that they have a physics-based meaning and it is possible to know or at least to approximate their true value.

Thus, if a model $\mathcal{M}$ is considered to be correct, then the true value $\theta^\star$ of the parameters can be approximated by solving parameter estimation problems with given data. We also assume that there is a direct link between the model's correctness and consistency:

Assumption 4.3. If a model $\mathcal{M}$ is correct, then it must be consistent, i.e., it should reproduce *all* measurement series obtained from *any* admissible inputs at *any* predefined sensor location with the *same* parameter values.

In practice, we relax this assumption to the claim that the parameter values must lie in a small confidence region due to the effects of measurement errors. In an optimally designed experiment, i.e., using optimal sensor locations and best input configurations, this confidence region becomes small. As discussed in Chapter 3, we offer both a probabilistic frequentist and a Bayesian framework to do so. Thus, the propagation of data uncertainty is controlled and its extent is minimized. If the model is correct, then repeated calibration and validation procedures should yield the *same* parameter values up to a small error. However, if a set of measurements leads to parameters that lie outside this confidence region, then the model is inconsistent and according to Assumption 4.3 incorrect. Algorithm 4.1 summarizes our approach. For a given mathematical model $\mathcal{M}$, a threshold level TOL for the Type I error, a given number of tests to be performed n_{tests} and a division strategy DS, the output is the assessment whether $\mathcal{M}$ should be rejected with probability TOL, e.g., 5%, for the Type I error. In the following, we explain each step in detail.

Algorithm 4.1 (Detection of Model Uncertainty in a Mathematical Model $\mathcal{M}$).

Input: Model $\mathcal{M}$, test level TOL, number of tests n_{tests}, division strategy DS.
Output: Does $\mathcal{M}$ need to be rejected? YES (1) or NO (0).

1: Initialize $i := 1$.
2: Generate initial data z_{ini} in all feasible sensor locations.
3: Compute θ_{ini} by PE using z_{ini}.
4: Solve OED problem using z_{ini} or θ_{ini}.
5: Acquire data z with the optimal setup of the experiment.
6: **if** measurement errors are *not* Gaussian **then**
7:     Go to line 5 or **exit**.
8: Divide z into one calibration set z_{cal} and one validation set z_{val} using DS.
9: Calculate $(\theta_{\text{cal}}, C_{\text{cal}})$ using z_{cal} by PE. Likewise, obtain $(\theta_{\text{val}}, C_{\text{val}})$ using z_{val}.
10: Find $\alpha_{\min} \in (0, 1)$, such that θ_{val} lies on boundary of $K(\alpha_{\min}, \theta_{\text{cal}}, C_{\text{cal}} + C_{\text{val}})$.
11: **if** $\alpha_{\min} \geq \text{TOL}/n_{\text{tests}}$ **then**
12:     **if** $i < n_{\text{tests}}$ **then**
13:         Set $i := i + 1$ and go to line 8.
14:     **else**
15:         **return** 0.
16: **else if** $\alpha_{\min} < \text{TOL}/n_{\text{tests}}$ **then**
17:     **return** 1.

4.5.1 Description of the Algorithm

First, we need to acquire initial data z_{ini} in all feasible sensor locations for some inputs in order to compute a first guess θ_{ini} of the model parameters by solving (3.9) or (3.10) depending on the probabilistic perspective (lines 2–3). Since the acquisition of data is costly, one can also generate artificial data via simulations to do so. If we consider a pure sensor placement problem, then one either uses z_{ini} to solve Problem 3.15 with $C = C_{\text{NLS/GN/AL/B/AB}}$ depending on the probabilistic framework or θ_{ini} to solve Problem 3.11 with $C = C_{\text{GN/B}}$. In this case, we obtain optimal sensor positions $\overline{\omega}$. If the experimental setup allows for a free choice of inputs, then one uses θ_{ini} to solve Problem 3.17 obtaining optimal sensor positions $\overline{\omega}$ and best input configurations $\overline{u}$. This is summarized in line 4.

The optimal experimental setup is used to collect the data z that are necessary for the upcoming hypothesis test (line 5). If the design of the experiment consists of optimal sensor positions and optimal input configurations $(\overline{\omega}, \overline{u})$, we allow little variations in the input choice around $\overline{u}$ in order to ensure a certain diversity in our data which will be exploited by the division strategy DS. The size of the confidence region around the parameters stays nearly the same because of continuity of the function $u \mapsto C(\overline{\omega}, u)$ provided that the variance of the measurement process Σ

stays constant. In case of a pure sensor placement problem, the inputs are fixed by the boundaries and conditions of the experiment – its setup consists of optimal sensor positions $\overline{\omega}$ only.

Since our approach is based on the assumption that observational noise is Gaussian, we check whether the measurement errors are indeed realizations of a normal distribution with zero mean and variance Σ. We use the Shapiro-Wilk goodness-of-fit test [32] to do so (line 6). If this test fails, we measure again or abort the procedure since our algorithm is not equipped to handle non-Gaussian observational noise.

Given a set of measurements z, we split the data into one calibration set z_{cal} and one validation set z_{val} according to a division rule DS (line 8). Let $u_1, \ldots, u_m \in U$ be pairwise distinct inputs for which data vectors $z(u_1), \ldots, z(u_m) \in Q_I$ are collected. The following strategies for DS have been suggested:

- **Monte-Carlo cross-validation** [37]. Let ς_m be a permutation of size $m \in \mathbb{N}$ and $k \in \mathbb{N}$ with $1 \leq k \leq m$ be arbitrary. Then $z_{\text{cal}} := \{z(u_{\varsigma_m(1)}), \ldots, z(u_{\varsigma_m(k)})\}$ and $z_{\text{val}} := \{z(u_{\varsigma_m(k+1)}), \ldots, z(u_{\varsigma_m(m)})\}$. In each run, ς_m and k are randomly changed.
- **m-fold cross-validation** [68]. Divide the data into m groups according to the inputs $u_1, \ldots, u_m$. Then use $m-1$ groups for the calibration set z_{cal} and the remaining group for the validation set z_{val}. After $n_{\text{tests}} = m$ runs, all combinations are passed through.
- **Expert judgment**. This strategy may be necessary if it is of interest whether a specific physical effect was considered in the modeling. It may help in finding the worst-case split to target the model uncertainty directly.

In any case, the division strategy DS must guarantee that the measurement errors stay Gaussian.

From line 9 onward, we perform a statistical test with Bonferroni correction for the test level, see Appendix A.6. Let θ_{cal} be the parameter estimate obtained from the calibration set z_{cal} and let θ_{val} be the one from the validation set z_{cal}. For convenience, define u_{cal} as the set of inputs used for calibration and let u_{val} be the one used for validation. Compute $C_{\text{cal}} := C(\overline{\omega}, u_{\text{cal}})$ and $C_{\text{val}} := C(\overline{\omega}, u_{\text{val}})$ according to one of the expressions given in (3.11)–(3.15) which depends on the probabilistic framework. Let $\theta^\star$ be the true but unknown parameter value. We test the following hypotheses:

$$H_0 : \theta^\star = \mathbb{E}_z[\theta_{\text{cal}}],$$
$$H_1 : \theta^\star \neq \mathbb{E}_z[\theta_{\text{cal}}].$$

First, we construct the acceptance region

$$K(\alpha, \theta_{\text{cal}}, C_{\text{cal}} + C_{\text{val}}) := \left\{ \theta \in \Theta \; : \; (\theta - \theta_{\text{cal}})^\top (C_{\text{cal}} + C_{\text{val}})^{-1} (\theta - \theta_{\text{cal}}) \leq \gamma_{n_\theta}(1-\alpha) \right\},$$

where $\gamma_{n_\theta}(\cdot)$ is the quantile function of the central $\chi^2_{n_\theta}$ distribution with n_θ degrees of freedom and $\alpha = \text{TOL}/n_{\text{tests}}$ is the corrected test level. Then the statistical test is defined as

$$\tau(z(u_1), \ldots, z(u_m)) := \begin{cases} 0, & \text{if } \theta_{\text{val}} \in K(\alpha, \theta_{\text{cal}}, C_{\text{cal}} + C_{\text{val}}), \\ 1, & \text{otherwise}, \end{cases}$$

meaning that the null hypothesis H_0 is rejected if $\tau(z(u_1), \ldots, z(u_m)) = 1$, and this is the case if $\theta_{\text{val}} \notin K(\alpha, \theta_{\text{cal}}, C_{\text{cal}} + C_{\text{val}})$. The test statistic is then defined as

$$T(z(u_1), \ldots, z(u_m)) := (\theta - \theta_{\text{cal}})^\top (C_{\text{cal}} + C_{\text{val}})^{-1} (\theta - \theta_{\text{cal}}).$$

and the critical value $c := \gamma_{n_\theta}(1 - \text{TOL})$. Let $Z_1, \ldots, Z_m$ be the random variables corresponding to the data samples $z(u_1), \ldots, z(u_m)$. Then the p-value of the statistical test is given by

$$\alpha_{\min} = P_{\theta=\theta^\star, T}[T(Z_1, \ldots, Z_m) > T(z(u_1), \ldots, z(u_m))],$$

which is computed in line 10. The null hypothesis H_0 is rejected if $\alpha_{\min} < \alpha$ (line 17). Otherwise, the algorithm returns to line 8 and divides the data again to perform the next hypothesis test.

The Bonferroni correction $\alpha = \text{TOL}/n_{\text{tests}}$ is a conservative method to account for the problem of multiple testing. We are interested in keeping the probability that at least one null hypothesis is falsely rejected small, i.e., we want to control the family-wise error rate (FWER). By adjusting the test level in each scenario, we achieve TOL as a bound for the FWER, see also Appendix A.6.

4.5.2 Main Theorems

In the following we prove that for correct linear models Algorithm 4.1 indeed returns 0 with probability $1 - \text{TOL}$, in a first order approximation. For this, recall that a linear model is defined as $\eta(\theta, u) = A(u)\,\theta$, where $u \subset U$ is a set of inputs and $A(u) \in \mathbb{R}^{n \times n_\theta}$ is the system matrix, see Section 2.1. The sensor accuracy $\Sigma \in \mathbb{R}^{n \times n}$ of the measurement process is a diagonal matrix with strictly positive diagonal entries σ_i^2. The sensor weights ω are again written in the diagonal matrix Ω which is dimension compatible to Σ. Furthermore, let $z^\star(u)$ be the true (but unknown)

value of the quantity of interest and $Z(u) \sim \mathcal{N}(z^\star(u), \Omega^{-1}\Sigma)$ be the corresponding random variable, where $\Omega^{-1}\Sigma$ is defined as in Section 3.2. We recall the definition of a chi-squared distribution:

Definition 4.4. Let $X_1, \ldots, X_k \sim \mathcal{N}(0, 1)$ be independently identically distributed univariate random variables. Then $\sum_{i=1}^{k} X_i^2$ is defined to be the *chi-squared distribution* χ_k^2 with $k \in \mathbb{N}$ degrees of freedom.

Without prior assumptions on the model parameters we have the following theorem:

Theorem 4.5. Let $u = \{u_1, \ldots, u_m\} \subset U$ be a set of inputs and let $\eta(\theta, u) = A(u)\,\theta$ be a linear model with $A(u_s)$ having full column-rank for any subset $\emptyset \neq u_s \subset u$. Let $\theta^\star$ be the true value of the model parameters, $Z(u) \sim \mathcal{N}(z^\star(u), \Omega^{-1}\Sigma)$ and let the model be correct, i.e., $z^\star(u) = \eta(\theta^\star, u)$ for all inputs $u \subset U$. Then, for any division strategy, Algorithm 4.1 returns 0 with probability $1 - \mathtt{TOL}$, in a first order approximation.

Proof. Let $z(u_1), \ldots, z(u_m)$ be the data which are acquired with the inputs $u_1, \ldots, u_m$. Let $Z_1 = Z(u_1), \ldots, Z_m = Z(u_m)$ be the corresponding independently distributed random variables. Let DS be any division strategy and without loss of generality we consider here an m-fold cross-validation. Set

$$z_{\mathrm{cal}} := \{z(u_1), \ldots, z(u_{m-1})\} \quad \text{and} \quad z_{\mathrm{val}} := \{z(u_m)\}.$$

For convenience, define $u_{\mathrm{cal}} = \{u_1, \ldots, u_{m-1}\}$ and $u_{\mathrm{val}} = \{u_m\}$. Since η is linear, the estimates $\theta_{\mathrm{cal}}, \theta_{\mathrm{val}}$ are given by solving Problem 2.1, and according to Proposition 2.2 we have

$$\theta_{\mathrm{cal}} = (A(u_{\mathrm{cal}})^\top \Omega \Sigma^{-1} A(u_{\mathrm{cal}}))^{-1} A(u_{\mathrm{cal}})^\top \Omega \Sigma^{-1} z(u_{\mathrm{cal}}),$$
$$\theta_{\mathrm{val}} = (A(u_{\mathrm{val}})^\top \Omega \Sigma^{-1} A(u_{\mathrm{val}}))^{-1} A(u_{\mathrm{val}})^\top \Omega \Sigma^{-1} z(u_{\mathrm{val}}).$$

By Proposition 2.3 these estimates are realizations of Gaussian random variables:

$$\theta_{\mathrm{cal}} \sim \mathcal{N}(\theta^\star, C_{\mathrm{cal}}) \quad \text{and} \quad \theta_{\mathrm{val}} \sim \mathcal{N}(\theta^\star, C_{\mathrm{val}}),$$

where

$$C_{\mathrm{cal}} := \left(A(u_{\mathrm{cal}})^\top \Omega \Sigma^{-1} A(u_{\mathrm{cal}})\right)^{-1} \quad \text{and} \quad C_{\mathrm{val}} := \left(A(u_{\mathrm{val}})^\top \Omega \Sigma^{-1} A(u_{\mathrm{val}})\right)^{-1}.$$

Since θ_{cal} and θ_{val} are independent, we have $\theta_{\mathrm{diff}} := \theta_{\mathrm{val}} - \theta_{\mathrm{cal}} \sim \mathcal{N}(0, C_{\mathrm{cal}} + C_{\mathrm{val}})$ by Lemma A.9. Using Proposition A.8, we know that $(C_{\mathrm{cal}} + C_{\mathrm{val}})^{-\frac{1}{2}} \theta_{\mathrm{diff}} \sim \mathcal{N}(0, I)$. Thus,

the test statistic in Algorithm 4.1

$$T(z(u_1),\dots,z(u_m)) := \left\|(C_{\text{cal}} + C_{\text{val}})^{-\frac{1}{2}} \theta_{\text{diff}}\right\|_2^2 = \theta_{\text{diff}}^\top (C_{\text{cal}} + C_{\text{val}})^{-1} \theta_{\text{diff}}$$

is a realization of a sum of squared $\mathcal{N}(0, I)$-distributed random variables, i.e., a chi-squared distribution with n_θ degrees of freedom. By the definition of its quantile function $\gamma_{n_\theta}(\cdot)$, we have

$$P\left[T(Z_1,\dots,Z_m) \le \gamma_{n_\theta}\left(1 - \frac{\text{TOL}}{n_{\text{tests}}}\right)\right] = 1 - \frac{\text{TOL}}{n_{\text{tests}}}$$

in each hypothesis test. The probability that Algorithm 4.1 returns 0 after n_{tests} tests have been performed is given by

$$\left(1 - \frac{\text{TOL}}{n_{\text{tests}}}\right)^{n_{\text{tests}}} \approx 1 - \text{TOL},$$

in a first order approximation. ∎

Remark 4.6. If a linear model is correct, then the probability that a null hypothesis is accepted in Algorithm 4.1 does not depend on the particular choice of $u_i \in U$ and $\omega_i \in \{0, 1\}$, i.e., it is independent of the experimental setup.

We now prove that in case of an incorrect linear model one can construct an instance such that Algorithm 4.1 returns 1 with probability greater than TOL. To do this, we review some properties of non-central chi-squared distributions.

Definition 4.7. Let $X_1 \sim \mathcal{N}(\mu_1, 1)$, $X_2 \sim \mathcal{N}(\mu_2, 1)$, ..., $X_k \sim \mathcal{N}(\mu_k, 1)$ be independently distributed univariate random variables. Then $\sum_{i=1}^k X_i^2$ is defined to be the *non-central chi-squared distribution* $\chi_{k,\lambda}^2$ with $k \in \mathbb{N}$ degrees of freedom and non-centrality parameter $\lambda := \sum_{i=1}^k \mu_i^2$.

Corollary 4.8. Let $X \sim \mathcal{N}(\mu, I)$ be a multivariate Gaussian random variable with mean $\mu \in \mathbb{R}^k$ and unit matrix variance. Then $X^\top X \sim \chi_{k,\lambda}^2$, where $\lambda = \mu^\top \mu$.

Proposition 4.9. Let $W(x; k)$ be the cumulative probability distribution function (CDF) of χ_k^2. Then the CDF of $\chi_{k,\lambda}^2$ is given by

$$Q(x; k, \lambda) = \exp\left(-\frac{\lambda}{2}\right) \sum_{i=0}^\infty \frac{\lambda^i}{i! \cdot 2^i} W(x; k + 2i).$$

Proof. See [70]. ∎

Lemma 4.10. In the setting of Proposition 4.9 the following holds:
  (a) Evidently, $Q(x; k, \lambda) < W(x; k)$ for all $k \in \mathbb{N}$ if $x > 0$ and $\lambda > 0$.
  (b) Let $\gamma_k(\cdot)$ be the quantile function of χ_k^2 and let $\alpha \in (0, 1)$. Then the function

$$(0, \infty) \ni \lambda \mapsto q(\lambda; \alpha, k) := 1 - Q(\gamma_k(1 - \alpha); k, \lambda) \in (0, 1)$$

  is increasing.

Proof. See [112, Lem. F.4]. ∎

We now have the following theorem:

Theorem 4.11. Let $\eta(\theta, u) = A(u)\theta$ be a linear model with $A(u)$ having full column-rank for all $u \in U$. Furthermore, let $u_1, u_2 \in U$ with $u_1 \neq u_2$, vectors $c_1 \in \text{range}\,(A(u_1))$, $c_2 \in \text{range}\,(A(u_2))$ with $c_1, c_2 \neq 0$ and $\bar{\theta} \in \Theta$ be given, such that

$$\mathbb{E}_z\left[\eta(\bar{\theta}, u_1) - Z(u_1)\right] = c_1 \quad \text{and} \quad \mathbb{E}_z\left[\eta(\bar{\theta}, u_2) - Z(u_2)\right] = -c_2,$$

where $Z(u_1) \sim \mathcal{N}(z^\star(u_1), \Omega^{-1}\Sigma)$ and $Z(u_2) \sim \mathcal{N}(z^\star(u_2), \Omega^{-1}\Sigma)$. Then there is a division strategy DS such that Algorithm 4.1 returns 1 after one test with probability

$$p = \text{TOL} + W(\gamma_{n_\theta}(1 - \text{TOL}); n_\theta) - Q(\gamma_{n_\theta}(1 - \text{TOL}); n_\theta, \lambda) > \text{TOL},$$

where $W(\cdot\,; n_\theta)$ is the CDF of $\chi_{n_\theta}^2$, $\gamma_{n_\theta}(\cdot)$ is its quantile function and $Q(\cdot\,; n_\theta, \lambda)$ is the CDF of $\chi_{n_\theta, \lambda}^2$ with

$$\lambda = (\theta_1 + \theta_2)^\top \left[(A(u_1)^\top \Omega \Sigma^{-1} A(u_1))^{-1} + (A(u_2)^\top \Omega \Sigma^{-1} A(u_2))^{-1}\right]^{-1} (\theta_1 + \theta_2),$$

for some $\theta_1, \theta_2 \in \Theta$ that satisfy $A(u_1)\theta_1 = c_1$ and $A(u_2)\theta_2 = c_2$.

Proof. Let $u_1, u_2 \in U$ with $u_1 \neq u_2$ be given such that

$$\eta(\bar{\theta}, u_1) = z^\star(u_1) + c_1, \qquad \eta(\bar{\theta}, u_2) = z^\star(u_2) - c_2,$$

where $z^\star(u_1) = \mathbb{E}[Z(u_1)]$ and $z^\star(u_2) = \mathbb{E}[Z(u_2)]$. Furthermore, let $z(u_1)$ and $z(u_2)$ be the corresponding data which are realizations of independently distributed random variables $Z_1 = Z(u_1)$ and $Z_2 = Z(u_2)$. We perform one test and choose the division strategy

$$z_{\text{cal}} := \{z(u_1)\} \quad \text{and} \quad z_{\text{val}} := \{z(u_2)\}.$$

Since η is linear, the estimates $\theta_{cal}, \theta_{val}$ are given by applying Proposition 2.2:

$$\theta_{cal} = (A(u_1)^\top \Omega \Sigma^{-1} A(u_1))^{-1} A(u_1)^\top \Omega \Sigma^{-1} z(u_1),$$
$$\theta_{val} = (A(u_2)^\top \Omega \Sigma^{-1} A(u_2))^{-1} A(u_2)^\top \Omega \Sigma^{-1} z(u_2).$$

According Proposition 2.3 these estimates are realizations of Gaussian variables:

$$\theta_{cal} \sim \mathcal{N}(\theta_{cal}^\star, C_{cal}) \quad \text{and} \quad \theta_{val} \sim \mathcal{N}(\theta_{val}^\star, C_{val}),$$

where the covariance matrices are given by

$$C_{cal} := (A(u_1)^\top \Omega \Sigma^{-1} A(u_1))^{-1} \quad \text{and} \quad C_{val} := (A(u_2)^\top \Omega \Sigma^{-1} A(u_2))^{-1},$$

and the means are given by

$$\begin{aligned}
\theta_{cal}^\star &= (A(u_1)^\top \Omega \Sigma^{-1} A(u_1))^{-1} A(u_1)^\top \Omega \Sigma^{-1} z^\star(u_1) \\
&= \bar\theta - C_{cal} A(u_1)^\top \Omega \Sigma^{-1} c_1 \\
&= \bar\theta - \theta_1, \\
\theta_{val}^\star &= (A(u_2)^\top \Omega \Sigma^{-1} A(u_2))^{-1} A(u_2)^\top \Omega \Sigma^{-1} z^\star(u_2) \\
&= \bar\theta + C_{val} A(u_2)^\top \Omega \Sigma^{-1} c_2 \\
&= \bar\theta + \theta_2,
\end{aligned}$$

for some $\theta_1, \theta_2 \in \Theta$. Define $\xi := (C_{cal} + C_{val})^{-\frac{1}{2}} (\theta_{val} - \theta_{cal})$. Using Lemma A.9 and Proposition A.8, we have $\xi \sim \mathcal{N}(s^\star(u_1, u_2), I)$, where

$$s^\star(u_1, u_2) := (C_{cal} + C_{val})^{-\frac{1}{2}} (\theta_{val}^\star - \theta_{cal}^\star) = (C_{cal} + C_{val})^{-\frac{1}{2}} (\theta_1 + \theta_2).$$

Thus, the test statistic $T(Z_1, Z_2) := \xi^\top \xi$ has a non-central chi-squared distribution $\chi^2_{n_\theta, \lambda}$ with non-centrality parameter

$$\lambda := s^\star(u_1, u_2)^\top s^\star(u_1, u_2) = (\theta_1 + \theta_2)^\top (C_{cal} + C_{val})^{-1} (\theta_1 + \theta_2) > 0,$$

and n_θ degrees of freedom. Following Lemma 4.10 (a), its cumulative distribution function (CDF) $Q(\cdot\,; n_\theta, \lambda)$ can be strictly bounded:

$$Q(x; n_\theta, \lambda) < W(x; n_\theta), \qquad \text{for } x > 0,$$

where $W(\cdot\,; n_\theta)$ is the CDF of $\chi^2_{n_\theta}$. Let $\gamma_{n_\theta}(\cdot)$ be the quantile function of $\chi^2_{n_\theta}$. Then

$$
\begin{aligned}
P\big[T(Z_1, Z_2) > \gamma_{n_\theta}(1 - \mathrm{TOL})\big] &= 1 - Q(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta, \lambda) \\
&= 1 - W(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta) \\
&\quad + W(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta) - Q(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta, \lambda) \\
&= \mathrm{TOL} + W(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta) - Q(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta, \lambda) \\
&> \mathrm{TOL}. \qquad \blacksquare
\end{aligned}
$$

Remark 4.12. In the proof above, the non-centrality parameter λ depends on the design (ω, u_1, u_2) of the experiment. We have

$$
s^\star(u_1, u_2)^\top s^\star(u_1, u_2) \geq \frac{1}{\lambda_{\max}(C_{\mathrm{cal}} + C_{\mathrm{val}})} \|\theta_1 + \theta_2\|_2^2.
$$

Let $(\overline{\omega}, \overline{u})$ be the optimally designed experiment and let either u_1 or u_2 be close to the optimal input $\overline{u}$. Then $\lambda_{\max}(C_{\mathrm{cal}})$ or $\lambda_{\max}(C_{\mathrm{val}})$ becomes small, and the lower bound of the non-centrality parameter $\lambda = s^\star(u_1, u_2)^\top s^\star(u_1, u_2)$ becomes large. According to Lemma 4.10 (b), the function

$$
\lambda \mapsto q(\lambda; \mathrm{TOL}, n_\theta) := 1 - Q(\gamma_{n_\theta}(1 - \mathrm{TOL}); n_\theta, \lambda)
$$

is increasing, and thus the lower bound of $P\big[T(Z_1, Z_2) > \gamma_{n_\theta}(1 - \mathrm{TOL})\big]$ becomes large as well. To summarize, an optimally designed experiment provokes the rejection of the null hypothesis for false models as intended.

Applications

Real technical systems encounter uncertainty in all phases of their life-cycle: from modeling and design to production, usage and recycling. If model uncertainty is present in the early stage of product development then it propagates to as far as the usage phase which can cause considerable problems. Therefore, for systems with high safety requirements, such as load-bearing structures, an early identification, quantification and later control of model uncertainty becomes a necessity. As mentioned in Chapter 4 models can only be validated and improved with actual measurements. We use these data to infer the true values of the model parameters, like material constants or geometrical properties, with minimal variance. Anticipating model uncertainty, it is reasonable to identify these parameters *after* the assembly of the load-bearing system. By conducting a hypothesis test we can evaluate the model's fitness to describe the behavior of the system in the usage phase.

In this chapter we apply our algorithm to detect model uncertainty in two load-bearing systems. Both played a major role in the Collaborative Research Center (CRC) 805. The first is a mechanical forming machine, the 3D Servo Press, which transmits torques and forces onto a component to be shaped according to a fixed design pattern. This machine was engineered by the Institute for Production Engineering and Forming Machines (PtU) in the working group of Prof. Dr.-Ing. Peter Groche. The second is a two-dimensional truss that represents the upper structure of an airplane's landing gear called Modular Active Spring-Damper System (MAFDS). The MAFDS was developed by the Department of Mechanical Engineering under Prof. Dr.-Ing. Peter Pelz and the Fraunhofer Institute for Structural Durability and System Reliability (LBF) under Prof. Dr.-Ing. Tobias Melz. These systems are described in much more detail in [104, Sec. 3.6].

This chapter is organized as follows. Section 5.1 introduces three models of the 3D Servo press which are distinguished by their friction part. We present numerical

results of our Algorithm 4.1 applied to these models for different probabilistic perspectives where the data have been collected on a small-scale prototype of the press. In Section 5.2 we apply our approach to identify model uncertainty in the dynamic linear-elastic model of vibrations in a truss structure. Since we could not perform our experiment as intended we use artificial data instead of real measurements to test our approach.

5.1 Identification of Model Uncertainty in Mathematical Models of the 3D Servo Press

We apply our method for identification of model uncertainty to a forming machine that transmits the torques and forces of its drives onto an elastic body to be shaped according to a target design. This technical system occurs in the literature under the name 3D Servo Press [104, Sec. 3.6.3] and the following exposition is based on the author's work in [50, 115]. This is joint work with the PtU, and the press models as well as the data were made available by Dr. Florian Hoppe.

5.1.1 Three Models of the 3D Servo Press

Forming machines are subject to high magnitudes of external forces. This causes various parts of the press to experience deformation. It has been shown in [59] that a rigid body model does not suffice to describe the behavior of the press in operation. However, an accurate model is crucial for a closed-loop control of the machine [66]. The best candidate model would be the time-variant PDE of linear elasticity coupled with a friction model. To keep the model computationally tractable, however, we use a mechanical substitute model in which the components of the press are modeled as bars, beams and spring elements depending on their stress under load. The notation used in this subsection is independent and may deviate from previous chapters.

The 3D Servo Press consists of three identical linkage mechanisms, one of which is shown in Figure 5.1. It shows several bars and beams connected via rotary joints. In the figure, each elastic component is represented by a spring or a thin beam and each mass component by a gray volume. The eccentric and spindle drives can move along three degrees of freedom: φ_{ecc} in rotation, y_{su} and y_{sl} in translation. We call this a gear unit. The drives cause all joints to perform a desired movement. The gear unit presses on point D and a linear pressure bar translates this force to the ram bearing R. Thus, the position of all marked points inside the press are determined by the angle of the eccentric drive φ_{ecc} and by the upper, as well as lower, spindle drive positions y_{su} and y_{sl}.
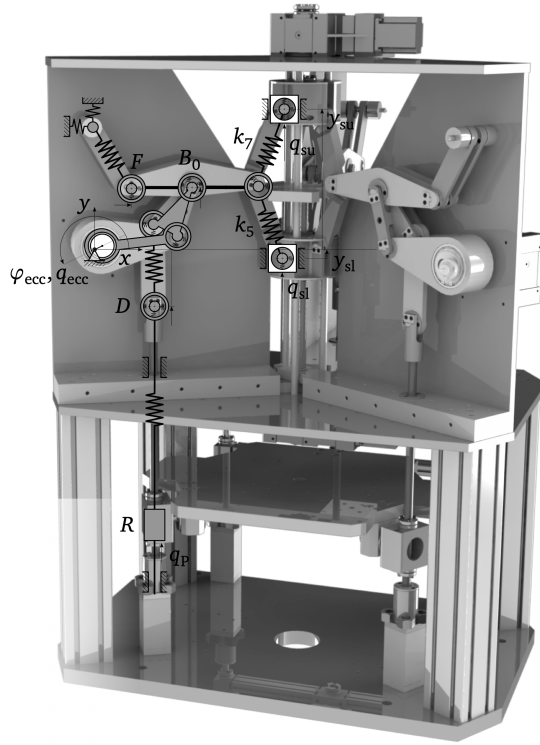
Figure 5.1 Linkage mechanism of the 3D Servo Press in profile view [50].

Let $\mathscr{L} = T - P$ be the Lagrangian, where T is the kinetic and P is the potential energy of the system. Moreover, let y be the state variables of the system and q be the non-conservative forces. The equations of motion are given by the Euler-Lagrange equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\partial_{\dot{y}}\mathscr{L}\right) - \partial_y \mathscr{L} = q. \tag{5.1}$$

All external forces applied to the 3D Servo Press, like the torque of the eccentric drive q_{ecc}, the forces of the upper and lower spindles q_{su}, q_{sl} and the reacting process force q_{P}, are summarized by the term *non-conservative* forces. In order to validate the elastic model, the position of the drives is fixed. Thus, all non-conservative forces are set to zero except for q_{P} which becomes the only applied force.

The press model consists of two rigid bodies, five bars, one beam, ten joints and

the elasticity of the press frame. The kinetic and potential energy of the system is equal to the sum of the kinetic and potential energy of each individual component:

$$T = \sum_{i=1}^{5} T_{\text{bar},i} + \sum_{i=1}^{1} T_{\text{beam},i} + \sum_{i=1}^{2} T_{\text{body},i},$$

$$P = \sum_{i=1}^{5} P_{\text{bar},i} + \sum_{i=1}^{1} P_{\text{beam},i} + \sum_{i=1}^{10} P_{\text{joint},i},$$

where the energy term of each individual component is defined as follows.

Bar model

A simplification of the theory of linear elasticity is the bar model. We use this model for coupling links which experience a very small stress under load. Each bar i is discretized using the Finite Element Method (FEM) and it is perceived as a pair of two mass elements and a spring between them. Let $m_{i,j}$ be the mass of element j of bar i, let $v_{i,j}$ be its translational velocity and let $k_{\text{bar},i}$ be the stiffness of the spring between the two mass elements. The total kinetic energy of each bar i is given by

$$T_{\text{bar},i} = \frac{1}{2} \sum_{j=1}^{2} m_{i,j} v_{i,j}^2 + \frac{1}{2} \Theta_i \dot{\varphi}_i^2,$$

where Θ_i is the mass moment of inertia and $\dot{\varphi}_i$ is the corresponding angular velocity. The total potential energy of each bar is the sum of the elastic energy of the springs and the gravitational energy of the masses

$$P_{\text{bar},i} = \frac{1}{2} k_{\text{bar},i} \xi_i^2 + \sum_{j=1}^{2} m_{i,j} g_0 y_{i,j},$$

where $g_0 = 9.81 \text{m/s}^2$, ξ_i is the relative displacement of the mass elements $m_{i,1}$ and $m_{i,2}$, and $y_{i,j}$ is the relative position of each mass element above the ground.

Note that the FEM is based on detailed knowledge of the elastic modulus and the geometry of the components where the latter comes from computer-aided design (CAD). However, due to fluctuations in the production process of the material the elastic modulus is not constant for all finite elements. Moreover, manufacturing tolerances compromise the accuracy of the CAD. Therefore, the stiffnesses are estimated by an a priori FEM simulation which needs to be tuned by solving parameter estimations problems with real data. This justifies our approach of estimating the stiffnesses of the two bars k_5 and k_7 *after* the assembly of the machine.
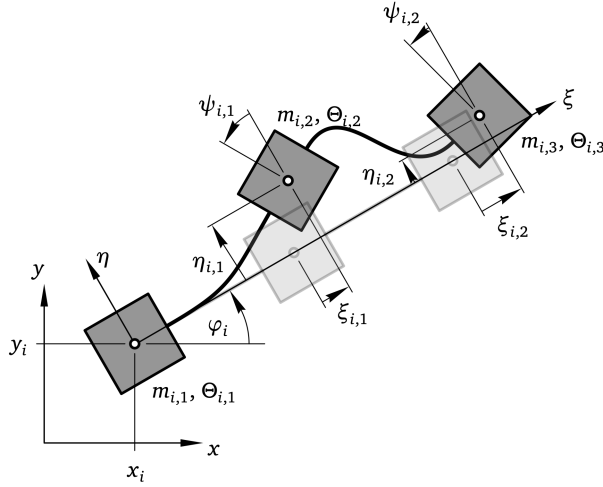
Figure 5.2 Beam model consisting of three masses and two springs [50].

Beam model

The remaining coupling links, whose stress under load is significantly higher, experience bending and thus, they are modeled as beams. This is certainly the case for the lever which connects the points F, B_0 and the common anchor point of the springs k_5 and k_7, see the thick gray line in Figure 5.1. The equations of motions for these coupling links are determined by the Euler-Bernoulli beam theory [53]. Each beam i is discretized by the FEM and the model is reduced to a mass-spring system with three masses and two springs, see Figure 5.2. The distribution of the total mass m_i of the beam upon the three masses is approximated in the following way:

$$m_{i,1} = \frac{m_i}{4}, \qquad m_{i,2} = \frac{m_i}{2} \qquad \text{and} \qquad m_{i,3} = \frac{m_i}{4}.$$

The kinetic energy of the beam has a similar form as for the bar:

$$T_{\text{beam},i} = \frac{1}{2} \sum_j^3 m_{i,j} v_{i,j}^2 + \frac{1}{2} \sum_j^3 \Theta_{i,j} \dot{\varphi}_i^2,$$

where $\Theta_{i,j}$ are the mass moments of inertia of the three masses and $\dot{\varphi}_i$ is the angular velocity of the whole beam. For the potential energy, again, the sum of the elastic

energy of the springs and the gravitational energy of the masses is given by

$$P_{\text{beam},i} = y_{\text{beam},i}^{\top} K_{\text{beam},i} y_{\text{beam},i} + \sum_{j=1}^{3} m_{i,j} g_0 y_{i,j},$$

where $K_{\text{beam},i}$ is the stiffness matrix obtained by the FEM discretization and

$$y_{\text{beam},i} = \left[x_{i,1}, y_{i,1}, \varphi_i, \xi_{i,1}, \eta_{i,1}, \psi_{i,1}, \xi_{i,2}, \eta_{i,2}, \psi_{i,2} \right]^{\top}$$

are the states of the beam, see Figure 5.2.

Bearing Model

We model the bearings as spring elements that are located between the joints of the couplers and whose stiffness $k_{\text{joint},i}$ is independent of the deflection. The radial force applied to the bearings is a function of the radial displacement Δr_i. Thus, the potential energy is given by

$$P_{\text{joint},i} = \frac{1}{2} k_{\text{joint},i} \, \Delta r_i^2.$$

The radial displacement Δr_i can be expressed in local coordinates (η, ξ) of the body which we omit for convenience.

Three Friction Models

All moving bearings experience friction which causes a hysteresis-type behavior of the whole press. We consider only the bearings guiding the pressure bar that connects the points D and R since the movement of the other bearings is insignificantly small. There exist a variety of approaches to model friction. In this application we consider three different models:

- **No friction**. We neglect the influence of friction and set

$$q_{\text{fric}}(t) = 0.$$

- **Discontinuous Coulomb model**. We assume that the sign of the velocity $\dot{R}_y$ at the point R in y-direction is identical to the sign of $\dot{q}_{\text{P}}$. Then

$$q_{\text{fric}}(t) = q_{\text{c}} \, \text{sign}(\dot{R}_y(t)) = q_{\text{c}} \, \text{sign}(\dot{q}_{\text{P}}(t)),$$

with q_c being a constant, gives a more accurate description of friction. This model is discontinuous with respect to $\dot{q}_P$.

- **Continuous model with rate-independent memory** [21]. Here, the force at the current time step t_i and the preceding one t_{i-1} is considered:

$$q_{\text{fric}}(t_i) = \mu(q_P(t_i), q_P(t_{i-1}), q_{P,\min}(t_i), q_{P,\max}(t_i)),$$

where μ is a memory function. We update the minimal and maximal process forces $q_{P,\min}(t_i)$, $q_{P,\max}(t_i)$ in each time iteration i according to the rule

$$q_{P,\min}(t_i) = \begin{cases} \min\{q_P(t_i), q_{P,\min}(t_{i-1})\} & \text{if } \dot{q}_P(t_i) \geq 0, \\ q_P(t_i) & \text{if } \dot{q}_P(t_i) < 0, \end{cases}$$

$$q_{P,\max}(t_i) = \begin{cases} \max\{q_P(t_i), q_{P,\max}(t_{i-1})\} & \text{if } \dot{q}_P(t_i) \leq 0, \\ q_P(t_i) & \text{if } \dot{q}_P(t_i) > 0. \end{cases}$$

The continuous friction model is comparable to an artificial neural network topology. In order to train this friction model we need the measured process force and the force estimated by the inverse model function. The inverse model maps the measured displacement and the estimated stiffness parameters (in absence of friction) to the causing force. When this procedure is applied to all measurements of a loading-unloading cycle, then the full hysteresis curve is identified by a functional relation. For more details we refer to [21, 94].

Summary

The press model incorporates two rigid bodies, five bars, one beam, ten joints and the elasticity of the press frame which represents support points relative to the environment. Thus, we have a 34-dimensional state vector y. Equation (5.1) yields

$$F(y, \dot{y}, \ddot{y}) + q_{\text{fric}}(t) - q_P(t) = 0,$$

with a 31-dimensional function F. We further simplify the model to the quasi-static case by requiring $\ddot{y} = \dot{y} = 0$. In the sequel, the parameters to be identified are the stiffness of the two bars k_5 and k_7. The inputs are the process forces q_P alone which are applied by an external pneumatic source. Recall that there are three candidate models for the 3D Servo Press:

$$\mathcal{M}_1 : \text{ press model with no friction,}$$

$$\mathcal{M}_2 \; : \; \text{press model with Coulomb's friction model,}$$

$$\mathcal{M}_3 \; : \; \text{press model with continuous friction behavior.}$$

In the following we apply Algorithm 4.1 to detect model uncertainty in the three models $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$ of the gear mechanism of the 3D Servo Press.

5.1.2 Numerical Results

First, we mention that we only perform measurements *on the small-scale prototype* of the 3D Servo Press. The data are collected for $m = 29$ different process forces q_{p} which we call inputs u in our setting. The press has a two-phase behavior: the first 15 applied forces are the *loading* scenario and the last 14 are the *unloading* scenario. For each input, the vertical displacement in point D, the horizontal displacement in point F and the vertical displacement in point B_0 is measured, see Figure 5.1. Moreover, we do not distinguish between initial and actual measurements but collect the data with sensors positioned at all $n_{\mathrm{s}} = 3$ previously introduced locations. Thus, line 5 in Algorithm 4.1 is skipped since all the relevant data are available in line 2. Our goal is to reduce the costs for obtaining new measurements in experiments that will be conducted *on the full-scale 160 tons press* by choosing two out of the three candidate sensor positions. The data obtained from the best two sensors are then used for the hypothesis test starting in line 8.

We repeat each measurement $n_{\mathrm{z}} = 6$ times. However, the experimental conditions do not allow for constant inputs u_{ij} to be applied in each measurement series $j = 1, \ldots, n_{\mathrm{z}}$ due to variations in the pneumatic pressure. This uncertainty leads to a violation of our assumption that the data series $z_1(u_i), \ldots, z_{n_{\mathrm{z}}}(u_i)$ are realizations of independently, identically distributed Gaussian random variables. Nevertheless, we know the *desired* setpoint values u_i^{d} for all $i = 1, \ldots, m$. A linear interpolation between the desired inputs u_i^{d} and the measured quantity of interest $z(u_{ij})$ is justified if the deviations $u_{ij} - u_i^{\mathrm{d}}$ are small. Since this is the case, we fix all inputs to these setpoint values and apply the linear interpolation:

$$z_j(u_i) := \frac{u_i^{\mathrm{d}}}{u^{ij}} \cdot z(u_{ij}) \in \mathbb{R}^{n_{\mathrm{s}}},$$

for all $j = 1, \ldots, n_{\mathrm{z}}$ and $i = 1, \ldots, m$. In the sequel, we only use these interpolated measurements.

Initially, we assumed that measurements $z_1(u_i), \ldots, z_{n_{\mathrm{z}}}(u_i)$ are realizations of independent, identically distributed Gaussian random variables. In order to verify this assumption, we test whether the measurement errors have a normal distribu-

tion, with mean zero and variance estimated from the data, using a Shapiro-Wilk goodness-of-fit-test [32]. Concretely, we consider the sample vector

$$\Delta z := \begin{pmatrix} z_2(u_i) - z_1(u_i) \\ z_4(u_i) - z_3(u_i) \\ z_6(u_i) - z_5(u_i) \end{pmatrix}$$

and set a threshold of 5% to the test level. The results for each sensor are summarized in Table 5.1. We clearly observe that the measurement errors Δz are indeed Gaussian with mean zero and variance estimated from the data.

The model parameters $\theta = (k_5, k_7)$ to be estimated describe the axial stiffness of important components inside the press, see Subsection 5.1.1 and Figure 5.1. In order to ensure that the principal part of the covariance matrix stays invertible, see Remark 3.14 (b), we choose two out of the three candidate sensor locations that were mentioned above. Since the overall number of combinations is small we solve the following discrete OED problem by enumeration:

$$
\begin{aligned}
\min_{\omega} \quad & \Psi[C(\overline{\theta}(z; \omega, u^{\mathrm{d}}), \omega, u^{\mathrm{d}})] \\
\text{s.t.} \quad & \overline{\theta}(z; \omega, u^{\mathrm{d}}) \text{ solves (3.9) or (3.10),} \\
& \omega \in \{0, 1\}^{n_s}, \quad \sum_{k=1}^{n_s} \omega_k = 2,
\end{aligned}
\tag{5.2}
$$

where $C(\overline{\theta}(z; \omega, u^{\mathrm{d}}), \omega, u^{\mathrm{d}}) = C_{\mathrm{NLS/GN/AB}}(\overline{\theta}(z; \omega, u^{\mathrm{d}}), \omega, u^{\mathrm{d}})$ depending on the probabilistic perspective, cf. (3.11), (3.12) and (3.15). We compare the outcome of these perspectives and compute all the design criteria that were examined in Section 3.4. The results for model $\mathcal{M}_3$ are summarized in Table 5.2–5.4. For the Bayesian perspective we analyze the influence of a weak and a strong prior which expresses the confidence we place in the expert knowledge:

$$\theta_0 = \begin{pmatrix} 5.82 \\ 5.82 \end{pmatrix} \cdot 10^7, \quad \Gamma_{\mathrm{weak}} = \begin{pmatrix} 6.25 & 0 \\ 0 & 6.25 \end{pmatrix} \cdot 10^{10}, \quad \Gamma_{\mathrm{strong}} = \begin{pmatrix} 2.5 & 0 \\ 0 & 2.5 \end{pmatrix} \cdot 10^9.$$

Table 5.1 Analysis of the measurement data [50].

Sensor	p-value in %	Standard deviation
1	60.11	$5.515 \cdot 10^{-6}$
2	79.64	$3.311 \cdot 10^{-6}$
3	60.26	$1.497 \cdot 10^{-6}$

For $C = C_{\text{NLS/GN}}$ we observe that omitting the second sensor leads to an increase in all design criteria by a factor of $\approx 2 \cdot 10^{19}$ in comparison to the initial design. This is a strong indication that the information matrix became singular. However, this is not the case in the Bayesian perspective $C = C_{\text{AB}}$ for both, a weak and a strong prior. The prior covariance matrix operates as a regularizer ensuring invertibility of the information matrix. A drawback of the Bayesian approach is the bias introduced by the prior. The values of the D-criterion are significantly higher than in the non-Bayesian cases and it seems that the stronger the prior the higher the difference. This is a hint that the estimated parameters $\bar{\theta}$ are not good approximations of the true values $\theta^\star$. Moreover, if one uses a strong prior, then the different sensor combinations do not have any impact on the design criterion, see the last column in Table 5.2–5.4, respectively.

We choose to be compliant to the E-criterion whereby effectively reducing the

Table 5.2 Results of the OED problem (5.2) for model $\mathcal{M}_3$, design criterion $\Psi = \Psi_{\text{A}}$ and different probabilistic perspectives (*w. p.* weak prior, *s. p.* strong prior), based on [50].

Sensor combination	$\Psi_{\text{A}}(C_{\text{NLS}})$	$\Psi_{\text{A}}(C_{\text{GN}})$	$\Psi_{\text{A}}(C_{\text{AB}})$ *w. p.*	$\Psi_{\text{A}}(C_{\text{AB}})$ *s. p.*
111 (initial)	$4.959 \cdot 10^9$	$9.221 \cdot 10^9$	$6.202 \cdot 10^{10}$	$4.982 \cdot 10^9$
101	$1.118 \cdot 10^{29}$	$1.137 \cdot 10^{29}$	$6.243 \cdot 10^{10}$	$4.994 \cdot 10^9$
011	$6.258 \cdot 10^9$	$9.478 \cdot 10^9$	$6.202 \cdot 10^{10}$	$4.986 \cdot 10^9$
110	$3.514 \cdot 10^9$	$3.495 \cdot 10^9$	$6.245 \cdot 10^{10}$	$4.989 \cdot 10^9$

Table 5.3 Results of the OED problem (5.2) for model $\mathcal{M}_3$, design criterion $\Psi = \Psi_{\text{D}}$ and different probabilistic perspectives (*w. p.* weak prior, *s. p.* strong prior), based on [50].

Sensor combination	$\Psi_{\text{D}}(C_{\text{NLS}})$	$\Psi_{\text{D}}(C_{\text{GN}})$	$\Psi_{\text{D}}(C_{\text{AB}})$ *w. p.*	$\Psi_{\text{D}}(C_{\text{AB}})$ *s. p.*
111 (initial)	$1.168 \cdot 10^{16}$	$4.184 \cdot 10^{16}$	$1.776 \cdot 10^{17}$	$6.206 \cdot 10^{18}$
101	$7.184 \cdot 10^{35}$	$7.736 \cdot 10^{35}$	$4.533 \cdot 10^{17}$	$6.235 \cdot 10^{18}$
011	$1.485 \cdot 10^{16}$	$4.795 \cdot 10^{16}$	$2.002 \cdot 10^{17}$	$6.215 \cdot 10^{18}$
110	$2.757 \cdot 10^{16}$	$6.570 \cdot 10^{16}$	$2.433 \cdot 10^{17}$	$6.222 \cdot 10^{18}$

Table 5.4 Results of the OED problem (5.2) for model $\mathcal{M}_3$, design criterion $\Psi = \Psi_{\text{E}}$ and different probabilistic perspectives (*w. p.* weak prior, *s. p.* strong prior), based on [50].

Sensor combination	$\Psi_{\text{E}}(C_{\text{NLS}})$	$\Psi_{\text{E}}(C_{\text{GN}})$	$\Psi_{\text{E}}(C_{\text{AB}})$ *w. p.*	$\Psi_{\text{E}}(C_{\text{AB}})$ *s. p.*
111 (initial)	$4.957 \cdot 10^9$	$9.217 \cdot 10^9$	$6.201 \cdot 10^{10}$	$2.499 \cdot 10^9$
101	$1.118 \cdot 10^{29}$	$1.137 \cdot 10^{29}$	$6.242 \cdot 10^{10}$	$2.500 \cdot 10^9$
011	$6.256 \cdot 10^9$	$9.473 \cdot 10^9$	$6.201 \cdot 10^{10}$	$2.499 \cdot 10^9$
110	$3.506 \cdot 10^9$	$3.476 \cdot 10^9$	$6.245 \cdot 10^{10}$	$2.500 \cdot 10^9$

size of the largest principal axis of the confidence ellipsoid, see Section 3.4. Since our sample size is small and the bias of the prior seems too strong, we evaluate the column $\Psi_E(C_{NLS})$ in Table 5.4 and conclude that omitting the last sensor, i.e., opting for the sensor combination 110, is the optimal solution to (5.2). This reduces the size of the largest principal axis of the confidence ellipsoid by $\approx 29\%$ and at the same time the costs are reduced by $\approx 33\%$. Note that when using C_{GN} the predicted reduction of the E-criterion is $\approx 62\%$ which may be too optimistic. To summarize, the optimal design of an experiment with the 3D Servo Press is to measure the vertical displacements in point D and the horizontal displacements in point F, see Figure 5.1.

We proceed with the hypothesis test described from line 8 onward in Algorithm 4.1. The fitness of the three models of the 3D Servo Press $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3$, which were introduced in Subsection 5.1.1, is now investigated. First, a graphical comparison is seen in Figure 5.3 where the model output and the measurement data are plotted against the input force. Second, the friction part of model $\mathcal{M}_3$ is constructed by an artificial neural network which needs to be trained by real measurement data. For this purpose we use four data series; the remaining two measurement series are used for the hypothesis test applied to all models. Third, the variance of the sensors is a critical value that significantly impacts the outcome of the hypothesis test. The accuracy of each sensor is determined by summing up the variance of the repeated measurement process, see Table 5.1, and the variance specified by the manufacturer. Thus, the total standard deviation of each sensor is given by

$$\sigma_1 = \sqrt{(5.515 \cdot 10^{-6})^2 + (1.414 \cdot 10^{-5})^2} \approx 1.518 \cdot 10^{-5},$$

$$\sigma_2 = \sqrt{(3.311 \cdot 10^{-6})^2 + (3.606 \cdot 10^{-6})^2} \approx 4.895 \cdot 10^{-6},$$

$$\sigma_3 = \sqrt{(1.497 \cdot 10^{-6})^2 + (3.606 \cdot 10^{-6})^2} \approx 3.904 \cdot 10^{-6}.$$

We split the data in four different ways. In all cases, we generate the calibration and validation sets with equal size and omit the first $u_1^d = 0$ and last $u_{29}^d = 0$ "applied" force since nothing is happening when the input is zero. First, we split the data into one loading $\mathscr{S}^l$ and one unloading $\mathscr{S}^u$ set according to the two-phase behavior of the press. The first case is obtained by splitting the data obtained from the loading scenario into one calibration $\mathscr{S}_c^{l_1}$ and one validation $\mathscr{S}_v^{l_2}$ set. The same is done for the data obtained from the unloading scenario $\mathscr{S}^u$ and thereby constructing the second case. The third way to split the data is to test loading versus unloading, i.e., we consider the data sets $\mathscr{S}_c^l$ for calibration and $\mathscr{S}_v^u$ for validation. The fourth and
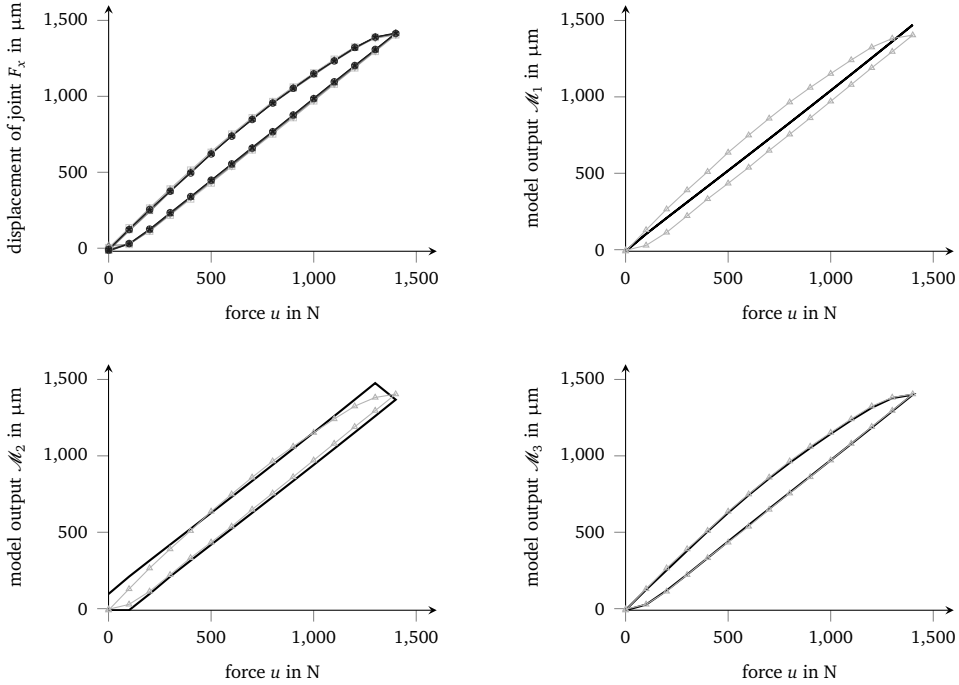
Figure 5.3 Measurement series of the force-displacement curve of the linkage mechanism and comparison with the model output for $\mathscr{M}_1$, $\mathscr{M}_2$ and $\mathscr{M}_3$, respectively, [50].

last case is a splitting where the loading and unloading sets are kept together and the distribution into one calibration set $\mathscr{S}_c^{lu}$ and one validation set $\mathscr{S}_v^{lu}$ alternates. This peculiar splitting maneuver was suggested by expert judgment in order to catch the worst test scenario, i.e., the most difficult test for a model to pass, which we believe is the third case. Our strategy is summarized in Table 5.5.

We now collect the outcome of the hypothesis test starting from line 9 in Algorithm 4.1 with $n_{tests} = 4$ test scenarios. Table 5.6–5.8 show the results. The value

Table 5.5 Summary of our four ways to split the data [50].

Force progression	Calibration	Validation
Loading	$\mathscr{S}_c^{l_1} = \{u_2^d, u_4^d, \ldots, u_{14}^d\}$	$\mathscr{S}_v^{l_2} = \{u_3^d, u_5^d, \ldots, u_{15}^d\}$
Unloading	$\mathscr{S}_c^{u_1} = \{u_{15}^d, u_{17}^d, \ldots, u_{27}^d\}$	$\mathscr{S}_v^{u_2} = \{u_{16}^d, u_{18}^d, \ldots, u_{28}^d\}$
Loading vs. unloading	$\mathscr{S}_c^{l} = \{u_2^d, \ldots, u_{14}^d\}$	$\mathscr{S}_v^{u} = \{u_{15}^d, \ldots, u_{28}^d\}$
Loading and unloading	$\mathscr{S}_c^{lu} = \{u_3^d, u_5^d, \ldots, u_{27}^d\}$	$\mathscr{S}_v^{lu} = \{u_2^d, u_4^d, \ldots, u_{28}^d\}$

$\alpha_{\min}$ is the minimal test level such that the null hypothesis can only just be rejected. We choose TOL = 5% as a bound for the family-wise error rate (FWER) and use the conservative Bonferroni correction $\text{TOL}/n_{\text{tests}} = 1.25\%$ for the individual level of each test, see Appendix A.6. From all three tables we infer that the Bayesian approach with strong prior convictions is misleading since model $\mathcal{M}_3$ is rejected but model $\mathcal{M}_1$ passes all tests! However, this is not the case when using weak prior convictions. Moreover, we clearly observe that in the classical approaches to probability (C_{NLS} and C_{GN}) the model $\mathcal{M}_3$ performs best as expected. In Table 5.6 we see that model $\mathcal{M}_1$ is mainly rejected in the second and third test scenario. This linear model does well in describing the loading and the whole two-phase process but is unable to explain unloading. Similarly, model $\mathcal{M}_2$ well describes the loading and the unloading phases separately but is unable to explain both of them with the same

Table 5.6 Test results for $\mathcal{M}_1$ from different probabilistic perspectives, based on [50].

Calibration	Validation	$\alpha_{\min}$ in %			
		C_{NLS}	C_{GN}	C_{AB} w. p.	C_{AB} s. p.
$\mathscr{S}_c^{l_1}$	$\mathscr{S}_v^{l_2}$	5.25	5.25	0.24	23.58
$\mathscr{S}_c^{u_1}$	$\mathscr{S}_v^{u_2}$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	28.25
$\mathscr{S}_c^{l}$	$\mathscr{S}_v^{u}$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	2.90
$\mathscr{S}_c^{lu}$	$\mathscr{S}_v^{lu}$	4.45	4.46	5.33	99.17

Table 5.7 Test results for $\mathcal{M}_2$ from different probabilistic perspectives, based on [50].

Calibration	Validation	$\alpha_{\min}$ in %			
		C_{NLS}	C_{GN}	C_{AB} w. p.	C_{AB} s. p.
$\mathscr{S}_c^{l_1}$	$\mathscr{S}_v^{l_2}$	86.40	86.40	$\ll 0.01$	28.74
$\mathscr{S}_c^{u_1}$	$\mathscr{S}_v^{u_2}$	77.01	77.01	36.89	32.27
$\mathscr{S}_c^{l}$	$\mathscr{S}_v^{u}$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$
$\mathscr{S}_c^{lu}$	$\mathscr{S}_v^{lu}$	$\ll 0.01$	$\ll 0.01$	$\ll 0.01$	99.67

Table 5.8 Test results for $\mathcal{M}_3$ from different probabilistic perspectives, based on [50].

Calibration	Validation	$\alpha_{\min}$ in %			
		C_{NLS}	C_{GN}	C_{AB} w. p.	C_{AB} s. p.
$\mathscr{S}_c^{l_1}$	$\mathscr{S}_v^{l_2}$	99.73	99.73	58.52	26.19
$\mathscr{S}_c^{u_1}$	$\mathscr{S}_v^{u_2}$	92.83	92.83	93.96	33.33
$\mathscr{S}_c^{l}$	$\mathscr{S}_v^{u}$	23.29	23.29	33.92	$\ll 0.01$
$\mathscr{S}_c^{lu}$	$\mathscr{S}_v^{lu}$	98.25	98.25	98.82	99.70

set of parameters, see Table 5.7. Thus, the 3D Servo Press model with a discontinuous Coulomb friction term does not reproduce all the data. Note that the difference between using C_{NLS} and C_{GN} is very small in this case. This should be considered when evaluating future experiments on the full-scale press since computing C_{NLS} is computationally more expensive. In conclusion, we see that Algorithm 4.1 is able to detect model uncertainty in $\mathcal{M}_1$ and $\mathcal{M}_2$ in the classical approaches to probability and in the Bayesian approach with weak prior convictions. For $\mathcal{M}_3$, the algorithm further provides a probability of the Type I error:

$$\alpha = \min\{\alpha_{\min,i} \in (0,1) : i = 1, \ldots, n_{\mathrm{tests}}\} = 23.29\%,$$

which is the necessary consequence if $\mathcal{M}_3$ is wrongly rejected.

5.2 Detection of Model Uncertainty in the Dynamic Linear-Elastic Model of a Truss Structure

In this section we apply Algorithm 4.1 to detect model uncertainty in the vibration equations of a linear-elastic truss. The truss is a two-dimensional model of the upper truss structure of the Modular Active Spring-Damper System [104, Sec. 3.6.1]. This part is based on the author's work in [93] which originated at the CRC 805 within Subproject A3. We first explain the model equations and recall the OED problem. Finally, we present numerical results.

5.2.1 Model Equations

Let $G \subset \mathbb{R}^2$ be a bounded Lipschitz domain with boundary $\partial G = \Gamma_{\mathrm{D}} \cup \Gamma_{\mathrm{F}} \cup \Gamma_{\mathrm{N}}$. The boundary parts $\Gamma_{\mathrm{D}}, \Gamma_{\mathrm{F}}, \Gamma_{\mathrm{N}}$ are pairwise disjoint and non-empty. Furthermore, let $(0, T)$ be an open and bounded time interval with some $T > 0$. We consider the parameter-dependent equations of motion for the linear-elastic body G which has a mass density $\varrho > 0$, see [67, Sec. 7.2]:

$$
\begin{aligned}
\varrho \, \partial_{tt}^2 y + a\varrho \, \partial_t y - \operatorname{div} \sigma &= 0, & &\text{in } (0, T) \times G, \\
y &= 0, & &\text{on } (0, T) \times \Gamma_{\mathrm{D}}, \\
\sigma \cdot n &= 0, & &\text{on } (0, T) \times \Gamma_{\mathrm{F}}, \\
\sigma \cdot n &= u, & &\text{on } (0, T) \times \Gamma_{\mathrm{N}}, \\
y &= 0, & &\text{on } \{0\} \times G, \\
\partial_t y &= 0, & &\text{on } \{0\} \times G.
\end{aligned}
\tag{5.3}
$$

Here, $a > 0$ is the weak damping constant. We add Rayleigh damping to our model by the generalized law of Hooke for the stress

$$\sigma := \sigma(y, \partial_t y; \theta) = \mathscr{C}_\theta\big(\varepsilon(y) + b\varepsilon(\partial_t y)\big),$$

where $b > 0$ is the strong damping constant, $\varepsilon(y) = \frac{1}{2}\big(\nabla y^\top + \nabla y\big)$ is the linearized strain and

$$\mathscr{C}_\theta \,:\, \varepsilon \mapsto \theta_1 \operatorname{tr}(\varepsilon)I + 2\theta_2\, \varepsilon,$$

is the fourth order elasticity tensor, see [29]. The model parameters that occur in this PDE are the Lamé constants $\theta = (\lambda_L, \mu_L) \in (0, \infty) \times (0, \infty)$. It is evident from (5.3) that the displacement $y : (0, T) \times G \mapsto \mathbb{R}^2$ is only caused by the force $u : (0, T) \times \Gamma_N \mapsto \mathbb{R}^2$ acting on the Neumann boundary.

We continue with the weak form of (5.3). Let $\mathscr{V} := \big\{v \in H^1(G)^2 : v = 0 \text{ on } \Gamma_D\big\}$ be the space of test functions and let $u \in H^1(0, T; L^2(\Gamma_N)^2)$. Furthermore, we define the parameter-dependent bilinear form $\mathscr{A}(\cdot, \cdot; \theta)$ as

$$\mathscr{A}(w, v; \theta) := \int_G \theta_1 \operatorname{tr}(\varepsilon(w)) \operatorname{tr}(\varepsilon(v)) + 2\theta_2\, \varepsilon(w) : \varepsilon(v)\, dx,$$

where $A : B = \operatorname{tr}\big(A^\top B\big)$ is the Frobenius scalar product of two matrices A and B. We call any $y \in L^2(0, T; \mathscr{V})$ with $\partial_t y \in L^2(0, T; H^1(G)^2)$ and $\partial_{tt}^2 y \in L^2(0, T; L^2(G)^2)$ a *weak solution* of the PDE (5.3) if it satisfies the variational equation

$$\varrho\left[\partial_{tt}^2 y(t), v\right]_G + a\varrho\left[\partial_t y(t), v\right]_G + b\mathscr{A}(\partial_t y(t), v; \theta) \\ + \mathscr{A}(y(t), v; \theta) = [u(t), v]_{\Gamma_N}, \tag{5.4}$$

for all $v \in \mathscr{V}$ and for almost all $t \in (0, T)$, and if it satisfies the initial conditions $y(0) = 0$, $\partial_t y(0) = 0$. Here, $[\cdot, \cdot]_G$ is the L^2 scalar product over the domain G. Equation (5.4) is obtained by testing (5.3) with $v \in \mathscr{V}$ and partial integration, cf. [29, 41]. Now, the following holds:

Proposition 5.1. Let $u \in H^1(0, T; L^2(\Gamma_N)^2)$. Then the PDE (5.3) admits a unique weak solution for almost all $t \in (0, T)$ which lies in the following function spaces:

$$y \in L^\infty(0, T; \mathscr{V}), \quad \partial_t y \in L^\infty(0, T; H^1(G)^2), \quad \partial_{tt}^2 y \in L^\infty(0, T; L^2(G)^2). \tag{5.5}$$

Proof. A more involved proof can be found in [83], see also [45, 135]. We give a simpler proof in Appendix A.7. ∎

This proposition ensures that there is no blow-up in the solution y provided that the inputs u are sufficiently regular.

In the following, we work with a finite-dimensional approximation of (5.4) known as the Galerkin ansatz. We employ standard quadratic finite elements and decompose the domain G via a piecewise affine parametrization as in [78, Sec. 5.2]. Let $d_y \in \mathbb{N}$ be the dimension of the space discretization. Then the Finite Element Method (FEM) applied to (5.4) yields

$$M\partial_{tt}^2 y(t) + C(\theta)\partial_t y(t) + A(\theta)y(t) - Nu(t) = 0, \tag{5.6}$$

where $A(\theta), M \in \mathbb{R}^{d_y \times d_y}$ are the stiffness and mass matrix, and $N \in \mathbb{R}^{d_y \times d_u}$ is the boundary mass matrix. The damping matrix $C(\theta) \in \mathbb{R}^{d_y \times d_y}$ originates from the weak damping term and the Rayleigh damping term as

$$C(\theta) := aM + bA(\theta), \qquad \text{where } a, b \geq 0. \tag{5.7}$$

The constants a and b are chosen in the way one wants to model damping. For $a > 0$, $b = 0$ we have pure mass damping, for $a = 0$, $b > 0$ we have pure stiffness damping and for $a, b > 0$ we have mixed damping. We shall mention here that both matrices, $A(\theta)$ and $C(\theta)$, have a linear dependence on the Lamé constants θ.

It is common to use a discrete numerical time update with a predefined step size $\Delta t > 0$ to solve (5.6). Renaming variables, (5.6) can be written as

$$Ma_n + C(\theta)v_n + A(\theta)d_n - Nu_n = 0,$$

with the acceleration vector $a_n = \partial_{tt}^2 y(t_n) \in \mathbb{R}^{d_y}$, the velocities $v_n = \partial_t y(t_n) \in \mathbb{R}^{d_y}$ and the displacements $d_n = y(t_n) \in \mathbb{R}^{d_y}$ for time steps $n = 1, \ldots, n_t$, respectively. The implicit Newmark scheme is a commonly used method to solve this equation, cf. [67]. We follow the implementation given by [139]. First, we choose the constants $\beta_N = \frac{1}{4}$ and $\gamma_N = \frac{1}{2}$, which makes the method equivalent to the implicit trapezoidal rule and thus unconditionally stable [67, Tab. 9.1.1], and define

$$\alpha_1 = \frac{1}{\beta_N \Delta t^2}, \qquad \alpha_2 = \frac{1}{\beta_N \Delta t}, \qquad \alpha_3 = \frac{1 - 2\beta_N}{2\beta_N},$$

$$\alpha_4 = \frac{\gamma_N}{\beta_N \Delta t}, \qquad \alpha_5 = 1 - \frac{\gamma_N}{\beta_N}, \qquad \alpha_6 = \left(1 - \frac{\gamma_N}{2\beta_N}\right)\Delta t.$$

Then the Newmark iteration scheme has the following form:

$$a_{n+1} = \alpha_1 (d_{n+1} - d_n) - \alpha_2 v_n - \alpha_3 a_n,$$
$$v_{n+1} = \alpha_4 (d_{n+1} - d_n) + \alpha_5 v_n + \alpha_6 a_n,$$
$$[\alpha_1 M + \alpha_4 C(\theta) + A(\theta)] d_{n+1} = N u_{n+1} + M(\alpha_1 d_n + \alpha_2 v_n + \alpha_3 a_n)$$
$$+ C(\theta)(\alpha_4 d_n - \alpha_5 v_n - \alpha_6 a_n),$$

(5.8)

for $n = 1, \ldots, n_t - 1$. We can write this update scheme as a matrix-vector equation:

$$L(\theta) y - F u = 0,$$

where $y = (y_1, \ldots, y_{n_t})^\top \in \mathbb{R}^{3 d_y n_t}$ are the states, with each $y_n = (a_n, v_n, d_n)^\top \in \mathbb{R}^{3 d_y}$ containing the acceleration, velocity and displacement, and $u = (u_1, \ldots, u_{n_t})^\top \in \mathbb{R}^{d_u n_t}$ are the boundary forces at all time points. Here, the matrices L and F have the block structure

$$L(\theta) = \begin{bmatrix} Q(\theta) & & & \\ P(\theta), & X(\theta) & & \\ & \ddots & \ddots & \\ & & P(\theta), & X(\theta) \end{bmatrix}, \qquad F = \begin{bmatrix} E_0 & & & \\ & E_1 & & \\ & & \ddots & \\ & & & E_1 \end{bmatrix},$$

where

$$Q(\theta) = \begin{bmatrix} M, & C(\theta), & A(\theta) \\ & I & \\ & & I \end{bmatrix}, \quad X(\theta) = \begin{bmatrix} I, & 0, & -\alpha_1 I \\ 0, & I, & -\alpha_4 I \\ 0, & 0, & D(\theta) \end{bmatrix}, \quad E_0 = \begin{bmatrix} N, \\ 0, \\ 0 \end{bmatrix}, \quad E_1 = \begin{bmatrix} 0, \\ 0, \\ N \end{bmatrix},$$

with $D(\theta) := \alpha_1 M + \alpha_4 C(\theta) + A(\theta)$ and

$$P(\theta) = \begin{bmatrix} \alpha_3 I, & \alpha_2 I, & \alpha_1 I \\ -\alpha_6 I, & -\alpha_5 I, & \alpha_4 I \\ -\alpha_3 M + \alpha_6 C(\theta), & -\alpha_2 M + \alpha_5 C(\theta), & -\alpha_1 M - \alpha_4 C(\theta) \end{bmatrix}.$$

Note that $Q(\theta)$, $X(\theta)$ and $P(\theta)$ are square matrices of order $3 d_y$, $L(\theta)$ is a square matrix of order $3 d_y n_t$ and F has $d_y n_t$ rows and $d_u n_t$ columns.

5.2.2 Problem Formulation and Adjoint Equations

Let $\Theta \subset (0, \infty) \times (0, \infty)$, $Y \subset \mathbb{R}^{3 d_y n_t}$ and $U \subset \mathbb{R}^{d_u n_t}$. The PDE constraint in the PE and OED problems is given by the operator $e : \Theta \times U \times Y \to Y$ which defines the state

equation:

$$e(\theta, u, y) := L(\theta)y - Fu = 0. \tag{5.9}$$

For given θ and u, we denote its unique solution by $y(\theta, u)$. Note that the operator e satisfies Assumption 2.12 and Assumption 3.16. We require the inputs to be constant along the Neumann boundary Γ_N such that it suffices to consider one time-dependent input vector. Thus, $d_u = 2$. Moreover, we set feasible box constraints to the input space which represent the limits of the experimenter. More concretely,

$$U := \left\{ u \in \mathbb{R}^{2n_t} : u_{\min} \le u \le u_{\max} \right\}.$$

In order to make the inputs smooth, we employ linear finite elements to discretize the time domain $(0, T)$ on which the inputs are defined. Here, we use the same grid distance Δt as in the time discretization of the state space. Let M_T and A_T be the resulting mass and stiffness matrices for this linear finite element time discretization, respectively. Then the discretized H^1-norm of an input $u \in U$ is given by

$$\|u\|_{H^1}^2 := u^\top (M_T + A_T) u.$$

We shall use this term as a regularizer in the objective function of Problem 3.17.

The sensitivity equations for $s_i = \partial_{\theta_i} y$ are given by

$$L(\theta)s_i + \partial_{\theta_i} L(\theta)y(\theta, u) = 0, \qquad i = 1, 2. \tag{5.10}$$

Equations (5.9) and (5.10) are solved using the iteration scheme (5.8). To demonstrate this, choose some $i \in \{1, 2\}$ and let $s := \partial_{\theta_i} y$ for convenience. The vector $s = (s_1, \ldots, s_{n_t})^\top \in \mathbb{R}^{3d_y n_t}$ consists of the components $s_n = (s_n^a, s_n^v, s_n^d)^\top \in \mathbb{R}^{3d_y}$, where the superscript a stands for the acceleration, v for the velocity and d for the displacement part. Then (5.10) is equivalent to

$$s_{n+1}^a = \alpha_1 \left(s_{n+1}^d - s_n^d \right) - \alpha_2 s_n^v - \alpha_3 s_n^a,$$
$$s_{n+1}^v = \alpha_4 \left(s_{n+1}^d - s_n^d \right) + \alpha_5 s_n^v + \alpha_6 s_n^a,$$
$$\left[\alpha_1 M + \alpha_4 C(\theta) + A(\theta) \right] s_{n+1}^d = M \left(\alpha_1 s_n^d + \alpha_2 s_n^v + \alpha_3 s_n^a \right) + C(\theta) \left(\alpha_4 s_n^d - \alpha_5 s_n^v - \alpha_6 s_n^a \right)$$
$$- \alpha_4 \partial_{\theta_i} C(\theta) d_{n+1} - \partial_{\theta_i} A(\theta) d_{n+1} + \partial_{\theta_i} C(\theta) \left[\alpha_4 d_n - \alpha_5 v_n - \alpha_6 a_n \right].$$

We solve Problem 3.17 with $\Psi = \Psi_E$, $C = C_{GN}$ and $R(u) = \|u\|_{H^1}^2$ to obtain optimal

sensor positions $\bar{\omega}$ and optimal input configurations $\bar{u}$:

$$\min_{\omega,u,y,s} \Psi[C_{\mathrm{GN}}(\theta,\omega,y,s)] + \kappa\mathscr{P}_\varepsilon(\omega) + \beta u^\top(M_{\mathrm{T}} + A_{\mathrm{T}})u$$

$$\begin{aligned} \text{s.t.} \quad & 0 = L(\theta)y - Fu, \\ & 0 = L(\theta)s_i + \partial_{\theta_i}L(\theta)y, \quad i = 1,2, \\ & \omega \in [0,1]^{n_\omega}, \ u \in U. \end{aligned} \tag{5.11}$$

The adjoint equations (3.26) are the following:

$$L(\theta)^\top\mu + \partial_{\theta_1}L(\theta)^\top\lambda_1 + \partial_{\theta_2}L(\theta)^\top\lambda_2 + \nabla_y\Psi = 0, \tag{5.12}$$

$$L(\theta)^\top\lambda_1 + \nabla_{s_1}\Psi = 0, \tag{5.13}$$

$$L(\theta)^\top\lambda_2 + \nabla_{s_2}\Psi = 0, \tag{5.14}$$

where $\nabla_y\Psi$ and $\nabla_{s_i}\Psi$ are understood as Clarke subdifferentials, see Proposition A.7 and Appendix A.5. These equations can be rewritten by using the iteration scheme above (5.8). The matrix $L(\theta)$ is transposed on the left hand side of (5.12)–(5.14) which leads to a backwards iteration. We show how to do this for equations (5.13) and (5.14), see also [78, Sec. 5.4]. For convenience, let $\lambda = \lambda_i$ and $r := -\nabla_{s_i}\Psi$ be the right-hand side in the equations for $i \in \{1,2\}$. In the terminal point $t_{n_{\mathrm{t}}}$ we solve

$$X(\theta)^\top\lambda_{n_{\mathrm{t}}} = r_{n_{\mathrm{t}}},$$

or equivalently $\lambda_{n_{\mathrm{t}}}^a = 0$, $\lambda_{n_{\mathrm{t}}}^v = 0$ and

$$-\alpha_1\lambda_{n_{\mathrm{t}}}^a - \alpha_4\lambda_{n_{\mathrm{t}}}^v + D(\theta)\lambda_{n_{\mathrm{t}}}^d = r_{n_{\mathrm{t}}}^d.$$

For points t_n, $n > 1$ the current iterate is computed using values from a step forward in time:

$$\left[X(\theta)^\top, P(\theta)^\top\right]\begin{pmatrix}\lambda_n \\ \lambda_{n+1}\end{pmatrix} = r_n,$$

or equivalently

$$\lambda_n^a = \alpha_3\lambda_{n+1}^a + \alpha_6\lambda_{n+1}^v + [\alpha_3 M - \alpha_6 C(\theta)]\lambda_{n+1}^d,$$

$$\lambda_n^v = \alpha_2\lambda_{n+1}^a + \alpha_5\lambda_{n+1}^v + [\alpha_2 M - \alpha_5 C(\theta)]\lambda_{n+1}^d,$$

$$-\alpha_1\lambda_n^a - \alpha_4\lambda_n^v + D(\theta)\lambda_n^d = r_n^d - \alpha_1\lambda_{n+1}^a - \alpha_4\lambda_{n+1}^v + [\alpha_1 M + \alpha_4 C(\theta)]\lambda_{n+1}^d.$$

In order to obtain the adjoint variable λ at the first time point t_1 we solve

$$\left[Q(\theta)^\top, P(\theta)^\top\right]\binom{\lambda_1}{\lambda_2} = r_1,$$

or equivalently

$$M\lambda_1^a = \alpha_3\lambda_2^a + \alpha_6\lambda_2^v + [\alpha_3 M - \alpha_6 C(\theta)]\lambda_2^d,$$
$$C(\theta)\lambda_1^a + \lambda_1^v = \alpha_2\lambda_2^a + \alpha_5\lambda_2^v + [\alpha_2 M - \alpha_5 C(\theta)]\lambda_2^d,$$
$$A(\theta)\lambda_1^a + \lambda_1^d = r_1^d - \alpha_1\lambda_2^a - \alpha_4\lambda_2^v + [\alpha_1 M + \alpha_4 C(\theta)]\lambda_2^d.$$

We solve the optimization problem (5.11) by an SQP-method with BFGS-updates [129] and initialize the part of the Hessian of the Lagrangian that belongs to the inputs with

$$\beta R''(u) = 2\beta(M_\mathrm{T} + A_\mathrm{T})$$

which is known to speed up convergence.

5.2.3 Numerical Results

Our test example is a two-dimensional truss consisting of nine beams and six connectors. The spatial finite element discretization has about 5 000 degrees of freedom. The Dirichlet boundary part Γ_D is positioned at the two outer top connectors and the Neumann boundary part Γ_N on the bottom left connector as seen in Figure 5.4. We employ pairs of strain gauges that can measure either the axial deflection or the displacement caused by bending of the beams, cf. [65] and [110]. The strain gauges are located on a subset of the free boundary part Γ_F, i.e., on the upper and lower boundaries of each beam, indicated as black bullets and connecting lines in the figure. The strain gauges measure the relative displacement of two adjacent finite element nodes:

$$\varepsilon_\mathrm{u} = y_\mathrm{N1} - y_\mathrm{N2} \qquad \text{and} \qquad \varepsilon_\ell = y_\mathrm{N3} - y_\mathrm{N4},$$

see Figure 5.4a.

However, It is easier to work with the square of the axial deflection $h_\mathrm{a}(y)$ and the square of the displacement caused by bending $h_\mathrm{b}(y)$ in order to keep the following computations simple:

$$h_\mathrm{a}(y) = \frac{1}{4}\|\varepsilon_\mathrm{u} + \varepsilon_\ell\|^2 \qquad \text{and} \qquad h_\mathrm{b}(y) = \frac{1}{4}\|\varepsilon_\mathrm{u} - \varepsilon_\ell\|^2.$$

(a)

(b)

(c)

(d)
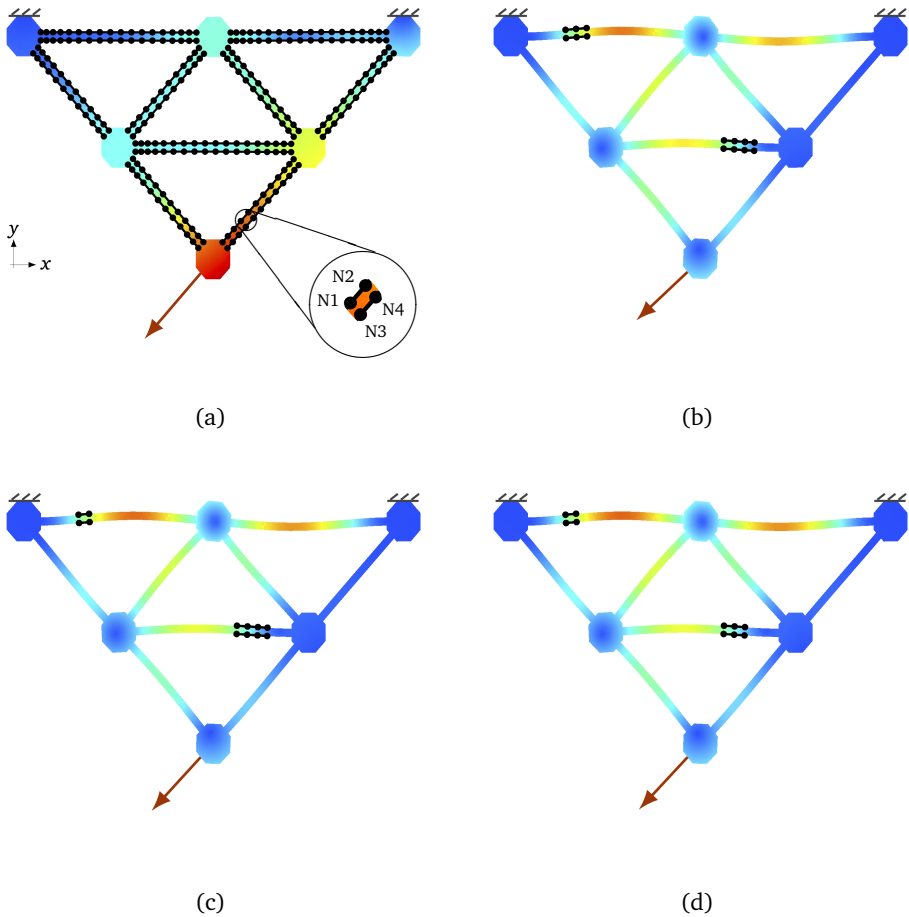
Figure 5.4 (a) snapshot of the dynamic behavior of the truss with all possible positions for strain gauges marked as bullets with connecting lines; the input boundary force is displayed by a red arrow, (b)–(d) snapshot at the time of maximal displacement after (5.11) has been solved with three different values for the penalty parameter: $\kappa_1 = 8, \kappa_2 = 12, \kappa_3 = 18$; based on [93].

The observation operator h is composed of h_a and h_b in a row, for all possible pairs of strain gauges and *for all time points*. We also create for each such sensor five weight variables which give us information about the importance of a strain gauge position. In our example we have 117 candidate sensor locations in total. Hence, there are $n_\omega = 117 \times 2 \times 5 = 1\,170$ weight variables ω_k.

In a study conducted by Alipour and Zareian [5], which makes reference to [101], it is suggested to perform such numerical simulations with pure stiffness

damping, i.e., $a = 0$ in (5.7). We adopt this approach since we did not collect real experimental data; we only simulate the acquisition of measurements. The accuracy of the strain gauge sensors is fixed to $\sigma_k = 1 \cdot 10^{-4} \, \text{mm}^2$ for $k = 1, \dots, n_\omega$. Note that the unit of the standard deviation σ_k is mm^2 since we consider the squared relative displacement in the observation operator.

In our numerical simulations of the dynamic truss, we choose $n_t = 600$ time steps with a step size of $\Delta t = 5 \, \text{ms}$. As a consequence, the solution of the state equation (5.9) involves 3 000 000 degrees of freedom. Our initial point $(\omega_0, u_0(t))$ for (5.11) is chosen in such a way that all weight variables ω_0 are set to one and $u_0(t) = u_0$ is the maximally feasible, constant-in-time boundary force within the box constraints.

Our model $\mathcal{M}$ consists of the operators $e(\theta, u, y), h(y)$ and the state equation (5.9). It operates under the assumption that all beams have equal cross-sectional area in the stress-free state. On the other hand, the real model $\mathcal{R}$ is almost identical to $\mathcal{M}$ except that two beams have a 5% and a 7% smaller diameter, respectively. Since we were unable to acquire actual measurements, we generated simulated Gaussian data on the computer with standard deviation σ_k and mean given by the output of the real model $\mathcal{R}$. In order to detect model uncertainty it is not important to know which beams differ from each other. Note that a beam diameter is not a single geometrical parameter but it significantly impacts the finite element terms in the mass, damping and stiffness matrices, see Equation (5.6). In the sequel, we apply Algorithm 4.1 to see if it detects model uncertainty in $\mathcal{M}$ when compared to data obtained from $\mathcal{R}$.

Since our data were generated on the computer as described above, line 6 in Algorithm 4.1 became obsolete. Moreover, we used textbook values for θ_{ini}, namely, the Lamé-constants for steel $\lambda_{\text{L}} = 121\,154 \, \text{N/mm}^2$ and $\mu_{\text{L}} = 80\,769 \, \text{N/mm}^2$ making z_{ini} in line 2 unnecessary. These values for $(\lambda_{\text{L}}, \mu_{\text{L}})$ are also used to generate the output of the real model $\mathcal{R}$. The optimization problem (5.11) has been solved after about 80 iterations where the computation time was observed to be less than eight hours on an AMD EPYC 48×2.8 GHz machine. As seen in Table 5.9, the value of the design criterion decreased by several magnitudes when compared to the initial setup. This improvement is significant and it is mainly caused by the smart choice of the input force $u(t)$ which excites the truss in such a way as to get the maximum information gain for the values of the model parameters. Moreover, the more sensors are involved in the final design the smaller becomes the largest principal axis of the confidence region which is achieved for small κ.

In Figure 5.4b–d the final position of the strain gauges is depicted in a snap shot. We observe that there are two main locations where a cluster of optimal sensors is

Table 5.9 Outcome of the OED problem (5.11) for $\theta = \theta_{\mathrm{ini}}$ and different $\kappa > 0$.

κ	β	$|\overline{\omega}|_0$	$\Psi_{\mathrm{E}}[C(\theta_{\mathrm{ini}}, \omega_0, u_0(t))]$	$\Psi_{\mathrm{E}}[C(\theta_{\mathrm{ini}}, \overline{\omega}, \overline{u}(t))]$
8	$8 \cdot 10^{-6}$	25	$2.177 \cdot 10^5$	$3.259 \cdot 10^2$
12	$8 \cdot 10^{-6}$	20	$2.177 \cdot 10^5$	$3.568 \cdot 10^2$
18	$8 \cdot 10^{-6}$	15	$2.177 \cdot 10^5$	$3.822 \cdot 10^2$

Table 5.10 Results for the hypothesis tests from Algorithm 4.1 with $C = C_{\mathrm{GN}}$ and $\kappa = 8$.

#	θ_{cal}	θ_{val}	$\|\theta_{\mathrm{cal}} - \theta_{\mathrm{val}}\|$	$\lambda_{\min}(C_{\mathrm{cal}} + C_{\mathrm{val}})^{-1}$	$\alpha_{\min}$ in %
1	$2.1373 \cdot 10^{-6}$	$1.2068 \cdot 10^5$	$1.254 \cdot 10^5$	$7.441 \cdot 10^3$	$\ll 0.001$
	$1.1257 \cdot 10^5$	$7.8682 \cdot 10^4$			
2	$7.8721 \cdot 10^{-5}$	$1.2069 \cdot 10^5$	$1.254 \cdot 10^5$	$7.431 \cdot 10^3$	$\ll 0.001$
	$1.1257 \cdot 10^5$	$7.8676 \cdot 10^4$			
3	$2.0892 \cdot 10^{-6}$	$1.2069 \cdot 10^5$	$1.254 \cdot 10^5$	$7.432 \cdot 10^3$	$\ll 0.001$
	$1.1257 \cdot 10^5$	$7.8676 \cdot 10^4$			
4	$2.0811 \cdot 10^{-6}$	$1.2069 \cdot 10^5$	$1.254 \cdot 10^5$	$7.429 \cdot 10^3$	$\ll 0.001$
	$1.1257 \cdot 10^5$	$7.8676 \cdot 10^4$			

Table 5.11 Results for the hypothesis tests from Algorithm 4.1 with $C = C_{\mathrm{GN}}$ and $\kappa = 12$.

#	θ_{cal}	θ_{val}	$\|\theta_{\mathrm{cal}} - \theta_{\mathrm{val}}\|$	$\lambda_{\min}(C_{\mathrm{cal}} + C_{\mathrm{val}})^{-1}$	$\alpha_{\min}$ in %
1	$5.0751 \cdot 10^{-4}$	$1.2075 \cdot 10^5$	$1.254 \cdot 10^5$	$6.419 \cdot 10^3$	$\ll 0.001$
	$1.1298 \cdot 10^5$	$7.9050 \cdot 10^4$			
2	$1.1156 \cdot 10^{-8}$	$1.2075 \cdot 10^5$	$1.254 \cdot 10^5$	$6.403 \cdot 10^3$	$\ll 0.001$
	$1.1298 \cdot 10^5$	$7.9056 \cdot 10^4$			
3	$8.8314 \cdot 10^{-9}$	$1.2075 \cdot 10^5$	$1.254 \cdot 10^5$	$6.404 \cdot 10^3$	$\ll 0.001$
	$1.1298 \cdot 10^5$	$7.9056 \cdot 10^4$			
4	$8.8215 \cdot 10^{-9}$	$1.2075 \cdot 10^5$	$1.254 \cdot 10^5$	$6.396 \cdot 10^3$	$\ll 0.001$
	$1.1298 \cdot 10^5$	$7.9056 \cdot 10^4$			

Table 5.12 Results for the hypothesis tests from Algorithm 4.1 with $C = C_{\mathrm{GN}}$ and $\kappa = 18$.

#	θ_{cal}	θ_{val}	$\|\theta_{\mathrm{cal}} - \theta_{\mathrm{val}}\|$	$\lambda_{\min}(C_{\mathrm{cal}} + C_{\mathrm{val}})^{-1}$	$\alpha_{\min}$ in %
1	$1.5694 \cdot 10^{-6}$	$1.2072 \cdot 10^5$	$1.254 \cdot 10^5$	$5.933 \cdot 10^3$	$\ll 0.001$
	$1.1268 \cdot 10^5$	$7.8769 \cdot 10^4$			
2	$1.5589 \cdot 10^{-6}$	$1.2072 \cdot 10^5$	$1.254 \cdot 10^5$	$5.938 \cdot 10^3$	$\ll 0.001$
	$1.1268 \cdot 10^5$	$7.8769 \cdot 10^4$			
3	$4.3249 \cdot 10^{-5}$	$1.2072 \cdot 10^5$	$1.254 \cdot 10^5$	$5.938 \cdot 10^3$	$\ll 0.001$
	$1.1268 \cdot 10^5$	$7.8769 \cdot 10^4$			
4	$3.2828 \cdot 10^{-5}$	$1.2072 \cdot 10^5$	$1.254 \cdot 10^5$	$5.932 \cdot 10^3$	$\ll 0.001$
	$1.1268 \cdot 10^5$	$7.8769 \cdot 10^4$			

(a)

(b)
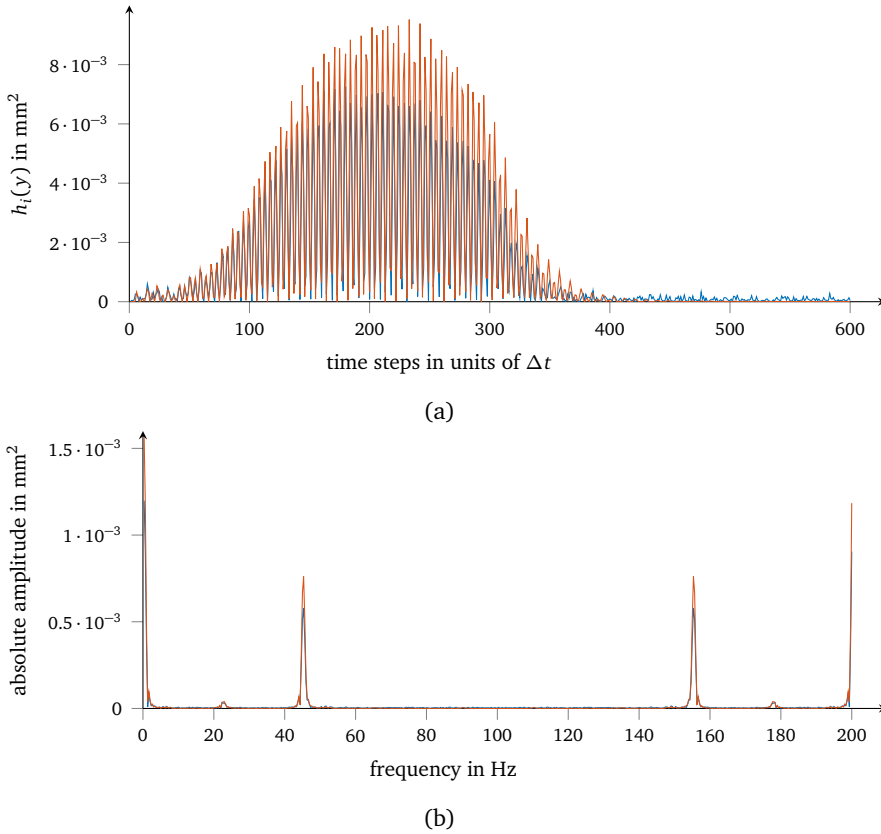
Figure 5.5 (a) comparison model output $h_i(y)$ in red for some sensor i and data z_i in blue, (b) amplitude spectra, i.e., fast Fourier transform, of the model output $h_i(y)$ and the data z_i.

formed: one on the upper left beam and one on the lower central beam. The cluster is only formed if all five weight variables for a sensor position are switched to one which gives evidence to the fact that there are certain positions that are clearly favorable. Clustering is thus a secondary effect and its extent is diminished if the value of κ is slowly increased. However, a cluster design may be valuable evidence for the engineer to couple the sensors right there or to use longer strain gauges if possible. This would even more improve the information gain.

For the hypothesis test in Algorithm 4.1, let $\bar{u}(t)$ be the optimal input force. In order to generate test instances, consider the following perturbed inputs:

$$u_1(t) = \bar{u}(t) + \delta_1, \qquad u_2(t) = \bar{u}(t) + 4\sin(2\pi t/n_t),$$
$$u_3(t) = \bar{u}(t) + 4\cos(2\pi t/n_t), \qquad u_4(t) = \bar{u}(t) + \delta_2(t),$$

where $\delta_1 \sim \mathcal{N}(0, 4 \cdot I)$ and $\delta_2(t) \sim \mathcal{N}(0, 4t/n_t \cdot I)$ are Gaussian processes with discrete time points $t \in \{k \Delta t \; : \; k = 1, \dots, n_t\}$ that have the step size Δt as introduced before.

These four inputs are used to generate four different data sets and we employ a 4-fold cross-validation as our division strategy. Thus, $n_{\text{tests}} = 4$ hypothesis tests were conducted for three values of κ, respectively, and the results are shown in Table 5.10–Table 5.12. We clearly observe that the suggested model $\mathcal{M}$ is rejected in all test instances when a threshold of $\text{TOL}/n_{\text{tests}} = 1.25\%$ is applied to $\alpha_{\min}$. According to Assumption 4.3 this is a significant indication of model uncertainty. What is more, the parameter estimates obtained from the calibration set have unreasonable values. We even observed bad convergence properties of the Gauss-Newton method in this case. Yet, this is evidence that the residuals are large, see Subsection 2.2.2.

A graphical comparison, however, either between the model output and the data as seen in Figure 5.5a or between their amplitude spectra via the fast Fourier transform, see Figure 5.5b, does not yield such clear evidence for model uncertainty. The difference only prevails in the amplitude's magnitude. A definite, quantifiable statement about model uncertainty is difficult to justify with this approach.

Conclusion and Outlook

In this thesis we developed an algorithm to identify model uncertainty which uses methods from parameter estimation (PE), optimum experimental design (OED) and statistical hypothesis testing. We showed that there are five different approaches based on two probabilistic perspectives to estimate model parameters from data. Each of these estimates and their covariance matrix can be used to define a confidence region in which it is very likely to find the true parameter values. We then introduced the classical theory of optimal design of experiments and its relaxation formulations. Since this approach is insufficient in the presence of PDE constraints, we introduced a modern approach to OED which returns optimal sensor weights and an optimal input configuration such that a design criterion of the parameter's covariance becomes minimal. We also incorporated a penalty term to control the costs of used sensors and a regularization to smooth the inputs. The solution to this problem can be found with an iterative solver scheme, such as an SQP-solver with BFGS-updates, for which we determined the adjoint equations. Moreover, we relied on the assumption that the true values of the model parameters lie in this small confidence region which was minimized by solving the OED problem. If the mathematical model is correct, then repeated calibration and validation procedures should yield parameter values in this confidence region. However, if a set of measurements leads to estimates that lie outside of this confidence region, we concluded that the model is incorrect.

We demonstrated the success of our algorithm to identify model uncertainty in mathematical models of the 3D Servo Press and in the linear-elastic model of vibrations in a truss. The press models were distinguished by their friction model for which we introduced three different approaches: no friction at all, a discontinuous Coulomb model and a continuous model with rate-independent memory trained by an artificial neural network. We saw that our algorithm correctly selects the third

model and rejects the others. Moreover, the Bayesian perspective with weak prior convictions is suitable to cope with regularity issues of the covariance matrix in the OED problem. However, we clearly suggest to abstain from strong prior convictions since these yielded erroneous conclusions. Our results are of strategic importance when conducting future experiments on the full-scale 3D Servo Press.

For the truss, we showed that our algorithm is able to detect a faulty model that is deficient in its geometrical description and quantify model uncertainty by the probability for the Type I error. Moreover, we significantly increased the information gain for the parameters to be estimated by finding optimal excitation forces and optimal sensor positions. A direct graphical comparison between the Fourier transform of the model's output and the data only results in a difference in amplitudes but not in frequencies. However, after an examination of the parameter estimates obtained from the calibration data set, one can also infer that the model must be inadequate.

To give an outlook, our approach to optimal input configuration with PDE constraints can be extended to large-scale problems where the dimension of the parameter space is large. Adopting the notation from Section 3.3 we follow [4] and use n_{tr} randomized trace estimators for the inverse Hessian C_{AB}. This Hessian depends on the data and the parameters are obtained by solving (2.16)–(2.18) as an inner optimization problem. An update of the input variables $u_{i+1} = u_i + \Delta u_i$ yields a model output that mismatches the data resulting in unreasonable parameter values. To deal with this difficulty one can update the data $z(u_{i+1})$ in the following way. First, extract the initial measurement error $\varepsilon_0 = h(y(\overline{\theta}_0, u_0), \overline{\theta}_0) - z(u_0)$ once, using the first fitted parameter value $\overline{\theta}_0$, and then update the data vector:

$$z(u_{i+1}) := h(y_z(\overline{\theta}_0, u_{i+1}), \overline{\theta}_0) + \varepsilon_0,$$

where $y_z(\overline{\theta}_0, u_{i+1})$ solves the state equation

$$e(\overline{\theta}_0, u_{i+1}, y_z) = 0.$$

Furthermore, $z(u_{i+1})$ is inserted into the objective function $g = g(\theta, y; z(u_{i+1}))$ of the PE problem. Let $\tau_i \sim \mathcal{N}(0, I)$ for $i = 1, \ldots, n_{\text{tr}}$ be given. Moreover, let $\kappa, \beta > 0$, $\mathscr{P}_\varepsilon(\omega)$ be a smooth approximation of the ℓ_0-"norm" and $R(u)$ be a smooth regularizer. We suggest to solve the following OED problem

$$\min_{\omega, u, y, y_z, \lambda, \theta, v_i, q_i, \xi_i} \sum_{i=1}^{n_{\text{tr}}} \tau_i^\top \xi_i + \kappa \mathscr{P}_\varepsilon(\omega) + \beta R(u),$$

where $y, y_z, \lambda, \theta, v_i, q_i, \xi_i$ satisfy the equality constraints

$$e(\theta, u, y) = 0,$$
$$e(\theta_0, u, y_z) = 0,$$
$$\nabla_y g + \partial_y e^\top \lambda = 0,$$
$$\nabla_\theta g + \partial_\theta e^\top \lambda = 0,$$
$$\partial_y e\, q_i + \partial_\theta e\, \xi_i = 0,$$
$$\partial_y e^\top v_i + \left\langle \partial_{yy}^2 e^\top, (\lambda, q_i) \right\rangle + \partial_{yy}^2 g\, q_i + \left\langle \partial_{y\theta}^2 e^\top, (\lambda, \xi_i) \right\rangle + \partial_{y\theta}^2 g\, \xi_i = 0,$$
$$\partial_\theta e^\top v_i + \partial_{\theta y}^2 g\, q_i + \left\langle \partial_{\theta y}^2 e^\top, (\lambda, q_i) \right\rangle + \partial_{\theta\theta}^2 g\, \xi_i + \left\langle \partial_{\theta\theta}^2 e^\top, (\lambda, \xi_i) \right\rangle = \tau_i,$$

for $i = 1, \ldots, n_{\text{tr}}$ and (ω, u) satisfy the inequality constraints

$$\omega \in [0, 1]^{n_\omega}, \qquad u \in U.$$

The number of PDEs to be solved in each step is $3 + 2n_{\text{tr}}$ which is independent of the dimension of the parameter space. As in [4], one can develop an adjoint calculus to apply gradient-based optimization algorithms. The success of this approach, however, still needs to be investigated in further research.

Appendix

A.1 Basic Matrix Calculus

We start by citing and proving a few results on matrix inversion.

Lemma A.1. The set of invertible matrices $\mathrm{GL}_n(\mathbb{R}) := \{A \in \mathbb{R}^{n \times n} : A \text{ invertible}\}$ is open in $\mathbb{R}^{n \times n}$ and the inversion mapping

$$\mathrm{inv} : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R}^{n \times n}, A \mapsto A^{-1}$$

is continuous.

Proof. We note that $\det(\cdot) : \mathrm{GL}_n(\mathbb{R}) \to \mathbb{R} \setminus \{0\}$ is continuous and that $\mathbb{R} \setminus \{0\}$ is open. Thus, $\det^{-1}(\mathbb{R} \setminus \{0\}) = \mathrm{GL}_n(\mathbb{R})$ is open in $\mathbb{R}^{n \times n}$ as well. For the continuity of the inversion mapping see [123]. ∎

Definition A.2. Let $D \subset \mathbb{R}^n$ be open and a function $f : D \to \mathbb{R}$ be given. The *directional derivative* of f at $x \in D$ in the direction $d \in \mathbb{R}^n$ is the function

$$\mathbb{R}^n \ni d \mapsto \langle \partial_x f(x), d \rangle := \lim_{t \searrow 0} \frac{f(x + td) - f(x)}{t},$$

if this limit exists.

Lemma A.3. Let $A \in \mathrm{GL}_n(\mathbb{R})$. Then the directional derivative of $\mathrm{inv} : A \mapsto A^{-1}$ in the direction of a matrix $\Delta A \in \mathbb{R}^{n \times n}$ is given by

$$\langle \partial_A \mathrm{inv}(A), \Delta A \rangle = -A^{-1} \cdot \Delta A \cdot A^{-1}.$$

Proof. We compute the limit

$$\langle \partial_A \operatorname{inv}(A), \Delta A \rangle := \lim_{t \searrow 0} \frac{\operatorname{inv}(A + t\Delta A) - \operatorname{inv}(A)}{t} = \lim_{t \searrow 0} \frac{(A + t\Delta A)^{-1} - A^{-1}}{t}$$

$$= A^{-1} \lim_{t \searrow 0} \frac{(I + t\Delta A \cdot A^{-1})^{-1} - I}{t}$$

directly. If t becomes sufficiently small, then the Neumann series $\sum_{k=0}^{\infty} (-t\Delta A \cdot A^{-1})^k$ converges to $(I + t\Delta A \cdot A^{-1})^{-1}$, see [136, Thm. II.1.11]. We thus have

$$(I + t\Delta A \cdot A^{-1})^{-1} - I = \sum_{k=0}^{\infty} (-t\Delta A \cdot A^{-1})^k - I = I - t\Delta A \cdot A^{-1} + o(t) - I$$

for small t and it follows

$$\langle \partial_A \operatorname{inv}(A), \Delta A \rangle = A^{-1} \lim_{t \searrow 0} \frac{-t\Delta A \cdot A^{-1} + o(t)}{t} = -A^{-1} \cdot \Delta A \cdot A^{-1}. \qquad \blacksquare$$

We also compute some directional derivatives of common matrix functions.

Lemma A.4. (a) Let $A, \Delta A \in \mathbb{R}^{n \times n}$. Then $\langle \partial_A \operatorname{tr}(A), \Delta A \rangle = \operatorname{tr}(\Delta A)$.
  (b) Let $A \in \operatorname{GL}_n(\mathbb{R})$ and $\Delta A \in \mathbb{R}^{n \times n}$. Then we have

$$\langle \partial_A \det(A), \Delta A \rangle = \det(A) \cdot \operatorname{tr}(A^{-1} \cdot \Delta A).$$

Proof. Statement (a) follows directly from the definition and the fact that $\operatorname{tr}(\cdot)$ is linear. For (b) we refer to [90]. $\qquad \blacksquare$

The next statements need some preparation; we follow [64]. Let $S_n(\mathbb{R})$ be the set of real symmetric matrices with inner product $\langle A, B \rangle = \operatorname{tr}(A \cdot B)$ for $A, B \in S_n(\mathbb{R})$. For $m = 1, \dots, n$ denote the m-th largest eigenvalue of $A \in S_n(\mathbb{R})$ by $\lambda_m(A)$. We further define

$$\widehat{m} := \min\{i : \lambda_i(A) = \lambda_m(A)\} \qquad \text{and} \qquad \overline{m} := \max\{i : \lambda_i(A) = \lambda_m(A)\}.$$

Thus, it is evident that

$$\lambda_{\widehat{m}-1}(A) > \lambda_{\widehat{m}}(A) = \dots = \lambda_m(A) = \dots = \lambda_{\overline{m}}(A) > \lambda_{\overline{m}+1}(A)$$

which can easily be reformulated for the cases $\widehat{m} = 1$ and $\overline{m} = n$. We now fix an

orthogonal matrix U that diagonalizes A such that

$$U^\top \cdot A \cdot U = \text{Diag}(\lambda_1(A), \dots, \lambda_n(A)).$$

Furthermore, let U_m be the submatrix of U consisting of the columns $\widehat{m}, \dots, \overline{m}$.

Lemma A.5. Let A, $\Delta A \in S_n(\mathbb{R})$. Then the following holds:

$$\langle \partial_A \lambda_m(A), \Delta A \rangle = \lambda_{m-\widehat{m}+1}(U_m^\top \cdot \Delta A \cdot U_m). \tag{A.1}$$

Proof. See [64, Thm. 3.3] and the references therein. ∎

In general, the function $\lambda_m(A)$ may be nonsmooth especially if there are multiple eigenvalues. This is a direct consequence of the implicit function theorem. It follows that the directional derivative in (A.1) may become instable. By introducing generalized subdifferentials this derivative can be regularized.

Definition A.6. Let $D \subset \mathbb{R}^n$ and $f : D \to \mathbb{R}$ be locally Lipschitz. Then the *Clarke directional derivative* at $x \in D$ is the mapping

$$\mathbb{R}^n \ni d \mapsto \langle f^\circ(x), d \rangle := \limsup_{t \searrow 0, \, y \to x} \frac{f(y + td) - f(y)}{t}.$$

If this function coincides with the regular directional derivative at x, then we say that f is *Clarke regular* at x. The *Clarke subdifferential* at x is the nonempty, convex and compact set

$$\partial^{\text{cl}} f(x) := \{s \in \mathbb{R}^n : s^\top d \leq \langle f^\circ(x), d \rangle \text{ for all } d \in \mathbb{R}^n\}.$$

Knowing that $\lambda_m(A)$ is Lipschitz, we consider the following result:

Proposition A.7. For any matrix $A \in S_n(\mathbb{R})$ with eigenspace $E_m(A) \subset \mathbb{R}^n$ corresponding to the m-th largest eigenvalue $\lambda_m(A)$ we have

$$\partial^{\text{cl}} \lambda_m(A) = \text{conv}\{x \cdot x^\top : x \in E_m(A), \|x\| = 1\}. \tag{A.2}$$

Proof. See [64, Thm. 5.3]. ∎

Following [86], Equation (A.2) can be used to write

$$\langle \lambda_m^\circ(A), \Delta A \rangle = \max\{x^\top \cdot \Delta A \cdot x : x \in E_m(A), \|x\| = 1\}. \tag{A.3}$$

A.2 Transformations of Multivariate Normal Distributions

In this section we briefly mention a few results on transformations of multivariate Gaussian random variables.

Proposition A.8. Let $X \sim \mathcal{N}(\mu, \Sigma)$ be a multivariate Gaussian random variable where Σ is non-singular, A be a matrix with full row-rank and c be a vector. Then the random variable $Z = AX + c$ is also Gaussian with $Z \sim \mathcal{N}(A\mu + c, A\Sigma A^\top)$.

Proof. See [126]. ∎

Lemma A.9. Let $X \sim \mathcal{N}(\mu_1, \Sigma_1)$ and $Y \sim \mathcal{N}(\mu_2, \Sigma_2)$ be two independent multivariate Gaussian random variables. Then

$$\mathbb{E}[X \pm Y] = \mu_1 \pm \mu_2 \quad \text{and} \quad \text{Var}[X \pm Y] = \Sigma_1 + \Sigma_2.$$

Proof. Let $Z = (X, Y)^\top$. Then $\mathbb{E}[Z] = (\mu_1, \mu_2)^\top$ and

$$\text{Var}[Z] = \begin{pmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{pmatrix},$$

since $\text{Cov}[X, Y] = 0$. Let $A = (I, \pm I)$ and $\xi := AZ$. By Proposition A.8 we have

$$\mathbb{E}[\xi] = \mu_1 \pm \mu_2 \quad \text{and} \quad \text{Var}[\xi] = \Sigma_1 + \Sigma_2.$$ ∎

A.3 A Few Concepts from Asymptotic Distribution Theory

We briefly present a few notions from asymptotic distribution theory following [112]. In this section, let $(X_n)_{n \in \mathbb{N}}$ be a sequence of independent, *identically* distributed univariate random variables.

Definition A.10. Let $F_{X_n}(\cdot)$ be the CDF of the random variable X_n. The sequence $(X_n)_{n \in \mathbb{N}}$ *converges in distribution* to X if $(F_{X_n}(y))_{n \in \mathbb{N}}$ converges to some CDF $F_X(y)$ at all points y where $F_X(y)$ is continuous.

The case where X is a constant often gets special attention:

Definition A.11. The sequence $(X_n)_{n \in \mathbb{N}}$ of random variables *converges in probability* to the constant X if for every $\varepsilon, \delta > 0$ there exists some $n_0(\varepsilon, \delta) \in \mathbb{N}$ such

that

$$P[|X_n - X| < \varepsilon] > 1 - \delta, \qquad \text{for all } n > n_0(\varepsilon, \delta).$$

The law of large numbers states that a sample average converges in probability to its mean under certain conditions:

Proposition A.12. Let $(X_n)_{n \in \mathbb{N}}$ be the sequence of random variables as before such that $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}[X_n] = \sigma^2$ exist for all $n \in \mathbb{N}$. Then $\left(\frac{1}{n}\sum_{i=1}^n X_i\right)_{n \in \mathbb{N}}$ converges in probability to μ as $n \to \infty$.

Proof. See [112, Thm. 8, p. 262]. ∎

We also mention the well-known central limit theorem:

Proposition A.13. Let $(X_n)_{n \in \mathbb{N}}$ be the sequence of random variables as before such that $\mathbb{E}[X_n] = \mu$ and $\mathrm{Var}[X_n] = \sigma^2 < \infty$ with $\sigma > 0$. Then $(Y_n)_{n \in \mathbb{N}}$ defined as

$$Y_n := \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^n X_i - \mu\right)$$

converges in distribution to $Y \sim \mathcal{N}(0,1)$ as $n \to \infty$.

Proof. See [112, Thm. 9, p. 265]. ∎

All these notions, the law of large numbers and the central limit theorem can be generalized to the case of multivariate random variables, see [112, Lem. 13.5] for example.

A.4 Derivatives in Newton's Method for PE

Following up Subsection 2.2.1, we compute $J(\theta; u)$ and $S(\theta; z, u)$ which are used in Equation (2.14) as we did in [50]. We derive the computation for $m = n_z = 1$ only in order to avoid complicated indices. Let $y(\theta; u)$ be the solution of the state equation (2.9) for given $\theta \in \Theta$ and $u \in U$ and let h be the observation operator. Then

$$\partial_\theta \eta_i(\theta, u) = \partial_y h_i(y(\theta; u), \theta)\partial_\theta y + \partial_\theta h_i(y(\theta; u), \theta),$$
$$\partial_{\theta\theta}^2 \eta_i(\theta, u) = \partial_\theta y^\top \partial_{yy}^2 h_i(y(\theta; u), \theta)\partial_\theta y + \partial_{\theta\theta}^2 h_i(y(\theta; u), \theta)$$
$$+ 2\partial_{y\theta}^2 h_i(y(\theta; u), \theta)\partial_\theta y + \partial_y h_i(y(\theta; u), \theta)\partial_{\theta\theta}^2 y.$$

To determine the expressions $\partial_\theta y$ and $\partial^2_{\theta\theta} y$, we apply the Implicit Function Theorem using Assumption 2.12:

$$\partial_y e(\theta, u, y) \partial_\theta y(\theta; u) = -\partial_\theta e(\theta, u, y).$$

In order to compute the vector-tensor product $\partial_y h_i(y(\theta; u), \theta) \partial^2_{\theta\theta} y$ efficiently we want to avoid the direct calculation of $\partial^2_{\theta\theta} y(\theta; u)$ whose directional derivatives in direction $(v_1, v_2) \in Y \times Y$ are given by

$$\partial_y e(\theta, u, y) \left\langle \partial^2_{\theta\theta} y(\theta; u), (v_1, v_2) \right\rangle = -\left\langle \partial^2_{yy} e(\theta, u, y), (\partial_\theta y(\theta; u) v_1, \partial_\theta y(\theta; u) v_2) \right\rangle$$
$$- 2\left\langle \partial^2_{y\theta} e(\theta, u, y), (\partial_\theta y(\theta; u) v_1, v_2) \right\rangle - \left\langle \partial^2_{\theta\theta} e(\theta, u, y), (v_1, v_2) \right\rangle.$$

In order to make the next lines more readable let $y = y(\theta; u)$, $e = e(\theta, u, y)$ and let the lower indices represent vector components. Let w^k be the solution of

$$\partial_y e^\top w^k = \nabla_y h_k.$$

where $h_k = h_k(y(\theta; u), \theta)$. Then

$$\partial_y h_k \partial^2_{\theta\theta} y = -\sum_{\ell,p=1}^{n_\theta} \left[\sum_{r=1}^{d_y} \left(\sum_{s=1}^{d_y} w^k_r (\partial_\theta y_k)_\ell \cdot \partial^2_{y_k y_s} e_r \cdot (\partial_\theta y_s)_p \right) \right.$$
$$\left. + 2\sum_{r=1}^{d_y} \left(\sum_{s=1}^{d_y} w^k_r \partial^2_{\theta_\ell y_s} e_r \cdot (\partial_\theta y_s)_p \right) + \sum_{r=1}^{d_y} w^k_r \partial^2_{\theta_\ell \theta_p} e_r \right] \mathbf{e}_\ell \otimes \mathbf{e}_p,$$

where $\mathbf{e}_\ell \otimes \mathbf{e}_p$ denotes the standard tensor product between unit vectors $\mathbf{e}_\ell$ and $\mathbf{e}_p$.

A.5 Derivatives in Adjoint Equations for OED

We derive the computation of the gradients $\nabla_y J(\omega, u, y, s)$, $\nabla_\omega J(\omega, u, y, s)$ and $\nabla_{s_i} J(\omega, u, y, s)$ from Section 3.3. The interesting parts are the gradients of the design criterion $\nabla_y \Psi[C(\bar\theta, \omega, y, s)]$, $\nabla_\omega \Psi[C(\bar\theta, \omega, y, s)]$ and $\nabla_{s_i} \Psi[C(\bar\theta, \omega, y, s)]$, where $C(\bar\theta, \omega, y, s) = M_{\mathrm{GN/B}}(\bar\theta, \omega, y, s)^{-1}$. We only compute directional derivatives in case of the D-criterion, i.e., $\Psi[C] = \Psi_D[C] = \det(C)$. With Lemma A.3 and Lemma A.4 we have

$$\langle \partial_\omega \Psi_D, \mathbf{e}_k \rangle = \det(C) \cdot \mathrm{tr}\left(C^{-1} \cdot \Delta C \right),$$
$$\Delta C = -C \cdot \Delta F \cdot C.$$

$$\Delta F = s^\top \cdot \partial_y h^\top \cdot \mathrm{Diag}\left(\mathrm{rep}(\mathbf{e}_k; n_t)\right) \cdot \Sigma^{-1} \cdot \partial_y h \cdot s,$$

for any unit vector $\mathbf{e}_k \in \mathbb{R}^{n_\omega}$, where $\mathrm{rep}(\mathbf{e}_k; n_t)$ provides n_t copies of $\mathbf{e}_k$. Furthermore, for $i = 1, \ldots, n_\theta$ we have

$$\begin{aligned}
\left\langle \partial_{s_i} \Psi_D, \tilde{s} \right\rangle &= \det(C) \cdot \mathrm{tr}\left(C^{-1} \cdot \Delta C\right), \\
\Delta C &= -C \cdot \Delta F \cdot C \\
\Delta F &= v_i^\top \cdot \partial_y h^\top \cdot \Omega \cdot \Sigma^{-1} \cdot \partial_y h \cdot s + s^\top \cdot \partial_y h^\top \cdot \Omega \cdot \Sigma^{-1} \cdot \partial_y h \cdot v_i, \\
v_i &= \left[\delta(1, i) \cdot \tilde{s}_1, \ldots, \delta(n_\theta, i) \cdot \tilde{s}_{n_\theta} \right],
\end{aligned}$$

for any $\tilde{s} = [\tilde{s}_1, \ldots, \tilde{s}_{n_\theta}] \in \mathbb{R}^{d_y \times n_\theta}$ with the well known Kronecker delta $\delta(k, l)$. We finally have

$$\begin{aligned}
\left\langle \partial_y \Psi_D, \tilde{y} \right\rangle &= \det(C) \cdot \mathrm{tr}\left(C^{-1} \cdot \Delta C\right), \\
\Delta C &= -C \cdot \Delta F \cdot C \\
\Delta F &= \left\langle \partial_{yy}^2 h, (s, \tilde{y}) \right\rangle^\top \cdot \Omega \cdot \Sigma^{-1} \cdot \partial_y h \cdot s + s^\top \cdot \partial_y h^\top \cdot \Omega \cdot \Sigma^{-1} \cdot \left\langle \partial_{yy}^2 h, (s, \tilde{y}) \right\rangle,
\end{aligned}$$

for all $\tilde{y} \in \mathbb{R}^{d_y}$, where

$$\left\langle \partial_{yy}^2 h, (s, \tilde{y}) \right\rangle = \sum_{\ell=1}^{d_y} \sum_{p=1}^{n_\theta} \sum_{i,j=1}^{d_y} \left(\tilde{y}_j \cdot \partial_{y_i y_j}^2 h_\ell \cdot s_{ip} \right) \mathbf{e}_\ell \otimes \mathbf{e}_p.$$

A.6 Statistical Hypothesis Testing

This section is based on [40] and [55, 118]. In statistical testing one is concerned with deciding whether to accept a *null hypothesis* $H_0 : \theta = \theta_0$, which makes a claim on the true value θ_0 of a parameter θ, or a *counter hypothesis* $H_1 : \theta \neq \theta_0$. Usually, observed measurements $z_1, \ldots, z_N \in \mathbb{R}$, which are realizations of independently, identically distributed random variables $Z_1, \ldots, Z_N$ with parameter-dependent probability distribution P_θ, are involved in this decision process.

> **Definition A.14.** Let two competing hypotheses H_0 and H_1 be given. Construct a mapping $\tau : \mathbb{R}^N \to \{0, 1\}$ with the following interpretation: $\tau(z_1, \ldots, z_N) = 0$ corresponds to the acceptance of H_0 and $\tau(z_1, \ldots, z_N) = 1$ corresponds to a decision in favor of H_1. Then τ is called a *statistical test*.

Alternatively, one can define a statistical test by introducing a *critical region* $K_c \subset \mathbb{R}^N$: the null hypothesis is rejected if $(z_1, \ldots, z_N) \in K_c$, otherwise H_0 is accepted. In this

setting

$$\tau(z_1, \ldots, z_N) = \begin{cases} 1, & \text{if } (z_1, \ldots, z_N) \in K_c, \\ 0, & \text{if } (z_1, \ldots, z_N) \notin K_c. \end{cases}$$

Clearly, it is also possible to define an *acceptance region K* instead of a critical region.

In such a test scenario one encounters two different kinds of errors: the incorrect rejection of H_0 is the *Type I* error and the incorrect rejection of H_1 is the *Type II* error. We denote the corresponding probabilities as

$$\alpha = P[\text{Type I error}] = P[\text{reject } H_0 \,|\, H_0 \text{ is true}],$$
$$\beta = P[\text{Type II error}] = P[\text{accept } H_0 \,|\, H_0 \text{ is false}].$$

Using the critical region, we derive a formula for the probability of the Type I error:

$$\alpha = P_{\theta=\theta_0}[(Z_1, \ldots, Z_N) \in K_c].$$

In general, it is not possible to minimize α and β at the same time. Therefore, an arbitrary *test level* TOL $\in (0, 1)$ is set as an upper bound to α and under this constraint one tries to minimize β. Usually, the hypotheses H_0 and H_1 are stated in such a way that H_1 contains the claim one wants to verify, i.e., one wants to reject H_0.

The critical region can often be written in the form

$$K_c = \{(z_1, \ldots, z_N) \in \mathbb{R}^N \,:\, T(z_1, \ldots, z_N) > c\},$$

where $c \in \mathbb{R}$ is the *critical value* and $T : \mathbb{R}^N \to \mathbb{R}$ is a function. The corresponding random variable $T(Z_1, \ldots, Z_N)$ is called *test statistic* and has a probability distribution $P_{\theta,T}$ depending also on the function T. We now have another way to determine the Type I error:

$$\alpha = P_{\theta=\theta_0,T}[T(Z_1, \ldots, Z_N) > c].$$

In order to evaluate a statistical test faster one considers the *p-value* of the test. For some given data, this is the smallest test level under which the null hypothesis can only just be rejected. Thus, the data $z_1, \ldots, z_N \in \mathbb{R}$ now determine the critical value in the definition of K_c which achieves a maximal Type I error probability:

$$p\text{-value} = P_{\theta=\theta_0,T}[T(Z_1, \ldots, Z_N) > T(z_1, \ldots, z_N)].$$

The evaluation of the statistical test is then simple: if p-value $>$ TOL, then H_0 cannot be rejected, otherwise H_0 can be rejected with a Type I error probability of less or equal than TOL. Often, the following guidelines are applied for choosing TOL: for very strong evidence choose TOL $= 0.01$, for strong evidence use TOL $= 0.05$. However, the use and abuse of p-values and their tolerances are richly debated in literature.

If we perform $m \in \mathbb{N}$ tests on the same data, the problem of *multiple testing* occurs. If the test level in each test is TOL, then the probability to reject at least one null hypothesis is $1 - (1 - \text{TOL})^m$. This is also called the family-wise error rate (FWER). A common way to deal with this problem is to adjust the individual test level α_i such that the desired Type I error probability for the FWER is TOL:

$$1 - (1 - \alpha_i)^m = \text{TOL} \quad \Longleftrightarrow \quad \alpha_i = 1 - (1 - \text{TOL})^{\frac{1}{m}}. \tag{A.4}$$

A first order Taylor approximation of (A.4) yields $\alpha_i = \frac{\text{TOL}}{m}$ which is called *Bonferroni correction* [38]. This is a very conservative way of controlling the FWER which is useful when one considers the rejection of at least one null hypothesis as significant. Other methods to deal with multiple testing are described in [19, 20, 52].

A.7 Existence of Solutions for the Rayleigh Damping Problem

We combine ideas from [39, Thm. 4.1] as well as from [41, Chap. 7.2, Thm. 5] and adapt their proofs such as to meet our setting in Section 5.2 which incorporates Rayleigh damping.

Since $\mathcal{V}$ is separable, there exists a sequence of smooth functions $(w_k)_{k \in \mathbb{N}}$ that forms an orthogonal basis of $\mathcal{V}$. We take the complete set of eigenfunctions of the negative Laplacian on $\mathcal{V}$, see [41, Chap. 6.5.1]. Furthermore, we define an *approximate solution* of (5.4) of order $m \in \mathbb{N}$ by $y_m(t) := \sum_{k=1}^{m} d_{k,m}(t) w_k$, with smooth coefficients $d_{k,m}(t)$, which solves the finite-dimensional system

$$\varrho \left[y_m''(t), v \right]_G + a\varrho \left[y_m'(t), v \right]_G + \mathcal{A}(y_m(t), v, \theta) \\ + b\mathcal{A}(y_m'(t), v, \theta) = [u(t), v]_{\Gamma_N}, \tag{A.5}$$

for all $v \in \text{Lin}(w_1, \ldots, w_m)$ with the initial conditions $y_m(0) = 0$ and $y_m'(0) = 0$. Since $w_1, \ldots, w_m$ are linearly independent and by standard arguments from the theory of ordinary differential equations, the linear system (A.5) has a unique solution y_m.

Lemma A.15. Let $u \in L^2(0, T; L^2(\Gamma_N)^2)$ and let y be the smooth approximate solution of (A.5) of order $m \in \mathbb{N}$. Then we have the following estimate:

$$\operatorname*{ess\,sup}_{t \in (0,T)} \left(\left\| y'(t) \right\|^2_{L^2(G)^2} + \|y(t)\|^2_{H^1(G)^2} \right) \leq c \, \|u\|^2_{L^2(0,T;L^2(\Gamma_N)^2)}$$

where c is a positive constant.

Proof. We test (A.5) with $y'(t)$ and thus arrive at

$$\frac{\varrho}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left\| y'(t) \right\|^2_{L^2(G)^2} + a\varrho \left\| y'(t) \right\|^2_{L^2(G)^2}$$
$$+ \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \mathscr{A}(y(t), y(t), \theta) + b\mathscr{A}(y'(t), y'(t), \theta) = \left[u(t), y'(t) \right]_{\Gamma_N}$$

and after time integration

$$\varrho \left\| y'(t) \right\|^2_{L^2(G)^2} + 2a\varrho \int_0^t \left\| y'(s) \right\|^2_{L^2(G)^2} \mathrm{d}s + \mathscr{A}(y(t), y(t), \theta)$$
$$+ 2b \int_0^t \mathscr{A}(y'(s), y'(s), \theta) \, \mathrm{d}s = 2 \int_0^t \left[u(s), y'(s) \right]_{\Gamma_N} \mathrm{d}s. \tag{A.6}$$

We know by Korn's inequality [29, Thm. 6.3-3] that the bilinear form $\mathscr{A}(\cdot, \cdot, \theta)$ satisfies

$$\mathscr{A}(v, v, \theta) \geq \alpha \|v\|^2_{H^1(G)^2} - \|v\|^2_{L^2(G)^2}, \qquad \text{for all } v \in \mathscr{V},$$

with a constant $\alpha > 0$. Thus, (A.6) becomes

$$\varrho \left\| y'(t) \right\|^2_{L^2(G)^2} + \alpha \|y(t)\|^2_{H^1(G)^2} - \|y(t)\|^2_{L^2(G)^2}$$
$$+ 2b \int_0^t \alpha \left\| y'(s) \right\|^2_{H^1(G)^2} - \left\| y'(s) \right\|^2_{L^2(G)^2} \mathrm{d}s \leq 2 \int_0^t \left[u(s), y'(s) \right]_{\Gamma_N} \mathrm{d}s. \tag{A.7}$$

Now, with $c > 0$ denoting various constants we have

$$2 \left[u(t), y'(t) \right]_{\Gamma_N} \leq 2b\alpha \left\| y'(t) \right\|^2_{H^1(G)^2} + c \, \|u(t)\|^2_{L^2(\Gamma_N)^2}$$

and

$$\|y(t)\|^2_{L^2(G)^2} \leq c \int_0^t \left\| y'(s) \right\|^2_{L^2(G)^2} \mathrm{d}s.$$

Thus, (A.7) becomes

$$\varrho \left\| y'(t) \right\|_{L^2(G)^2}^2 + \alpha \left\| y(t) \right\|_{H^1(G)^2}^2 \leq c \int_0^t \|u(s)\|_{L^2(\Gamma_N)^2}^2 \, \mathrm{d}s + c \int_0^t \left\| y'(s) \right\|_{L^2(G)^2}^2 \, \mathrm{d}s. \qquad \text{(A.8)}$$

Because of $u \in L^2(0, T; L^2(\Gamma_N)^2)$, the first term on the right side remains bounded for all $t \in (0, T)$ and with $\varphi(t) := \|y'(t)\|_{L^2(G)^2}^2 + \|y(t)\|_{H^1(G)^2}^2$ the inequality (A.8) becomes

$$\varphi(t) \leq c \int_0^t \|u(s)\|_{L^2(\Gamma_N)^2}^2 \, \mathrm{d}s + c \int_0^t \varphi(s) \, \mathrm{d}s.$$

From Gronwall's inequality it follows

$$\varphi(t) \leq c \left(\int_0^t \|u(s)\|_{L^2(\Gamma_N)^2}^2 \, \mathrm{d}s \right) \cdot \exp(ct). \qquad \blacksquare$$

Lemma A.16. Let $u \in H^1(0, T; L^2(\Gamma_N)^2)$ and let y be the smooth approximate solution of (A.5) of order $m \in \mathbb{N}$. Then we have the additional estimate:

$$\operatorname*{ess\,sup}_{t \in (0,T)} \left(\left\| y''(t) \right\|_{L^2(G)^2}^2 + \left\| y'(t) \right\|_{H^1(G)^2}^2 \right) \leq c \left\| u' \right\|_{L^2(0,T;L^2(\Gamma_N)^2)}^2$$

where c is a positive constant.

Proof. We differentiate (A.5) with respect to t, write $\tilde{y}(t) := y'(t)$ and test with $\tilde{y}'(t)$. Thus, we arrive at

$$\frac{\varrho}{2} \frac{\mathrm{d}}{\mathrm{d}t} \left\| \tilde{y}'(t) \right\|_{L^2(G)^2}^2 + a\varrho \left\| \tilde{y}'(t) \right\|_{L^2(G)^2}^2$$
$$+ \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}t} \mathscr{A}(\tilde{y}(t), \tilde{y}(t), \theta) + b \mathscr{A}(\tilde{y}'(t), \tilde{y}'(t), \theta) = \left[u'(t), \tilde{y}'(t) \right]_{\Gamma_N}$$

and after time integration

$$\varrho \left\| \tilde{y}'(t) \right\|_{L^2(G)^2}^2 + 2a\varrho \int_0^t \left\| \tilde{y}'(s) \right\|_{L^2(G)^2}^2 \, \mathrm{d}s + \mathscr{A}(\tilde{y}(t), \tilde{y}(t), \theta)$$
$$+ 2b \int_0^t \mathscr{A}(\tilde{y}'(s), \tilde{y}'(s), \theta) \, \mathrm{d}s = 2 \int_0^t \left[u'(s), \tilde{y}'(s) \right]_{\Gamma_N} \, \mathrm{d}s. \qquad \text{(A.9)}$$

As in the proof of Lemma A.15 we can argue that the right hand side of (A.9) is

bounded from below:

$$
\varrho \left\| \tilde{y}'(t) \right\|^2_{L^2(G)^2} + \alpha \left\| \tilde{y}(t) \right\|^2_{H^1(G)^2} - \left\| \tilde{y}(t) \right\|^2_{L^2(G)^2}
$$
$$
+ 2b \int_0^t \alpha \left\| \tilde{y}'(s) \right\|^2_{H^1(G)^2} - \left\| \tilde{y}'(s) \right\|^2_{L^2(G)^2} \, ds \le 2 \int_0^t \left[u'(s), \tilde{y}'(s) \right]_{\Gamma_N} ds. \tag{A.10}
$$

From here it follows again that

$$
2 \left[u'(t), \tilde{y}'(t) \right]_{\Gamma_N} \le 2b\alpha \left\| \tilde{y}'(t) \right\|^2_{H^1(G)^2} + c \left\| u'(t) \right\|^2_{L^2(\Gamma_N)^2}
$$

and

$$
\| \tilde{y}(t) \|^2_{L^2(G)^2} \le c \int_0^t \left\| \tilde{y}'(s) \right\|^2_{L^2(G)^2} ds,
$$

where $c > 0$ denotes various constants. Thus, (A.10) becomes

$$
\varrho \left\| \tilde{y}'(t) \right\|^2_{L^2(G)^2} + \alpha \| \tilde{y}(t) \|^2_{H^1(G)^2} \le c \int_0^t \left\| u'(s) \right\|^2_{L^2(\Gamma_N)^2} ds + c \int_0^t \left\| \tilde{y}'(s) \right\|^2_{L^2(G)^2} ds. \tag{A.11}
$$

Because of $u \in H^1(0, T; L^2(\Gamma_N)^2)$, the first term on the right side remains bounded for all $t \in (0, T)$ and with $\varphi(t) := \| \tilde{y}'(t) \|^2_{L^2(G)^2} + \| \tilde{y}(t) \|^2_{H^1(G)^2}$ the inequality (A.11) becomes

$$
\varphi(t) \le c \int_0^t \left\| u'(s) \right\|^2_{L^2(\Gamma_N)^2} ds + c \int_0^t \varphi(s) \, ds.
$$

From Gronwall's inequality it follows

$$
\varphi(t) \le c \left(\int_0^t \left\| u'(s) \right\|^2_{L^2(\Gamma_N)^2} ds \right) \cdot \exp(ct). \qquad \blacksquare
$$

We conclude that the sequences $(y_m)_{m \in \mathbb{N}}$ and $(y_m')_{m \in \mathbb{N}}$ both are bounded in $L^\infty(0, T; H^1(G)^2)$ and that the sequence $(y_m'')_{m \in \mathbb{N}}$ is bounded in $L^\infty(0, T; L^2(G)^2)$ for $m \to \infty$, respectively. It is now possible to choose a subsequence y_k, such that $y_k \to y$ weakly star and $y_k' \to y'$ weakly star in $L^\infty(0, T; H^1(G)^2)$, respectively, and $y_k'' \to y''$ weakly star in $L^\infty(0, T; L^2(G)^2)$.

To prove that this y is indeed the unique solution of the original problem (5.4) one can proceed in a similar way as in the proof found in [39, Thm. 4.1].

Bibliography

[1] Alefeld, G. and Mayer, G. (2000): *Interval analysis: Theory and applications*. In: Journal of Computational and Applied Mathematics, **121**(1–2): 421–464.

[2] Alexanderian, A. (2021): *Optimal experimental design for infinite-dimensional Bayesian inverse problems governed by PDEs: A review*. In: Inverse Problems, **37**(4): 043001.

[3] Alexanderian, A., Petra, N., Stadler, G. and Ghattas, O. (2014): *A-optimal design of experiments for infinite-dimensional Bayesian linear inverse problems with regularized ℓ_0-sparsification*. In: SIAM Journal on Scientific Computing, **36**(5): A2122–A2148.

[4] Alexanderian, A., Petra, N., Stadler, G. and Ghattas, O. (2016): *A fast and scalable method for A-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems*. In: SIAM Journal on Scientific Computing, **38**(1): A243–A272.

[5] Alipour, A. and Zareian, F. (2008): *Study Rayleigh damping in structures; uncertainties and treatments*. In: Proceedings of 14th World Conference on Earthquake Engineering, Beijing, China, 1–8.

[6] Allen-Zhu, Z., Li, Y., Singh, A. and Wang, Y. (2021): *Near-optimal discrete optimization for experimental design: a regret minimization approach*. In: Mathematical Programming, **186**(1): 439–478.

[7] Alvin, K. F., Oberkampf, W. L., Diegert, K. V. and Rutherford, B. M. (1998): *Uncertainty quantification in computational structural dynamics: a new paradigm for model validation*. In: Society for Experimental Mechanics, Inc, 16th International Modal Analysis Conference, volume 2, 1191–1198. Citeseer.

[8] Arendt, P. D., Apley, D. W. and Chen, W. (2012): *Quantification of model uncertainty: Calibration, model discrepancy, and identifiability*. In: Journal of Mechanical Design, **134**(10): 100908.

[9] Asprey, S. P. and Macchietto, S. (2002): *Designing robust optimal dynamic experiments*. In: Journal of Process Control, **12**(4): 545–556.

[10] Aster, R. C., Borchers, B. and Thurber, C. H. (2013): *Parameter Estimation and Inverse Problems*. 2nd edition. Academic Press, Boston.

[11] Bard, Y. (1974): *Nonlinear parameter estimation*. Academic press, New York.

[12] Bauer, I., Bock, H. G., Körkel, S. and Schlöder, J. P. (2000): *Numerical methods for optimum experimental design in DAE systems*. In: Journal of Computational and Applied Mathematics, **120**(1-2): 1–25.

[13] Bayarri, M. J., Berger, J. O., Cafeo, J., Garcia-Donato, G., Liu, F., Palomo, J., Parthasarathy, R. J., Paulo, R., Sacks, J. and Walsh, D. (2007): *Computer model validation with functional output*. In: Annals of Statistics, **35**(5): 1874–1906.

[14] Beale, E. M. L. (1960): *Confidence Regions in Non-Linear Estimation*. In: Journal of the Royal Statistical Society. Series B (Methodological), **22**(1): 41–88.

[15] Bellman, R. (1997): *Introduction to Matrix Analysis*. SIAM.

[16] Ben-Tal, A., Ghaoui, L. E. and Nemirovski, A. (2009): *Robust Optimization*. Princeton University Press.

[17] Bendsøe, M. P. and Sigmund, O. (1995): *Optimization of structural topology, shape, and material*. Springer.

[18] Bendsøe, M. P. and Sigmund, O. (2003): *Topology optimization: theory, methods, and applications*. Springer Science & Business Media.

[19] Benjamini, Y. and Hochberg, Y. (1995): *Controlling the false discovery rate: a practical and powerful approach to multiple testing*. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology), **57**(1): 289–300.

[20] Benjamini, Y., Krieger, A. M. and Yekutieli, D. (2006): *Adaptive linear step-up procedures that control the false discovery rate*. In: Biometrika, **93**(3): 491–507.

[21] Bertotti, G. and Mayergoyz, I. D. (2006): *The science of hysteresis*. 1st edition. Academic Press, Oxford.

[22] Björck, A. (1996): *Numerical Methods for Least Square Problems*. SIAM, Philadelphia.

[23] Bock, H. G., Carraro, T., Jäger, W., Körkel, S., Rannacher, R. and Schlöder, J. P. (2013): *Model Based Parameter Estimation: Theory and Applications*, volume 4. Springer Science & Business Media.

[24] Brynjarsdóttir, J. and O'Hagan, A. (2014): *Learning about physical parameters: The importance of model discrepancy*. In: Inverse Problems, **30**(114007): 1–24.

[25] Carlen, E. (2010): *Trace Inequalities and Quantum Entropy: An Introductory Course*. In: Entropy and the Quantum, *Contemporary Mathematics*, volume 529, 73–140. AMS, Providence, RI.

[26] Castro-Triguero, R., Murugan, S., Gallego, R. and Friswell, M. I. (2013): *Robustness of optimal sensor placement under parametric uncertainty*. In: Mechanical Systems and Signal Processing, **41**(1-2): 268–287.

[27] Chianeh, H. A., Stigter, J. and Keesman, K. J. (2011): *Optimal input design for parameter estimation in a single and double tank system through direct control of parametric output sensitivities*. In: Journal of Process Control, **21**(1): 111–118.

[28] Chib, S. and Jeliazkov, I. (2006): *Inference in semiparametric dynamic models for binary longitudinal data*. In: Journal of the American Statistical Association, **101**(474): 685–700.

[29] Ciarlet, P. G. (1988): *Mathematical elasticity. Vol. I*, *Studies in Mathematics and its Applications*, volume 20. North-Holland Publishing Co., Amsterdam.

[30] Cramer, H. (1946): *Mathematical Methods of Statistics*. Princeton University Press, New Jersey.

[31] Czado, C. and Schmidt, T. (2011): *Mathematische Statistik*. Statistik und ihre Anwendungen. Springer-Verlag.

[32] D'Agostino, R. B. (1986): *Goodness-of-fit-techniques*, volume 68. CRC press.

[33] Demtröder, W. (2016): *Experimentalphysik 3: Atome, Moleküle und Festkörper*. 5th edition. Springer-Verlag.

[34] Dennis Jr., J. E., Gay, D. M. and Walsh, R. E. (1981): *An Adaptive Nonlinear Least-Squares Algorithm*. In: ACM Transactions on Mathematical Software, **7**(3): 348–368.

[35] Der Kiureghian, A. and Ditlevsen, O. (2009): *Aleatory or epistemic? Does it matter?* In: Structural Safety, **31**(2): 105–112.

[36] Donaldson, J. R. and Schnabel, R. B. (1987): *Computational experience with confidence regions and confidence intervals for nonlinear least squares*. In: Technometrics, **29**(1): 67–82.

[37] Dubitzky, W., Granzow, M. and Berrar, D. P. (2007): *Fundamentals of data mining in genomics and proteomics*. Springer Science & Business Media.

[38] Dunn, O. J. (1961): *Multiple comparisons among means*. In: Journal of the American Statistical Association, **56**(293): 52–64.

[39] Duvaut, G. and Lions, J.-L. (1976): *Inequalities in mechanics and physics*. Springer-Verlag, Berlin-New York. Translated from the French by C. W. John, Grundlehren der Mathematischen Wissenschaften, 219.

[40] Eckle-Kohler, J. and Kohler, M. (2017): *Eine Einführung in die Statistik und ihre Anwendungen*. Springer-Verlag.

[41] Evans, L. C. (2010): *Partial differential equations*, *Graduate Studies in Mathematics*, volume 19. 2nd edition. American Mathematical Society, Providence, RI.

[42] Farajpour, I. and Atamturktur, S. (2012): *Error and uncertainty analysis of inexact and imprecise computer models*. In: Journal of Computing in Civil Engineering, **27**(4): 407–418.

[43] Fedorov, V. V. and Leonov, S. L. (2013): *Optimal Design for Nonlinear Response Models*. CRC Press.

[44] Ferguson, T. S. (1996): *A Course in Large Sample Theory*. Chapman and Hall.

[45] Fitzgibbon, W. E. (1981): *Strongly damped quasilinear evolution equations*. In: Journal of Mathematical Analysis and Applications, **79**(2): 536–550.

[46] Flynn, E. B. and Todd, M. D. (2010): *A Bayesian approach to optimal sensor placement for structural health monitoring with application to active sensing*. In: Mechanical Systems and Signal Processing, **24**(4): 891–903.

[47] Franceschini, G. and Macchietto, S. (2008): *Model-based design of experiments for parameter precision: State of the art*. In: Chemical Engineering Science, **63**(19): 4846–4872.

[48] Franceschini, G. and Macchietto, S. (2008): *Novel anticorrelation criteria for model-based experiment design: Theory and formulations*. In: American Institute of Chemical Engineers Journal, **54**(4): 1009–1024.

[49] Gally, T., Gehb, C. M., Kolvenbach, P., Kuttich, A., Pfetsch, M. E. and Ulbrich, S. (2015): *Robust Truss Topology Design with Beam Elements via Mixed Integer Nonlinear Semidefinite Programming*. In: Pelz, P. F. and Groche, P. (Editors): Uncertainty in Mechanical Engineering II, *Applied Mechanics and Materials*, volume 807, 229–238. Trans Tech Publications.

[50] Gally, T., Groche, P., Hoppe, F., Kuttich, A., Matei, A., Pfetsch, M. E., Rakowitsch, M. and Ulbrich, S. (2021): *Identification of model uncertainty via optimal design of experiments applied to a mechanical press*. In: Optimization and Engineering.

[51] Galvanin, F., Barolo, M., Bezzo, F. and Macchietto, S. (2010): *A backoff strategy for model-based experiment design under parametric uncertainty*. In: American Institute of Chemical Engineers Journal, **56**(8): 2088–2102.

[52] Genovese, C. R. and Wasserman, L. (2006): *Exceedance control of the false discovery proportion*. In: Journal of the American Statistical Association, **101**(476): 1408–1417.

[53] Gere, J. M. and Timoshenko, S. P. (1997): *Mechanics of Materials*. 4th edition. PWS Publishing.

[54] Ghahramani, Z. and Roweis, S. (1999): *A Unifying Review of Linear Gaussian Models*. In: Neural Computation, **11**(4): 305–345.

[55] Gillard, J. (2020): *A First Course in Statistical Inference*. Springer International Publishing, Cham.

[56] Götz, B., Kersting, S. and Kohler, M. (2021): *Estimation of an improved surrogate model in uncertainty quantification by neural networks*. In: Annals of the Institute of Statistical Mathematics, **73**(2): 249–281.

[57] Graham, C. and Talay, D. (2013): *Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation*, volume 68. Springer.

[58] Green, P. L. (2015): *Bayesian system identification of a nonlinear dynamical system using a novel variant of simulated annealing*. In: Mechanical Systems and Signal Processing, **52**: 133–146.

[59] Groche, P., Hoppe, F. and Sinz, J. (2017): *Stiffness of multipoint servo presses: Mechanics vs. control*. In: CIRP Annals, **66**(1): 373–376.

[60] Gu, M. and Wang, L. (2018): *Scaled Gaussian stochastic process for computer model calibration and prediction*. In: SIAM/ASA Journal on Uncertainty Quantification, **6**(4): 1555–1583.

[61] Hanss, M. (2005): *Applied Fuzzy Arithmetic: An Introduction With Engineering Applications*. Springer.

[62] Heisenberg, W. (1927): *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik*. In: Zeitschrift für Physik, **43**: 172–198.

[63] Hiramoto, K., Doki, H. and Obinata, G. (2000): *Optimal sensor/actuator placement for active vibration control using explicit solution of algebraic Riccati equation*. In: Journal of Sound and Vibration, **229**(5): 1057–1075.

[64] Hiriart-Urruty, J.-B. and Lewis, A. S. (1999): *The Clarke and Michel-Penot subdifferentials of the eigenvalues of a symmetric matrix*. In: Computational Optimization and Applications, **13**(1): 13–23.

[65] Hoffmann, K. (1987): *An Introduction to Stress Analysis using Strain Gauges*. Hottinger-Baldwin-Messtechnik-GmbH.

[66] Hoppe, F., Pihan, C. and Groche, P. (2019): *Closed-loop control of eccentric presses based on inverse kinematic models*. In: Procedia Manufacturing, **29**: 240–247. 18th International Conference on Sheet Metal, SHEMET 2019.

[67] Hughes, T. J. R. (1987): *The finite element method*. Prentice Hall, Inc.

[68] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013): *An introduction to statistical learning*, volume 112. Springer.

[69] Jauberthie, C., Bournonville, F., Coton, P. and Rendell, F. (2006): *Optimal input design for aircraft parameter estimation*. In: Aerospace science and technology, **10**(4): 331–337.

[70] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995): *Continuous univariate distributions*, volume 2. John Wiley & Sons.

[71] Kalbfleisch, J. G. (1979): *Probability and Statistical Inference*, volume 2. Springer, New York.

[72] Kalman, R. E. (1960): *A New Approach to Linear Filtering and Prediction Problems*. In: Journal of Basic Engineering, **82**(1): 35–45.

[73] Kass, R. E. and Raftery, A. E. (1995): *Bayes factors*. In: Journal of the American Statistical Association, **90**(430): 773–795.

[74] Kennedy, M. C. and O'Hagan, A. (2001): *Bayesian calibration of computer models*. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology), **63**(3): 425–464.

[75] Kinne, M., Schneider, R. and Thöns, S. (2021): *Reconstructing Stress Resultants in Wind Turbine Towers Based on Strain Measurements*. In: Pelz, P. F. and Groche, P. (Editors): Uncertainty in Mechanical Engineering. ICUME 2021. Lecture Notes in Mechanical Engineering, 224–235. Springer, Cham.

[76] Kohler, M. and Krzyżak, A. (2019): *Estimation of extreme quantiles in a simulation model*. In: Journal of Nonparametric Statistics, **31**(2): 393–419.

[77] Kohler, M. and Krzyżak, A. (2020): *Estimating quantiles in imperfect simulation models using conditional density estimation*. In: Annals of the Institute of Statistical Mathematics, **72**(1): 123–155.

[78] Kolvenbach, P. (2018): *Robust optimization of PDE-constrained problems using second-order models and nonsmooth approaches*. PhD thesis, TU Darmstadt.

[79] Kolvenbach, P, Lass, O. and Ulbrich, S. (2018): *An approach for robust PDE-constrained optimization with application to shape optimization of electrical engines and of dynamic elastic structures under uncertainty*. In: Optimization and Engineering, 697–731.

[80] Körkel, S. (2002): *Numerische Methoden fuür optimale Versuchsplanungsprobleme bei nichtlinearen DAE-Modellen*. PhD thesis, Universität Heidelberg.

[81] Körkel, S., Kostina, E., Bock, H. G. and Schlöder, J. P. (2004): *Numerical methods for optimal control problems in design of robust optimal experiments for nonlinear dynamic processes*. In: Optimization Methods and Software, **19**(3-4): 327–338.

[82] Koval, K., Alexanderian, A. and Stadler, G. (2020): *Optimal experimental design under irreducible uncertainty for linear inverse problems governed by PDEs*. In: Inverse Problems, **36**(7).

Bibliography

[83] Kröner, A., Kunisch, K. and Vexler, B. (2011): *Semismooth Newton methods for optimal control of the wave equation with control constraints*. In: SIAM Journal on Control and Optimization, **49**(2): 830–858.

[84] Lehmich, S., Neff, P. and Lankeit, J. (2014): *On the convexity of the function $C \mapsto f(\det(C))$ on positive-definite matrices*. In: Mathematics and Mechanics of Solids, **19**(4): 369–375.

[85] Lemaire, M. (2014): *Mechanics and Uncertainty*. Wiley Online Library.

[86] Lewis, A. S. (1996): *Derivatives of spectral functions*. In: Mathematics of Operations Research, **21**(3): 576–588.

[87] Li, S. and Platz, R. (2017): *Observations by Evaluating the Uncertainty of Stress Distribution in Truss Structures Based on Probabilistic and Possibilistic Methods*. In: Journal of Verification, Validation and Uncertainty Quantification, **2**(3): 031006.

[88] Lima, E. A. B. F., Oden, J. T., Wohlmuth, B., Shahmoradi, A., Hormuth II, D. A., Yankeelov, T. E., Scarabosio, L. and Horger, T. (2017): *Selection and validation of predictive models of radiation effects on tumor growth based on noninvasive imaging data*. In: Computer Methods in Applied Mechanics and Engineering, **327**: 277–305.

[89] Liu, Y., Chen, W., Arendt, P. and Huang, H.-Z. (2011): *Toward a better understanding of model validation metrics*. In: Journal of Mechanical Design, **133**(7): 071005.

[90] Magnus, J. R. and Neudecker, H. (2019): *Matrix differential calculus with applications in statistics and econometrics*. 3rd edition. John Wiley & Sons.

[91] Mallapur, S. and Platz, R. (2018): *Quantification of Uncertainty in the Mathematical Modelling of a Multivariable Suspension Strut Using Bayesian Interval Hypothesis-Based Approach*. In: Pelz, P. F. and Groche, P. (Editors): Uncertainty in Mechanical Engineering III, *Applied Mechanics and Materials*, volume 885, 3–17. Trans Tech Publications.

[92] Mallapur, S. and Platz, R. (2019): *Uncertainty quantification in the mathematical modelling of a suspension strut using Bayesian inference*. In: Mechanical Systems and Signal Processing, **118**: 158–170.

[93] Matei, A. and Ulbrich, S. (2021): *Detection of Model Uncertainty in the Dynamic Linear-Elastic Model of Vibrations in a Truss*. In: Pelz, P. F. and Groche,

P. (Editors): Uncertainty in Mechanical Engineering, 281–295. Springer International Publishing, Cham.

[94] Mayergoyz, I. D. (2003): *Mathematical Models of Hysteresis and Their Applications*. Elsevier.

[95] Mehra, R. (1974): *Optimal input signals for parameter estimation in dynamic systems–Survey and new results*. In: IEEE Transactions on Automatic Control, **19**(6): 753–768.

[96] Mohri, M., Rostamizadeh, A. and Talwalkar, A. (2018): *Foundations of machine learning*. 2nd edition. MIT press.

[97] Morelli, E. A. and Klein, V. (1990): *Optimal input design for aircraft parameter estimation using dynamic programming principles*. In: Proceedings of the 17th Atmospheric Flight Mechanics Conference.

[98] Neitzel, I., Pieper, K., Vexler, B. and Walter, D. (2019): *A sparse control approach to optimal sensor placement in PDE-constrained parameter estimation problems*. In: Numerische Mathematik, **143**(4): 943–984.

[99] Nečas, J., Málek, J., Rokyta, M. and Růžička, M. (1996): *Weak and measure-valued solutions to evolutionary PDEs*, volume 13. CRC Press.

[100] Oden, T., Moser, R. and Ghattas, O. (2010): *Computer predictions with quantified uncertainty, part I*. In: SIAM News, **43**(9): 1–3.

[101] Otani, S. (1980): *Nonlinear dynamic analysis of reinforced concrete building structures*. In: Canadian Journal of Civil Engineering, **7**(2): 333–344.

[102] Papadopoulos, M. and Garcia, E. (1998): *Sensor placement methodologies for dynamic testing*. In: AIAA journal, **36**(2): 256–263.

[103] Parzen, E. (1962): *On estimation of a probability density function and mode*. In: Annals of Mathematical Statistics, **33**: 1065–1076.

[104] Pelz, P. F., Feldmann, R., Gehb, C. M., Groche, P., Hoppe, F., Knoll, M., Lenz, J., Melz, T., Pfetsch, M. E., Rexer, M. and Schaeffner, M. (2021): *Our Specific Approach on Mastering Uncertainty*. In: Mastering Uncertainty in Mechanical Engineering, 43–111. Springer.

[105] Pelz, P. F., Pfetsch, M. E., Kersting, S., Kohler, M., Matei, A., Melz, T., Platz, R., Schaeffner, M. and Ulbrich, S. (2021): *Types of Uncertainty*. In: Mastering Uncertainty in Mechanical Engineering, 25–42. Springer.

[106] Popper, K. (2005): *The logic of scientific discovery*. Routledge.

[107] Puig, V., Quevedo, J., Escobet, T., Nejjari, F. and de las Heras, S. (2008): *Passive robust fault detection of dynamic processes using interval models*. In: IEEE Transactions on Control Systems Technology, **16**(5): 1083–1089.

[108] Pukelsheim, F. (2006): *Optimal Design of Experiments*. SIAM.

[109] Repin, S. I. and Sauter, S. A. (2020): *Accuracy of Mathematical Models*, *EMS Tracts in Mathematics*, volume 33. European Mathematical Society Publishing House.

[110] Rohrbach, C. (1967): *Handbuch für elektrisches Messen mechanischer Größen*. VDI-Verlag, Düsseldorf.

[111] Roy, C. J. and Oberkampf, W. L. (2011): *A comprehensive framework for verification, validation, and uncertainty quantification in scientific computing*. In: Computer Methods in Applied Mechanics and Engineering, **200**(25): 2131–2144.

[112] Ruud, P. A. (2000): *An Introduction to Classical Econometric Theory*. Oxford University Press.

[113] Saad, Y. (2003): *Iterative methods for sparse linear systems*. SIAM.

[114] Sankararaman, S. and Mahadevan, S. (2011): *Model validation under epistemic uncertainty*. In: Reliability Engineering & System Safety, **96**(9): 1232–1241.

[115] Schaeffner, M., Abele, E., Anderl, R., Bölling, C., Brötz, J., Dietrich, I., Feldmann, R., Gehb, C. M., Geßner, F., Hartig, J., Hedrich, P., Hoppe, F., Kersting, S., Kohler, M., Lenz, J., Martin, D., Matei, A. et al. (2021): *Analysis, Quantification and Evaluation of Uncertainty*. In: Mastering Uncertainty in Mechanical Engineering, 113–207. Springer.

[116] Schueller, G. I. and Pradlwarter, H. J. (2009): *Uncertainty analysis of complex structural systems*. In: International Journal for Numerical Methods in Engineering, **80**(6–7): 881–913.

[117] Seber, G. A. F. (2008): *A Matrix Handbook for Statisticians*, volume 15. John Wiley & Sons.

[118] Selvamuthu, D. and Das, D. (2018): *Introduction to Statistical Methods, Design of Experiments and Statistical Quality Control*. Springer Singapore.

[119] Shahmoradi, A. (2017): *Multilevel Bayesian Parameter Estimation in the Presence of Model Inadequacy and Data Uncertainty*. arXiv:1711.10599.

[120] Simani, S., Fantuzzi, C. and Patton, R. J. (2003): *Model-based fault diagnosis techniques*. In: Model-based Fault Diagnosis in Dynamic Systems Using Identification Techniques, 19–60. Springer.

[121] Smith, R. C. (2014): *Uncertainty Quantification; Theory, Implementation, and Applications, Computational Science & Engineering*, volume 12. SIAM, Philadelphia, PA.

[122] Stegmaier, J., Skanda, D. and Lebiedz, D. (2013): *Robust optimal design of experiments for model discrimination using an interactive software tool*. In: PLOS ONE, **8**(2): e55723.

[123] Stewart, G. W. (1969): *On the Continuity of the Generalized Inverse*. In: SIAM Journal on Applied Mathematics, **17**(1): 33–45.

[124] Stigter, J. D. and Keesman, K. J. (2004): *Optimal parametric sensitivity control of a fed-batch reactor*. In: Automatica, **40**(8): 1459–1464.

[125] Stuart, A. M. (2010): *Inverse problems: a Bayesian perspective*. In: Acta Numerica, **19**: 451–559.

[126] Taboga, M. (2017): *Lectures on probability theory and mathematical statistics*. Createspace.

[127] Teng, Y., Qi, S., Xiao, D., Xu, L., Li, J. and Kang, Y. (2016): *A General Solution to Least Squares Problems with Box Constraints and Its Applications*. In: Mathematical Problems in Engineering, **2016**.

[128] Tuomi, M., Pinfield, D. and Jones, H. R. A. (2011): *Application of Bayesian model inadequacy criterion for multiple data sets to radial velocity models of exoplanet systems*. In: Astronomy & Astrophysics, **532**: A116.

[129] Ulbrich, M. and Ulbrich, S. (2012): *Nichtlineare Optimierung*. Birkhäuser Basel.

[130] van der Vaart, A. W. (1998): *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

[131] Vandepitte, D. and Moens, D. (2011): *Quantification of uncertain and variable model parameters in non-deterministic analysis*. In: IUTAM symposium on the vibration analysis of structures with uncertainties, 15–28. Springer.

[132] Walter, É. and Pronzato, L. (1987): *Optimal experiment design for nonlinear models subject to large prior uncertainties*. In: American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, **253**(3): R530–R534.

[133] Walter, É. and Pronzato, L. (1990): *Qualitative and quantitative experiment design for phenomenological models–a survey*. In: Automatica, **26**(2): 195–213.

[134] Wang, S., Chen, W. and Tsui, K.-L. (2009): *Bayesian validation of computer models*. In: Technometrics, **51**(4): 439–451.

[135] Webb, G. F. (1980): *Existence and asymptotic behavior for a strongly damped nonlinear wave equation*. In: Canadian Journal of Mathematics, **32**(3): 631–643.

[136] Werner, D. (2006): *Funktionalanalysis*. Springer.

[137] Willsky, A. S. (1985): *Detection of abrupt changes in dynamic systems*. In: Detection of abrupt changes in signals and dynamical systems, 27–49. Springer.

[138] Wong, R. K. W., Storlie, C. B. and Lee, T. C. M. (2017): *A frequentist approach to computer model calibration*. In: Journal of the Royal Statistical Society: Series B (Statistical Methodology), **79**(2): 635–648.

[139] Wriggers, P. (2008): *Nonlinear finite element methods*. Springer Science & Business Media.

[140] Yuen, K.-V. and Kuok, S.-C. (2011): *Bayesian Methods for Updating Dynamic Models*. In: Applied Mechanics Reviews, **64**(1).

[141] Zadeh, L. A. (1965): *Fuzzy sets*. In: Information and Control, **8**(3): 338–353.

[142] Zang, T. A., Hemsch, M. J., Hilburger, M. W., Kenny, S. P., Luckring, J. M., Maghami, P, Padula, S. L. and Stroud, W. J. (2002): *Needs and Opportunities for Uncertainty-Based Multidisciplinary Design Methods for Aerospace Vehicles*. Tech. Rep. TM-2002-211462, NASA.

[143] Zhao, L., Lu, Z., Yun, W. and Wang, W. (2017): *Validation metric based on Mahalanobis distance for models with multiple correlated responses*. In: Reliability Engineering & System Safety, **159**: 80–89.

List of Figures

List of Tables

Index

Wissenschaftlicher Werdegang

Alexander Matei, geboren am 19. August 1992 in Cottbus, Deutschland.

10/2017 – 07/2022	Wissenschaftlicher Mitarbeiter am Fachbereich Mathematik der Technischen Universität Darmstadt, Arbeitsgruppe Nichtlineare Optimierung
04/2019 – 12/2020	Mitglied im Sonderforschungsbereich 805: *Beherrschung von Unsicherheit in lasttragenden Systemen des Maschinenbaus*
09/2018	Gastaufenthalt am Fraunhofer ITWM, Kaiserslautern
07/2018 – 03/2019	Mitglied im Sonderforschungsbereich 1194: *Wechselseitige Beeinflussung von Transport- und Benetzungsvorgängen*
10/2015 – 09/2017	Studium der Mathematik Master of Science, Technische Universität Darmstadt
10/2012 – 09/2015	Studium der Mathematik Bachelor of Science (bilingual), Technische Universität Darmstadt
07/2012	Abitur, Ricarda Huch Schule, Dreieich