TECHNISCHE UNIVERSITÄT MÜNCHEN

Professur für Mechanik auf Höchstleistungsrechnern

# Bayesian Calibration of Nonlinear Cardiovascular Models for the Predictive Simulation of Arterial Growth

Sebastian Kehl

Vollständiger Abdruck der von der Fakultät für Maschinenwesen der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktor-Ingenieurs (Dr.-Ing.)

genehmigten Dissertation.

| | |
|---|---|
| Vorsitzender: | Prof. Dr.-Ing. Markus Lienkamp |
| Prüfer der Dissertation: | 1. Prof. Dr.-Ing. Michael W. Gee |
| | 2. Prof. Phaedon-Stelios Koutsourelakis, Ph.D. |

Die Dissertation wurde am 27. Juni 2017 bei der Technischen Universität München eingereicht und durch die Fakultät für Maschinenwesen am 5. Dezember 2017 angenommen.

# Abstract

The onset of screening programs for abdominal aortic aneurysms (AAAs) (a balloon-like, pathological dilation of the abdominal aorta) suggested a beneficial effect on overall mortality rates associated to the risk of rupture of a AAA. Due to the associated increase in asymptomatic findings, these programs raised the need for the definition of policies for surveillance intervals of patients with small aneurysms ($<$ 5 cm) that are not subject to direct surgical intervention. But due to the high patient-specific character of AAA expansion, common policies for surveillance intervals have not been established yet. In this context, the use of computational models that are able to predict AAA growth on a patient-specific basis is expected to contribute in the clinical management of patients and to help in the determination of guidelines for surveillance. This expectation is nourished by the availability of calibration techniques that enable a high degree of personalization of parametrized continuum mechanical models of arterial growth. The calibration of input parameters of such models is possible through measurements associated to output quantities of the computational model. However, measurements are subject to an observation error that will necessarily be reflected in the calibrated model parameters and consequently in the predictive quality of the model output. Whereas this error can be induced by systematic effects that require a careful processing of measured data, it is also subject to random effects. The quantification of this source of aleatoric uncertainty that is reflected in the calibrated model parameters and in the model output is inevitable for a reliable assessment of the predictive quality of a model.

To this end, the work presented in this thesis deals with the development of a Bayesian calibration framework for parameters of computational models of arterial growth. Such a formulation results in a description of the calibrated input parameters in terms of a probability distribution. The propagation of this distribution to the model output – i.e., a patient-specific prediction of growth – allows a quantitative assessment of the quality of the model output in terms of statistical quantities such as variance or credible intervals. In this context, the main contribution of this thesis is given by the following aspects.

Measured data showing AAA growth is often given by longitudinal or follow-up image data. The mapping between the model output in terms of a deformation and a corresponding measured deformation encoded in a sequence of images is not straightforward. On the contrary, such a mapping is associated to sources of significant systematic errors, e.g., introduced through the application of image registration techniques to decode the deformation from the sequence of image data. To avoid the propagation of this systematic error to the predictive quality of the model, the presented framework enables the use of measurements defined as surfaces that can be directly extracted from image data by means of segmentation. Moreover, the formalism of surface currents is applied as a mathematically convenient description of surfaces and allows for a seamless integration of measurements given as surfaces into a Bayesian formulation.

As a second aspect, a novel dimensionality reduction approach is proposed that ac-

counts for a priori assumptions on the spatial structure of the input parameters given in terms of functions with bounded total variation. This approach allows for the solution of the Bayesian calibration problem via the application of advanced sampling techniques that result in particle approximations of the distribution of the calibrated input parameters. Such approximations can be directly propagated through the model enabling the probabilistic representation of the model output.

Finally, the application of the proposed framework to a large-scale, nonlinear, patient-specific model of AAA growth is shown. For the first time, the present work demonstrates the feasibility of a quantitative assessment of the quality of models for arterial growth based on a probabilistic formulation. As an important consequence, the proposed framework provides a basis for an information theoretic comparison of models for AAA growth, which is able to guide future model development. Furthermore, the presented framework enables the statistical testing of the predictive quality of models for AAA growth, which is an inevitable prerequisite for the prospective application of the model in the clinical management routine.

# Zusammenfassung

Das Aufkommen von Programmen zur Früherkennung abdomineller Aortenaneurysmen (AAAs) (einer krankhaften ballonartigen Erweiterung der abdominellen Aorta) belegte einen vorteilhaften Effekt auf die allgemeine Sterberate, die mit dem Rupturrisko eines AAA einhergeht. Aufgrund des begleitenden Anstiegs von asymptomatischen Befunden bedingten diese Programme die Notwendigkeit der Festlegung von Richtlinien zur Bestimmung von Untersuchungsintervallen von Patienten mit kleinen AAAs ($< 5$ cm), die keinen direkten operativen Eingriff erfordern. Allerdings konnten derartige Richtlinien aufgrund des hohen patientenspezifischen Charakters des Wachstums von AAAs bisher nicht ermittelt werden. In diesem Zusammenhang ist zu erwarten, dass der Einsatz computergestützter Berechnungsmodelle zur patientenspezifischen Vorhersage von AAA-Wachstum einen Beitrag zum klinischen Patientenmanagement und zur Festlegung von Untersuchungsintervallen leisten kann. Diese Erwartung wird durch die hohe Anpassungsfähigkeit parametrisierter kontinuumsmechanischer Modelle des arteriellen Wachstums durch Kalibrierung genährt. Die Kalibrierung solcher Modelle hinsichtlich ihrer Eingangsgrößen wird durch Messungen von Modellausgangsgrößen ermöglicht. Allerdings unterliegen Messungen immer einem Messfehler, der sich zwangsläufig in den kalibrierten Parametern und daher auch in der Vorhersagekraft des Modells niederschlägt. Neben systematischen Effekten, die eine sorgfältige Behandlung der Messdaten erfordern, spielen zufällige Effekte eine große Rolle. Die Quantifizierung des Einflusses dieser Quelle zufälliger Unsicherheit auf die kalibrierten Modellparameter sowie auf die Modellausgangsgrößen stellt eine unabdingbare Notwendigkeit bei der zuverlässigen Einschätzung der Vorhersagekraft eines Modells dar.

Die vorliegende Arbeit befasst sich daher mit der Entwicklung eines bayesschen Ansatzes zur Parameterkalibrierung computergestützter Modelle des arteriellen Wachstums. Ein derartiger Ansatz resultiert in der Beschreibung der kalibrierten Parameters anhand einer Wahrscheinlichkeitsverteilung. Die Fortpflanzung dieser Verteilung auf die Modellausgangsgrößen – das heißt konkret: auf eine patientenspezifische Wachstumsvorhersage – ermöglicht eine quantifizierbare Einschätzung der Qualität der Modellausgangsgrößen anhand statistischer Kenngrößen wie Varianz oder Intervallangaben. In diesem Zusammenhang liegt der Schwerpunkt der vorliegenden Arbeit auf den folgenden Aspekten.

Messdaten zu AAA-Wachstum liegen oftmals als Nachfolgeuntersuchungen bzw. als Zeitreihenstudien in Form von Bilddaten vor. Die Abbildung zwischen Modellausgangsgrößen in Form von Deformationen und einer zugehörigen gemessenen Deformation, die in einer Sequenz von Bildaten verschlüsselt ist, ist nicht trivial. Im Gegenteil: Eine derartige Abbildung unterliegt dem Einfluss systematischer Fehlerquellen, die zum Beispiel durch die Anwendung von Methoden der Bildregistrierung zur Dekodierung der Deformation aus sequentiellen Bilddaten hervorgerufen wird. Um die Fortpflanzung dieser systematischen Fehler auf die Vorhersagekraft des Modells zu verhindern, werden in der vorliegenden Arbeit Messungen in Form von Oberflächen definiert, die mittels Segmentierung direkt aus den Bilddaten gewonnen werden können. Darüberhinaus erlaubt die Verwendung des mathematischen Formalismus der *surface currents* zur Beschreibung von Oberflächen eine nathlose Einbettung von Oberflächenmessungen in den bayesschen Ansatz.

Als zweiter Aspekt wird in dieser Arbeit ein neuartiger Ansatz zur Dimensionsreduktion vorgeschlagen, der a priori Annahmen über die räumlische Struktur der Modelleingangsparameter, die durch eine Beschränktheit der totalen Variation charakterisiert ist, berücksichtigt. Dieser Ansatz ermöglicht die Lösung des bayesschen Kalibrierungsproblems durch die Anwendung moderner Stichprobenverfahren, die in einer partikelbasierten Approximation der Wahrscheinlichkeitsverteilung der Modelleingangsparameter resultiert. Durch die direkte Fortpflanzung durch das Modell ermöglichen derartige Approximationen eine probabilistische Darstellung der Modellausgangsgrößen.

Abschließend erfolgt die Anwendung des vorgeschlagenen Ansatzes auf ein großes, nichtlineares, patientenspezifisches Modell. Dies zeigt erstmals die Durchführbarkeit einer quantitativen Abschätzung der Qualität von Modellen des arteriellen Wachstums basierend auf einer probabilistischen Formulierung. Als wichtige Konsequenz bietet der vorgeschlagene Ansatz daher die Grundlage für einen informationstheoretischen Vergleich von Wachstumsmodellen, der als Richtlinie für die zukünftige Modellentwicklung dienen kann. Darüberhinaus ermöglicht der vorgeschlagene Ansatz die Möglichkeit des statistischen Testens der Vorhersagekraft des Wachstumsmodells, welches eine unabdingbare Voraussetzung für den zukünftigen Einsatz des Modells in der klinischen Praxis darstellt.

# Contents

# 1. Introduction

The field of biomechanics, as the application of mechanical principles in the analysis of the function of biological processes, enabled a new perspective on human health [69]. The contributions of these principles to the understanding of diseases have been intensified in the last decades due to the immense acceleration in the development and the availability of computational resources. In particular, continuum mechanical approaches to the field of human soft tissue and cardiovascular diseases [101] have developed rapidly owing to the high burden of cardiovascular diseases (e.g., atherosclerosis, cerebrovascular diseases, aneurysms) to the modern society [see e.g. 75]. Thereby, computational modeling is used to study a wide range of different aspects of the cardiovascular system. This is reflected in the variety of modeling approaches ranging from fully resolved patient-specific fluid structure interaction (FSI) models of the arterial network [228] over particular aspects on the organ level, such as the heart [98, 125, 164, 198], the carotid artery [80, 120] or aneurysms [61, 80, 105, 126, 136] (cerebral and abdominal), to reduced dimensional representations of the cardiovascular system [64, 172].

An important aspect of model development is the eventual application of the model as a *predictive* tool. A predictive model is thereby understood as a model which allows for a quantification of the quality of the model output. Continuum mechanical models are usually characterized by generic *input* → *output* relations. Thus, the quality of the output depends on the case specific adaption of the input parameters. Due to the complexity of the applied models, a direct measurement of the input properties is usually not possible. This requires the determination of input parameters by *calibration* techniques enabled by auxiliary measurements associated to the model output. Depending on the field of research, this process is also referred to as *data assimilation*, *model personalization* or *parameter identification*. The inherent uncertainty included in this process will have a direct impact on the reliability and variability of the model output. A quantification of the quality of the model is thus directly connected to the determination of the propagation of the uncertainty in the input parameters to the model output.

A particular field of cardiovascular modeling, which has attracted much attention in the last decades, is the modeling of abdominal aortic aneurysms (AAAs). The clinical management of this condition centers around the assessment of the risk associated to prophylactic interventions and the risk of fatal incidents. For example, the mortality rate in the case of rupture of a AAA is in the range of 65% to 85% in men above 65 years of age [210]. Whereas overall mortality rates can be reduced by elective repair, it is reported that early elective repair does not necessarily save lives [212]. In this difficult conflict of interest, the use of computational models has already shown to be able to add valuable information on the risk of rupture to the process of decision making [72, 148]. These approaches are based on deterministic criteria relating stress or strain information at a specific point in time to a patient-specific risk of rupture. Therefore, these models are not suited to predict the progression of a AAA on a patient-specific basis. However, such information is an essential measure for the assessment of surveillance intervals and the associated development of the risk of rupture. Thus, models representing the transient behavior of arterial growth are expected to be able to contribute to the clinical management of AAA.

In this context, the presented work aims at the development of predictive models of AAA growth with an application towards the use in the clinical management routine. Given the direct impact such models can eventually have on decision making in a patient-specific context, a rigorous quantification of the involved uncertainties is an inevitable factor towards the application in clinical practice.

## 1.1. Medical background

An abdominal aortic aneurysm (AAA) is characterized by a permanent local distention of the abdominal aorta. Whereas the precise cause for this phenomenon is still unclear, there is some agreement that it is the result of a complex interplay of different factors: degenerative biological processes in the arterial wall due to structural effects, aging, diseases, infections and specific changes in hemodynamic conditions. Beside certain genetic disorders being known to cause aneurysms [227] and a clear correlation of the prevalence of AAA and age [117, 210], statistical evaluations revealed several *risk factors* associated to the formation of a AAA, e.g.: smoking, hypertension, alcohol consumption, atherosclerosis [132].

**Clinical management and the risk assessment** The rupture of a AAA is often lethal and is associated to mortality rates up to 90% [117]. *Symptomatic* AAAs, characterized by harbingers of rupture such as the sudden onset of pain in the abdomen or back, should be immediately repaired [117]. However, most AAAs are *asymptomatic* and the result of an incidental finding. Since elective prophylactic repair is also associated to significant mortality [60, 133], physicians have to balance the risk of rupture with the risks associated to elective repair. Thereby, decision making is mainly based on measurements of the maximum diameter. This choice is justified by large trial data revealing a marked increase in rates of yearly incidents of rupture for aneurysms with a maximum diameter $> 5.5$ cm [133, 177]. Despite studies indicating more efficient indices for the potential of rupture of a AAA [72, 148, 201], the maximum diameter criterion is the dominant index used in the current clinical routine.

**Assessment of AAA growth rates** The use of the 5.5 cm criterion raises the need for strategies and guidelines for small aneurysms $< 5.5$ cm. The finding of small aneurysms has become more common since the introduction of screening programs which suggested to have a beneficial effect on overall mortality [210, 212]. In this scenario, clinical research centers around rates of expansion of AAAs. On the one hand, as a direct index for the risk of rupture associated to rapid rates of expansion measured in terms of maximum diameter: rates $> 1$ cm/yr are associated to a high risk of rupture [29]. On the other hand, as a measure in the clinical management to allow for the assessment of surveillance intervals. To this end, beside a plain quantification of growth rates in small aneurysms, studies try to identify risk factors as predictors for growth [45, 195].

In contrast to the well established guideline towards aneurysm repair in the case of a maximum diameter $> 5.5$ cm, clear guidelines for the definition of surveillance intervals have not been established. On the contrary, suggested intervals vary widely from a few months to several years [211]. In addition, the quantification of AAA expansion by means of longitudinal measurements of the maximum diameter in time is also being questioned [154]. Furthermore, growth cannot necessarily be predicted by linear extrapolation of the maximum diameter [130] and it is also subject to a high inter-patient variability [28].

With the intention of the patient's safety, these issues indicate the need for a much more patient-specific assessment of AAA progression. By means of calibration and model personalization, continuum mechanical approaches provide an opportunity towards a more patient-specific treatment in the clinical management routine.

## 1.2. Computer methods towards clinical application

Despite great advances in the computational modeling of cardiovascular problems, a broad application of these approaches in the clinical practice has not been established so far. This fact is mostly due to the complexity of the questions to be answered in a clinical setting and the expected reliability and validity of the answers provided by a computational model. Nevertheless, efforts and advances towards a clinical application have recently shown great potential. For instance, the remarkable progress made in the context of AAA modeling has already shown to provide indices for the risk of rupture superior to the 5.5 cm criterion used as a clinical standard [72, 148].

**AAA and the risk of rupture** From a mechanical point of view, the diameter criterion can be justified from the representation of an AAA as a thin walled cylindrical structure. This perspective enables estimations of the arterial wall stress by means of measurements of pressure and diameter. It was shown that these estimations provide more sensitive predictors on the rupture potential than the maximum diameter alone [90]. But given the complex geometrical structure of a AAA, the cylindrical assumption is potentially over simplifying. In fact, it was shown by Fillinger et al. [61] that patient-specific computational models based on finite element (FE) analysis are able to provide estimations of maximum wall stresses which are more decisive for the risk of rupture than the maximum diameter or the stresses estimated from the cylindrical assumption. Since then, FE models with increasing complexity have been applied. Whereas the impact of increasingly complex models is still under discussion [182], there is some agreement on the fact that an aneurysm ruptures when wall stress exceeds wall strength [222]. Unfortunately, wall strength is not a quantity which can be assessed in-vivo. Nevertheless, independent studies found that statistical models of wall strength in combination with patient-specific wall stress analysis outperforms pure stress based models [72, 148]. In this context, the term *rupture potential index (RPI)*, as some measure of the relation between wall stress and wall strength, was coined.

A common basis in the assessment of the risk of rupture based on stress or strength is a static point of view. In this sense, loading conditions and geometry for a specific point in time are related to a 'snap-shot' estimation of a rupture potential. The patient-specific transient behavior of growth of a AAA is not covered. Thus, a direct application of these models in the assessment of surveillance intervals is not possible.

### 1.2.1. Modeling of AAA growth

As a consequence of the unclear biological causes leading to the formation of an aneurysm, also the continuum mechanical modeling of arterial growth is still a controversially discussed topic [see e.g. 5, 101]. It is generally acknowledged that arterial expansion has to be modeled as a combination of *growth and remodeling (G&R)*. Whereas growth is referring to the increasing mass of the tissue, remodeling describes the reorganization of constituents in the arterial wall. Currently there are two main approaches to the modeling of these phenomena.

*Kinematic* formulations [187] attempt a phenomenological description of the consequences of the biological processes leading to growth. Although this approach is very versatile and can be conveniently incorporated into existing continuum mechanical formulations, the phenomenological character represents a major cause of criticism. In this regard, the *constrained mixture* theory [102], and its numerous extensions and variants [see e.g. 100], claims superiority by accounting for the composite structure of the arterial wall and the steady turnover of the associated constituents in healthy tissue. As a consequence, the modeling of arterial growth by means of the constrained mixture formulation is always related to a ‘healthy’ state. Furthermore, the initiation of the process of dilation of an artery is not intrinsically contained in this formulation but has to be triggered manually [100].

Despite some illustrative applications to patient-specific data and claims to the applicability to patient-specific AAA data, the development of both approaches is mainly carried out with respect to conceptual issues and their demonstration on simplified AAA-like geometries. Consequently, neither of the two approaches has been validated with respect to its predictive capabilities on the extrapolation of AAA expansion so far. Validation is thereby understood as the statistical testing of a hypothesis such as ‘the model is able to predict AAA growth’ for a certain level of confidence. Beside a significant cohort of patient-specific data, such a validation requires models with a certain maturity with respect to their predictive capabilities. In this regard, the approaches by Baek et al. [12], Zeinali-Davarani et al. [232] and Zeinali-Davarani et al. [231] provide major steps towards a patient-specific estimation of AAA growth. These studies try to establish patient-specific homeostatic conditions in a constrained mixture approach for models of healthy aortas by an inverse approach. But, in a clinical setting, a corresponding ‘healthy’ state is never available. Rather, an already diseased aorta (i.e., an aneurysm) is monitored by means of longitudinal or data. Computational modeling in such a setting was first attempted by Tinkl [218].

### 1.2.2. Model personalization

Continuum mechanical formulations represent a generic and phenomenological approach to biomechanical problems. To be used as a means in patient-specific modeling, application specific specializations of the applied formulations have to be utilized. The issue of model personalization is represented in many different aspects of the model. Thereby, the use of patient-specific geometries is enabled by routinely available imaging technologies such as ultrasound (US), magnetic resonance imaging (MRI) or computed tomography (CT). Other components of a model such as boundary conditions, constitutive descriptions of elastic or inelastic properties (i.e., growth) are often intricate to define on a patient-specific basis. This is mainly caused by the fact that in the cardiovascular field associated parameters cannot be measured in-vivo. This issue resulted in the development of elaborate methods for model calibration.

These calibration methods invert the standard *input* → *output* relation of a model by trying to identify model input from measurements associated to the model output. Such *inverse problems* play an important role in almost every field of engineering or science such as, e.g., geophysics [32, 175], weather forecast [203] or astrophysics [214]. In the field of biomechanics, (model-based) elastography, i.e., the detection of normal and diseased tissue from medical image data, is a very popular application of inverse problems [55, 66, 170]. In the more specific situation of cardiovascular applications, the estimation of elastic properties of arteries is also called *vascular elastography.* Among many different topics, this field covers applications towards models of the healthy aorta [16, 20, 232] and aneurysms (cerebral and abdominal) [14, 48, 126]. However, the general concept of elastography is not constrained to an elasticity setting and the general approach is, e.g., also applied to electromechanical properties in cardiac mechanics [98, 198, 229]. This general field of vascular elastography has recently been considerably influenced by the *sequential approach* proposed by Moireau et al. [163]. This approach has been applied to estimation problems in cardiac (electro-)mechanics [37, 150] or FSI models of the aorta [20, 161].

In a clinical situation, measurements are often given by image data. The association of such measurements to the output quantities of a computational model is not straightforward and requires careful treatment [162]. It has recently been shown that a proper treatment of the image data can highly influence the efficiency and the results of a parameter identification framework [104, 112].

**Calibration of models for AAA growth** To be able to identify parameters for AAA growth, the image data must capture the transient behavior of growth in time. Such data is given by *longitudinal* data which provides images of at least two distinct points in time showing growth of an aneurysm. In clinical practice however, neither of these images shows a healthy reference state which could be used to assess the homeostatic conditions necessary for a constrained mixture approach [see 12, 231, 232]. This identifiability issue is accounted for by Tinkl [218] by introducing modeling assumptions imposed on the initialization of growth from an already diseased state. This approach allows for an identification of growth parameters and accordingly also a prediction of future growth. However, the predictive use of growth models in the sense of a quantification of the uncertainty associated to the prediction has not been considered so far.

### 1.2.3. Incorporation of uncertain and probabilistic effects

In the field of biomechanical applications and the intention of clinical applicability, the reliability and quality of predictive models is of high importance. In practice, the modeling process is subject to various sources of uncertainty. This uncertainty will necessarily result in a variability of the predictive model output. Thus, the assessment of the quality of a model necessitates a quantification of the variability of the output quantities of the model. Conceptually, different sources of uncertainty can be distinguished [116]:

- The most obvious source of uncertainty is given by the lack of precise knowledge or randomness of the input parameters of a model referred to as *parameter uncertainty.* In the context of models of AAA growth, this uncertainty can by reduced by exploiting medical image data. Random and systematic errors in the imaging machinery and the bounded resolution of medical images are reflected in the definition of the model geometry and the boundary conditions. Further sources of uncertainties are given through the imperfect knowledge of constitutive properties or growth parameters.

- In a calibration setting, the effects of parameter uncertainty typically interfere with the *observation error.* A measurement process is usually subject to systematic and random errors which will be reflected in the measured data. The use of these measurements in the identification of input parameters will additionally be propagated to the model output. In the context of AAA growth, the measurement is again given by image data. It is therefore subject to the same sources of error as the initial model geometry. However, the data must not necessarily be given by the same image modality or the same spatial resolution. Thus, the specification of measurement noise and geometric uncertainty of the initial model can differ.

- From an epistemological perspective, every model is subject to model limitations bounding the predictive quality [219]. This source of error is referred to as *model inadequacy.* Model limitations might also be reflected by the inability of a model to capture a seemingly randomly fluctuating process. Thereby, the fluctuation can either be truly random or the result of aspects of the physical process not covered by the model. In this setting, model inadequacy is defined by the process mean, and the remaining uncertainty is referred to as *residual variability.*

- In principle, a computer program represents a deterministic functional representation of the model. But due to the complexity of the implementation, a computer program is never tested for arbitrary combinations of input parameters. Thus, there is some de facto uncertainty introduced through the application of software, which is referred to as *code uncertainty.*

Whereas these effects are characterized by distinct features, in practical situations, the identification and quantification of a particular source of uncertainty is difficult. Thus, to date, most approaches concentrate on a subset of possible uncertain effects. Thereby, a very successful class of approaches adopts a Bayesian point of view. In this sense, the concept of probability is also interpreted as the substantiation of a degree of belief. Beside truly random relations, also the absence of accurate knowledge is modeled by means of probabilistic formulations.

The implementation of probabilistic concepts in standard engineering processes is made difficult by the complexity of the applied models. This complexity is reflected in two different but often associated effects. On the one hand, the complexity of a model is given by sophisticated and possibly nonlinear functional descriptions. In many situations, this requires a significant amount of computational effort in terms of time and resources for a single evaluation of a model. Given that reliable methods for uncertainty quantification (UQ) require a considerable number of model evaluations, the computational effort associated to a single model evaluation represents a decisive bottleneck factor in practical applications. On the other hand, the complexity of a model refers to the high dimension of the input parameter space. If the input parameters are considered random, this induces a high stochastic dimension for probabilistic formulations. But as a consequence of the *curse of dimensionality* [see e.g. 25], many methods for UQ perform poorly in a high dimensional setup. This poor performance is thereby often related to a drastic increase in the number of model evaluations with increasing dimension of the input parameter space.

Continuum mechanical cardiovascular models are often associated to both aspects of complexity. The accurate representation of the physical processes requires highly resolved nonlinear models. Consequently, the applied numerical techniques result in expensive computational evaluations. Furthermore, input parameters, such as constitutive parameters or growth parameters, are subject to a priori unknown spatial distributions. An accurate modeling of these distributions by means of finite dimensional representations implies a high dimension of the input parameter space. This renders the application of standard Bayesian methods [see e.g. 184] a great challenge and is a reason for the restraint of a general transition towards probabilistic formulations in the field of cardiovascular modeling.

**Model complexity** In principle however, Bayesian methods and techniques that are able to handle complex models efficiently are being established. Thereby, issues related to the computational effort associated to the model evaluation are approached by *advanced sampling* techniques or the use of *surrogate models*.

- Advanced sampling: in order to overcome the inefficiency of standard Markov chain Monte Carlo (MCMC) methods, *hybrid monte carlo* techniques try to improve the convergence, and thereby the number of model evaluations, by employing insight into the physical properties of models [56, 82, 141, 166]. Such schemes can also be enriched by second order information [152], which is enabled by the availability of efficient approaches to gradient and hessian computations based on the well established adjoint theory [224]. Nevertheless, the necessary number of sequential model evaluations might still render these approaches infeasible for large-scale models. In this regard, techniques of *sequential filtering* try to exploit the sequential nature of models [42, 50, 54]. This results in efficient schemes for transient physical systems with an inherent sequential nature. However, the same strategy can also be applied to static systems whereby a sequential nature is introduced artificially [165]. The *Kalman filter* and its various variants represent a particular but very popular approach to sequential filtering problems [202].

- Surrogate modeling: whereas advanced sampling techniques try to reduce the number of necessary model evaluations, the use of surrogate models tries to reduce the computational cost associated to a single model evaluation. A surrogate model or *reduced order model* is given by an approximation to the full order model (FOM) such that a significant reduction in computational time is achieved. A popular approach are nonlinear model reduction techniques such as proper orthogonal decomposition (POD) that employ a snapshot database of FOM solutions created *offline* to construct a reduced basis for the model under certain optimality criteria [23, 33, 70, 149, 167]. Another approach to the construction of reduced order models is given by coarse grid representations of FOMs based on spatial discretization techniques like the classical finite element method (FEM). In contrast to approaches based on POD, these techniques are not claiming optimality of the reduced order representations but build on a statistical relation between coarse grid and fine grid solutions [22, 122, 123].

**Input dimensionality** The aspect of high dimensional input spaces has recently been influenced by advances in the field of *machine learning* and *pattern recognition*. With the growth in digitization and storage capacities, methods and applications in these fields are faced by the immense increase in data to be processed. The demand for inference from this large amount of data requires the application of techniques for the robust representation of the available data [18, 108, 138, 197, 215]. All these techniques are applied to data sets being already available. However, in the setup of model calibration, input data is not available but to be inferred from measurements associated to the model output. In certain situations, this inference process can be directly perfomed on sparse representations of the input space. Such sparse representations are available for certain a priori assumptions on the structure of the input space [139]. But in general, dimension reduction in the setting of inverse problems is significantly more difficult. Recently, this problem was addressed by Franck and Koutsourelakis [66] in a framework for approximate inference.

**Uncertainty quantification in AAA modeling** The biomechanics community has only recently started to adopt probabilistic principles in the development of models. And despite the influential work of Moireau et al. [163] in the general field of cardiovascular (electro-)mechanics, the field of AAA modeling is mainly approached in a deterministic setting. But it was pointed out that the consideration of parameter uncertainty can increase the predictive capabilities of models applied in the computation of the risk of rupture [176]. Furthermore, it has been shown by Biehler et al. [22] that UQ can be efficiently performed for large-scale biomechanical problems and for models applied in the assessment of the risk of AAA rupture.

Towards a clinical application of models for AAA growth, a probabilistic formulation is inevitable. On the one hand, a rigorous quantification of uncertainty greatly increases the reliability of models and allows for a proper assessment of their predictive capabilities. On the other hand, given the ambiguity in the mathematical formulation of growth, a probabilistic formulation allows for an information theoretic comparison of existing models [111]. Thus, by incorporating longitudinal data and the associated observation error, a decision towards the kinematic or the constrained mixture approach can be performed on a formally sound basis. Despite these clear advantages, approaches towards a probabilistic calibration of models for AAA growth have not been developed so far and are thus subject to the work presented in this thesis.

## 1.3. Objective of this thesis

The goal of this thesis is the development of a Bayesian parameter calibration framework for predictive models of AAA growth. The requirements for such models to be of predictive quality in a patient-specific setting result in large-scale nonlinear continuum mechanical models. The patient-specific calibration of such models with respect to parameters describing the arterial growth necessitates high dimensional input spaces that attend the need for a flexible spatial representation of parameters describing arterial growth. In this setting, the main focus of the thesis is twofold: one aspect is the efficient and accurate treatment of measurement data given by medical imaging technologies; the other aspect is the handling of the high input dimensionality.

**Efficient and accurate treatment of measurement data** In the context of cardiovascular modeling, data is often given by medical imaging technologies. To represent the transient process of growth, longitudinal data showing at least two snapshots in time are necessary. As a result, the information of growth is encoded in a series of images. In contrast, *primal variables* in continuum mechanical formulations are often given as displacement or velocity. In order to provide meaningful measurements for such models, the deformation encoded in longitudinal image data has to be extracted. The process of decoding is often performed by *image registration*. However, the information provided by images usually does not allow for the determination of a unique deformation. Thus, image registration itself is an inverse problem controlled by modeling assumptions and regularization [160]. But these assumptions are not necessarily capable of representing the true underlying physical processes depicted in the sequence of images. Consequently, the application of image registration results in a measured deformation incorporating a potential systematic error. In a parameter calibration framework, this error will be reflected in the calibrated model and thus it will affect the predictive quality of the model. Since the predictive quality represents a decisive factor towards the clinical applicability, a sensible treatment of the available image data is indispensable. To this end, this thesis presents a calibration framework based on measurements given by surface data. Surface data can be extracted from medical images by means of *segmentation* without the use of image registration techniques. Thus, the systematic error accompanying image registration is effectively avoided. To close the association of the model output in terms of displacements and a corresponding measurement given by surface information, the mathematical formalism of *surface currents* is employed [221]. This formalism can be seamlessly integrated into a Bayesian calibration framework since it enables a perspective on surfaces as random variables.

**Dimensionality reduction** The need to represent parameters as spatially varying functions is associated to a high dimension of the input space. But input parameters such as constitutive or growth parameters are rarely associated to arbitrary fluctuating functions. Rather, certain structural assumptions on the spatial variability are actually reflected in sparse representations. However, these sparse representations are not necessarily available a priori. In the context of this thesis, parameters for AAA growth are assumed to be represented by functions of *bounded variation* [83]. Such functions are globally characterized by a bound in *total variation (TV)*. Nevertheless, this bound does not restrict a distinct local variability of a function. Thus, functions of bounded variation are able to represent ‘smoothly’ varying functions with distinct and locally bounded features. In combination with FE models, such functions can be conveniently approximated by element-wise constant functions resulting in the dimension of the input space given by the number of elements in the discretization or in the respective partition of the discretization where the parameter is physically defined. However, such a high dimensional representation renders a probabilistic treatment of the calibration problem by the application of advanced sampling techniques highly inefficient. Furthermore, although not a priori accessible, the setup of bounded variation suggest some sparsity of solutions in the patient-specific case, which is not utilized by the element-wise basis. To this end, a novel dimensionality reduction approach that is able to exploit the effective sparsity of parameters on a patient-specific basis is proposed. It is shown that this approach is capable to reflect the main covariance structure contained in the corresponding full dimensional solution.

The numerical simulations shown throughout this thesis were realized using the parallel in-house research code BACI which is jointly developed at the Institute for Computational Mechanics and the Mechanics and High Performance Computing Group of the Technische Universität München [223]. Building on existing modules for the FE simulation in the context of nonlinear solid mechanics and the simulation of arterial G&R, a generic probabilistic calibration framework was implemented and used to conduct the simulations presented in this thesis.

## 1.4. Organization of the thesis

Chapter 2 begins with the introduction of the continuum mechanical framework of quasi-static nonlinear structural elasticity. The presentation covers all relevant aspects of nonlinear continuum mechanics as well as the incorporation of the transient process of arterial growth into the quasi-static framework such that the final problem can be formulated as a sytem of partial differential equations (PDEs). After that, the FEM as a numerical approach for the discrete representation of this system as a set of algebraic nonlinear equations is briefly summarized and approaches for the solutions of this nonlinear system are presented. The chapter is concluded by the introduction of a parametrization of the system and some aspects specific to cardiovascular modeling.

Given a properly formulated model for the prediction of AAA growth, chapter 3 provides the Bayesian formulation of the parameter calibration problem. Beside a very brief presentation of the concepts of Bayesian statistics, the main building blocks of the Bayesian formulation are introduced and analyzed in detail. Thereby, the choice of the representation of measured data as surfaces is motivated and a statistical model of surfaces in the context of the surface current formalism is provided. Furthermore, the definition of a prior model based on functions of bounded variation is given and analyzed with respect to other popular prior models. Finally, a Lagrangian formulation of the inverse problem is presented which allows for the efficient computation of gradient information by means of *adjoint equations*.

The solution of the calibration problem is outlined in chapter 4 covering the definition of optimal solutions as well as the associated numerical computations in a Bayesian context. Thereby a self-contained presentation is targeted that completes the setup and definition of the calibration problem given in the preceding chapters and enables the concentration on the relevant and new aspects in the following chapters.

The main aspect of chapter 5 is the description and analysis of the proposed approach to dimensionality reduction under total variation (TV) prior assumptions. The validity of these assumptions in the case of AAA growth is motivated by a Bayesian argument and a review on existing approaches to dimensionality reduction in the field of biomechanical simulations is provided. The functionality of the approach is demonstrated on a small-scale problem based on synthetic data before it is applied to a large-scale simulation in chapter 7. The analysis in the small-scale regime thereby enables a quantitative assessment of the predictive quality of the reduced basis approach in comparison to a solution based on a full dimensional representation of the input space. This quality is also shown in comparison to an often applied a priori dimensionality reduction technique based on a patch-wise approximation of parameters.

Chapter 6 highlights the beneficial properties of the surface current formalism in the context of systematic errors introduced through registration techniques. By using synthetic data with a quantifiable source of systematic error representing a typical application of image registration, the efficiency of the surface current formalism is demonstrated. This chapter is partly based on the work presented by Kehl and Gee [112].

Chapter 7 shows the application of the dimensionality reduction approach to patient-specific data. This allows the Bayesian calibration of a large-scale patient-specific model of AAA growth based on longitudinal image data. By using the statistical model for surface measurements, a predictive use of the model that accounts for the uncertainty incorporated in the data is enabled. The validity of this model under various aspects of modeling is discussed: beside the patient-specific character of the model, the alignment of the image data and the validity of the statistical model for surface data are reviewed.

The summary of main results and achievements and possible directions for future research are provided in chapter 8.

# 2. Computational solid mechanics

This chapter introduces the general mechanical and mathematical concepts used to model AAA growth in this thesis. The modeling of the complex physiological process of arterial growth is thereby embedded in the continuum mechanical formulation of nonlinear elasticity of solid structures. The approach pursued in this thesis follows the assumption that the timescale of the process of vascular growth (months to years) is orders of magnitude larger than the timescale of the pulsatile blood flow (seconds). Resolving the accurate dynamical behavior of an artery during a heart cycle is therefore neither necessary nor reasonable. Instead, the elastic properties of an artery are modeled by a quasi-static description. After the presentation of the incorporation of the irreversible processes of arterial growth in the elasticity framework, the chapter concludes with the introduction of the numerical treatment of the mechanical system by means of the FEM and some aspects relevant to patient-specific cardiovascular modeling.

The presentation of these concepts is not meant to be comprehensive, but aims at a consistent introduction of the quantities of interest needed to arrive at a computationally solvable set of equations. However, these equations are not the main target of the work presented in this thesis but meant to be incorporated into a parameter identification framework. Hence, the presentation also tries to back up certain concepts helpful in the setup and the understanding of the associated inverse problem introduced in chapter 3. In particular, this includes the nonlinear solution process by means of continuous incremental formulations which enable a consistent derivation of the adjoint equations resulting from a Lagrangian formulation of the inverse problem, see chapter 3.5.

For details, the reader is referred to the respective literature cited throughout this chapter. Generally, a comprehensive introduction to nonlinear solid mechanics is, e.g., given by Bonet and Wood [27], Holzapfel [99] or Truesdell and Noll [220]. An introduction to the variational concepts used in the nonlinear elastic theory is given by Pedregal [174] and a reference for the continuum formulations of kinematic relations is provided by Donea et al. [53]. For the theoretical foundations of the FEM the reader is referred to Larson and Bengzon [131] or Zienkiewicz et al. [233].

## 2.1. Nonlinear solid mechanics

Solid mechanics is the analysis of the deformation of solid bodies subject to certain forces. The nonlinearity of this deformation is introduced through the potential nonlinearity of the relations in the chain

$$stress \leftrightarrow strain \leftrightarrow deformation. \tag{2.1}$$

The functional description of these relations follows a continuum mechanical approach since the body of interest is modeled as a continuous mass. To account for the possible occurrence of large deformations, a theory for finite deformations is pursued.

The body of interest is represented by a certain *reference configuration* $\Omega_0 \subset \mathbb{R}^3$ which is deformed reversibly under prescribed loads and/or irreversibly, e.g., due to growth, into a *current configuration* $\Omega_t \subset \mathbb{R}^3$. Since the configurations are associated to a specific set of mass, they are also termed *material configurations*. In contrast, configurations that are associated to a fixed volumetric subset of the ambient euclidean space are referred to as *spatial configurations*. In the context of medical applications, these material configurations can be monitored by medical imaging technologies. Since medical images span a larger space than the configuration of interest, e.g., in *whole-body-scans*, it is reasonable to define $\Omega_0$ and $\Omega_t$ as subsets in the space of images: $\Omega_0 \subset \Omega_0^{\mathrm{I}}$ and $\Omega_t \subset \Omega_t^{\mathrm{I}}$, see figure 2.1 for an illustration. The domain boundaries $\partial\Omega_i, i \in (0, t)$, are uniquely partitioned into the subsets $\partial\Omega_{i,u}$ and $\partial\Omega_{i,\sigma}$, with

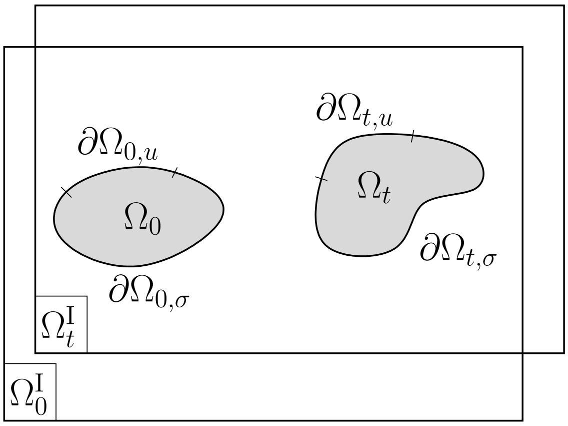

Figure 2.1.: Illustration of the involved physical domains of interest $\Omega_0 \subset \Omega_0^{\mathrm{I}}$ and $\Omega_t \subset \Omega_t^{\mathrm{I}}$, embedded in the euclidean space $\mathbb{R}^3$. As material domains they are labeled by material coordinates $\mathbf{X} \in \Omega_0$ and $\mathbf{x} \in \Omega_t$.

$$\partial\Omega_{i,u} \cup \partial\Omega_{i,\sigma} = \partial\Omega_i \quad \text{and} \quad \partial\Omega_{i,u} \cap \partial\Omega_{i,\sigma} = \varnothing. \tag{2.2}$$

To allow for a description of the configurations $\Omega_0$ and $\Omega_t$, a *Lagrangian formulation* is applied which associates coordinates $\mathbf{X}$ and $\mathbf{x}$ to material points with $\mathbf{X} \in \Omega_0$ and $\mathbf{x} \in \Omega_t$.

Given certain loads on the structure $\Omega_0$ and certain constraints on the deformation, the identification of a current configuration $\Omega_t$ is the main subject of interest in the solid mechanical treatment. To this end, a kinematical description of the deformation (chapter 2.1.1) and a kinetic description of the current configuration (chapter 2.1.2) are coupled by the definition of stress-strain relations (chapter 2.1.3).

## A note on notational conventions

In the following chapters, a common Cartesian reference frame will be assumed such that all tensorial quantities share a common vector-/matrix-representation. Given the vectors $\mathbf{a} \in \mathbb{R}^m$ and $\mathbf{b} \in \mathbb{R}^m$, with components $a_i$ and $b_i$, as representation of 1-tensors, and the matrices $\mathbf{A} \in \mathbb{R}^{m \times m}$ and $\mathbf{B} \in \mathbb{R}^{m \times m}$, with components $A_{ij}$ and $B_{ij}$, as representation of 2-tensors, the following conventions are used: the inner and outer vector products are given by

$$\mathbf{a}\cdot\mathbf{b}=\sum_{i=1}^{m}a_ib_i,\quad [\mathbf{ab}]_{ij}=a_ib_j. \tag{2.3}$$

In particular, the definition of the outer product holds also for the vectorial representation of $\nabla_{\mathbf{a}} := \{\frac{\partial}{\partial \mathrm{a}_i}\}$. For the matrix-vector product the convention

$$\mathbf{a}\cdot\mathbf{B}=\sum_{i=1}^{m}B_{ij}a_i,\quad \mathbf{B}\cdot\mathbf{a}=\sum_{j=1}^{m}B_{ij}a_j, \tag{2.4}$$

is used. These definitions render a specific differentiation between column- and row-vectors irrelevant. Matrix-Matrix products are defined by

$$[\mathbf{AB}]_{ij}=\sum_{k=1}^{m}A_{ik}B_{kj},\quad \mathbf{A}:\mathbf{B}=\mathrm{tr}(\mathbf{AB}),\quad \mathrm{tr}(\mathbf{A})=\sum_{i=1}^{m}A_{ii}. \tag{2.5}$$

### 2.1.1. The law of motion

One of the most fundamental principles in the continuum mechanical description of a body is the definition of a deformation called the *law of motion*:

**Definition 2.1.1** (Law of motion)**.** *Given a reference configuration $\Omega_0$ and a current configuration $\Omega_t$, the diffeomorphic mapping*

$$\varphi(\mathbf{X},t):\mathbf{X}\mapsto\mathbf{x}\quad\forall\quad \mathbf{X}\in\Omega_0,\mathbf{x}\in\Omega_t, \tag{2.6}$$

$$s.t.\quad \mathfrak{c}(\varphi)=0, \tag{2.7}$$

*is called law of motion. $\mathfrak{c}$ represents a constraint on the deformation.*

This mapping provides the basis for the kinematic description of the body of interest, see figure 2.2. Most importantly, it implies the mapping between the tangent spaces called *deformation gradient*:

**Definition 2.1.2** (Deformation gradient)**.** *Given a mapping according to definition 2.1.1, the associated mapping $\mathbf{F}(\mathbf{X}) : T_{\mathbf{X}}\Omega_0 \to T_{\varphi(\mathbf{X})}\Omega_t$ is given by*

$$\mathbf{F}(\mathbf{X}):=\frac{\partial\varphi(\mathbf{X})}{\partial\mathbf{X}}=\frac{\partial\mathbf{x}}{\partial\mathbf{X}}=\hat{\mathbf{F}}(\mathbf{x}). \tag{2.8}$$

Given the mapping (2.6) in terms of the displacement field $\mathbf{U}(\mathbf{X})$ as $\mathbf{x}=\mathbf{X}+\mathbf{U}(\mathbf{X})$, the deformation gradient can be written as

$$\mathbf{F}(\mathbf{X})=\mathbf{I}+\frac{\partial\mathbf{U}(\mathbf{X})}{\partial\mathbf{X}}=\mathbf{I}+\nabla_{\mathbf{X}}\mathbf{U}(\mathbf{X}). \tag{2.9}$$

Furthermore, the constraint (2.7) can also be formulated in terms of the displacements. Throughout the work presented in this thesis, this constraint is given by fixing the boundaries $\partial\Omega_{0,u}=\partial\Omega_{t,u}$ resulting in the *Dirichlet boundary condition*

$$\mathfrak{c}(\varphi)=\varphi(\mathbf{X})-\mathbf{X}=0\iff\mathbf{U}(\mathbf{X})=0\quad\forall\mathbf{X}\in\partial\Omega_{0,u}. \tag{2.10}$$

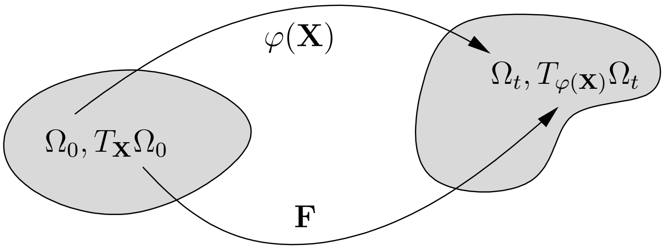

Figure 2.2.: Illustration of the fundamental relations between the reference configurations $\Omega_0$ and the current configuration $\Omega_t$, and their respective tangent spaces $T_{\mathbf{X}}\Omega_0$ and $T_{\varphi(\mathbf{X})}\Omega_t = T_{\mathbf{x}}\Omega_t$.

Since $\varphi^{-1}(\mathbf{x})$ can be represented by $\varphi^{-1}(\mathbf{x}) = \mathbf{X} = \mathbf{x} - \mathbf{u}(\mathbf{x})$ with $\mathbf{u}(\mathbf{x}) = \mathbf{U}(\varphi^{-1}(\mathbf{x}))$, this boundary constraint is equivalently expressed in the current configuration by

$$\mathbf{u}(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in \partial\Omega_{t,u}. \tag{2.11}$$

The deformation gradient plays the role of the fundamental measure of strain by relating infinitesimal line elements $\mathrm{d}\mathbf{x} \in T_{\mathbf{x}}\Omega_t$ and $\mathrm{d}\mathbf{X} \in T_{\mathbf{X}}\Omega_0$. According to definition 2.1.2, this relation is given by

$$\mathrm{d}\mathbf{x} = \mathbf{F}(\mathbf{X}) \cdot \mathrm{d}\mathbf{X}. \tag{2.12}$$

Since the elements $\mathrm{d}\mathbf{x}$ and $\mathrm{d}\mathbf{X}$ imply a fixed spatial association, the relation (2.12) is often abbreviated to $\mathrm{d}\mathbf{x} = \mathbf{F} \cdot \mathrm{d}\mathbf{X}$.

Despite the fundamental character of the deformation gradient, general stress-strain relations (see chapter 2.1.3) are more conveniently described in quantities derived from the basic relation (2.12). For instance, the shear of an infinitesimal area in terms of the change in angle between infinitesimal line segments $\mathrm{d}\mathbf{x}_1, \mathrm{d}\mathbf{x}_2$ and $\mathrm{d}\mathbf{X}_1, \mathrm{d}\mathbf{X}_2$ can be expressed in terms of the *right Cauchy-Green* tensor $\mathbf{C}(\mathbf{X}) \in T_{\mathbf{X}}\Omega_0 \times T_{\mathbf{X}}\Omega_0$ via

$$\mathrm{d}\mathbf{X}_1 \cdot \mathbf{C} \cdot \mathrm{d}\mathbf{X}_2 = \mathrm{d}\mathbf{x}_1 \cdot \mathrm{d}\mathbf{x}_2. \tag{2.13}$$

Inserting (2.12) results in $\mathbf{C} = \mathbf{F}^\top\mathbf{F}$. Thus, the right Cauchy-Green tensor captures the relation of the scalar product in the spatial and the reference configuration. This can be used to define the difference in angle as

$$\mathrm{d}\mathbf{x}_1 \cdot \mathrm{d}\mathbf{x}_2 - \mathrm{d}\mathbf{X}_1 \cdot \mathrm{d}\mathbf{X}_2 = \mathrm{d}\mathbf{X}_1 \cdot (\mathbf{C} - \mathbf{I}) \cdot \mathrm{d}\mathbf{X}_2. \tag{2.14}$$

The tensor $\mathbf{E}(\mathbf{X}) = \frac{1}{2}(\mathbf{C}(\mathbf{X}) - \mathbf{I})$ is called *Green-Lagrange strain* tensor with $\mathbf{E} \in T_{\mathbf{X}}\Omega_0 \times T_{\mathbf{X}}\Omega_0$.

Restarting from (2.12) but expressing $\mathrm{d}\mathbf{X}$ in terms of $\mathrm{d}\mathbf{x}$ via $\mathrm{d}\mathbf{X} = \mathbf{F}^{-1} \cdot \mathrm{d}\mathbf{x}$ leads to the definition of the *left Cauchy-Green* tensor $\boldsymbol{\ell}(\mathbf{x}) = \mathbf{F}(\mathbf{x})\mathbf{F}^\top(\mathbf{x}) \in T_{\mathbf{x}}\Omega_t \times T_{\mathbf{x}}\Omega_t$ and accordingly the so called *Euler-Almansi strain* tensor $\mathbf{e}(\mathbf{x}) = \frac{1}{2}(1 - \boldsymbol{\ell}^{-1}(\mathbf{x}))$. As a consequence, it can then be seen from equation (2.14) that the relation between the Green-Lagrange tensor and the Euler-Almansi tensor is given in terms of the *push-forward* $\varphi_*[\bullet]$ and *pull-back* $\varphi_*^{-1}[\bullet]$ operations:

$$\mathbf{e}(\mathbf{x}) = \varphi_*[\mathbf{E}] = \mathbf{F}^{-\top}(\mathbf{x})\mathbf{E}(\varphi^{-1}(\mathbf{x}))\mathbf{F}^{-1}(\mathbf{x}), \tag{2.15}$$

$$\mathbf{E}(\mathbf{X}) = \varphi_*^{-1}[\mathbf{e}] = \mathbf{F}^\top(\mathbf{X})\mathbf{e}(\varphi(\mathbf{X}))\mathbf{F}(\mathbf{X}). \tag{2.16}$$

**Variation of strain measures** In anticipation of the kinetic description of balance equations in the sense of variational formulations (chapter 2.1.2), the variation of these measures of strain under an admissible variation $\delta\mathbf{U}$ of the displacement field $\mathbf{U}$ is of particular importance. This variation is given in terms of the *Gâteaux differential* [see e.g. 193]. For the variation of the deformation gradient $\delta\mathbf{F}(\mathbf{X})[\delta\mathbf{U}]$, the application of the directional derivative results in

$$\delta\mathbf{F}(\mathbf{X})[\delta\mathbf{U}] = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\left(\mathbf{F}[\mathbf{U}+\varepsilon\delta\mathbf{U}]\right)|_{\varepsilon=0} = \nabla_{\mathbf{X}}\delta\mathbf{U}. \tag{2.17}$$

Based on this result, the variation of the Green-Lagrange strains $\delta\mathbf{E}$ is obtained as

$$\delta\mathbf{E}(\mathbf{X})[\delta\mathbf{U}] = \frac{1}{2}\left((\nabla_{\mathbf{X}}\delta\mathbf{U})^{\top}\mathbf{F} + \mathbf{F}^{\top}\nabla_{\mathbf{X}}\delta\mathbf{U}\right). \tag{2.18}$$

The variation of the Euler-Almansi strain $\delta\mathbf{e}(\mathbf{x})[\delta\mathbf{u}]$ is given according to (2.15) by

$$\delta\mathbf{e}(\mathbf{x})[\delta\mathbf{u}] = \frac{1}{2}\left(\nabla_{\mathbf{x}}\delta\mathbf{u} + (\nabla_{\mathbf{x}}\delta\mathbf{u})^{\top}\right), \tag{2.19}$$

whereby the variations $\delta\mathbf{u}$ and $\delta\mathbf{U}$ are simply related via $\delta\mathbf{u}(\mathbf{x}) = \delta\mathbf{U}(\varphi^{-1}(\mathbf{x}))$. The notation $\nabla_{\mathbf{x}}(\bullet)$ is thereby introduced in contrast to $\nabla_{\mathbf{X}}(\bullet)$ to represent $\nabla_{\mathbf{x}}(\bullet) = \frac{\partial\bullet}{\partial\mathbf{x}}$.

## 2.1.2. Balance of momentum

In classical mechanics, one possibility to arrive at a solvable set of the equations of motion for the system of interest is given by the balance of momentum.

**Balance of linear momentum** In the case of quasi-static analysis, the balance of linear momentum reduces to the equilibrium of forces. Since it must hold for arbitrary $\mathcal{I} \subset \Omega_t$ with boundary $\partial\mathcal{I}$, it can be written as

$$\int_{\partial\mathcal{I}} \mathbf{t}(\mathbf{x})\,\mathrm{d}\Gamma + \int_{\mathcal{I}} \mathbf{b}(\mathbf{x})\,\mathrm{dV} = 0. \tag{2.20}$$

Therein, the vector field $\mathbf{b}$ represents a body force (force/volume) and the vector field $\mathbf{t}$ is assumed to be related to the so called *Cauchy stress* tensor $\boldsymbol{\sigma} \in T^{*}_{\mathbf{x}}\Omega_t \times T^{*}_{\mathbf{x}}\Omega_t$ via the relation

$$\mathbf{t}(\mathbf{x}) = \mathbf{n}(\mathbf{x}) \cdot \boldsymbol{\sigma}(\mathbf{x}), \tag{2.21}$$

with $\mathbf{n}$ being the field of outwards-pointing normals on $\partial\mathcal{I}$. The assumption (2.21) is a fundamental assumption in mechanics referred to as *Cauchy theorem.* It is further assumed that the vector field $\mathbf{t}(\mathbf{x})$ is given at the boundary by

$$\mathbf{t}(\mathbf{x}) = \hat{\mathbf{t}}(\mathbf{x}) \quad \forall\mathbf{x} \in \partial\mathcal{I} \cap \partial\Omega_{t,\sigma}. \tag{2.22}$$

With the application of the divergence theorem to each of the 3 components of the vector equation (2.20), the equilibrium of forces can be written as

$$\int_{\mathcal{I}} (\mathrm{div}\boldsymbol{\sigma}(\mathbf{x}) + \mathbf{b}(\mathbf{x}))\,\mathrm{dV} = 0. \tag{2.23}$$

Therein, $\mathrm{div}\,\bullet$ denotes the divergence operation $\mathrm{div}\,\bullet = \nabla_{\mathbf{x}} \cdot \bullet$. In anticipation, it is already noted that the divergence operator with respect to material coordinates will be denoted by $\mathrm{Div}\,\bullet = \nabla_{\mathbf{X}} \cdot \bullet$.

**Remark.** *For the evaluation of (2.20) and for the application of the divergence theorem, it is implicitly assumed that the involved vector fields* $\mathbf{t}$ *and* $\mathbf{b}$ *posses the necessary regularity.*

Since equation (2.23) holds for arbitrary volumes $\mathcal{I}$, it also holds for arbitrarily small volumes and thus it also holds in a point-wise sense:

$$\mathrm{div}\boldsymbol{\sigma}(\mathbf{x}) + \mathbf{b}(\mathbf{x}) = 0 \quad \forall \mathbf{x} \in \Omega_t. \tag{2.24}$$

**Balance of angular momentum** In the quasi-static case, the balance of angular momentum for the arbitrary set $\mathcal{I}$ reduces to the equilibrium of torque. It is given with respect to the origin of the Cartesian reference frame by

$$\int_{\partial\mathcal{I}} (\mathbf{x} \times \mathbf{t}) \, \mathrm{d}\Gamma + \int_{\mathcal{I}} (\mathbf{x} \times \mathbf{b}) \, \mathrm{dV} = 0. \tag{2.25}$$

Using the divergence theorem and some algebraic rearrangements, it can be shown that this equilibrium is obtained by setting

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}^\top, \tag{2.26}$$

see appendix A. Under this constraint and in combination with the boundary conditions (2.22) and (2.11), (2.24) constitutes the complete set of the equations of motion for the body of interest. Due to the point-wise evaluation and the hard requirements on the differentiability of the solution, these equations are also referred to as *strong equations*.

**Variational formulation** The FEM relies on a formulation with weaker differentiability requirements given by the *principle of virtual work*. This can be obtained by a *weighted residual formulation* of (2.24) given by

$$\int_{\Omega_t} (\mathrm{div}\boldsymbol{\sigma}(\mathbf{x}) + \mathbf{b}(\mathbf{x})) \cdot \delta\mathbf{u}(\mathbf{x}) \, \mathrm{dV} = 0 \quad \forall \delta\mathbf{u}. \tag{2.27}$$

The test functions $\delta\mathbf{u}$ must thereby comply with the concept of *admissible variations* [see e.g. 193]. Formally, this can be expressed as

$$\forall\, \mathbf{u}(\mathbf{x}) \in U, \quad \hat{\mathbf{u}}(\mathbf{x}) = \mathbf{u}(\mathbf{x}) + \delta\mathbf{u}(\mathbf{x}) \in U, \tag{2.28}$$

for some vector space $U$. A specific implication of this definition is that the test-functions $\delta\mathbf{u}$ must comply with the Dirichlet boundary condition (2.11).

Using the identity $\mathrm{div}(\boldsymbol{\sigma} \cdot \delta\mathbf{u}) = \mathrm{div}(\boldsymbol{\sigma}) \cdot \delta\mathbf{u} + \boldsymbol{\sigma} : \nabla\delta\mathbf{u}$ in combination with the divergence theorem allows to write (2.27) as

$$\int_{\partial\Omega_t} (\boldsymbol{\sigma}(\mathbf{x}) \cdot \delta\mathbf{u}(\mathbf{x})) \cdot \mathbf{n} \, \mathrm{d}\Gamma - \int_{\Omega_t} \boldsymbol{\sigma}(\mathbf{x}) : \nabla_{\mathbf{x}}\delta\mathbf{u}(\mathbf{x}) \, \mathrm{dV} + \int_{\Omega_t} \mathbf{b}(\mathbf{x}) \cdot \delta\mathbf{u}(\mathbf{x}) \, \mathrm{dV} = 0. \tag{2.29}$$

Due to the symmetry $\boldsymbol{\sigma}^\top = \boldsymbol{\sigma}$, it is possible to identify

$$\rightarrow \quad (\boldsymbol{\sigma} \cdot \delta\mathbf{u}) \cdot \mathbf{n} = (\boldsymbol{\sigma} \cdot \mathbf{n}) \cdot \delta\mathbf{u} = \mathbf{t} \cdot \delta\mathbf{u} \tag{2.30}$$

$$\rightarrow \quad \boldsymbol{\sigma} : \nabla\delta\mathbf{u} = \boldsymbol{\sigma} : \frac{1}{2}(\nabla\delta\mathbf{u} + (\nabla\delta\mathbf{u})^\top) = \boldsymbol{\sigma} : \delta\mathbf{e}. \tag{2.31}$$

The virtual work in the spatial configuration is therefore given by

$$\delta W = \int_{\Omega_t} \boldsymbol{\sigma}(\mathbf{x}) : \delta \mathbf{e}(\mathbf{x}) \, \mathrm{dV} - \int_{\Omega_t} \mathbf{b}(\mathbf{x}) \cdot \delta \mathbf{u}(\mathbf{x}) \, \mathrm{dV} - \int_{\partial\Omega_{t,\sigma}} \hat{\mathbf{t}}(\mathbf{x}) \cdot \delta \mathbf{u}(\mathbf{x}) \, \mathrm{d}\Gamma = 0. \tag{2.32}$$

Such a formulation in the spatial configuration, i.e., an a priori unknown configuration, is not the most convenient formulation from a computational point of view. More often, a so called *Total Lagrangian* formulation is applied. In the Total Lagrangian formulation, equation (2.32) is transformed as an integral over the reference configuration $\Omega_0$. The necessary transformations of the infinitesimal measures of volume and surface are established by the deformation gradient and are given by

$$\mathrm{dV} = det(\mathbf{F}(\mathbf{X})) \, \mathrm{dV}_0 = J \, \mathrm{dV}_0, \tag{2.33}$$

$$\mathrm{d}\Gamma = J \sqrt{\mathbf{N}(\mathbf{X}) \cdot \mathbf{C}(\mathbf{X})^{-1} \cdot \mathbf{N}(\mathbf{X})} \, \mathrm{d}\Gamma_0. \tag{2.34}$$

The virtual work can thus be written in the reference configuration as

$$\begin{aligned} \delta W = & \int_{\Omega_0} \boldsymbol{\sigma}^*(\mathbf{X}) : \delta \mathbf{e}^*(\mathbf{X}) J \, \mathrm{dV}_0 - \int_{\Omega_0} \mathbf{b}(\mathbf{X})^* \cdot \delta \mathbf{U}(\mathbf{X}) J \, \mathrm{dV}_0 \\ & - \int_{\partial\Omega_{0,\sigma}} \hat{\mathbf{t}}(\mathbf{X})^* \cdot \delta \mathbf{U} J \sqrt{\mathbf{N} \cdot \mathbf{C}^{-1} \cdot \mathbf{N}} \, \mathrm{d}\Gamma_0 = 0. \end{aligned} \tag{2.35}$$

Thereby, the $^*$-notation was used to highlight the change in functional dependency $f(\varphi(\mathbf{X})) = f^*(\mathbf{X})$. The Cauchy stress $\boldsymbol{\sigma}^*(\mathbf{X})$ and the Euler-Almansi strain $\mathbf{e}^*(\mathbf{X})$ still represent quantities being related to the spatial configuration and thus it is more consistent to express the internal virtual work in terms of stress/strain measures being defined directly in the reference configuration. This is achieved by the definition of the second Piola-Kirchoff stress tensor $\mathbf{S} \in T^*_{\mathbf{X}}\Omega_0 \times T^*_{\mathbf{X}}\Omega_t$ such that the energy conjugate pairing $\mathbf{S} : \mathbf{E}$ maintains the invariance of the infinitesimal virtual work:

$$\mathrm{d}\delta W(\mathbf{X}) = \boldsymbol{\sigma}^*(\mathbf{X}) : \delta \mathbf{e}^*(\mathbf{X}) \, \mathrm{dV} = \mathbf{S}(\mathbf{X}) : \delta \mathbf{E}(\mathbf{X}) \, \mathrm{dV}_0. \tag{2.36}$$

This invariance is achieved by the definition of the second Piola-Kirchoff tensor via the pull-back of the cauchy-stress according to

$$\mathbf{S}(\mathbf{X}) := \varphi_*^{-1}(\boldsymbol{\sigma}) = J \mathbf{F}(\mathbf{X})^{-1} \boldsymbol{\sigma}^*(\mathbf{X}) \mathbf{F}(\mathbf{X})^{-\top}. \tag{2.37}$$

Finally, using $\mathbf{b}_0 := J\mathbf{b}^*$ and $\mathbf{t}_0 := J\sqrt{\mathbf{N} \cdot \mathbf{C}\mathbf{N}} \, \hat{\mathbf{t}}^*$, the principle of virtual work expressed in the reference configuration is given as

$$\begin{aligned} \delta W = & \int_{\Omega_0} \mathbf{S}(\mathbf{X}) : \delta \mathbf{E}(\mathbf{X}) \, \mathrm{dV}_0 - \int_{\Omega_0} \mathbf{b}_0(\mathbf{X}) \cdot \delta \mathbf{U}(\mathbf{X}) \, \mathrm{dV}_0 \\ & - \int_{\partial\Omega_{0,\sigma}} \mathbf{t}_0(\mathbf{X}) \cdot \delta \mathbf{U}(\mathbf{X}) \, \mathrm{d}\Gamma_0 = 0. \end{aligned} \tag{2.38}$$

This is the equation subject to the numerical solution approach presented in section 2.3. For the sake of completeness, the corresponding PDE expressed in the reference configuration – obtained by the reversion of the procedure to obtain the virtual work from the strong form – is given:

$$\mathrm{Div}(\mathbf{F}\mathbf{S}) + \mathbf{b}_0 = 0 \quad \forall \mathbf{X} \in \Omega_0, \tag{2.39}$$

$$(\mathbf{F}\mathbf{S}) \cdot \mathbf{N} = \mathbf{t}_0 \quad \forall \mathbf{X} \in \partial\Omega_{0,\sigma}, \tag{2.40}$$

$$\mathbf{U} = 0 \quad \forall \mathbf{X} \in \partial\Omega_{0,u}. \tag{2.41}$$

### 2.1.3. Hyperelasticity

The missing link to solve (2.38) is given by the relation between stresses and deformation. In the context of elasticity, such a relation is characterized by a path-independence property. This property allows for the definition of an elastic potential $\tilde{\Psi}(\mathbf{F}(\mathbf{X}), \mathbf{X})$ called strain energy function (SEF). Observing the polar decomposition of the deformation gradient $\mathbf{F} = \mathbf{RU}$ into a rotational component $\mathbf{R}$ and a stretch component $\mathbf{U}$, with $\mathbf{UU} = \mathbf{C}$ [see e.g. 27], the SEF is also conveniently expressed in terms of the Cauchy-Green tensor $\mathbf{C}$:

$$\tilde{\Psi}(\mathbf{F}(\mathbf{X}), \mathbf{X}) = \Psi(\mathbf{C}(\mathbf{X}), \mathbf{X}), \tag{2.42}$$

$$\dot{\Psi} = \frac{\partial \Psi}{\partial \mathbf{C}} : \dot{\mathbf{C}}. \tag{2.43}$$

With the work conjugate pairing $\mathbf{S} : \mathbf{E}$ and $\dot{\mathbf{E}} = \frac{1}{2}\dot{\mathbf{C}}$, the work done by the stresses $\mathbf{S}$ can be written as

$$\Psi(\mathbf{C}(\mathbf{X}), \mathbf{X}) = \int_{t_0}^{t} \mathbf{S} : \dot{\mathbf{E}} \mathrm{d}t \quad \rightarrow \quad \dot{\Psi} = \mathbf{S} : \frac{1}{2}\dot{\mathbf{C}}. \tag{2.44}$$

The comparison of (2.44) and (2.43) reveals the relation

$$\left(\frac{1}{2}\mathbf{S} - \frac{\partial \Psi}{\partial \mathbf{C}}\right) : \dot{\mathbf{C}} = 0. \tag{2.45}$$

If $\dot{\mathbf{C}}$ is not subject to particular constraints, this implies that the desired stress-strain relation is given by

$$\mathbf{S} = 2\frac{\partial \Psi}{\partial \mathbf{C}} = \frac{\partial \Psi}{\partial \mathbf{E}}. \tag{2.46}$$

**Isotropy** Isotropic materials represent an important subclass of hyperelastic materials. Isotropy is thereby characterized by the invariance of the constitutive behavior with respect to the spatial direction in the reference configuration. This implies that the SEF can be written in terms of the invariants $\mathrm{I}_1 = tr(\mathbf{C}), \mathrm{I}_2 = tr(\mathbf{CC}), \mathrm{I}_3 = \det(\mathbf{C}) = J^2$ of the right Cauchy-Green tensor $\mathbf{C}$:

$$\Psi(\mathbf{C}(\mathbf{X}), \mathbf{X}) = \Psi_{\mathrm{I}_1,\mathrm{I}_2,\mathrm{I}_3}(\mathrm{I}_1, \mathrm{I}_2, \mathrm{I}_3, \mathbf{X}). \tag{2.47}$$

**Incompressibility** Incompressible materials are characterized by conservation of volume as a body undergoes a deformation. This is formally expressed by the condition $J = 1$ and accordingly $\dot{J} = 0$. Using the identity $\frac{\partial \det(A)}{\partial A} = \det(A)A^{-1}$, this results in the additional constraint on $\dot{\mathbf{C}}$ given as

$$J\mathbf{C}^{-1} : \dot{\mathbf{C}} = 0. \tag{2.48}$$

Compliance with (2.45) implies the relation

$$\forall \gamma \quad \frac{1}{2}\mathbf{S} - \frac{\partial \Psi}{\partial \mathbf{C}} = \gamma J \mathbf{C}^{-1}. \tag{2.49}$$

By reformulation of the SEF with respect to the distortional component of the right Cauchy-Green tensor $\hat{\mathbf{C}} = J^{-1/3}\mathbf{C}$ as $\hat{\Psi}(\mathbf{C}) := \Psi(\hat{\mathbf{C}})$, it can be shown that the scalar $\gamma$ can be identified with the hydrostatic pressure p [see e.g. 27]. The general stress-strain relation for an incompressible material hence emerges as

$$\mathbf{S} = 2\frac{\partial \hat{\Psi}(\mathbf{C})}{\partial \mathbf{C}} + \mathrm{p}J\mathbf{C}^{-1}. \tag{2.50}$$

**Near incompressibility** Nearly incompressible materials are often used to approximate incompressible material behavior. In the modeling of this sort of materials, the constraint $J = 1$ is not strictly enforced. Instead the overall SEF is defined as

$$\Psi(\mathbf{C}) = \hat{\Psi}(\mathbf{C}) + \Psi_{vol}(J), \tag{2.51}$$

whereby the volumetric strain energy function $\Psi_{vol}$ acts as a penalty for volumetric deformations. A very simple choice for the volumetric function is $\Psi_{vol} = \frac{1}{2}\kappa(J-1)^2$. This choice reflects the solution of the variational formulation (2.38) subject to the incompressibility constraint by means of a perturbed Lagrangian formulation [233].

**Polyconvexity** The existence of solutions of the weak formulation (2.38) depends on some specific properties of the SEF $\tilde{\Psi}$. In fact, it can be shown that the solution of (2.38) corresponds to the minimization of the functional

$$W(\mathbf{U}) = \int_{\Omega_0} \tilde{\Psi}(\mathbf{X}, \mathbf{F}(\mathbf{X}))\,\mathrm{dV} \tag{2.52}$$

with respect to $\mathbf{U}$ [see e.g. 174]. The existence of a minimizer of the functional (2.52) is generally obtained under restrictions of the solution space and the requirement that the SEF be *polyconvex*. Abridging the introduction of the theory, the concept of polyconvexity in 3 dimensions can be reduced to the requirement that the SEF can be expressed in terms of a convex function $\Psi_{pcx}$ via

$$\tilde{\Psi}(\mathbf{F}) = \Psi_{pcx}(\mathbf{F}, \operatorname{cof}\mathbf{F}, \det\mathbf{F}). \tag{2.53}$$

A general introduction to this topic can be found in Dacorogna [43]. An analysis of some common nearly incompressible formulations is given by Hartmann and Neff [92].

## 2.2. Incorporation of arterial growth

The continuum mechanical description of a material introduced in section 2.1.3 materialized from an elasticity argument. Functional adaption, i.e., the adaption of living tissue to external stimuli [68], clearly is a process which cannot be cast in the framework of elasticity. Thus, the incorporation of such irreversible processes necessitates an extension of the presented continuum mechanical description.

Modern formulations of growth are often referring back to Thompson [216] who opined on a mechanistic view on growth and the inherent connection between growth and form. The reformulation of this line of thoughts in terms of a kinematic description of growth is often credited to Skalak et al. [204]. Motivated by this work, Rodriguez et al. [187] introduced the concept of the multiplicative split of the deformation gradient to allow for a continuum mechanical treatment of growth. The theoretical foundation of this approach is based in the field of finite strain plasticity [134]. The multiplicative split

$$\mathbf{F} = \mathbf{F}_g \mathbf{F}_e \tag{2.54}$$

as the decomposition of the deformation gradient into an elastic component $\mathbf{F}_e$ and an inelastic/growth component $\mathbf{F}_g$ implies the existence of an intermediate configuration $\Omega_g$, see figure 2.3. It is important to highlight that this intermediate configuration need not be kinematically compatible to still obtain a compatible spatial configuration. As a consequence, the mapping $\varphi_g(\mathbf{X}) : \Omega_0 \rightarrow \Omega_g$ is not necessarily differentiable. Thus, $\mathbf{F}_g$ cannot be defined via differentiation of $\varphi_g$. The concept of incompatible configurations is also utilized in the modeling of residual stresses in living tissue [205].

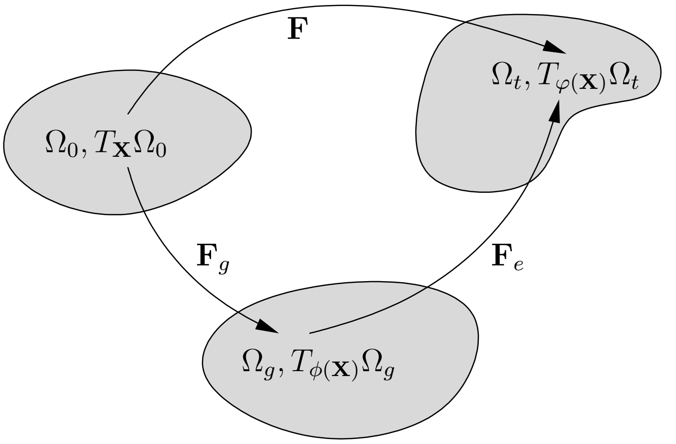

Figure 2.3.: Illustration of the multiplicative split of the deformation gradient into a growth component $\mathbf{F}_g$ and an elastic component $\mathbf{F}_e$ giving rise to an intermediate configuration $\Omega_g$ that need not necessarily be compatible (as might erroneously be deduced from the drawing).

Generally, the continuum mechanical treatment of growth of living tissue is a current topic of research and is still controversially discussed. Whereas there has been some agreement on the applicability of the multiplicative framework [see e.g. 71, 97, 128, 143, 208], it has also been questioned due to its phenomenological approach. Particularly with respect to the accurate modeling of constant mass turnover, i.e., the constant formation and degradation of mass of particular constituents of the tissue, the constrained mixture theory [102] claims superiority at the cost of increased model complexity. Since the focus of the work presented here is on parameter identification and predictive modeling, the reader is referred to Ambrosi et al. [5] for a more detailed review on the current state of growth modeling. Here instead, the incorporation and specification of a growth model based on the multiplicative split (2.54) into the continuum mechanical framework is pursued. A detailed introduction to this approach is also provided by Tinkl [218].

**Isotropic growth** Isotropic growth is a subclass of the multiplicative formulation obtained by specifying growth in terms of the so called *growth stretch* $\vartheta(\mathbf{X}) \in \mathbb{R}$. The growth stretch is used to uniformly define the deformation gradient $\mathbf{F}_g = \vartheta \mathbf{I}$ such that the overall deformation gradient is given by

$$\mathbf{F} = \vartheta \mathbf{F}_e. \tag{2.55}$$

Since growth itself is not supposed to induce stresses directly, the SEF is formally reformulated with respect to a growth-free deformation according to

$$\Psi_e(\mathbf{C}) := \Psi(\mathbf{C}_e), \tag{2.56}$$

with $\mathbf{C}_e = \mathbf{F}_e^\top \mathbf{F}_e = \frac{1}{\vartheta^2}\mathbf{F}^\top \mathbf{F} = \frac{1}{\vartheta^2}\mathbf{C}$. For a fixed $\vartheta$, the SEF $\Psi_e$ admits the path independent property necessary to allow for the definition according to (2.44). The stress-strain relation resulting from an elastic deformation only is then given by

$$\mathbf{S}_e = 2\frac{\partial \Psi_e}{\partial \mathbf{C}_e}. \tag{2.57}$$

And the relation between the second Piola-Kirchhoff stress $\mathbf{S}$ and the elastic component $\mathbf{S}_e$ is given according to (2.56) by

$$\mathbf{S}_e = 2\frac{\partial \Psi_e}{\partial \mathbf{C}_e} = 2\frac{\partial \Psi}{\partial \mathbf{C}} : \frac{\partial \mathbf{C}}{\partial \mathbf{C}_e} = \mathbf{S}\vartheta^2. \tag{2.58}$$

To close the stress-strain relation, it remains to define the growth stretch $\vartheta$. Again there is no unified theory available, and the application of the multiplicative framework is mainly based on *ad-hoc* formulae for the evolution of $\vartheta$. These can be motivated by analyzing the change of mass $\dot{m}$ induced by the growth. The mass $m$ of the body can be expressed by

$$m = \int_{\Omega_t} \rho_t \,\mathrm{dV} = \int_{\Omega_g} \rho_t^{**} \,\mathrm{dV}_g = \int_{\Omega_0} \rho_t^{*} J_g \,\mathrm{dV}_0 = \int_{\Omega_0} \rho_0 \,\mathrm{dV}_0, \tag{2.59}$$

whereby incompressibility of the elastic deformation, i.e., $J_e = \det(\mathbf{F}_e) = 1$, was assumed. Using $J_g = \vartheta^3$, the change of mass $\dot{m}$ is computed as

$$\dot{m} = \int_{\Omega_0} \dot{\rho}_0 \,\mathrm{dV}_0 \implies \dot{\rho}_0 = 3\vartheta^2 \dot{\vartheta} \rho_t. \tag{2.60}$$

Thus, with the elastic part of the overall deformation being incompressible, the change in density $\dot{\rho}_0$ in the reference configuration is driven by the evolution of the stretch ratio $\dot{\vartheta}$, motivating relations of the form

$$\dot{\vartheta} = f(\vartheta, ...). \tag{2.61}$$

A popular choice is to express the evolution of $\vartheta$ on some measure of the pressure. E.g.,

$$\dot{\vartheta} = f(\vartheta, \mathbf{C}_e : \mathbf{S}_e) = k_\vartheta(\vartheta)\mathbf{C}_e : \mathbf{S}_e \tag{2.62}$$

is a popular approach [97, 128]. Therein, the function $k_\vartheta$ is defined by

$$k_\vartheta(\vartheta) = \begin{cases} k_\vartheta^+ \left[\frac{\vartheta^+ - \vartheta}{\vartheta^+ - 1}\right]^{m_\vartheta^+} & \text{for } \mathbf{C}_e : \mathbf{S}_e > 0 \\ k_\vartheta^- \left[\frac{\vartheta - \vartheta^-}{1 - \vartheta^-}\right]^{m_\vartheta^-} & \text{for } \mathbf{C}_e : \mathbf{S}_e < 0 \end{cases} \tag{2.63}$$

with the constants $k_\vartheta^+, k_\vartheta^-, \vartheta^+, \vartheta^-, m_\vartheta^+$ and $m_\vartheta^-$, which have to be defined on a patient-specific basis. The predictive capabilities of such a model are highly related to the identifiability of the parameters involved. Furthermore, these formulae don't model the biochemical processes in the living tissue but are merely the result of a top-down modeling approach. Thus, with the goal of identifying parameters, it is reasonable to accumulate the function $k_\vartheta(\vartheta)$ from (2.63) in a single parameter $c_\vartheta$ resulting in

$$k_\vartheta(\vartheta) = c_\vartheta. \tag{2.64}$$

The resulting relation (2.62) can be further simplified by ignoring the dependency of the measure of pressure such that the evolution of the growth stretch is directly given by

$$\dot{\vartheta} = c_\vartheta. \tag{2.65}$$

With respect to the accurate modeling of the physical processes of growth, the simple growth law (2.65) is clearly degenerated. However, with regard to predictive modeling in the context of a parameter identification framework, it offers clear advantages. By defining the parameter $c_\vartheta$ as the parameter to be identified, the identification process has a direct flexible control over the volumetric growth stretch $\vartheta$. In contrast to complicated formulations such as (2.63), this increases the flexibility and can therefore have a beneficial effect on the nature of the identification problem. Furthermore, the high number of parameters in the formulation (2.63) can have a negative effect on the identifiability of the parameters. Another important advantage of the simple formulation (2.65) is given by its independence of a homeostatic (i.e., healthy) state.

## 2.3. Numerical approximation

The equations of motion (2.39)-(2.41) do generally not allow for analytic solutions, except for very specific and simplified cases. Instead, approximate solutions can be obtained by the application of numerical techniques. A well established numerical approach in the field of structural mechanics to obtain approximate solutions is the FEM. Beside its popularity in this particular field, it is a very general method to obtain so called *weak* solutions of partial differential equations. This chapter aims at a brief introduction of the method and the particular discretization techniques applied throughout the work presented in this thesis. Detailed introductions to the topic are given by Zienkiewicz et al. [233] or Larson and Bengzon [131]. For the application in the context of nonlinear structural mechanics, the reader is referred to Bonet and Wood [27].

### 2.3.1. Weak formulation

Starting point for the FEM is a variational formulation of a PDE. In the context of nonlinear structural elasticity, this is given by (2.38). In order for (2.38) to be valid, it must be possible to evaluate the integrals, i.e., there are certain requirements on the functions $\mathbf{U}$ and $\delta\mathbf{U}$. The natural choice $\mathbf{U}, \delta\mathbf{U} \in C^2(\Omega_0)$ generally constitutes infeasible requirements from a numerical point of view. With respect to the existence of solutions of PDEs, the theory of Sobolev spaces provides a bases for the mathematical analysis [see e.g. 30]. A fundamental result in the analysis of linear elliptic PDEs – the Lemma of Lax-Milgram [see e.g. 3] – is based on the theory of Sobolev spaces. But they are also a key aspect in the mathematical analysis of the nonlinear elasticity problem (2.38). For compactness of the presentation, in the following only the essential *tools* necessary to arrive at a valid mathematical formulation are stated.

Using the multi-index notation $\alpha = (\alpha_i, ..., \alpha_d)$ and $|\alpha| = \sum_{i=1}^{d} \alpha_i$, the notation

$$D^\alpha \varphi = \prod_{i=1}^{d} (\frac{\partial}{\partial x_i})^{\alpha_i} \varphi \tag{2.66}$$

can be used to denote the classical partial derivative of any sufficiently smooth function $\varphi$. The Sobolev space $W_k^p(\Omega_0)$ is defined as

$$W_k^p(\Omega_0) := \{v \in \mathrm{L}^p(\Omega_0) : D^\alpha v \in \mathrm{L}^p(\Omega_0), \forall\, |\alpha| \leq k\} \tag{2.67}$$

The subclass of spaces $W_k^2(\Omega_0) =: \mathrm{H}^k(\Omega_0)$ represents an important family due to the Hilbert space structure. For $k = 1$, the familiar space $\mathrm{H}^1(\Omega_0)$ is obtained. This space is a Hilbert space with inner product and norm

$$\langle u, v \rangle_{\mathrm{H}^1} = \langle u, v \rangle_{\mathrm{L}^2} + \langle \nabla u, \nabla v \rangle_{\mathrm{L}^2}, \tag{2.68}$$

$$\|u\|_{\mathrm{H}^1}^2 = \|u\|_{\mathrm{L}^2}^2 + \|\nabla u\|_{\mathrm{L}^2}^2 . \tag{2.69}$$

With the trace operator $\gamma : \mathrm{H}^1(\Omega_0) \to \mathrm{L}^2(\partial\Omega_0)$, the constrained space $\mathrm{H}_0^1(\Omega_0)$ can be defined for sufficiently smooth boundaries $\partial\Omega_0$ as

$$\mathrm{H}_0^1(\Omega_0) = \left\{v \in \mathrm{H}^1(\Omega_0) : (\gamma v)|_{\partial\Omega_{0,u}} = 0\right\}. \tag{2.70}$$

In the following, the shorthand notation $\mathbf{v} \in (\mathrm{H}_0^1(\Omega_0))^3$ is used to state that for a vector field $\mathbf{v}$ each spatial component satisfies $v_i \in \mathrm{H}_0^1(\Omega_0)$. Finally, in the hyperelastic regime, it can be shown that the integrals in the weak form (2.38) can be properly evaluated for $\mathbf{U}, \delta\mathbf{U} \in (\mathrm{H}_0^1(\Omega_0))^3$ [see e.g. 174]. The variational problem therefore takes the form: find $\mathbf{U} \in \mathcal{V} := (\mathrm{H}_0^1(\Omega_0))^3$ such that

$$\delta W(\mathbf{U}, \delta\mathbf{U}) = 0 \quad \forall \delta\mathbf{U} \in \mathcal{V} \tag{2.71}$$

### 2.3.2. The finite element method

To render the solution of the weak form (2.71) feasible, it is necessary to represent the infinite-dimensional space $\mathcal{V}$ by a finite-dimensional approximation $\mathcal{V}_h$. To this end, the FEM uses a mesh $\mathcal{K} = \{\mathcal{E}\}$ of *elements* $\mathcal{E}$ such that $\Omega_0 = \bigcup_{\mathcal{E}\in\mathcal{K}} \mathcal{E}$. The intersection of two elements is either a face, an edge, a vertex or the empty set. The reasoning behind this meshing is to enable the simple construction of piecewise polynomial function spaces via the polynomials $\mathcal{P}_\kappa(\mathcal{E})$. By restricting the order $\kappa$ of this polynomial spaces, a finite dimensional approximation $\mathcal{V}_h$ is obtained. E.g., with the restriction to linear polynomials ($\kappa = 1$), the space of piecewise linear continuous functions in $\Omega_0$ is given as

$$\mathcal{V}_h := \left\{\mathbf{v} : v_i \in C^0(\Omega_0), v_i|_{\mathcal{E}} \in \mathcal{P}_1(\mathcal{E}), \forall \mathcal{E} \in \mathcal{K}\right\}. \tag{2.72}$$

Introducing a *nodal basis* $\{N_i(\mathbf{X})\}_{i=1}^{n_{nod}^{\mathcal{E}}}$ (the *shape functions*), functions $\mathbf{v} \in (\mathcal{P}_1(\mathcal{E}))^3$ can be represented by

$$\mathbf{v}(\mathbf{X}) = \sum_{i=1}^{n_{nod}^{\mathcal{E}}} N_i(\mathbf{X})\mathbf{d}_i, \forall \mathbf{X} \in \mathcal{E}. \tag{2.73}$$

The $\mathbf{d}_i \in \mathbb{R}^3$ are referred to as *degrees of freedom (dofs)* and represent the nodal values of $\mathbf{v}(\mathbf{X}_i)$, with the coordinate $\mathbf{X}_i$ of node $i$ and the number $n_{nod}^{\mathcal{E}}$ of nodes of the element $\mathcal{E}$. Using this *ansatz*, the displacements $\mathbf{U}$ as well as the virtual displacements $\delta\mathbf{U}$ can be approximated on the elemental level via

$$\left.\begin{aligned} \mathbf{U}(\mathbf{X}) \approx \mathbf{U}_h(\mathbf{X}) &= \sum_{i=1}^{n_{nod}^{\mathcal{E}}} N_i(\mathbf{X})\mathbf{d}_i, \\ \delta\mathbf{U}(\mathbf{X}) \approx \delta\mathbf{U}_h(\mathbf{X}) &= \sum_{i=1}^{n_{nod}^{\mathcal{E}}} N_i(\mathbf{X})\delta\mathbf{d}_i, \end{aligned}\right\} \quad \forall \mathbf{X} \in \mathcal{E}. \tag{2.74}$$

Using such an ansatz, continuity is enforced by the nodal connectivity of the mesh and the association of a unique dof per node and direction. Since the boundary condition (2.10) is not explicitly modeled by the polynomial spaces (2.72), this has to be taken into account specifically. A convenient possibility in practice is to set

$$\mathbf{d}_i = 0, \delta\mathbf{d}_i = 0, \forall \mathbf{X}_i \in \partial\Omega_{0,u}. \tag{2.75}$$

The explicit assignment of the dofs describing the variation $\delta\mathbf{U}$ is not necessary for a solution of (2.71). Although, it is formally correct to comply with the admissibility of the variation $\delta\mathbf{U}$ given by (2.28), which it is relevant for a well-posed adjoint formulation of the inverse problem, see chapter 3.5. Inserting (2.74) with (2.75) into the weak formulation (2.71) results in the finite dimensional approximation called *Galerkin approximation*: find $\mathbf{U}_h \in \mathcal{V}_h$ such that

$$\delta W(\mathbf{U}_h, \delta\mathbf{U}_h) = 0 \quad \forall \delta\mathbf{U} \in \mathcal{V}_h. \tag{2.76}$$

**Integration** The integration of (2.76) can now be carried out on an elemental level according to

$$\int_{\Omega_0} (\bullet)\, \mathrm{dV}_0 = \sum_{\mathcal{E}} \int_{\mathcal{E}} (\bullet)\, \mathrm{dV}_0. \tag{2.77}$$

This approach represents a particular advantage of the FEM. Despite not being compulsory, the *isoparametric concept* allows for a generic and elegant way do perform this task: the geometry is represented in terms of the same shape functions used for the approximation of the displacements via

$$\mathbf{X}(\boldsymbol{\xi}) = \sum_{i=1}^{n_{nod}^{\mathcal{E}}} N_i(\boldsymbol{\xi})\mathbf{X}_i \quad \forall \boldsymbol{\xi} \in \mathcal{E}_g. \tag{2.78}$$

The element-wise integration can therefore be performed with respect to the generic element $\mathcal{E}_g$ according to

$$\int_{\mathcal{E}} f(\mathbf{X})\,\mathrm{dV}_0 = \int_{\mathcal{E}_g} f(\mathbf{X}(\boldsymbol{\xi}))\det(\frac{\partial \mathbf{X}}{\partial \boldsymbol{\xi}})\,\mathrm{dV}_{\mathcal{E}_g}. \tag{2.79}$$

The numerical integration [see e.g. 233] is now evaluated generically for every element $\mathcal{E}$, whereby the mapping between the volume forms $\mathrm{dV}_0$ and $\mathrm{dV}_{\mathcal{E}_g}$, given by $\det(\frac{\partial \mathbf{X}}{\partial \boldsymbol{\xi}})$, can be pre-evaluated for a fixed reference configuration $\Omega_0$.

**Assembly** Assembling all the element-wise degrees of freedom $\mathbf{d}_i$ and $\delta\mathbf{d}_i$ in the system vectors $\mathbf{D}$ and $\delta\mathbf{D} \in \mathbb{R}^{n_{dof}}$, the solution of (2.76) can finally be expressed in terms of the nonlinear form

$$\delta\mathbf{D} \cdot \mathbf{f}(\mathbf{D}) = 0. \tag{2.80}$$

Since this form has to hold for arbitrary $\delta\mathbf{D}$ complying with $\delta\mathbf{U} \in \mathcal{V}_h$, its solution is equivalently given by the solution of the nonlinear system

$$\mathbf{f}(\mathbf{D}) = \mathbf{0}. \tag{2.81}$$

### 2.3.3. Nonlinear solution

The solution of (2.81) is usually obtained using iterative methods such as *Newton's method* [see e.g. 113]. It is given by the iteration rule

$$\mathbf{D}^{k+1} = \mathbf{D}^k + \Delta\mathbf{D}^k, \tag{2.82}$$

$$\Delta\mathbf{D}^k = -(\mathbf{f}'(\mathbf{D}^k))^{-1}\mathbf{f}(\mathbf{D}^k). \tag{2.83}$$

Therein $\mathbf{f}'(\mathbf{D})$ denotes the matrix $[\mathbf{f}]_{ij} = \frac{\partial f_i}{\partial D_j}$. In the context of the FEM, this matrix is called *stiffness matrix*. Restating Newton's method as fix-point iteration

$$\mathbf{D}^{k+1} = \mathbf{D}^k - (\mathbf{f}'(\mathbf{D}^k))^{-1}\mathbf{f}(\mathbf{D}^k) =: \mathbf{h}(\mathbf{D}^k), \tag{2.84}$$

it can be seen that the iteration converges if $\mathbf{h}$ represents a contraction, i.e., there exists a constant $L < 1$ such that $\|\mathbf{h}(\mathbf{a}_1) - \mathbf{h}(\mathbf{a}_2)\| \leq L\,\|\mathbf{a}_1 - \mathbf{a}_2\|$. In the general setup of nonlinear elasticity, this condition is only fulfilled for the initial guess $\mathbf{D}^0$ 'close' enough to the solution $\mathbf{D}^*$. Since such an initial guess is usually not available, a common heuristic approach is to solve a sequence of nonlinear equations $\mathbf{f}^\alpha(\mathbf{D}) = \mathbf{0}$ with the solution of one step of the sequence being used as the initial guess for the next step. A popular example is the *load control* scheme. Thereby, the sequence $\mathbf{f}^\alpha$ is generated by sequentially increasing the traction load $\hat{\mathbf{t}}$ according to

$$\hat{\mathbf{t}}^\alpha = \alpha\hat{\mathbf{t}}, \quad \alpha = 0, ..., 1. \tag{2.85}$$

Convergence for each subproblem $\mathbf{f}^\alpha$ thereby depends on the number of load-levels. Parametrizing a load sequence represents only one instance of a general class called *parameter continuation* methods [2].

One such approach, which enables a more formal approach to the stability of nonlinear solutions, is obtained by observing that the solution of a nonlinear system (2.81) coincides with the steady state solution of the ordinary differential equation

$$\frac{\mathrm{d}\mathbf{D}}{\mathrm{d}t} = -\mathbf{f}(\mathbf{D}). \tag{2.86}$$

Using an implicit time discretization and one step of Newton's method per pseudo time step $\tau$ results in the update scheme

$$\Delta\mathbf{D}^k = -(\tau_k^{-1} + \mathbf{f}'(\mathbf{D}^k))^{-1}\mathbf{f}(\mathbf{D}^k). \tag{2.87}$$

In combination with certain update strategies for the time step $\tau_k$, e.g.,

$$\tau_k = \tau_{k-1}\frac{\left\|\mathbf{f}(\mathbf{D}^{k-1})\right\|}{\left\|\mathbf{f}(\mathbf{D}^k)\right\|}, \tag{2.88}$$

this approach is known as *pseudotransient continuation (PTC)* [114]. The initial time step $\tau_0$ thereby controls the convergence properties of the scheme. The Newton update (2.83) is recovered for the limit $\tau \to \infty$.

**Newton-Galerkin method** Whereas the update equations (2.83) and (2.87) were obtained from a *discretize-linearize* strategy, the same results can be obtained from a *linearize-discretize* strategy. In the context of weak solutions of PDEs, this approach is called *Newton-Galerkin* method and it is used in this work for the Lagrangian formulation of the inverse problem, see chapter 3.5. The Newton-Galerkin method is obtained by linearizing the weak form (2.71) at a current displacement $\mathbf{U}^k$ as

$$\delta W(\mathbf{U}^k, \delta\mathbf{U}) + \Delta\delta W(\mathbf{U}^k, \delta\mathbf{U})[\Delta\mathbf{U}] = 0 \quad \forall \delta\mathbf{U} \in \mathcal{V}. \tag{2.89}$$

For notational reasons, the symbol $\Delta$ is thereby used to avoid the expression of the variation of the virtual work of the form $\delta\delta W$.

The variation of the virtual work $\Delta\delta W(\mathbf{U}^k, \delta\mathbf{U})[\Delta\mathbf{U}]$ with respect to the direction $\Delta\mathbf{U}$ is given by

$$\begin{aligned}\Delta\delta W(\mathbf{U}^k, \delta\mathbf{U})[\Delta\mathbf{U}] = &\int_{\Omega_0} \delta\mathbf{S}[\Delta\mathbf{U}] : \delta\mathbf{E}\,\mathrm{dV}_0 \\ &+ \int_{\Omega_0} \mathbf{S} : \Delta\delta\mathbf{E}[\Delta\mathbf{U}]\,\mathrm{dV}_0 \\ &+ \int_{\partial\Omega_0} \delta\hat{\mathbf{t}}_0[\Delta\mathbf{U}] \cdot \delta\mathbf{U}\,\mathrm{d}\Gamma_0. \end{aligned} \tag{2.90}$$

Assuming the stresses $\mathbf{S}$, strains $\mathbf{E}$ and the traction $\hat{\mathbf{t}}$ to be smooth functions of $\mathbf{U}$, it is possible to identify

$$\Delta\delta\mathbf{E}[\Delta\mathbf{U}] = \frac{1}{2}\left((\nabla_{\mathbf{X}}\delta\mathbf{U})^\top\nabla_{\mathbf{X}}\Delta\mathbf{U} + (\nabla_{\mathbf{X}}\Delta\mathbf{U})^\top\nabla_{\mathbf{X}}\delta\mathbf{U}\right), \tag{2.91}$$

$$\delta\mathbf{S}[\Delta\mathbf{U}] = \frac{\mathrm{d}\mathbf{S}}{\mathrm{d}\mathbf{E}} : \delta\mathbf{E}[\Delta\mathbf{U}] =: \mathcal{C} : \delta\mathbf{E}[\Delta\mathbf{U}], \tag{2.92}$$

$$\delta\hat{\mathbf{t}}_0[\Delta\mathbf{U}] = \frac{\mathrm{d}\hat{\mathbf{t}}}{\mathrm{d}\mathbf{U}}\Delta\mathbf{U}. \tag{2.93}$$

In the general case of growth, the functional dependency $\mathbf{E} \mapsto \mathbf{S}$ necessary to compute (2.92) can be more intricate than in the standard elastic case. However, in the simple case (2.65), the growth factor $\vartheta$ is not a function of kinematic quantities and thus the variation $\delta\mathbf{S}$ reduces to the standard evaluation known from the hyperelastic regime, see appendix C.2.1 for details on this differentiation. A detailed description of the incorporation of more complex growth laws is provided in [218]. The body force $\mathbf{b}_0$ was ignored in the above considerations but can be treated in the same manner as the traction force. The identities (2.91)-(2.93) reveal that the differential virtual work (2.90) is a symmetric bilinear form:

$$\Delta\delta W(\mathbf{U}^k, \delta\mathbf{U})[\Delta\mathbf{U}] = \Delta\delta W(\mathbf{U}^k, \Delta\mathbf{U})[\delta\mathbf{U}]. \tag{2.94}$$

Close enough to the solution, it allows the representation in terms of the positive definite bounded linear operator $\mathsf{K} : \mathcal{V} \to \mathcal{V}$ via

$$\Delta\delta W(\mathbf{U}^k, \delta\mathbf{U})[\Delta\mathbf{U}] = \langle \mathsf{K}\Delta\mathbf{U}, \delta\mathbf{U}\rangle_{\mathrm{L}^2}. \tag{2.95}$$

In anticipation of the Lagrangian formulation of the inverse problem (see chapter 3.5), it is noted that the operator $\mathsf{K}$ admits an adjoint $\mathsf{K}^\dagger$ such that

$$\langle \mathsf{K}\Delta\mathbf{U}, \delta\mathbf{U}\rangle_{\mathrm{L}^2} = \langle \Delta\mathbf{U}, \mathsf{K}^\dagger\delta\mathbf{U}\rangle_{\mathrm{L}^2}, \tag{2.96}$$

whereby due to the symmetry (2.94) $\mathsf{K}^\dagger = \mathsf{K}$.

Using the Galerkin-approximation for $\delta\mathbf{U} \approx \delta\mathbf{U}_h$ and $\Delta\mathbf{U} \approx \Delta\mathbf{U}_h$ and assuming that $\mathbf{U}^k$ can be represented by a function from $\mathcal{V}_h$, the solution of the incremental virtual work (2.89) reads: find $\Delta\mathbf{U}_h \in \mathcal{V}_h$ such that

$$\delta W(\mathbf{U}^k, \delta\mathbf{U}_h) + \Delta\delta W(\mathbf{U}^k, \delta\mathbf{U}_h)[\Delta\mathbf{U}_h] = 0 \quad \forall \delta\mathbf{U}_h \in \mathcal{V}_h. \tag{2.97}$$

Reformulating this in terms of the nodal degrees of freedom $\mathbf{D}^k$, $\delta\mathbf{D}^k$ and $\Delta\mathbf{D}^k$, the incremental solution $\Delta\mathbf{D}^k$ is given by solution of the linear system

$$\mathbf{K}\Delta\mathbf{D}^k = -\mathbf{f}(\mathbf{D}^k). \tag{2.98}$$

The $n_{dof} \times n_{dof}$-matrix $\mathbf{K}$ is thereby identified as the stiffness matrix $\mathbf{f}'$ from (2.83).

**Newton-Krylov method** In the context of the FEM, the matrix $\mathbf{K}$ is often sparsely populated suggesting the use of Krylov-subspace methods for the solution of (2.83) or (2.98) [see e.g. 192]. The resulting update scheme is then called *Newton-Krylov* method.

## 2.4. Parametrization

Since the modeling equations introduced in the chapters 2.1 and 2.2 are to be used within a parameter identification framework, it is necessary to define a parametrization $\theta$. To accurately represent the real physical processes, these parameters are assumed to vary spatially such that $\theta = \theta(\mathbf{X}) \in \mathcal{X}(\Omega_0)$. Throughout the work presented here, the parameters are assumed to enter the formulations either directly via a parametrized strain energy function

$$\Psi(\mathbf{C}(\mathbf{X}), \mathbf{X}) = \Psi_\theta(\mathbf{C}(\mathbf{X}), \mathbf{X}, \theta(\mathbf{X})), \tag{2.99}$$

or indirectly as in the case of the kinematic growth modeling given by (2.65). By introducing these dependencies into the weak formulation (2.71), the nonlinear problem is formally given as: find $\mathbf{U} \in \mathcal{V}$ such that

$$\delta W(\mathbf{U}, \delta\mathbf{U}, \theta) = 0 \quad \forall \delta\mathbf{U} \in \mathcal{V}. \tag{2.100}$$

The physical requirement on the spatial variability of $\theta$ is often expressed in terms of certain smoothness assumptions on $\theta(\mathbf{X})$. In some applications, these requirements might even include the possibility of relatively sharp changes (or even discontinuities) at a limited amount of locations. These requirements must be reflected in the choice of space $\mathcal{X}(\Omega_0)$ which is used to model the functions $\theta$. A common choice is $\mathcal{X}(\Omega_0) = \mathrm{L}^2(\Omega_0)$. The parametrization of the SEF then has to comply with the polyconvexity conditions on $\Psi$. Given such a parametrization, the solution of (2.100) can be expressed in terms of the nonlinear mapping $A : \mathcal{X}(\Omega_0) \rightarrow \mathcal{V}$ such that

$$\mathbf{U} = A(\theta). \tag{2.101}$$

Equivalently, there exists a nonlinear mapping $A_h : \mathcal{X}(\Omega_0) \rightarrow \mathcal{V}_h$ for the discrete version (2.76) such that

$$\mathbf{U}_h = A_h(\theta). \tag{2.102}$$

In order for the function $\theta(\mathbf{X})$ to be identifiable, i.e., solve (2.102) for $\theta$ given a solution $\mathbf{U}_h$, the infinite dimensional space $\mathcal{X}(\Omega_0)$ is infeasible. To this end, a finite dimensional approximation has to be introduced whereby it has proven convenient to choose a space of piecewise-constant functions. Given a mesh $\mathcal{K}$ this can be achieved by the definition:

$$\mathcal{X}_h(\Omega_0) := \left\{ v \in \mathrm{L}^2(\Omega_0) : v|_{\mathcal{E}} \in \mathcal{P}_0(\mathcal{E}), \forall \mathcal{E} \in \mathcal{K} \right\}. \tag{2.103}$$

This is a convenient approach since it does not complicate the numerical integration on the elemental level. Functions $\theta_h \in \mathcal{X}_h(\Omega_0) \subset \mathcal{X}(\Omega_0)$ can then be represented in terms of the basis $\{g_{\mathcal{E}}(\mathbf{X})\}$ with

$$g_{\mathcal{E}}(\mathbf{X}) = \begin{cases} 1 & \text{if } \mathbf{X} \in \mathcal{E} \\ 0 & \text{otherwise.} \end{cases} \tag{2.104}$$

Assembling the $n_{ele}$ constants $\beta_i$ defining the polynomials $\mathcal{P}_0(\mathcal{E})$ in the vector $\boldsymbol{\theta} \in \mathbb{R}^{n_{ele}}$ and the basis functions in the vector $\mathbf{G}(\mathbf{X}) \in \mathbb{R}^{n_{ele}}$, $\theta_h(\mathbf{X})$ can be written as

$$\theta_h(\mathbf{X}) = \mathbf{G}^{\top}(\mathbf{X})\boldsymbol{\theta} = \sum_{\mathcal{E} \in \mathcal{K}}^{n_{ele}} g_{\mathcal{E}}(\mathbf{X})\beta_i. \tag{2.105}$$

This approximation can then be inserted into the nonlinear mapping (2.101) and (2.102):

$$\mathbf{U} = A(\theta_h(\boldsymbol{\theta})) =: \mathbf{A}(\boldsymbol{\theta}), \tag{2.106}$$

$$\mathbf{U}_h = A_h(\theta_h(\boldsymbol{\theta})) =: \mathbf{A}_h(\boldsymbol{\theta}). \tag{2.107}$$

Thereby, the mappings $\mathbf{A} : \mathbb{R}^{n_{ele}} \rightarrow \mathcal{V}$ and $\mathbf{A}_h : \mathbb{R}^{n_{ele}} \rightarrow \mathcal{V}_h^d$ have been defined implicitly.

## 2.5. Special features in AAA modeling

Beside the incorporation of arterial growth, the computational mechanical framework introduced above generally holds for the modeling of quasi-static elasticity. To be used as a physiologically meaningful macroscopic model for the description of the mechanics of AAAs, particular features have to be considered.

### 2.5.1. Orthopressure load

Generally, the traction-load $\hat{\mathbf{t}}(\mathbf{x})$ onto the luminal surface of an AAA is generated by the dynamics of the blood-flow. The blood-flow itself is caused by the pulsatile ejection of blood from the heart. Thus, an accurate determination of the resulting traction-load requires to consider the dynamical response of the full cardiovascular system. However, the timescale of the process of vascular growth (months to years) is orders of magnitude larger than the timescale of the pulsatile dynamics of blood flow (seconds). Hence, with respect to a particular state within the heart-cycle, the dynamics of growth is assumed not to be affected by the small timescale of blood flow dynamics. Consequently, for the any particular state, it is assumed that the bulk of the load is given by the static pressure $\mathrm{p}$ caused by the ejection of blood from the heart and the *windkessel* effect generated by the elastic arteries. Assuming further that the blood pressure remains constant during the simulated period of growth, the traction load is given by

$$\hat{\mathbf{t}}(\mathbf{x}) = -\mathrm{p}\mathbf{n}(\mathbf{x}) \quad \forall \mathbf{x} \in \partial\Omega_{t,\sigma}. \tag{2.108}$$

### 2.5.2. Prestressing

The formulation of hyperelasticity in terms of an elastic potential usually refers to a stress-free reference configuration such that

$$\left.\frac{\partial \Psi}{\partial \mathbf{E}}\right|_{\mathbf{E}=\mathbf{0}} = \mathbf{0}. \tag{2.109}$$

In cardiovascular applications where the configuration $\Omega_0$ is created from medical images, this assumption does not necessarily hold. Particularly with respect to longitudinal imaging of AAAs, the imaged state usually corresponds to end-diastolic conditions. Thus, the image represents an equilibrium state, the so called *prestressed* state, subject to approximately diastolic pressure $\mathrm{p}_{\mathrm{dia}}$.

In principle, the computation of a stress-free reference configuration $\Omega_z$ can be realized by the same approach as pursued for the standard forward analysis. But with the reference configuration $\Omega_0$ being known, an inverse deformation $\varphi_z^{-1} : \Omega_0 \mapsto \Omega_z$ is sought instead of a deformation $\varphi : \Omega_0 \mapsto \Omega_t$, see figure 2.4. This problem is also known as *inverse design* or *inverse elasticity* problem. However, the formulation of stress-strain relations with respect to the stress-free reference configuration $\Omega_z$ renders the computation of the inverse deformation more intricate than the standard forward analysis [see e.g. 77, 142].

To avoid these intricacies, a number of approaches have been developed to approximate the inverse elasticity problem [46, 77, 225]. These approaches have in common that, instead of computing the deformation $\varphi_z$ explicitly, they only try to build up its

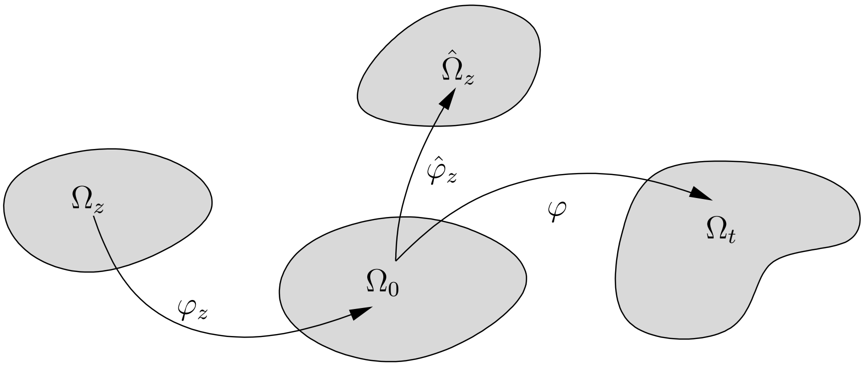

Figure 2.4.: Illustration of the relations between a stress-free configuration $\Omega_z$, the reference configuration $\Omega_0$, the current configuration $\Omega_t$ and a virtual configuration $\hat{\Omega}_z$ used to approximate $\varphi_z$ by $\hat{\varphi}_z$.

associated deformation gradient $\mathbf{F}_z$. To enable forward simulation from the prestressed state using the balance equation formulated with respect to the stress-free configuration $\Omega_z$, the total deformation gradient $\mathbf{F}_t : T_{\mathbf{X}_z}\Omega_z \to T_{\mathbf{x}}\Omega_t$ is now given by

$$\mathbf{F}_t = \frac{\partial \varphi(\mathbf{X})}{\partial \mathbf{X}} \frac{\partial \varphi_z(\mathbf{X}_z)}{\partial \mathbf{X}_z} = \mathbf{F}\mathbf{F}_z \tag{2.110}$$

with the coordinates $\mathbf{X}_z \in \Omega_z$ being the material coordinates associated to the stress-free configuration $\Omega_z$. Given an approximation $\tilde{\mathbf{F}}_z \approx \mathbf{F}_z$, the effect of the true deformation $\varphi_z$ can be approximated.

For the work presented in this thesis, the modified updated Lagrangian formulation (MULF) [77] is used to compute the approximated deformation gradient $\tilde{\mathbf{F}}_z$. This method postulates that the mapping $\varphi_z$ is similar to the mapping $\varphi$ obtained from a standard forward analysis whereby the stress-strain relation is formulated stress-free with respect to the reference configuration $\Omega_0$ under a constant pressure load

$$\mathbf{t}_0(\mathbf{X}) = \mathrm{p}_{\mathrm{dia}}\mathbf{N}(\mathbf{X}), \forall \mathbf{X} \in \partial\Omega_{0,\sigma}. \tag{2.111}$$

For cylindrical geometries and specifically also for simulations of AAAs, this approach has shown to yield stresses which are in good agreement with the exact inverse elasticity solution [76].

# 3. Mathematical formulation of the identification problem

The objective of this chapter is the mathematical formulation of the parameter identification problem associated to the nonlinear problem formulated in chapter 2 and made explicit in chapter 2.4. It is defined by the *inversion* of the mapping given by (2.101) as

$$\theta \mapsto A(\theta). \tag{3.1}$$

The basis for the inversion of this equation is given by some observation $\mathbf{Z} \in \mathcal{Z}$ of the actual state $\hat{\mathbf{U}}$ of the system. Thereby the relation between the state and the observation is expressed in terms of the *state-to-observation* mapping (also called observation operator) $C : \mathcal{V} \to \mathcal{Z}$. The *naive* approach of solving the identification problem directly via

$$\delta W(C^{-1}(\mathbf{Z}), \delta\mathbf{U}, \theta) = 0 \quad \forall \delta\mathbf{U} \tag{3.2}$$

suffers from severe disadvantages. These disadvantages are induced by the model inadequacy and the observation error. On the one hand, the modeling set forth in chapter 2 describes the real physical processes of arterial tissue in an approximate manner. Thus, given an ‘exact’ distribution $\theta^*$ of the parameters $\theta$, the modeling error causes a potential discrepancy between the model output and the real system state resulting in

$$\hat{\mathbf{U}} \neq A(\theta^*). \tag{3.3}$$

On the other hand, errors in the measurement process result in a discrepancy between the measurement $\mathbf{Z}$ and the state $\hat{\mathbf{U}}$ to be measured such that

$$\mathbf{Z} \neq C(\hat{\mathbf{U}}). \tag{3.4}$$

This measurement error can be so severe that the measurement is not even compatible with neither the system state $\hat{\mathbf{U}}$ nor the model output $A(\theta^*)$. Furthermore, the direct solution of (3.2) is only feasible if a unique relation between the measurements $\mathbf{Z}$ and the corresponding state $\hat{\mathbf{U}}$ is available. Formally, this requirement results in the availability of the inversion of the observation operator. In the specific application presented in this thesis, such a setting would be given by measurements of full-field displacement data. However, as outlined in chapter 3.2, such data is usually difficult to obtain.

These circumstances are the reason why classical identification problems are more conveniently cast in an optimization framework. A popular approach is the so called *output least squares* formulation given by

$$\underset{\theta}{\operatorname{argmin}} \, \|\mathbf{Z} - C(A(\theta))\|^2_{\mathcal{Z}} \tag{3.5}$$

[see e.g. 15]. Whereas this formulation resolves the issue of incompatibility of the measurements **Z**, it still suffers from *ill-conditioning* in the sense that small changes in **Z** might lead to dramatic changes in the best fit parameters $\theta^*$. To control these variations, the output least squares formulation is usually stated in the regularized form:

$$\underset{\theta}{\operatorname{argmin}} \left( \|\mathbf{Z} - C(A(\theta))\|_{\mathcal{Z}}^2 + R(\theta) \right). \tag{3.6}$$

The regularization $R(\theta)$ is a functional $R(\theta) : \mathcal{X} \to \mathbb{R}_0^+$ controlling the norm of the solution $\theta^*$. Classical regularization methods often refer back to A. N. Tikhonov [see e.g. 217]. In the context of least-squares optimization, the choice

$$R(\theta) = \alpha \, \|\theta\|^2 \tag{3.7}$$

is often referred to as *Tikhonov functional.* A variety of different problems can be cast into the least-squares framework (3.6) [see e.g. 26]. These sort of problems are mathematically well understood and there exists a broad range of algorithms and solution techniques for a variety of specific problems [see e.g. 169].

The identification problem formulated in terms of an optimization problem results in a so called *point-estimate* as *optimal* solution. Such a formulation neglects the variability in the optimal solution induced by measurement noise and inherent variability in the population. To take these effects into account, the identification problem has to be formulated in a probabilistic sense. Such a formulation is available through the application of the *Bayesian paradigm* that is utilized in the remainder of this thesis. After introducing the statistical notation to allow for the Bayesian formulation in chapter 3.1, chapter 3.1.2 introduces the application of the Bayesian formulation to the specific parameter identification problem of interest. The main building blocks of this setup are covered in chapter 3.2 and chapter 3.4. This encompasses the definition of the statistical relation between the computational model $\mathbf{A}(\boldsymbol{\theta})$ and the measurements **Z**. This definition is closely related to the mathematical description of the measurement space $\mathcal{Z}$ and a measure of distance in this space. Furthermore, it covers the formulation of a priori assumptions on the statistical universe of the parameters $\boldsymbol{\theta}$. Finally, chapter 3.5 introduces a reformulation of the identification problem allowing for a convenient computation of gradient information.

## 3.1. General setting and assumptions

The implementation of the Bayesian paradigm with respect to statistical inference is given by *Bayes' theorem* which is closely linked to the concept of conditional probability. To be able to cast the identification problem in a statistical setting, these concepts are briefly introduced and the necessary notation is fixed. For a detailed introduction to the applied concepts, the reader is referred to general literature on probability theory and statistics such as Billingsley [24], Rohatgi and Saleh [188] or Schay [194].

In the following, a probability space $(\mathcal{S}, \mathcal{F}, P)$ is employed. It is constituted by a sample space $\mathcal{S}$, a $\sigma$-algebra $\mathcal{F}$ of events and a probability measure $P$ such that the probability measure $P$ satisfies

(i) $0 \leq P(\mathrm{A}) \leq 1, \forall \mathrm{A} \in \mathcal{F}$,

(ii) $P(\mathcal{S}) = 1$,

(iii) $P(\bigcup_{i=1}^{\infty} \mathrm{A}_i) = \sum_{i=1}^{\infty} P(\mathrm{A}_i)$ for the mutually disjoint sets $\mathrm{A}_i \in \mathcal{F}$.

A fundamental concept in probability theory is that of *conditional probability.*

**Definition 3.1.1** (Conditional probability)**.** *On the probability space* $(\mathcal{S}, \mathcal{F}, P)$*, with* $\mathrm{A}, \mathrm{B} \in \mathcal{F}$ *and* $P(\mathrm{B}) > 0$*, the conditional probability* $P(\mathrm{A}|\mathrm{B})$ *is given by*

$$P(\mathrm{A}|\mathrm{B}) = \frac{P(\mathrm{A} \cap \mathrm{B})}{P(\mathrm{B})}. \tag{3.8}$$

By the symmetry $P(\mathrm{A} \cap \mathrm{B}) = P(\mathrm{B} \cap \mathrm{A})$, this definition directly implies *Bayes' theorem:*

$$P(\mathrm{B}|\mathrm{A}) = \frac{P(\mathrm{A}|\mathrm{B})P(\mathrm{B})}{P(\mathrm{A})}. \tag{3.9}$$

### 3.1.1. Random variables

A random variable $\mathsf{X}$ is defined as a measurable function on the sample space $\mathcal{S}$ such that events $\in \mathcal{F}$ can be described in terms of sets such as

(a) $\{s \in \mathcal{S} : \mathsf{X}(s) = x\}$

(b) $\{s \in \mathcal{S} : \mathsf{X}(s) \leq x\}$

(c) $\{s \in \mathcal{S} : \mathsf{X}(s) \in \mathcal{B}\}$

for some real value $x \in \mathbb{R}^n$ and the Borel set $\mathcal{B}$ on $\mathbb{R}^n$. These sets are often stated in short as $\{\mathsf{X} = x\}$, $\{\mathsf{X} \leq x\}$ or $\{\mathsf{X} \in \mathcal{B}\}$. Throughout this thesis, random variables are indicated by the use of a serif-free front.

**The discrete case** In discrete sample spaces, the association of a function $p_{\mathsf{X}}(x) = P(\{\mathsf{X} = x\}) =: P(\mathsf{X} = x)$ is straightforward and usually $p_{\mathsf{X}}$ is referred to as *probability-(mass)* function of the random variable $\mathsf{X}$. Since in finite dimensions also (a) represents a proper set, the conditional probability $P(\mathsf{X} = x|\mathsf{Y} = y)$ is well defined as long as $P(\mathsf{Y} = y) \neq 0$.

**The continuous case** In the continuous case, the probability $P(\mathsf{X} = x) = 0$ due to compliance with the axiom (ii). Anyhow, for a proper set $\mathrm{A} \in \mathcal{F}$, e.g., $\mathrm{A} = \{\mathsf{X} \in \mathcal{B}\}$, the probability $P(\mathsf{X} \in \mathcal{B})$ can be expressed in terms of the Lebesgue integral on $\mathbb{R}^n$ according to

$$P(\mathsf{X} \in \mathcal{B}) = \int_{\mathcal{B}} P(dx) = \int_{\mathcal{B}} p_{\mathsf{X}}(x)dx \tag{3.10}$$

In this setup, the function $p_{\mathsf{X}} : \mathcal{S} \to [0, \infty[$ is referred to as *probability density* function and must comply with the 3 axioms of probability (i)-(iii). Equivalently, the *joint probability density* $p_{\mathsf{X},\mathsf{Y}}(x, y)$ describes the joint probability $P(\mathsf{X} \in \mathcal{B}, \mathsf{Y} \in \mathcal{D}) = P(\mathsf{X} \in \mathcal{B} \cap \mathsf{Y} \in \mathcal{D})$ via

$$P(\mathsf{X} \in \mathcal{B}, \mathsf{Y} \in \mathcal{D}) = \int_{\mathcal{B}} \int_{\mathcal{D}} p_{\mathsf{X},\mathsf{Y}}(x, y)dxdy. \tag{3.11}$$

In this setting, the conditional probability $P(\mathsf{X} \in \mathcal{B} | \mathsf{Y} = y)$ is undefined in terms of definition 3.1.1 since $P(\mathsf{Y} = y) = 0$. Instead, it is defined as the limit

$$P(\mathsf{X} \in \mathcal{B} | \mathsf{Y} = y) := \lim_{\epsilon \to 0^+} P(\mathsf{X} \in \mathcal{B} | \mathsf{Y} = ]y - \epsilon, y + \epsilon]), \tag{3.12}$$

provided it exists [see 188, p. 109]. By inserting this into the definition 3.1.1, it can be shown that a conditional probability density function $p_{\mathsf{X}|\mathsf{Y}}$ is given by

$$p_{\mathsf{X}|\mathsf{Y}}(x|y) = \frac{p_{\mathsf{X},\mathsf{Y}}(x, y)}{p_{\mathsf{Y}}(y)} \tag{3.13}$$

such that the probability $P(\mathsf{X} \in \mathcal{B} | \mathsf{Y} = y)$ can be evaluated according to

$$P(\mathsf{X} \in \mathcal{B} | \mathsf{Y} = y) = \int_{\mathcal{B}} p_{\mathsf{X}|\mathsf{Y}}(x|y) \mathrm{d}x. \tag{3.14}$$

The function $p_{\mathsf{X}|\mathsf{Y}}(x|y)$ represents a probability density with respect to the random variable $\mathsf{X}$. Seen as a function of $y$ it is called *likelihood* since it does not represent a density. Noticing that the conditional density (3.13) can be constructed similarly for $p_{\mathsf{Y}|\mathsf{X}}(y|x)$ and that the joint probability density is symmetric, Bayes' theorem in terms of densities emerges as a direct consequence as

$$p_{\mathsf{X}|\mathsf{Y}}(x|y) = \frac{p_{\mathsf{Y}|\mathsf{X}}(y|x) p_{\mathsf{X}}(x)}{p_{\mathsf{Y}}(y)}. \tag{3.15}$$

**Expected value and variance** A random continuous variable $\mathsf{X}$ is entirely defined by the associated density $p_{\mathsf{X}}(x)$. However, this description is too elusive to identify the characteristic properties of a random variable. Therefore, often summarizing *statistics* are used to characterize random variables. The most prominent summary is the *expected value*

$$\mathbb{E}_x[\mathsf{X}] := \int_{\mathcal{S}} x p_{\mathsf{X}}(x) \mathrm{d}x, \tag{3.16}$$

provided the integral converges absolutely. It can be seen as a measure of the center of the distribution $p_{\mathsf{X}}$. A measure of the dispersion of the distribution is given by the covariance

$$\mathbb{COV}_x[\mathsf{X}] := \mathbb{E}_x[(x - \mathbb{E}_x[\mathsf{X}])(x - \mathbb{E}_x[\mathsf{X}])]. \tag{3.17}$$

In a multidimensional setting, $\mathsf{X} = \{\mathsf{X}_1, \dots, \mathsf{X}_n\}$, this information is difficult to visualize. Results on the variability of a distribution are therefore more conveniently reported in terms of marginal densities $p_{\mathsf{X}_i}(x_i)$. Thus, the variance of $\mathsf{X}_i$ is given as the one-dimensional covariance

$$\mathbb{V}_{x_i}[\mathsf{X}_i] = \mathbb{E}_{x_i}[(x_i - \mathbb{E}_{x_i}[\mathsf{X}_i])^2] = [\mathbb{COV}_x[\mathsf{X}]]_{ii}. \tag{3.18}$$

To quantify this dispersion in the same units as the random variable itself, a specification in terms of the standard deviation

$$\mathbb{SD}_x[\mathsf{X}] = \sqrt{\mathbb{V}_x[\mathsf{X}]} \tag{3.19}$$

is more convenient. Furthermore, given a random variable $\mathsf{X}$ with probability density $p_{\mathsf{X}}$ and a function $f : \mathsf{X} \mapsto f(\mathsf{X})$ with unknown density $p_{f(\mathsf{X})}$, the expected value $\mathbb{E}_{f(x)}[f(\mathsf{X})]$ is given by

$$\mathbb{E}_{f(x)}[f(\mathsf{X})] = \int_{\mathcal{S}} f(x) p_{\mathsf{X}}(x) \mathrm{d}x. \tag{3.20}$$

Variance $\mathbb{V}_{f(x)}[f(\mathsf{X})]$ and standard deviation $\mathbb{SD}_{f(x)}[f(\mathsf{X})]$ are obtained equivalently.

**Remark.** *In the probabilistic literature, it is common to ambiguously use the symbol $p$ to denote different probability distributions such that Bayes' theorem (3.15) is also stated as*

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \tag{3.21}$$

*Thereby, the specific functional relation of a distribution has to be derived from the argument. Whenever misinterpretation is impossible, this convention is also applied throughout this thesis.*

### 3.1.2. Statistical identification problem

Bayes' theorem (3.15) provides the basis for the statistical formulation of the identification problem. To this end, the observations $\mathbf{Z} \sim p(\mathbf{Z})$ and the parameters $\boldsymbol{\theta} \sim p(\boldsymbol{\theta})$ are interpreted as random variables with $\mathbf{Z} \in \mathcal{Z}$ and $\boldsymbol{\theta} \in \mathbb{R}^{n_p}$. The dimension $n_p$ of the identification problem is thereby introduced to decouple the strict association of the parameters $\boldsymbol{\theta}$ to the coefficients of the element-wise basis (2.105). This allows for a more flexible interpretation of the parameters $\boldsymbol{\theta}$ in terms of a different basis or in the sense of an assembly of different physical parameters. In this setting, the identification problem is given by the *posterior* density $p(\boldsymbol{\theta}|\mathbf{Z})$ which is defined by the application of Bayes' theorem:

$$p(\boldsymbol{\theta}|\mathbf{Z}) = \frac{p(\mathbf{Z}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{Z})}. \tag{3.22}$$

In the context of statistical inference, the densities $p(\mathbf{Z}|\boldsymbol{\theta})$, $p(\boldsymbol{\theta})$ and $p(\mathbf{Z})$ are often referred to as *likelihood* (as a function of $\boldsymbol{\theta}$), *prior* and *evidence*, respectively.

The practical application of (3.22) requires the formulation of a statistical relation $\boldsymbol{\theta} \mapsto \mathbf{Z}$ such that the likelihood function $p(\mathbf{Z}|\boldsymbol{\theta})$ can be defined. Furthermore, knowledge on the statistical universe of the parameters $\boldsymbol{\theta}$ must be available in terms of a prior $p(\boldsymbol{\theta})$. In the limit, this knowledge can also be *non-informative*. In this case, the data $\mathbf{Z}$ must provide enough information such that (3.22) represents a probability density.

In contrast to the classical optimization problem (3.5), the solution of the identification problem is now given as a probability density. This fact directly reveals the ambiguity in the definition of a single optimal solution. In order to define some notion of optimality (see chapter 4) it is often sufficient to observe the proportionality

$$p(\boldsymbol{\theta}|\mathbf{Z}) \propto p(\mathbf{Z}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{3.23}$$

The remainder of this chapter is concerned with the specific definitions of the likelihood function $p(\mathbf{Z}|\boldsymbol{\theta})$ and the prior-density $p(\boldsymbol{\theta})$ such that the relation equation (3.23) is well defined. To this end, it is assumed that the *randomness* in the observations $\mathbf{Z}$ is entirely created by noise in the measurement process. Furthermore, it is assumed that this noise is represented by a normally distributed random variable $\boldsymbol{\zeta}$ with zero mean and covariance $\boldsymbol{\Sigma}_{\mathbf{Z}}$:

$$\boldsymbol{\zeta} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{Z}}). \tag{3.24}$$

Throughout this thesis, the measurement covariance $\boldsymbol{\Sigma}_{\mathbf{Z}}$ is assumed to be known or to be inferred from the measurement process a priori. Therefore, it represents a user input to the problem usually given in the form $\boldsymbol{\Sigma}_{\mathbf{Z}} = \sigma^2 \mathbf{I}$.

Given the specific definition of the measurement noise (3.24), the observation $\mathsf{Z}$ is related to the computational model via the *additive noise model*

$$\mathsf{Z} = C(\mathbf{A}(\boldsymbol{\theta})) + \boldsymbol{\zeta}. \tag{3.25}$$

Whenever the compound application $F(\boldsymbol{\theta}) := (C \circ \mathbf{A})(\boldsymbol{\theta})$ is not of particular interest in a specific context, this relation is often stated more concisely as

$$\mathsf{Z} = F(\boldsymbol{\theta}) + \boldsymbol{\zeta}. \tag{3.26}$$

The generic function $F : \mathbb{R}^{n_p} \to \mathcal{Z}$ is also referred to as *model* (and also computational model) but not to be confused with the parameter-to-state mapping (3.1). Despite the simplified form, it has to be kept in mind that the functional dependence $\boldsymbol{\theta} \mapsto F(\boldsymbol{\theta})$ is not explicitly available. A rearrangement of (3.26) gives rise to the likelihood in terms of a Gaussian probability density

$$p(\mathbf{Z}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^m \, |\boldsymbol{\Sigma}_{\mathbf{Z}}|}} \exp(-\frac{1}{2} \left\| \mathbf{Z} - F(\boldsymbol{\theta}) \right\|^2_{\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}}). \tag{3.27}$$

Using the simplification $\boldsymbol{\Sigma}_{\mathbf{Z}} = \sigma^2 \mathbf{I}$ and a generic similarity measure $D : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}_0^+$, the likelihood can be written as

$$p(\mathbf{Z}|\boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi\sigma^2)^m}} \exp(-\frac{1}{2\sigma^2} D(\mathbf{Z}, F(\boldsymbol{\theta}))). \tag{3.28}$$

In order for this generic formulation to represent a Gaussian probability distribution, the similarity measure $D$ must be induced by a scalar-product in $\mathbb{R}^{m<\infty}$. The presentation throughout this thesis is therefore restricted to a setup defined by

$$D(\mathbf{Z}, F(\boldsymbol{\theta})) := \langle \mathbf{Z} - F(\boldsymbol{\theta}), \mathbf{Z} - F(\boldsymbol{\theta}) \rangle_{\mathcal{Z}} \equiv \langle f_{\mathcal{Z}}(\mathbf{Z} - F(\boldsymbol{\theta})), f_{\mathcal{Z}}(\mathbf{Z} - F(\boldsymbol{\theta})) \rangle_{\mathbb{R}^m}, \tag{3.29}$$

whereby $f_{\mathcal{Z}} : \mathcal{Z} \to \mathbb{R}^m$ is seen as a generic function that maps from a potentially continuous space $\mathcal{Z}$ to some discrete space $\mathbb{R}^m$. Its existence is implied by proper choices of similarity measures $D$ and vice versa.

**Remark.** *It is to be mentioned, that the additive noise model (3.26) is not the only one and more possibilities exist [110, 116]. The choice for a specific model has to be made in combination with an experimental setup.*

The assumption on the availability of a priori knowledge of parameters of the measurement noise is not a necessary assumption and does not affect the validity of Bayes' theorem (3.23). In fact, from a *fully Bayesian perspective* noise-parameters such as the measurement variance $\sigma^2$ are seen as additional unknown random variables. They are modeled by an additional prior $p(\sigma)$ such that Bayes' theorem is given by

$$p(\boldsymbol{\theta}, \sigma|\mathbf{Z}) = \frac{p(\mathbf{Z}|\boldsymbol{\theta}, \sigma)p(\boldsymbol{\theta})p(\sigma)}{p(\mathbf{Z})}. \tag{3.30}$$

Furthermore, in case the model-parameters $\boldsymbol{\theta}$ represent some physically meaningful quantity, they can usually be applied through different models. The particular mapping $F(\boldsymbol{\theta})$ only represents one single possible model $\mathbb{M}$ (epistemologically wrong) that enters the likelihood as an additional argument to be conditioned on. Assembling model- and noise-parameters into a single vector

$$\boldsymbol{\Theta} = (\boldsymbol{\theta}, \sigma^2), \tag{3.31}$$

Bayes' theorem conditioned on the model $\mathbb{M}$ reads

$$p(\boldsymbol{\Theta}|\mathbf{Z}, \mathbb{M}) = \frac{p(\mathbf{Z}|\boldsymbol{\Theta}, \mathbb{M})p(\boldsymbol{\Theta})}{p(\mathbf{Z}|\mathbb{M})}. \tag{3.32}$$

The evidence $p(\mathbf{Z}|\mathbb{M})$ plays a key role in Bayesian model selection, e.g., via Bayes factors [111]. Due to the fact that the measurement noise is not modeled as a random variable throughout this thesis, $\boldsymbol{\theta}$ is used instead of $\boldsymbol{\Theta}$ for notational convenience in the following. But noticing the similarity between (3.22) and (3.32), $\boldsymbol{\theta}$ can always be understood as an assembly in the sense of (3.31).

**Remark.** *The description of a priori knowledge through prior probabilities usually introduces additional parameters, so called hyper-parameters. Within the Bayesian paradigm, these can again be modeled by priors, so called hyper-priors. The resulting formulation is then referred to as 'hierarchical Bayes'.*

The remaining ingredients to complete the setup of the identification problem are twofold. In chapter 3.2, the definition of an observation space $\mathcal{Z}$ and accordingly the definition of a similarity measure on the elements of that space completes the definition of the likelihood. Chapter 3.4 introduces the modeling of the prior density to account for the statistical universe of the parameters $\boldsymbol{\theta}$.

## 3.2. Definition of similarity

In practical applications, observations are rarely available in terms of the primal variables of some model to be fitted. In the particular setting of AAA growth, the raw measurements are given in terms of medical images. Thereby different technologies such as MRI, CT or US are commonly used.

**Information from medical imaging** A typical example of a longitudinal CT study showing the same patient with a time-lag of approximately 2 year is given in figure 3.1. An image $\mathrm{I}_0$ can formally be defined as a function $\mathrm{I}_0 : \Omega_0^{\mathrm{I}} \to \mathbb{R}$ such that $\mathrm{I}_0(\mathbf{X})$ is the (gray-scale) value at the pixel being located at the position $\mathbf{X} \in \Omega_0^{\mathrm{I}}$. Similarly, an image $\mathrm{I}_t$ is defined as function $\mathrm{I}_t : \Omega_t^{\mathrm{I}} \to \mathbb{R}$. Beside the raw pixel-value information, the additional

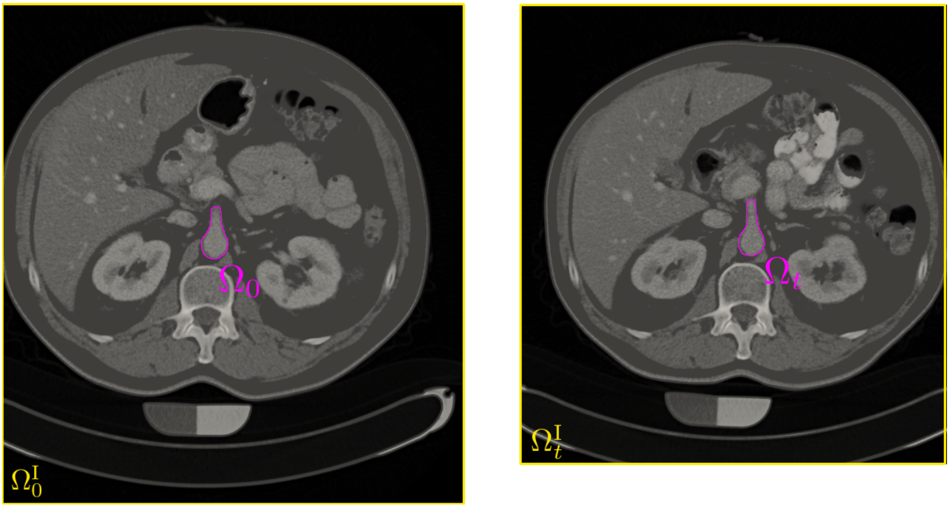

Figure 3.1.: A transverse slice of a patient's longitudinal CT study[1]with a time-lag in between the left and the right scan of 2 years. The boundaries of the computational domains $\Omega_0$ and $\Omega_t$ are highlighted in magenta. The corresponding boundaries of the domains $\Omega_0^{\mathrm{I}}$ and $\Omega_t^{\mathrm{I}}$ covered by the images are highlighted in yellow.

information encoded in longitudinal image data is given by a deformation depicted by the changes of the morphological structures captured in the images. Assuming, for the time being, the existence of this deformation as a smooth invertible mapping $\varphi_{\mathrm{I}}(\mathbf{X}) : \mathbf{X} \mapsto \mathbf{x}$ $\forall \mathbf{X} \in \Omega_0^{\mathrm{I}}$ and $\forall \mathbf{x} \in \Omega_t^{\mathrm{I}}$ given by

$$\mathbf{x} = \varphi_{\mathrm{I}}(\mathbf{X}) = \mathbf{X} + \mathbf{V}(\mathbf{X}), \tag{3.33}$$

two images $\mathrm{I}_0$ and $\mathrm{I}_t$ could be related via interpolation according to

$$\mathrm{I}_0(\varphi_{\mathrm{I}}^{-1}(\mathbf{x})) = \mathrm{I}_t(\mathbf{x}). \tag{3.34}$$

This assumption is referred to as *brightness constancy assumption* [see e.g. 160]. The task of finding the mapping $\varphi_{\mathrm{I}}$ is covered by the large mathematical field of *image registration*. This field is centered around the fact that the desired mapping $\varphi_{\mathrm{I}}$ is not uniquely given by the image data in almost all practical applications. As a consequence, the direct interpolation problem (3.34) is not well-posed. This issue is addressed by a vast amount of different formulations and numerical approaches. A common element of all these approaches is given by the application of certain modeling assumptions or regularization techniques to render the interpolation problem well-posed. But these assumptions are rarely capable of accurately representing the true physical processes being imaged. On the contrary, usually over-simplistic models are applied. Hence, the application of image registration results in inherent systematic deviations of the estimated mapping $\hat{\varphi}_{\mathrm{I}}$ from the true mapping $\varphi_{\mathrm{I}}$. By using the estimation $\hat{\varphi}_{\mathrm{I}}$ as a measurement $\mathsf{\mathbf{Z}}$ in the parameter identification problem (3.23), it is therefore expected that the systematic errors are propagated to the parameters $\boldsymbol{\theta}$.

---

[1] Medical images are courtesy of the Institut für diagnostische und interventionelle Radiologie, Klinikum rechts der Isar, Technische Universität München. The use of clinically induced CT-imaging was approved by the ethics commission of the TUM School of medicine; 'M-AAA-rker Studie' (Nr. 1897/07, 27.9.2007, Amendment 19.3.2012).

**Images as measurements** In an attempt to avoid the propagation of the systematic errors in $\hat{\varphi}_{\mathrm{I}}$, it seems appealing to tackle the parameter identification problem directly in defining the space of measurements $\mathcal{Z}$ to be constructed by the set of functions representing the images in $\Omega_0^{\mathrm{I}}$ and $\Omega_t^{\mathrm{I}}$. Assuming it is possible to evaluate $\mathrm{I}_t(\varphi(X))$ by using the deformation $\varphi$ from equation (2.6) associated to the solution of the nonlinear forward problem (2.101), the parameter identification problem could be stated as: Find $\boldsymbol{\theta}$ such that

$$\int_{\Omega_0^{\mathrm{I}}} (\mathrm{I}_0(\mathbf{X}) - \mathrm{I}_t(\varphi(\mathbf{X})))^2 \,\mathrm{dV}_0 \to \min\,. \tag{3.35}$$

This approach results in an optimal control formulation where the controlling parameters are given by the parametrization of the forward problem according to chapter 2.4. Despite its appeal, the need to be able to perform the interpolation $\mathrm{I}_t(\varphi(\mathbf{X}))$ directly reveals that this approach is constrained to situations in which $\Omega_0^{\mathrm{I}} = \Omega_0$, i.e., the image domain and the computational domain match exactly. This is clearly not the situation presented in figure 3.1. In principle, this constraint can be bypassed by the definition of an extension operator $\mathrm{Ext} : \partial\Omega_0 \to \Omega_0^{\mathrm{I}} \setminus \Omega_0$. But this extension operator clearly involves modeling assumptions which will again be reflected in the optimal parameters $\boldsymbol{\theta}$. Thus, the initial intention is undermined. In the setup of conforming domains of the images and the computational model, this approach was first proposed by Miga [158] but using the concept of *mutual information* as a similarity measure on the space of images [see e.g. 226].

**Issues with longitudinal data** Due to the limitations of the formulation (3.35), applying image registration as a tool to provide measurements from image data is still predominant in practical applications. But analyzing the images showing AAA growth in figure 3.1, it becomes clear that the interpolation assumption (3.34) is not just slightly violated but heavily corrupted. This fact alone is not an issue since mappings between images from different modalities inherently violating the brightness constancy assumption can very well be obtained by state-of-the-art image registration technologies such as *mutual information image registration*. It is the change in morphology in parts of the image that makes any attempt to obtain a meaningful transformation between the two raw images unreasonable. Thus, the basic assumption on the mapping $\varphi_{\mathrm{I}}$ is not fulfilled. A remedy to this dilemma is provided by performing a segmentation of the structure of interest and applying the registration step to these segmented images, see figure 3.2 for an illustration. Representing the estimated deformation $\hat{\varphi}(\mathbf{X})$ via a smooth displacement field $\hat{\mathbf{V}} \in \mathcal{V}_{\mathrm{I}}$ according to

$$\mathbf{x} = \hat{\varphi}(\mathbf{X}) = \mathbf{X} + \hat{\mathbf{V}}(\mathbf{X}), \tag{3.36}$$

it is principally possible to define this displacement field on the domain $\Omega_0$ as the measurement $\mathbf{Z}$. More practically, by defining the operator $C_{\mathrm{I}} : \mathcal{V}_{\mathrm{I}} \to \mathbb{R}^{n_{dof}}$ to map the measured displacement field to the degrees of freedom of the mesh $\mathcal{K}$ of $\Omega_0$, a similarity measure to be used in the likelihood (3.27) is given by

$$D(\hat{\mathbf{V}}, \mathbf{A}(\boldsymbol{\theta})) = \left\| C_{\mathrm{I}}(\hat{\mathbf{V}}) - C(\mathbf{A}(\boldsymbol{\theta})) \right\|^2_{\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}}. \tag{3.37}$$

In some sense, a segmentation as presented in figure 3.2 provides a compression of information. Thereby the information inside segmented areas is clearly reduced, whereas at

Figure 3.2.: 2D-illustration of image registration on a segmented set of 3D-images showing an aneurysm with intraluminal thrombus (ILT) (light gray), luminal area (white) and surrounding tissue (dark gray). Left: initial aneurysm represented by a segmented image $\mathrm{I}_0^s$. Middle: Grown aneurysm represented by a segmentation $\mathrm{I}_t^s$. Right: One possible mapping $\hat{\varphi}^{-1}$ instantiating the relation $\mathrm{I}_t^s(\mathbf{x}) \approx \mathrm{I}_0^s(\hat{\varphi}^{-1}(\mathbf{x}))$.

the transitions between different structures, i.e., surfaces in a 3D image, the information of the original images is almost completely preserved. This variation in accuracy of the information is also reflected in the registered deformation and it is often more convenient to define

$$D(\hat{\mathbf{V}}, \mathbf{A}(\boldsymbol{\theta})) = \left\| C_{\mathrm{I},\gamma}(\hat{\mathbf{V}}) - C(\mathbf{A}(\boldsymbol{\theta})) \right\|^2_{\boldsymbol{\Sigma}_{\mathbf{Z}}^{-1}} . \tag{3.38}$$

The operator $C_{\mathrm{I},\gamma} : \mathcal{V}_{\mathrm{I}} \to \mathbb{R}^{n_{dof}^{\gamma}}$ as well as the state-to-observation operator $C : \mathbb{R}^{n_{dof}} \to \mathbb{R}^{n_{dof}^{\gamma}}$ represent a mapping to the discrete degrees of freedom of the boundary $\partial\Omega_0$.

**Surfaces as measurements** Although being restricted to the boundary, the measurement $C_{\mathrm{I},\gamma}(\hat{\mathbf{V}})$ might still be incompatible with the model output $\mathbf{A}(\boldsymbol{\theta})$. Due to the inflexibility of the measure (3.38) with respect to the hard point-wise correspondences $C_{\mathrm{I},\gamma}(\hat{\mathbf{V}}) \leftrightarrow C(\mathbf{A}(\boldsymbol{\theta}))$, this incompatibility can lead to highly *ill-conditioned* identification problems. Furthermore, this incompatibility will be reflected in the posterior distribution $p(\boldsymbol{\theta}|\mathbf{Z})$ leading to a misinterpretation of the variability of the parameters $\boldsymbol{\theta}$. Analyzing medical image data of AAAs more closely, it is revealed that the structural information encoded by the image is actually not given in terms of volume information. Rather, different structures are distinguished by sharp transitions in the pixel values. In a 3D image, these transitions are identified as surfaces, see figure 3.3. Since surface data is naturally decoded from medical images by means of segmentation during the creation of some sort of computational model, it is natural to try to compare different models directly in terms of surfaces. In the area of parameter identification, this concept was brought forward by Moireau et al. [162].

Since a mathematically rigorous formulation of a distance in spaces created by surfaces can be rather involved, chapter 3.3 is entirely devoted to this topic.

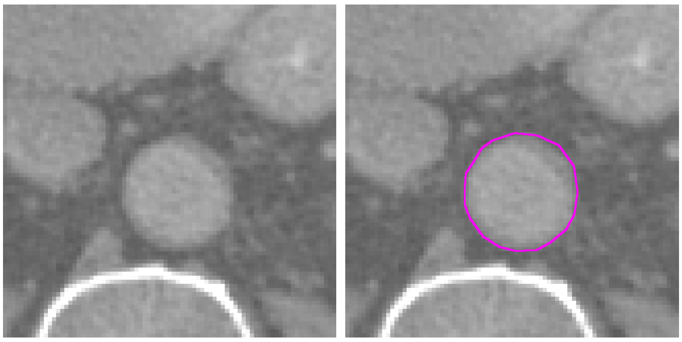

Figure 3.3.: Zoom into patient-specific CT data showing a small aneurysm with its enclosing surface colored in magenta. Different structures can clearly be distinguished only by means of sharp transitions/gradients. E.g., the aneurysm can clearly be distinguished from the surrounding tissue.

## 3.3. Surfaces as measurements

In order to define measures of distance in a *space of surfaces*, it is necessary to fix a notion of the concept of surfaces. A mathematically precise definition of the concept of surface is ambiguous. In this thesis, a surface is understood in the context of topological structures and differential geometry. More precisely, surfaces have already been used throughout this thesis without further specification as the boundaries $\partial\Omega_0$ or $\partial\Omega_t$ of the domains $\Omega_0$ and $\Omega_t$. The notion of surface will be solely used in this setup and can thus be defined as a two-dimensional differentiable manifold embedded in the 3-dimensional euclidean space $\mathbb{R}^3$. These particular manifolds are also referred to as *regular surfaces*. Though sufficient for the applications presented here, it is to be noted that the concept of surfaces is not constrained to this setting. For a broader perspective, the reader is referred to the respective literature [see e.g. 11, 52].

**Definition 3.3.1** (Regular surface, from do Carmo [52])**.** *A subset $S \subset \mathbb{R}^3$ is a regular surface if for every point $x \in S$ there exists a neighborhood $V_x$ and a mapping $\phi_\alpha : U_\alpha \subset \mathbb{R}^2 \to V_x \cup S$ such that*

*(1) $\phi_\alpha$ is a homeomorphism*

*(2) the differential $(\mathrm{d}\phi_\alpha)_q : T_qU_\alpha \to \mathbb{R}^3$ is injective for all $q \in U_\alpha$.*

The pairing $(U_\alpha, \phi_\alpha)$ is called a *parametrization* and the whole family of parametrizations $(U_\alpha, \phi_\alpha)$ is called a differentiable structure on $S$ if $\bigcup_\alpha \phi_\alpha(U_\alpha) = S$. This technical definition has important implications: given two parametrizations $(U_\alpha, \phi_\alpha)$ and $(U_\beta, \phi_\beta)$, the *change of coordinates* $\phi_\beta^{-1} \circ \phi_\alpha$ is differentiable. This differentiability provides a basis for the field of differential geometry.

As far as concerned in this thesis, the most important implication of the properties of the parametrizations $\phi_\alpha$ is given by the definition of a metric structure: let the mapping $\phi_\alpha$ be given by $\phi_\alpha : (s,t) \in U \subset \mathbb{R}^2 \to S$, the so called *metric tensor* $\mathbf{g}$ is defined as

$$\mathbf{g} := \begin{pmatrix} \langle \frac{\partial\phi}{\partial s}, \frac{\partial\phi}{\partial s} \rangle & \langle \frac{\partial\phi}{\partial s}, \frac{\partial\phi}{\partial t} \rangle \\ \langle \frac{\partial\phi}{\partial t}, \frac{\partial\phi}{\partial s} \rangle & \langle \frac{\partial\phi}{\partial t}, \frac{\partial\phi}{\partial t} \rangle \end{pmatrix}. \tag{3.39}$$

Given such a parametrization, the integral of a function $f$ along the surface $S$, the *surface integral*, is given by

$$\int_S f(\mathbf{X})\,\mathrm{d}\Gamma := \iint_U f(\phi_\alpha(s,t))\sqrt{\det(\mathbf{g})}\,\mathrm{d}s\mathrm{d}t, \tag{3.40}$$

[see e.g. 127]. The determinant $\det(\mathbf{g})$ can also be expressed in terms of the tangent vectors $\boldsymbol{\nu} := \frac{\partial \phi}{\partial s}$ and $\boldsymbol{\eta} := \frac{\partial \phi}{\partial t}$ as

$$\det(\mathbf{g}) = \|\boldsymbol{\nu} \times \boldsymbol{\eta}\|^2 . \tag{3.41}$$

By definition of the surface normal $\mathbf{n} := (\boldsymbol{\nu} \times \boldsymbol{\eta})/\sqrt{\det(\mathbf{g})}$, the integration of a vector field $\mathbf{w}$ across the surface $S$ is given according to

$$\int_S (\mathbf{w} \cdot \mathbf{n})\,\mathrm{d}\Gamma = \iint_U \mathbf{w}(\phi(s,t)) \cdot (\boldsymbol{\nu} \times \boldsymbol{\eta})\,\mathrm{d}s\mathrm{d}t. \tag{3.42}$$

From the right hand side, it can be seen that the integrand represents a linear form $\omega$ on the tangent vectors $\boldsymbol{\nu}$ and $\boldsymbol{\eta}$. Furthermore, if the vector field $\mathbf{w}$ varies smoothly on $S$, $\omega$ is a so called *differential 2-form*. As a function of $\phi$, such forms represent densities on a 2-dimensional manifold such as the surface $S$. Thus, they are naturally integrated by

$$\int_S (\mathbf{w} \cdot \mathbf{n})\,\mathrm{d}\Gamma =: \int_S \omega. \tag{3.43}$$

This provides a basis for the formalism of *surface currents* that enables to see surfaces as elements of a Hilbert space. With the encompassing benefit of providing norms and thus metrics, the Hilbert space structure can be naturally used to define the likelihood 3.27. But before delving into this topic, a conceptually easier measure of similarity on surfaces is presented.

### 3.3.1. Distance as closest point projection

A classical measure to compute a distance between two surfaces is given by the so called *closest point projection*, see figure 3.4. This measure can be computed for 2 surfaces $S$

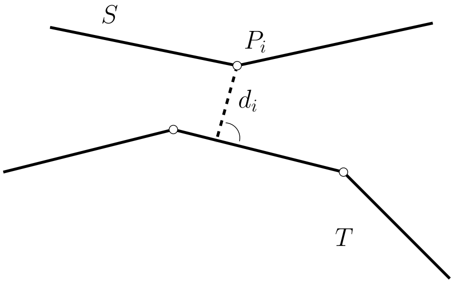

Figure 3.4.: Illustration of the closest point projection.

and $T$ represented by their respective triangulations.

**Definition 3.3.2** (Triangulation)**.** *Given a surface $S$, a triangulation $\mathcal{T}_S$ of $S$ is given as the set $\mathcal{T}_S = \{\boldsymbol{\tau}_i^S\}_{i=1}^{N_S}$ defined on the set of point $\{\boldsymbol{\varkappa}_i^S\}_{i=1}^{N_S^p}$, whereby a triangle $\boldsymbol{\tau}_i^S$ is given as the triplet $\boldsymbol{\tau}_i^S := (\boldsymbol{\varkappa}_{i,1}^S, \boldsymbol{\varkappa}_{i,2}^S, \boldsymbol{\varkappa}_{i,3}^S)$ such that a triangle's center is computed as $\mathbf{c}_i^S := \frac{1}{3}\sum_{j=1}^{3}\boldsymbol{\varkappa}_{i,j}^S$ and a triangle's (unnormalized) normal is computed as $\tilde{\mathbf{n}}_i^S := (\boldsymbol{\varkappa}_{i,2}^S - \boldsymbol{\varkappa}_{i,1}^S) \times (\boldsymbol{\varkappa}_{i,3}^S - \boldsymbol{\varkappa}_{i,1}^S)$.*

Given a triangulation $\mathcal{T}_S$ for the surface $S$ and a triangulation $\mathcal{T}_T$ for the surface $T$, the minimal projected distance for each point $\boldsymbol{\varkappa}_i^S \in \mathcal{T}_S$ to the triangulation $\mathcal{T}_T$ is given by

$$d_i = \min_{\boldsymbol{\tau}_j^T} \frac{|\tilde{\mathbf{n}}_j^T \cdot \boldsymbol{\varkappa}_i^S - \tilde{\mathbf{n}}_j^T \cdot \boldsymbol{\varkappa}_{j,1}|}{\sqrt{\|\tilde{\mathbf{n}}_j^T\|}} \quad \text{s.t.} \quad \boldsymbol{\varkappa}_i^S - d_i \frac{\tilde{\mathbf{n}}_j^T}{\sqrt{\|\tilde{\mathbf{n}}_j^T\|}} \in \boldsymbol{\tau}_j. \tag{3.44}$$

A global distance $D(S,T)$ can then be defined as the sum

$$D(S,T) = \sum_{i=1}^{N_S^p} d_i. \tag{3.45}$$

Despite its simplicity, the projected distance measure suffers from several disadvantages hindering its practical applicability. Beside the non-uniqueness and the irrelevancy of the projection point, see figure 3.5, it can also happen that the solution to 3.44 is undefined due to incompatibilities of the constraint. Given that algorithmic measures are applied to resist these issues, (3.45) is a popular choice of measuring similarity between surfaces. In the area of parameter identification, it was made popular by Moireau et al. [162] and has been variously applied [see e.g. 21, 37]. Anyhow, it was observed by Imperiale et al. [104] that the inherent disadvantages of the closest point projection can lead to inefficient estimators and that the usage of the surface current formalism is capable to improve on this problem.

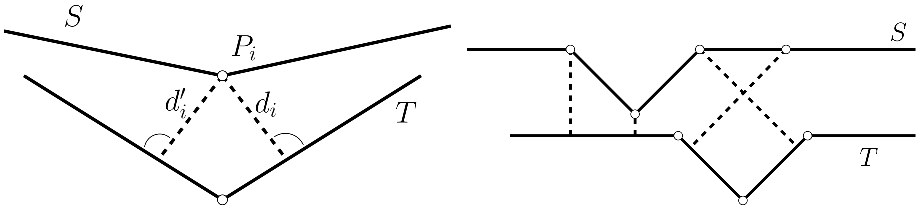

Figure 3.5.: The two main disadvantages of the closest point projection on discretized surfaces [see also 104]. Left: the non-uniqueness of the projection point. Right: Capturing of irrelevant shape characteristics.

### 3.3.2. Distance in terms of surface currents

Surface currents represent a distributional perspective on surfaces. More specifically, a surface is seen in the sense of distributions of de Rham called *currents* [47]. These currents are classical distributions acting on the space of differential forms. Formally, an $m$-current is a linear functional on the space of $m$-forms. In particular, a 2-current is called surface current and it is naturally defined by a 2-form $\omega$ via $S = \int_S \omega$. Surface currents have been introduced in the context of the large deformation diffeomorphic metric mapping (LDDMM) framework [see e.g. 57] for surface matching by Vaillant and Glaunès [221] extending the work of Glaunès et al. [84] on the pairing of classical distributions. In the field of parameter identification, they have been introduced as a measure by Imperiale et al. [104]. For a mathematically profound introduction to the general topic of matching of distributions and surface currents, the reader is referred to Glaunès [85] and Durrleman [58].

A surface current $S(\omega) \in \mathcal{W}^*$ is a linear functional on the space of test functions $\mathbf{w} \in \mathcal{W}$ defining the differential 2-form $\omega$ given by (3.42). It is given by the integral

$$S(\omega) := \int_S \mathbf{w} \cdot \mathbf{n} \, \mathrm{d}\Gamma. \tag{3.46}$$

With $\mathcal{W}$ being given as a Hilbert space and by the application of the Riesz representation theorem, every current $S$ admits a dual representation $\boldsymbol{K}_{\mathbf{n}}^S \in \mathcal{W}$ such that

$$\forall \mathbf{w} \in \mathcal{W} \quad S(\omega) = \langle \boldsymbol{K}_{\mathbf{n}}^S, \mathbf{w} \rangle_{\mathcal{W}}. \tag{3.47}$$

Thus, a norm in the space of currents $\mathcal{W}^*$ is defined via the dual space norm

$$\|S\|_{\mathcal{W}^*}^2 = \langle \boldsymbol{K}_{\mathbf{n}}^S, \boldsymbol{K}_{\mathbf{n}}^S \rangle_{\mathcal{W}} = \int_S \boldsymbol{K}_{\mathbf{n}}^S \cdot \mathbf{n} \, \mathrm{d}\Gamma. \tag{3.48}$$

Fixing a space $\mathcal{W}$ such that (3.46) is well defined, the dual space norm can be used to construct a similarity measure on surfaces. Considering the surface $S$ to be given by the model $S = C(\mathbf{A}(\boldsymbol{\theta}))$ and a measured surface $\mathbf{Z}$, a distance in the space of currents is given by

$$D(\mathbf{Z}, S) = \frac{1}{2} \|\mathbf{Z} - S\|_{\mathcal{W}^*}^2 . \tag{3.49}$$

**Push-forward action on currents** Given the diffeomorphism $\varphi : \Omega_0 \to \Omega_t$ associated to the solution of the nonlinear problem (2.102) and a surface $S \subset \Omega_0$, the relation between the natural push-forward of the surface $\varphi(S) \subset \Omega_t$ and the push-forward current $\varphi(S)(\omega)$ is given in terms of the pull-back on the 2-form $\varphi_*^{-1}(\omega)$ via

$$\varphi(S)(\omega) = S(\varphi_*^{-1}(\omega)). \tag{3.50}$$

This results as a direct consequence of (3.40) and the integration under a change of coordinates, see appendix B for details. Given a model boundary $\partial\Omega_0$, or a subset $\Gamma$ thereof, the push-forward action (3.50) formally defines the compound action $C \circ \mathbf{A}$ on currents via

$$C \circ \mathbf{A}(\boldsymbol{\theta}) = \varphi(\Gamma)(\omega). \tag{3.51}$$

**Reproducing kernel Hilbert spaces** Following Vaillant and Glaunès [221], the explicit computation of the integral in (3.48) is based on a formulation of the space $\mathcal{W}$ as a reproducing kernel Hilbert space (RKHS). The theory of RKHSs can be traced back to Aronszajn [8] and has a huge range of applications in modern computer science, especially in the area of machine learning [181].

**Definition 3.3.3** (Reproducing kernel Hilbert space)**.** *A reproducing kernel Hilbert space* $\mathcal{H}_k$ *is a Hilbert space of functions* $f : X \mapsto \mathbb{R}$ *over a set* $X$ *such that for each* $x \in X$ *the evaluation functional* $\delta_x(f) = f(x)$ *is a bounded linear functional, i.e.,*

$$\forall f \in \mathcal{H}_k \quad |\delta_x| \leq M \|f\| . \tag{3.52}$$

By the application of the Riesz representation theorem, this definition implies the existence of a unique representer $K_x \in \mathcal{H}_k$ establishing the so called reproducing property

$$f(x) = \langle K_x, f \rangle_{\mathcal{H}_k} \quad \forall f \in \mathcal{H}_k. \tag{3.53}$$

Since $K_x \in \mathcal{H}_k$, it holds that $K_x(y) = \langle K_x, K_y \rangle = k(x,y)$. The symmetric positive definite function $k$ is called a kernel and the unique relation between a kernel and it's RKHS was established by Aronszajn [8]. It is stated here as given by Rasmussen and Williams [181]:

**Theorem 3.3.1** (Moore-Aronszajn theorem)**.** *For every positive definite function* $k : X \times X \mapsto \mathbb{R}$ *there exists a unique RKHS, and vice versa.*

The application of theorem 3.3.1 and the reproducing property (3.53) finally allow for an evaluation of the integral (3.48) via choosing a positive definite function $k$ to represent the space $\mathcal{W}$ of test functions $\mathbf{w}$. Since the test functions are vector-valued, the reproducing property has to be stated for a vector-valued RKHS [see e.g. 35]: For every $\mathbf{v} \in \mathbb{R}^3$ and $x \in X$

$$\langle \mathbf{w}(x), \mathbf{v} \rangle_{\mathbb{R}^3} = \langle \mathbf{w}, K_x \mathbf{v} \rangle_{\mathcal{W}} \quad \forall \mathbf{w} \in \mathcal{W}. \tag{3.54}$$

**Remark.** *The application of the reproducing property to vector-valued spaces actually requires matrix-valued representers* $\Pi(x, \cdot)$*. Throughout this thesis, these representers are chosen to be separable in the sense that* $\Pi(x, \cdot)\mathbf{v} = K_x \mathbf{v}$ *[see e.g. 4].*

Applying the reproducing property to the integrand in (3.48) results in

$$\boldsymbol{K}^S_{\mathbf{n}} \cdot \hat{\mathbf{n}} = \langle \boldsymbol{K}^S_{\mathbf{n}}, K_x \hat{\mathbf{n}} \rangle_{\mathcal{W}} = \int_S K_x \hat{\mathbf{n}} \cdot \mathbf{n} \, \mathrm{d}\Gamma = \int_S K_x \mathbf{n} \, \mathrm{d}\Gamma \cdot \hat{\mathbf{n}} \quad \forall \hat{\mathbf{n}} \in \mathbb{R}^3. \tag{3.55}$$

Thus, the dual representation $\boldsymbol{K}^S_{\mathbf{n}}$ of a current $S$ is given by

$$\boldsymbol{K}^S_{\mathbf{n}} = \int_S K_x \mathbf{n} \, \mathrm{d}\Gamma = \int_S k(x, \cdot) \mathbf{n} \, \mathrm{d}\Gamma. \tag{3.56}$$

This enables the computation of the inner product in the space of currents between two surfaces $S_1$ and $S_2$. By using (3.56) in (3.48), this inner product is given by

$$\langle S_1, S_2 \rangle_{\mathcal{W}^*} = \langle \boldsymbol{K}^{S_1}_{\mathbf{n}}, \boldsymbol{K}^{S_2}_{\mathbf{n}} \rangle_{\mathcal{W}} = \int_{S_1} \int_{S_2} \mathbf{n}^{S_1}(x) \cdot k(x,y) \mathbf{n}^{S_2}(y) \, \mathrm{d}\Gamma \, \mathrm{d}\Gamma. \tag{3.57}$$

Accordingly, the computation of a distance between the two elements $S_1$ and $S_2$ is given by

$$\begin{aligned}\|S_1 - S_2\|^2_{\mathcal{W}^*} = &\int_{S_1}\int_{S_1} \mathbf{n}^{S_1}(x) \cdot k(x,y)\mathbf{n}^{S_1}(y)\,\mathrm{d}\Gamma\,\mathrm{d}\Gamma \\ &- 2\int_{S_1}\int_{S_2} \mathbf{n}^{S_1}(x) \cdot k(x,y)\mathbf{n}^{S_2}(y)\,\mathrm{d}\Gamma\,\mathrm{d}\Gamma \\ &+ \int_{S_2}\int_{S_2} \mathbf{n}^{S_2}(x) \cdot k(x,y)\mathbf{n}^{S_2}(y)\,\mathrm{d}\Gamma\,\mathrm{d}\Gamma. \end{aligned} \tag{3.58}$$

**Discretization in space** For the numerical computation of the distance in the space of currents, (3.58) has to be approximated by a discretization. Following Vaillant and Glaunès [221], a one point integration is employed by using a triangulation $\mathcal{T}_S$. As a consequence, a current $S(\omega)$ given by the dense span of Dirac currents $\delta_x^{\mathbf{n}} = \mathbf{w}(x)\cdot\mathbf{n}(x)\,\mathrm{d}\Gamma$ can be approximated by a finite sum:

$$S(\omega) = \sum_{x \in S} \delta_x^{\mathbf{n}} \approx S_h(\omega) = \sum_{i=1}^{N_S} \delta_{\mathbf{c}_i}^{\tilde{\mathbf{n}}_i} = \sum_{i=1}^{N_S} \mathbf{w}(\mathbf{c}_i) \cdot \tilde{\mathbf{n}}_i, \tag{3.59}$$

where $\tilde{\mathbf{n}}_i$ is the $i$-th triangles normal with length equal to the triangle's area. A norm on the discrete current $S_h(\omega)$ is then given as the approximation to (3.57) as

$$\|S\|_{\mathcal{W}^*} \approx \|S_h\| = \sum_{i=1}^{N_S}\sum_{j=1}^{N_S} \tilde{\mathbf{n}}_i \cdot k(\mathbf{c}_i, \mathbf{c}_j)\tilde{\mathbf{n}}_j. \tag{3.60}$$

Assembling each of the 3 spatial components of the normals $\tilde{\mathbf{n}}_i$ in the $N_S$-vectors $\underline{\tilde{\mathbf{n}}}_1, \underline{\tilde{\mathbf{n}}}_2$ and $\underline{\tilde{\mathbf{n}}}_3$, the notation can be simplified to

$$\|S_h\| := \sum_{i=1}^{3} \underline{\tilde{\mathbf{n}}}_i \cdot \underline{K} \cdot \underline{\tilde{\mathbf{n}}}_i, \tag{3.61}$$

whereby the elements of the $N_S \times N_S$ matrix $\underline{K}$ are given by $[\underline{K}]_{ij} = k(\mathbf{c}_i, \mathbf{c}_j)$.

**Random variables in the space of discrete currents** To be able to setup the statistical model (3.23), it is necessary to define random variables with associated probability density in the space of currents. Formally, a definition of a current as a random Gaussian variable is given through the projection of a current as an element of an infinite dimensional Hilbert space to a finite dimensional subspace. Whereas this definition is substantiated by the concept of Gaussian processes, see chapter 3.4.1, a practical approach is pursued at this point. This approach is enabled by the finite dimensional representation (3.59) of a current. For a detailed introduction to random variables in the space of currents, the reader is referred to Durrleman [58].

Initially, it is assumed that the normals $\tilde{\mathbf{n}}_i$ are random variables with uncorrelated spatial components. The joint probability density of the random vectors $\boldsymbol{\mathsf{n}}_1$, $\boldsymbol{\mathsf{n}}_2$ and $\boldsymbol{\mathsf{n}}_3$ associated to the component-wise vectors $\underline{\tilde{\mathbf{n}}}_1, \underline{\tilde{\mathbf{n}}}_2, \underline{\tilde{\mathbf{n}}}_3$ can consequently be given as

$$p_{\boldsymbol{\mathsf{n}}_1,\boldsymbol{\mathsf{n}}_2,\boldsymbol{\mathsf{n}}_3}(\underline{\mathbf{n}}_1, \underline{\mathbf{n}}_2, \underline{\mathbf{n}}_3) = p_{\boldsymbol{\mathsf{n}}_1}(\underline{\mathbf{n}}_1)p_{\boldsymbol{\mathsf{n}}_2}(\underline{\mathbf{n}}_2)p_{\boldsymbol{\mathsf{n}}_3}(\underline{\mathbf{n}}_3). \tag{3.62}$$

If the elements in $\mathsf{n}_i$ are further assumed to be normally distributed variables with mean $\underline{\tilde{\mathbf{n}}}_i$ and covariance $\mathbf{\Sigma_Z}$, the probability density $p_{\mathsf{n}_i}$ is given by

$$p_{\mathsf{n}_i}(\underline{\mathbf{n}}_i) \propto \exp(-\frac{1}{2}(\underline{\mathbf{n}}_i - \underline{\tilde{\mathbf{n}}}_i) \cdot \mathbf{\Sigma}_{\mathbf{Z}}^{-1} \cdot (\underline{\mathbf{n}}_i - \underline{\tilde{\mathbf{n}}}_i)). \tag{3.63}$$

Accordingly, the joint probability is computed by

$$p_{\mathsf{n}_1,\mathsf{n}_2,\mathsf{n}_3}(\underline{\mathbf{n}}_1, \underline{\mathbf{n}}_2, \underline{\mathbf{n}}_3) \propto \exp(-\frac{1}{2}\sum_{i=1}^{3}(\underline{\mathbf{n}}_i - \underline{\tilde{\mathbf{n}}}_i) \cdot \mathbf{\Sigma}_{\mathbf{Z}}^{-1} \cdot (\underline{\mathbf{n}}_i - \underline{\tilde{\mathbf{n}}}_i)). \tag{3.64}$$

Setting $\mathbf{\Sigma}_{\mathbf{Z}}^{-1} = \underline{K}$, it is immediately identified from (3.61) that

$$p_{\mathsf{n}_1,\mathsf{n}_2,\mathsf{n}_3}(\underline{\mathbf{n}}_1, \underline{\mathbf{n}}_2, \underline{\mathbf{n}}_3) = p_{\mathsf{S}_h}(S) \propto \exp(\frac{1}{2}\left\|S - S_h\right\|^2). \tag{3.65}$$

As a consequence, a discrete random current $\mathsf{S}_h$, which is distributed according to

$$\mathsf{S}_h \propto \exp(\frac{1}{2}\left\|S - S_h\right\|^2), \tag{3.66}$$

is in fact a normally distributed random variable entirely defined by the field of normals $\tilde{\mathbf{n}}_i$ and a kernel function $k$ (the kernel of the associated RKHS of test functions $\mathbf{w}$).

**Remark.** *To directly control the variance of the random currents, it is useful to use kernels with*

$$k(x, y) = 1 \iff x = y. \tag{3.67}$$

*Scaling such kernels by a factor* $\sigma_N^{-2}$, *the probability density of the associated current is given by* $p(S) \propto \exp(-\frac{\|S\|^2_{\mathcal{W}^*}}{2\sigma_N^2})$.

**Fixing a kernel** Definition 3.3.3 and theorem 3.3.1 allow the use of a variety of different kernel functions [see e.g. 181]. Generally, specific applications require specific choices of kernels. The application in the context of the current formalism utilizes the RKHS theory as a means to define the space $\mathcal{W}$ of test functions $\mathbf{w}$. Thus, the choice of a kernel specifies a particular way of *testing* a surface. Throughout this thesis, this testing is defined in terms of the Gaussian kernel

$$k(x, y) = \sigma_N^{-2} \exp(-\frac{\left\|x - y\right\|^2}{2\sigma_W^2}). \tag{3.68}$$

In the setting of random variables in the space of currents, the factor $\sigma_N$ specifies the variance of a random current. It can directly be interpreted as the variance of the field of normals defining the random current. The scale $\sigma_W$ characterizes the spatial scale of the covariance structure on the field of normals. Accordingly, $\sigma_W$ can be seen as a spatial scale at which the space of test functions distinguishes features of a surface. For large scales $\sigma_W$, the matrix $\underline{K}$ defining a discrete current's probability distribution becomes increasingly ill-conditioned. This issue is intensified for currents defined by fine triangulations. From the current's perspective, the discretization size implicitly defines the spatial scale of distinct surface features. Thus, the spatial scale $\sigma_W$ of the kernel and the discretization size should not differ by orders of magnitude.

**Likelihood in terms of currents** The definition of Gaussian random variables in the space of currents and the transformation of currents under some push-forward action $\varphi(\bullet)(\omega)$ finally allows to substantiate a likelihood-function in terms of measurements of surfaces. Given a measured surface **Z** corresponding to some model boundary $\Gamma$, the likelihood $p(\mathbf{Z}|\boldsymbol{\theta})$ is given by

$$p(\mathbf{Z}|\boldsymbol{\theta}) \propto \exp(-\frac{\|\mathbf{Z} - \varphi(\Gamma)\|^2_{\mathcal{W}^*}}{2\sigma_N^2}). \tag{3.69}$$

Given a triangulation $\mathcal{T}_Z$ of the surface $\mathbf{Z}$ and a triangulation $\mathcal{T}_\Gamma$ of the surface $\varphi(\Gamma)$ according to definition 3.3.2, the norm defining the likelihood (3.69) is approximated according to (3.59) by

$$\begin{aligned}\|\mathbf{Z} - \varphi(\Gamma)\|^2_{\mathcal{W}^*} \approx & \sum_{\tau_i^Z}\sum_{\tau_j^Z} \tilde{\mathbf{n}}_i^Z \cdot k(\mathbf{c}_i^Z, \mathbf{c}_j^Z)\tilde{\mathbf{n}}_j^Z \\ & -2\sum_{\tau_i^Z}\sum_{\tau_j^\Gamma} \tilde{\mathbf{n}}_i^Z \cdot k(\mathbf{c}_i^Z, \mathbf{c}_j^\Gamma)\tilde{\mathbf{n}}_j^\Gamma \\ & +\sum_{\tau_i^\Gamma}\sum_{\tau_j^\Gamma} \tilde{\mathbf{n}}_i^\Gamma \cdot k(\mathbf{c}_i^\Gamma, \mathbf{c}_j^\Gamma)\tilde{\mathbf{n}}_j^\Gamma .\end{aligned} \tag{3.70}$$

Throughout the work presented in this thesis, the triangulation $\mathcal{T}_\Gamma$ is directly obtained from the mesh associated to the FE discretization. Therefore, the triangles are either given directly as faces of linear tetrahedral elements or by subdivision of linear hexahedral elements. Thus, the coordinates defining the triangles in $\mathcal{T}_\Gamma$ are coupled to the degrees of freedom $\mathbf{D}$ defining the approximate solution $\mathbf{U}_h$ given by (2.102) in a *one-to-one* relation.

**Visualization of random currents** In order to get an intuition of the concepts introduced above, a geometrically simple example given by a cylindrical oriented surface is analyzed. For the purpose of visualization, only a 2-dimensional slice is depicted. Two different discretizations (coarse and fine) of this surface are shown in figure 3.6. The current representations of these discretizations are entirely defined by the kernel (3.68) with $\sigma_N = 0.02$ mm and $\sigma_W = 0.5$ mm. Figure 3.7 shows a random realization for each of the two discretizations. It can be seen by inspection of the respective dual representation $\boldsymbol{K}_\mathbf{n}$ that the spatial scale of the action $S(\mathbf{w})$ of the current on some test function is defined by the scale $\sigma_W$ of the kernel $k$. This action is independent of the discretization of the surface. The magnitude of this action is defined by the variance $\sigma_N^2$. However, it is important to note that the physical interpretation of the variance $\sigma_N^2$ on real data is only meaningful with respect to a discretization size that is able to reflect the random variation of the surface. I.e., $\sigma_N^2$ describes the variance of normal vectors corresponding to a discretization whose randomness can be described in the sense of (3.66). This issue is important when the discretization size is defined by the requirements of the computational model and not just seen as a means of data representation. The effects of this aspect in a situation using real data are discussed in detail in chapter 7.

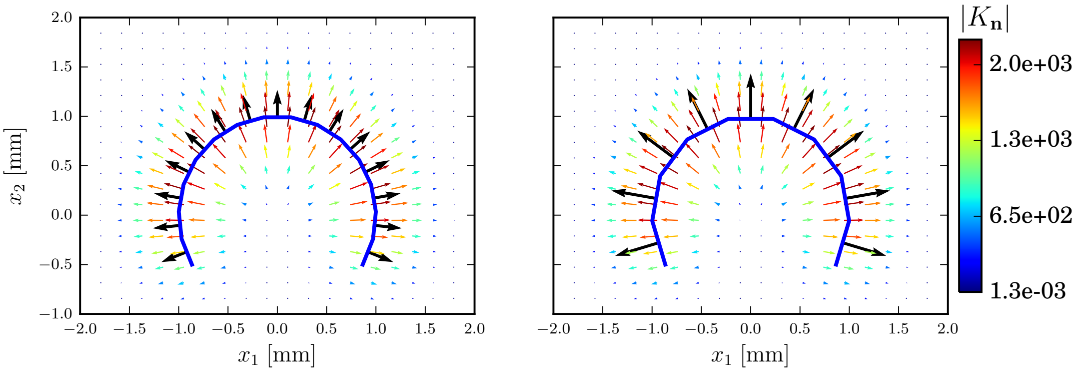

Figure 3.6.: A 2-dimensional slice of an open cylindrical surface and its discretization (blue). Left: fine discretization with 15 'triangles' (or line elements). Right: Coarse discretization with 7 'triangles' (or line elements). Additionally, the respective currents are visualized in terms of the field of normals $\tilde{\mathbf{n}}$ (black) and the dual representation $\boldsymbol{K}_{\mathbf{n}}$. Arrow lengths of the vector field $\boldsymbol{K}_{\mathbf{n}}$ are scaled for the purpose of visualization. It can be seen that the spatial scale of the current is not determined by the discretization but by the kernel $k$.

**Computational aspects** For a dense kernel $k(x, y)$, the computation of (3.70) is a $\mathcal{O}(N_\Gamma N_Z + N_\Gamma^2)$ operation, with $N_Z$ and $N_\Gamma$ being the size of $\mathcal{T}_Z$ and $\mathcal{T}_\Gamma$. Since it is to be evaluated frequently as part of the computational model, the efficiency of this evaluation is inevitable.

A straight forward approach is given by truncation of the kernel. Although 'density' of the kernel implies infinite support, depending on the spatial scale of the kernel, this support does not necessarily contribute significantly to the summations in (3.70). However, a truncation $k(x, y) < \epsilon \rightarrow k(x, y) = 0$ reduces the accuracy of the linearization of (3.70), see appendix C.1. On the other hand, for dense kernels with small spatial scale, the accuracy is affected due to cancellation and round-off errors anyway. The impact of such effects can in principle be ameliorated by using sparse kernels, i.e., kernels with bounded support.

Nevertheless, the approach pursued in this thesis is based on the use of modern computer architecture. In this respect, the efficiency of the evaluation of (C.11) is improved by the use of modern shared-memory programming models such as Pthreads, open multiprocessing (OpenMP) or CUDA. To this end, the implementation used for the work in this thesis utilizes the Kokkos library from the Trilinos project [95] that provides a portable abstraction layer.

## 3.4. A priori assumptions and regularization

In order for Bayes theorem (3.22) to be applicable, it is necessary to define the 'a priori'-knowledge on the parameters $\boldsymbol{\theta}$ in a functional relation $p(\boldsymbol{\theta})$, the prior. As a probability density, this functional relation represents the statistical universe of the parameters. In many practical situations, a mathematically rigorous formulation of this relation is difficult. On the one hand, in a multi-dimensional setting, prior knowledge is often vaguely

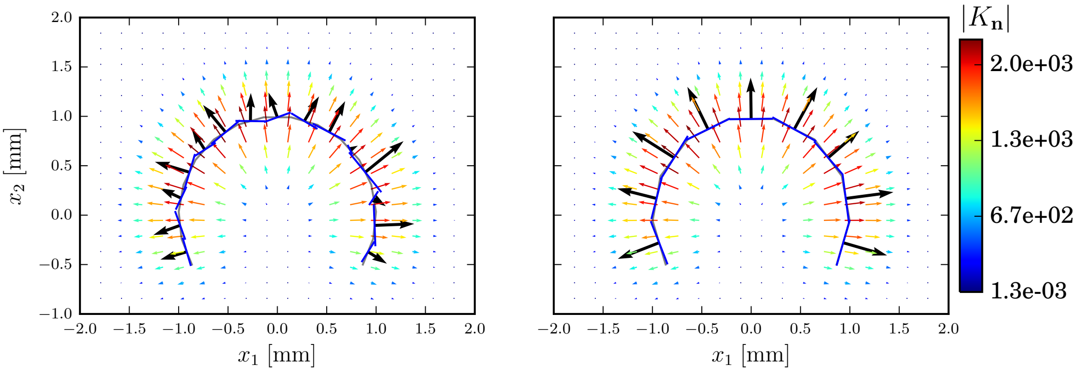

Figure 3.7.: Currents from figure 3.6 corrupted with normally distributed noise $\propto \mathcal{N}(0, \sigma_N^2)$. Comparing the two noisy samples from the coarse and the fine discretization, it can be seen, as for the uncorrupted currents from figure 3.6, that also the noise in the space of currents is determined by the spatial scale $\sigma_W$ and the magnitude $\sigma_N$ of the kernel $k$. However, the effect on the normals only has a physically meaningful interpretation where the discretization size fits into the noise model (3.66).

expressed in terms of certain structural requirements on the solution. These requirements can be given by assumptions on the heterogeneity (e.g., in case of anisotropy) or smoothness of the solution. On the other hand, committing to a certain functional dependency of the prior might introduce additional parameters, so called *hyperparameters*, which have to be modeled again by (hyper-)priors. In this light, it has been suggested that the construction of prior models has to put the focus on the gross value of the probability density [44]. Given a set E of expectable values and a set U of unexpectable values, the prior should suffice to the relation

$$p(x) \gg p(x') \quad \forall x \in \mathrm{E}, x' \in \mathrm{U}. \tag{3.71}$$

In the following, several popular choices of prior models are presented. For a more detailed introduction to prior models used in inverse problems, the reader is referred to Kaipio and Somersalo [110]. Insight into the general topic of prior modeling is, e.g., given by Berger [19] or D'Agostini [44].

### 3.4.1. Gaussian prior

A very popular class of priors is generated by the normal distribution $\mathcal{N}$. A random variable $\mathbf{x} \in \mathbb{R}^{n_p}$ is said to be normally distributed if its probability density is given by the multivariate normal distribution

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\left((2\pi)^{n_p} |\boldsymbol{\Sigma}|\right)^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}) \cdot \boldsymbol{\Sigma}^{-1} \cdot (\boldsymbol{x} - \boldsymbol{\mu})\right), \tag{3.72}$$

with mean $\boldsymbol{\mu}$, the symmetric positive definite (SPD) $n_p \times n_p$ covariance matrix $\boldsymbol{\Sigma}$ and the determinant of the covariance $|\boldsymbol{\Sigma}|$. Beside being represented by a relatively simple explicit formula, its popularity stems mainly from two important facts:

1. For large data sets of independently and identically distributed variables, the *Central Limit theorem* establishes the important relation between the arithmetic mean and the normal distribution [see e.g. 59]. This implies that Gaussian densities can often be used to approximately describe statistics of a large amount of independent observations.

2. Under certain circumstances, the use of Gaussian priors results in explicitly tractable formulas for the posterior. In cases where, for a fixed likelihood, the posterior falls in the same family of functions as the prior, the prior is called a *conjugate prior.* For certain families of functions these relations are tabulated.

**Gaussian process** In the absence of specific knowledge on the covariance structure of the signal $\mathsf{x}$, the covariance matrix is often simply given by $\mathbf{\Sigma} = \sigma_g^2\mathbf{I}$. Confidence in the scale of variability is thereby given by the variance $\sigma_g^2$. However, a signal can also be associated to a certain spatial structure which will potentially be reflected in the covariance structure. E.g., this is the case when the random signal is given by the model parameters $\mathsf{x} = \boldsymbol{\theta}$ that are defined as the spatial distribution of growth parameters on the domain $\Omega_0$. If in line with prior knowledge, it might be reasonable to expect that the notion of distance in the space will be reflected in a correlation of parameters defined on that space. Generally, given a finite index set J and a $d$-dimensional space $\mathbb{R}^d$, the spatial association of the signal $\mathsf{x}$ can be given in terms of the random function $\mathsf{y}(s)$ as

$$\mathsf{x} = \{\mathsf{y}(s_i)\}_{i\in \mathrm{J}}, s_i \in \mathbb{R}^d. \tag{3.73}$$

Thereby, the symbol $s$ is used as a spatial coordinate to avoid confusion with $\mathsf{x}$ as a general random variable. Given a *covariance function* $k_\Sigma : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$, the covariance matrix can be defined by

$$[\mathbf{\Sigma}]_{ij} = k_\Sigma(s_i, s_j). \tag{3.74}$$

If this definition holds for any finite index set J, the function $\mathsf{y}(s)$ is referred to as *Gaussian process.* For $d \geq 2$, it is also more specifically referred to as *Gaussian random field.* As can be seen from the definition of the covariance matrix, a Gaussian process represents an infinite-dimensional extension of a multivariate normally distributed random variable. The specific choice of the covariance function $k_\Sigma$ depends on the assumed spatial correlation structure. Beside the constraint that the covariance matrix $\mathbf{\Sigma}$ has to remain SPD for any set J, there exists a multitude of possible choices [see e.g. 181]. A very popular covariance function is the *squared exponential kernel*

$$k_\Sigma(s_1, s_2) := \sigma_g^2 \exp(-\frac{\|s_1 - s_2\|^2_{\mathbb{R}^d}}{2\ell^2}), \tag{3.75}$$

with the spatial scale of the kernel given by $\ell$.

**Remark.** *The extension of a scalar valued Gaussian process/field to the vector valued case reveals that the considerations (3.62) - (3.66), which lead to the definition of discrete random currents, are formalized by modeling the field of normals* $\mathbf{n}$ *as a vector valued Gaussian random field with the precision matrix* $\mathbf{\Sigma}^{-1}$ *defined by the kernel (3.68).*

### 3.4.2. Gaussian smoothness priors

The term *Gaussian smoothness prior* is used in a context where a certain *structural information* such as smoothness or anisotropy is encoded in the a priori knowledge. These priors are typically defined by

$$p_{\mathsf{x}}(\boldsymbol{x}) \propto \exp\left(-\frac{1}{2\sigma_s^2}\left\|\mathbf{L}\cdot\boldsymbol{x}\right\|^2\right), \tag{3.76}$$

where the differential operator $\mathbf{L} : \mathbb{R}^{n_p} \mapsto \mathbb{R}^k$ incorporates the structural information. In situations where $k < n_p$ or $\ker(\mathbf{L}) \neq \{0\}$, the matrix $\mathbf{L}^\top\mathbf{L}$ is not necessarily invertible and a direct interpretation in terms of a Gaussian density is therefore not straightforward. In case of $\ker(\mathbf{L}) \neq \{0\}$, the smoothness prior (3.76) does not even represent a probability density since

$$\int_{\mathbb{R}^{n_p}} \exp\left(-\frac{1}{2\sigma_s^2}\left\|\mathbf{L}\cdot\boldsymbol{x}\right\|^2\right)\mathrm{d}\boldsymbol{x} = \infty. \tag{3.77}$$

A normalization is therefore not generally possible rendering this sort of priors *improper*. The rationale behind the name *Gaussian smoothness* is given by the fact that (3.76) emerges as a limiting process of some Gaussian density.

Let an orthonormal basis for $\ker(\mathbf{L})$ be given by the columns of the $n_p \times n_{ker}$-matrix $\mathbf{Q} = [\mathbf{q}_i, \ldots, \mathbf{q}_{n_{ker}}]$ with $n_{ker} = \dim(\ker(\mathbf{L}))$. Then the factor space $\mathbb{R}^{n_p}_{\backslash\mathbf{Q}} := \mathbb{R}^{n_p} \setminus \ker(\mathbf{L})$ is given by $\mathbb{R}^{n_p}_{\backslash\mathbf{Q}} = \{\mathbf{v} \in \mathbb{R}^{n_p} | \mathbf{v}\cdot\mathbf{Q} = \mathbf{0}\}$. The integration of (3.77) over this factor space is bounded:

$$\int_{\mathbb{R}^{n_p}_{\backslash\mathbf{Q}}} p_{\mathsf{x}}(\boldsymbol{x})\mathrm{d}\boldsymbol{x} < \infty. \tag{3.78}$$

Furthermore, let a signal $\mathsf{y}$ be given in terms of the basis $\mathbf{Q}$, a vector of random coefficients $\mathsf{h}_i$ and a constant vector $\boldsymbol{y}_0 \in \ker(\mathbf{L})$ such that

$$\mathsf{y} = \sum_{i=1}^{n_{ker}} \mathbf{q}_i\mathsf{h}_i + \boldsymbol{y}_0 = \mathbf{Q}\cdot\mathsf{h} + \boldsymbol{y}_0. \tag{3.79}$$

If the coefficients $\mathsf{h}$ are chosen to satisfy $\mathsf{h} \sim \mathcal{N}(\mathbf{0}, a^2\mathbf{I})$, then $\mathbb{E}[\mathsf{y}] = \boldsymbol{y}_0$ and $\mathrm{Cov}(\mathsf{y}) = a^2\mathbf{Q}\mathbf{Q}^\top$. An (improper) prior $p_{\mathsf{y}}(\boldsymbol{y})$ is then given by

$$p_{\mathsf{y}}(\boldsymbol{y}) \propto \exp\left(-\frac{1}{2a^2}\boldsymbol{y}\cdot\mathbf{Q}\mathbf{Q}^\top\cdot\boldsymbol{y},\right). \tag{3.80}$$

Finally, a re-parametrization of the random signal $\mathsf{x}$ can be considered such that

$$\mathsf{z} = \mathsf{x} + \mathsf{y}, \tag{3.81}$$

whereby $\mathsf{x} \sim p_{\mathsf{x}}(\boldsymbol{x})$ and $\mathsf{y} \sim p_{\mathsf{y}}(\boldsymbol{y})$. Since $\mathsf{x}$ and $\mathsf{y}$ are mutually independent, a new prior $p_{\mathsf{z}}(\boldsymbol{z})$ is given by

$$p_{\mathsf{z}}(\boldsymbol{z}) = p_{\mathsf{x},\mathsf{y}}(\boldsymbol{x}, \boldsymbol{y}) = p_{\mathsf{x}}(\boldsymbol{x})p_{\mathsf{y}}(\boldsymbol{y}). \tag{3.82}$$

Inserting the definitions of $p_{\mathsf{x}}(\boldsymbol{x})$ and $p_{\mathsf{y}}(\mathbf{y})$ into (3.82) results in

$$\begin{aligned} p_{\mathsf{z}}(\boldsymbol{z}) &\propto \exp\left(-\frac{1}{2}\left(\frac{1}{\sigma_s^2}\boldsymbol{x}\cdot\mathbf{L}^\top\mathbf{L}\cdot\boldsymbol{x} + \frac{1}{a^2}\boldsymbol{y}\cdot\mathbf{Q}\mathbf{Q}^\top\cdot\boldsymbol{y}\right)\right) \\ &= \exp\left(-\frac{1}{2}\boldsymbol{z}\cdot\left(\frac{1}{\sigma_s^2}\mathbf{L}^\top\mathbf{L} + \frac{1}{a^2}\mathbf{Q}\mathbf{Q}^\top\right)\cdot\boldsymbol{z}\right). \end{aligned} \tag{3.83}$$

The matrix $\mathbf{M} := \left(\frac{1}{\sigma_s^2}\mathbf{L}^\top\mathbf{L} + \frac{1}{a^2}\mathbf{Q}\mathbf{Q}^\top\right)$ is invertible as a direct consequence of the analysis provided in appendix D.1. Thus, (3.83) represents a Gaussian probability density justifying the naming *Gaussian smoothness.*

One popular example for an a priori assumed structural information is smooth differentiability. In this case, the operator $\mathbf{L}$ is given by a discretized version of the Laplace-operator $\nabla \cdot (\nabla \bullet)$. Since fixing data on some part of the boundary is usually not justified by a priori knowledge, the operator is left with the natural boundary condition resulting in the mentioned singular structure. The kernel of $\mathbf{L}$ is then represented by a constant mode. If the formulation (3.83) is used, confidence in an a priori chosen absolute level $\boldsymbol{y}_0$ of this constant mode is expressed by the variance $a^2$.

**Remark.** *The limiting case $a \to \infty$ recovers the original definition of the Gaussian smoothness prior, in which properness of the posterior is left to be imposed entirely by the data. Depending on the data, this approach can be justified from an application point of view.*

### 3.4.3. Markov random field priors

The prior models introduced so far impose the assumption of a regularly varying spatial variation of a signal. In some applications, this regularity is not necessarily justified. For instance, signals might be expected to be bounded, but this boundedness is not expected to be spatially uniform. In light of (3.71), the priors introduced in chapter 3.4.1 and chapter 3.4.2 might impose too strict assumptions. A very popular approach to model prior knowledge in these scenarios is the concept of total variation (TV) made popular by Rudin et al. [190]. For the application of image denoising the TV functional was used to account for the irregular and 'blocky' character occurring in images. Whereas in image denoising the basic problem to solve is mostly linear, TV regularization has also been applied to nonlinear problems in PDE-constrained optimization [see e.g. 38].

The concept of total variation is defined in terms of functions of bounded variation [83]:

**Definition 3.4.1** (Total variation)**.** *Let $\Omega \subseteq \mathbb{R}^n$ be an open set and let $f \in \mathrm{L}^1(\Omega)$. Let $C_0^1$ be the differentiability class $C^1$ of functions with vanishing support at $\partial\Omega$:*

$$TV(f) = \sup\left\{\int_\Omega f\nabla \cdot g \mathrm{d}x \;:\; g \in (C_0^1(\Omega))^n,\ \|g(x)\| \leq 1, \forall x \in \Omega\right\}. \tag{3.84}$$

A function $f$ is said to have bounded variation if $TV(f) < \infty$. Explicit formulas exist for the computation of $TV(f)$ if more restrictive spaces are considered. E.g., for $f \in C^1(\Omega)$ it follows from the integration by parts that

$$TV(f) = \int_\Omega |\nabla f| \,\mathrm{d}x. \tag{3.85}$$

In order to render this metric applicable in gradient based solution methods, i.e., to account for the non-differentiability, a smoothed approximation can be applied, see figure 3.8:

$$TV_\epsilon(f) = \int_\Omega \sqrt{|\nabla f|^2 + \epsilon}\,\mathrm{d}x. \tag{3.86}$$

Although well-defined for $C^1$-functions, it is convenient to degenerate the original definition of TV even further to allow for a more flexible setup of functions. Mainly this is motivated by the fact that it is useful to represent functions by piece-wise constant approximations for computational convenience. Particularly in terms of FE discretization of the forward problem, an element-wise constant parametrization is appealing. Thereby the connectivity between elements defines a unidirectional graph structure and the solution to the inverse problem (3.22) can be represented as discrete signal defined on the vertices of this graph.

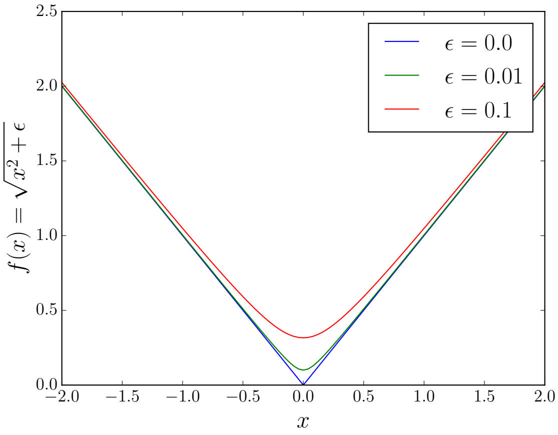

Figure 3.8.: Illustration of the smoothing properties of the parameter $\epsilon$ in the smoothed TV approximation (3.86). With increasing $\epsilon$, the singularity in $\frac{\mathrm{d}f}{\mathrm{d}x}|_{x=0}$ is increasingly smoothed out. Choosing $\epsilon$ small is desirable with respect to the exactness of the approximation. On the other hand, finding $\min f(x)$ in the context of Quasi-Newton methods (especially with inexact line searches) becomes an increasingly oscillatory process with decreasing $\epsilon$. Choosing a value for $\epsilon$ is a trade off between the exact representation of the TV functional (3.85) and an affordable numerical solution process.

**Definition 3.4.2.** *Let* $\mathrm{J}$ *be an index set with* $\dim(\mathrm{J}) = n_V$*. Let* $G = (V, E, w)$ *be a weighted unidirectional graph with the indexed set of vertices* $\{V_i\}_{i \in \mathrm{J}}$*, edges* $E \subseteq \mathrm{J} \times \mathrm{J}$ *and the non negative symmetric weight function* $w^G(i,j) = w^G(j,i)\,\forall\,(i,j) \in E$*. A graph structure is then defined by the adjacency matrix* $\mathbf{W}^G$ *with entries*

$$[\mathbf{W}^G]_{ij} = \begin{cases} w^G(i,j) & \textit{if } (i,j) \in E, \\ 0 & \textit{otherwise.} \end{cases} \tag{3.87}$$

For a signal $\boldsymbol{x} \in \mathbb{R}^{n_V}$ defined on the vertices of the graph $G$, a graph-based approximation of the TV functional is given by

$$TV_w(\boldsymbol{x}) = \|\nabla_w \boldsymbol{x}\|_{1,2} = \sum_i \left( \sum_j w^G(i,j)(x_j - x_i)^2 + \epsilon^2 \right)^{1/2} \tag{3.88}$$

[see e.g. 96]. This formulation is very flexible with respect to the parametrization of the signal to be represented while still allowing for the beneficial properties of the original TV functional [38].

The definition of a prior probability density over functions of bounded variation is available via the interpretation of (3.88) as the energy of a Markov random field (MRF) [see e.g. 206]. Formally, an MRF on the graph $G$ defines a probability density that can be factorized over the *maximal cliques* of the graph. Given the vertex set $\{V_i\}_{i \in \mathrm{J}}$, the set of maximal cliques $\{\boldsymbol{\iota}_i\}_{i=1}^{n_{cli}}$ is comprised of the maximal cliques $\boldsymbol{\iota}_i = \{k \in \mathrm{J} \,|\, w(i,k) \neq 0\}$. For a signal $\boldsymbol{x} = \{x_i, \ldots, x_n\}$ defined on the vertex set $V$, these cliques define the overlapping partitions

$$\boldsymbol{x}_i = \{x_j : j \in \boldsymbol{\iota}_i\}. \tag{3.89}$$

The signal $\boldsymbol{x}$ is called an MRF if the conditional probability density $\pi(x_j|\boldsymbol{x}\backslash x_j)$ can be written as

$$\pi(x_j \,|\, \boldsymbol{x}\backslash x_j) = \pi(x_j \,|\, \boldsymbol{x}_j). \tag{3.90}$$

Given a potential $\psi(\boldsymbol{x}_i)$ defined over the cliques $\boldsymbol{\iota}_i$, the probability density of the MRF is given according to the Hammersley-Clifford theorem [91] as product over the potential function $\psi(\boldsymbol{x}_i) \geq 0$

$$p(\boldsymbol{x}) = \frac{1}{Z} \prod_{i=1}^{n_{cli}} \psi(\boldsymbol{x}_i). \tag{3.91}$$

The partition function $Z$ accounts for the normalization of $p(\boldsymbol{x})$. Since the potential $\psi$ is positive, it is conveniently given in terms of an *energy* $H$, with $\psi = \exp(-H)$, such that the density can be written as the sum over the energy

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left( -\sum_{i=1}^{n_{cli}} H(\boldsymbol{x}_i) \right). \tag{3.92}$$

Identifying the energy $H(\boldsymbol{x}_i) = (\sum_{j \in \boldsymbol{\iota}_i} w^G(i,j)(x_j - x_i)^2 + \epsilon^2)^{1/2}$, it can be seen that the probability density associated to the TV functional is given by

$$p(\boldsymbol{x}) = \frac{1}{Z} \exp\left( -\alpha_{tv} \left\| \nabla_w \boldsymbol{x} \right\|_{1,2} \right). \tag{3.93}$$

So the TV prior represents one instance of a much more versatile class of priors [see e.g. 34].

By inspection of the energy $H$, it can be seen that it is invariant with respect to translational modes of the signal $\boldsymbol{x}$. Consequently, the TV prior (3.93) suffers from the same improperness as the Gaussian smoothness prior (3.76). Due to the nonlinearity, an interpretation as the limit of a Gaussian density is not possible. But then non-integrability can be approached by the same subspace integration as introduced in chapter 3.4.2. So if the data is not invariant with respect to arbitrary constant modes of the parameters, the inverse problem represented by (3.22) is still well-posed.

### 3.4.4. Comparison of priors

To get an intuition of the effects the priors presented above will have on the posterior, it is instructive to visualize samples drawn randomly from the respective prior distributions. For the Gaussian probability densities (3.72) and (3.83) this can be accurately achieved by means of state-of-the-art random number generating algorithms such a the *Mersenne-Twister* engine [155] and methods of non-uniform random variate generation [51]. These methods are available in many modern software packages and libraries such as MATLAB, Python or C++ (since C++11), just to name a few. For the TV prior (3.88), standard methods for variate generation are not at hand and methods for sampling such as MCMC methods, see chapter 4.4.3, have to be applied.

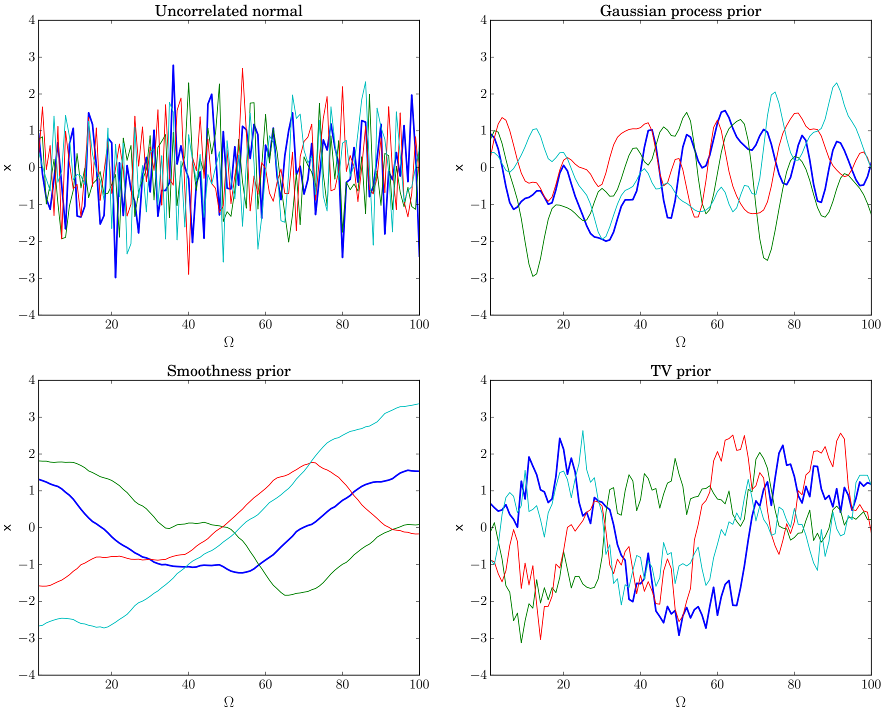

Figure 3.9.: 4 samples drawn from each of the prior models presented in the chapters 3.4.1, 3.4.2 and 3.4.3. *Upper Left*: Gaussian prior without any structural information, i.e., $\mathbf{\Sigma} = \sigma_g^2\mathbf{I}$ ($\sigma_g = 1.0$). *Upper right*: Gaussian prior with covariance structure given by the squared exponential kernel ($\ell = 3.0$). *Bottom left*: Smoothness prior given by a discrete Laplace operator with Neumann boundary conditions. *Bottom right*: Total variation prior with the structural information encoded by the graph structure of the signal.

For the purpose of visualization, a signal $\mathbf{x} \in \mathbb{R}^{100}$ on a simple 1-dimensional domain $\Omega = [1, 100]$ is considered, see figure 3.9. Given a uniform discretization of $\Omega$ into 99 sub-intervals $[s_i, s_j]$, the signal is defined as a function of the coordinates $s_i$. This allows for the definition of the discrete operator $\mathbf{L}$ as a finite-difference version of the Laplace operator. Furthermore, it defines a graph structure of neighboring coordinates necessary for the computation of the energy of the random field in the TV prior. 4 samples of each of the priors (3.72) (uncorrelated and correlated), (3.83) and (3.88) are shown in figure 3.9. For visual comparison, all scale factors have been chosen such that the resulting variates reside in the interval $[-4, 4]$.

Obviously, the Gaussian prior (with $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{\Sigma} = \sigma_g^2\mathbf{I}, \sigma_g = 1.0$) does not favor any particular structural information. Signals are only expected to be drawn in a certain range around the mean more likely than outside. E.g., the probability of the signal $\mathbf{x}$ to reside inside the interval $\mathrm{A} = [\boldsymbol{\mu} - 3\sigma_g, \boldsymbol{\mu} + 3\sigma_g]$ is $P(\mathbf{x} \in \mathrm{A}) \approx 0.997^{100} \approx 0.74$.

In contrast, the variates from the Gaussian process prior exhibit a clear correlation. This correlation is induced by the spatial correlation of the squared exponential kernel ($\ell = 3.0$). The resulting signals are fluctuating with a certain spatial frequency controlled by the scale $\ell$. The smoothness prior and the TV prior represent a different interpretation of 'smoothness'. The smoothness-prior ($\sigma_s^2 = 0.025$) imposes a smooth structure in the sense of smooth differentiability. Since it favors signals with low 'second derivative', likely variates will vary smoothly and at low frequency. In contrast, the TV prior ($\alpha_{tv} = 1.0$, $\epsilon^2 = 0.001$) favors signals with bounded variation. This allows variates to have a much more flexible and irregular structure. In particular, it results in likely samples to be allowed to have certain distinct features, but the amount of such features is bounded.

## 3.5. Lagrangian formulation

Popular approaches for the numerical exploration of the posterior are introduced in detail in chapter 4. Nevertheless, one aspect of the so called maximum a posteriori (MAP) solution is already introduced at this point since it involves a reformulation of the inverse problem. It is therefore rather associated to the setup of the inverse problem than to its numerical solution. By defining

$$\mathcal{J}(\mathbf{U}(\boldsymbol{\theta}), \boldsymbol{\theta}) := \frac{1}{2\sigma^2} D(\mathbf{Z}, F(\boldsymbol{\theta})) - \log p(\boldsymbol{\theta}), \tag{3.94}$$

c.f. chapter 3.1.2, it can be seen that the maximization of the posterior (3.23) is obtained by

$$\underset{\boldsymbol{\theta}}{\operatorname{argmax}}\, p(\mathbf{Z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \equiv \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{J}(\mathbf{U}(\boldsymbol{\theta}), \boldsymbol{\theta}). \tag{3.95}$$

Given enough regularity of the posterior, efficient numerical approaches almost always make use of gradient information $\frac{\mathrm{d}J}{\mathrm{d}\boldsymbol{\theta}}$ to compute the solution to (3.95). But due to the complex relation $\boldsymbol{\theta} \mapsto F(\boldsymbol{\theta})$, the computation of this gradient information is usually accompanied by a significant computational cost. In the low-dimensional regime, it is often feasible to compute approximations of the gradient $\frac{\mathrm{d}\mathcal{J}}{\mathrm{d}\boldsymbol{\theta}}$ by means of finite differences (FDs) via

$$\begin{aligned} \frac{\mathrm{d}\mathcal{J}}{\mathrm{d}\theta_i} &= \frac{\mathrm{d}(-\log p(\theta_1, \dots, \theta_i, \dots, \theta_n|\mathbf{Z}))}{\mathrm{d}\theta_i} && (3.96) \\ &\approx \frac{-\log p(\dots, \theta_i + \Delta_i, \dots |\mathbf{Z}) + \log p(\boldsymbol{\theta}|\mathbf{Z})}{\Delta_i}, && (3.97) \end{aligned}$$

whereby $\Delta_i$ represents an incremental variation in the $i$-th component of the vector of parameters $\boldsymbol{\theta}$. Beside the fact that the accuracy of FD methods is bounded by cancellation and round-off errors in the floating-point arithmetic [137], it is highly inefficient in a high-dimensional regime. This inefficiency results from the fact that the necessary number of evaluations $n_{eval}$ of the posterior is directly coupled to the dimension of the parameters according to

$$n_{eval} = n_p + 1. \tag{3.98}$$

Given that gradient based optimization schemes can easily take in the order of 100-1000 iterations to converge to an optimum, such a computational cost per iteration is prohibitive for many applications.

In the context of PDE-constrained optimization, a remedy to this problem is provided by *adjoint* approaches. These can be derived from a Lagrangian formulation of the inverse problem. To highlight its properties, the derivation starts from a description using a continuous formulation:

$$\tilde{\mathcal{L}}(\mathbf{U}(\theta), \theta) := \frac{1}{2\sigma^2} D(\mathbf{Z}, \tilde{F}(\theta)) + R(\theta), \tag{3.99}$$

whereby the model $\tilde{F} := (C \circ A)(\theta)$, with $\tilde{F} : \mathcal{X} \to \mathcal{Z}$, is the continuous counterpart to the model $F$. This formulation can be extended by the weak form (2.71) in the sense of a constraint to give the *Lagrangian*:

$$\mathcal{L}(\mathbf{U}(\theta), \boldsymbol{\lambda}, \theta) := \tilde{\mathcal{L}}(\mathbf{U}(\theta), \theta) + \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta). \tag{3.100}$$

Thereby, the test functions $\delta\mathbf{U}$ are replaced with the symbol $\boldsymbol{\lambda}$, to avoid confusion of the test functions $\delta\mathbf{U}$ with the variation of the displacements $\delta_{\boldsymbol{\theta}}\mathbf{U}$ due to a variation in the parameters. Given the solution $\mathbf{U} = A(\theta)$, the weak form is always evaluated to

$$\delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta) = 0 \quad \forall \theta. \tag{3.101}$$

Consequently, the following equivalence holds:

$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \tilde{\mathcal{L}}(\mathbf{U}(\theta), \theta) \equiv \underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, \mathcal{L}(\mathbf{U}(\boldsymbol{\theta}), \boldsymbol{\lambda}, \theta), \quad \forall \boldsymbol{\lambda}. \tag{3.102}$$

Given (3.101), it must further hold that

$$\Delta \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta)[\delta\theta] = 0 \quad \forall \delta\theta. \tag{3.103}$$

This guarantees the equivalence

$$\delta\tilde{\mathcal{L}}(\mathbf{U}(\theta), \theta)[\delta\theta] = \delta\mathcal{L}(\mathbf{U}(\theta), \boldsymbol{\lambda}, \theta)[\delta\theta] \quad \forall \boldsymbol{\lambda}, \delta\theta. \tag{3.104}$$

With the definition of the Gateaux derivative $\frac{\delta\Phi}{\delta v}$ of some functional $\Phi(v)$ defined by the duality pairing $\langle \frac{\delta\Phi}{\delta v}, \delta v \rangle := \delta\Phi(v)[\delta v]$ [see e.g. 89], the equivalence (3.104) directly implies

$$\frac{\delta\tilde{\mathcal{L}}}{\delta\theta} = \frac{\delta\mathcal{L}}{\delta\theta}. \tag{3.105}$$

In contrast to classical constrained-optimization, the Lagrange parameters $\boldsymbol{\lambda}$ are free parameters in the Lagrangian formulation (3.100). This degree of freedom can be utilized to efficiently compute the derivative $\frac{\delta\mathcal{L}}{\delta\theta}$. To this end, the variation $\delta\mathcal{L}(\mathbf{U}(\theta), \boldsymbol{\lambda}, \theta)[\delta\theta]$ is expanded to

$$\begin{aligned}
\delta\mathcal{L}(\mathbf{U}(\theta), \boldsymbol{\lambda}, \theta)[\delta\theta] = {} & \frac{1}{2\sigma^2} \langle \frac{\delta D}{\delta_\theta \mathbf{U}}, \delta_\theta \mathbf{U} \rangle \\
& + \langle \frac{\delta R}{\delta\theta}, \delta\theta \rangle \\
& + \Delta_{\mathbf{U}} \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta)[\delta_\theta \mathbf{U}] \\
& + \Delta_\theta \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta)[\delta\theta] \quad \forall \boldsymbol{\lambda}, \delta\theta.
\end{aligned} \tag{3.106}$$

Since the variation $\delta_\theta \mathbf{U}$ can only be computed through the nonlinear solution of (2.101), a straight-forward computation of (3.106) is difficult. However, by using the free choice in $\boldsymbol{\lambda}$, the computation of $\delta_\theta \mathbf{U}$ can be avoided by choosing $\boldsymbol{\lambda}$ such that

$$\frac{1}{2\sigma^2}\langle \frac{\delta D}{\delta_\theta \mathbf{U}}, \delta_\theta \mathbf{U}\rangle + \Delta_{\mathbf{U}} \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta)[\delta_\theta \mathbf{U}] = 0 \quad \forall\, \delta_\theta \mathbf{U}. \tag{3.107}$$

Therein, the second term is readily identified with the differential virtual work (2.90). Since the use of the test functions $\boldsymbol{\lambda}$ and the variation $\delta_{\boldsymbol{\theta}} \mathbf{U}$ is exchanged with respect to the differential virtual work (2.90), (3.107) is referred to as *adjoint equation* in the sense of (2.96). The solution is usually obtained under the same discretization as the incremental forward problem. In this context, well-posedness is given by the application of the boundary conditions (2.75) for the dofs $\delta \mathbf{d}$ describing the Lagrange parameters $\boldsymbol{\lambda}$. For the evaluation of the first term in (3.107), it is noted that the dependency of the similarity measure on the displacements is given by $D(\mathbf{Z}, C(\mathbf{U})) = D(\mathbf{Z}, \tilde{F}(\theta))$. Using (3.29), it can be seen that the right hand side of the linear adjoint equation is given by

$$-\frac{1}{2\sigma^2}\langle \frac{\delta D}{\delta_\theta \mathbf{U}}, \delta_\theta \mathbf{U}\rangle = \frac{1}{\sigma^2}\langle \mathbf{Z} - C(\mathbf{U}), \delta C(\mathbf{U})[\delta_\theta \mathbf{U}]\rangle_{\mathcal{Z}}. \tag{3.108}$$

For similarity measures in the sense of point-wise displacement measurements such as (3.37) or (3.38), the observer variation $\delta C(\mathbf{U})[\delta_\theta \mathbf{U}]$ is trivial since the observation operator is just a linear operator and thus

$$\delta C(\mathbf{U})[\delta_\theta \mathbf{U}] = C \delta_\theta \mathbf{U}. \tag{3.109}$$

For similarity measures in terms of surface currents such as (3.58), the observation is defined through the push-forward action (3.50). This leads to a more involved definition which is detailed in appendix C.1.

Upon inserting the solution $\boldsymbol{\lambda}$ of the adjoint equation (3.107) into (3.106), the variation of the Lagrangian $\delta \mathcal{L}$ is obtained as

$$\delta \mathcal{L}(\mathbf{U}(\theta), \boldsymbol{\lambda}, \theta)[\delta \boldsymbol{\theta}] = \langle \frac{\delta R}{\delta \theta}, \delta \theta \rangle + \Delta_\theta \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta)[\delta \theta] \quad \forall\, \delta \theta. \tag{3.110}$$

These remaining terms can usually be evaluated economically due to the explicit dependencies on the parameters $\theta$. Whereas for the regularization this evaluation depends on the specific choice of prior, for the virtual work the formal application of the directional derivative results in

$$\Delta_\theta \delta W(\mathbf{U}, \boldsymbol{\lambda}, \theta)[\delta \theta] = \int_{\Omega_0} \delta_\theta \mathbf{S}[\delta \theta] : \delta \mathbf{E} \,\mathrm{dV}_0 + \int_{\Omega_0} \mathbf{S} : \Delta_\theta \delta \mathbf{E}[\delta \theta] \,\mathrm{dV}_0. \tag{3.111}$$

Due to the explicit parametrization, see chapter 2.4, the variation of the stresses $\delta_\theta \mathbf{S}[\delta \theta]$ and the variation of the strains $\Delta_\theta \delta \mathbf{E}[\delta \theta]$ are given by

$$\delta_\theta \mathbf{S}[\delta \theta] = \frac{\partial \mathbf{S}}{\partial \theta} \delta \theta, \tag{3.112}$$

$$\Delta_\theta \delta \mathbf{E}[\delta \theta] = \frac{\partial \delta \mathbf{E}}{\partial \theta} \delta \theta. \tag{3.113}$$

The functional description of these terms is specific to the physical meaning modeled by the parameters $\theta$. With respect to the application of arterial growth modeled by (2.65), the detailed linearization is provided in appendix C.2.2.

Finally, to arrive at a discrete gradient of the negative log-posterior (3.94), the continuous parameters $\theta$ are represented in terms of the element-wise basis (2.105) such that $\delta\theta = \mathbf{G}^\top \delta\boldsymbol{\theta}$. The differential $\delta\hat{L}[\delta\boldsymbol{\theta}]$ is then given by

$$\begin{aligned}\delta\hat{\mathcal{L}}(\mathbf{U}(\boldsymbol{\theta}), \boldsymbol{\lambda}, \boldsymbol{\theta})[\delta\boldsymbol{\theta}] &= \delta\mathcal{L}(\mathbf{U}(\theta), \boldsymbol{\lambda}, \mathbf{G}^\top\boldsymbol{\theta})[\mathbf{G}^\top\delta\boldsymbol{\theta}] \\ &= \frac{\mathrm{d}R}{\mathrm{d}\theta}\mathbf{G}^\top\delta\boldsymbol{\theta} + \int_{\Omega_0} (\frac{\partial\mathbf{S}}{\partial\theta} : \delta\mathbf{E} + \mathbf{S} : \frac{\partial\delta\mathbf{E}}{\partial\theta})\mathbf{G}^\top \,\mathrm{dV}_0 \cdot \delta\boldsymbol{\theta}.\end{aligned} \tag{3.114}$$

Comparing this to the definition of the Gateaux derivative, the discrete gradient is obtained as

$$\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\delta\mathcal{L}}{\delta\boldsymbol{\theta}} = \frac{\mathrm{d}R}{\mathrm{d}\theta}\mathbf{G}^\top + \int_{\Omega_0} (\frac{\partial\mathbf{S}}{\partial\theta} : \delta\mathbf{E} + \mathbf{S} : \frac{\partial\delta\mathbf{E}}{\partial\theta})\mathbf{G}^\top \,\mathrm{dV}_0. \tag{3.115}$$

Assuming that the regularization was defined by any of the prior-models $p(\boldsymbol{\theta})$ presented in chapter 3.4 such that

$$\frac{\mathrm{d}R}{\mathrm{d}\theta}\mathbf{G}^\top = -\frac{\mathrm{d}\log p(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}}, \tag{3.116}$$

the desired equivalence

$$\frac{\mathrm{d}\mathcal{J}}{\mathrm{d}\boldsymbol{\theta}} = \frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}} \tag{3.117}$$

is established.

In contrast to the approximation (3.96), the gradient (3.115) can be computed by the solution of one additional linear problem (3.107). This process is independent of the dimension of the parameter vector $\boldsymbol{\theta}$. This independence represents a huge advantage compared to the FD approximation. Furthermore, for a certain discretization of the forward problem, the gradient (3.115) can in principle be evaluated exact up to machine precision. Prerequisite is the exact solution of the adjoint equation which is dependent on the solution $\mathbf{U} = \mathbf{A}_h(\boldsymbol{\theta})$ obtained through the nonlinear solution (2.84). In contrast to the nonlinear solution ipso facto, the convergence tolerance of the nonlinear solution process in combination with an adjoint formulation is not governed by the desired accuracy of the primal solution $\mathbf{U}$ but by the accuracy of the gradient $\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\boldsymbol{\theta}}$. Whereas an analytic assessment of this accuracy is difficult, it is revealed by practical application that the accuracy of the nonlinear solution of the forward problem can have a significant influence on the accuracy of the gradient. In general, the convergence tolerance for the nonlinear solution has to be chosen with respect to machine precision to obtain the gradient as accurate as possible. If the adjoint equation is solved by means of iterative methods, the same arguments hold for the convergence of this iterative solution.

# 4. Numerical solution of the identification problem

This chapter is concerned with the introduction to formulations and numerical approaches of *statistical inference* based on the posterior (3.22). Since the posterior is defined in terms of the computational model $F(\boldsymbol{\theta})$, the functional relation $\boldsymbol{\theta} \mapsto p(\boldsymbol{\theta}|\mathbf{Z})$ is not given explicitly. However, a point-wise evaluation for a specific realization $\boldsymbol{\theta} = \boldsymbol{\theta}_1$ is always possible through the evaluation of the computational model $F(\boldsymbol{\theta}_1)$. Given that this evaluation is associated with a considerable computational cost, an exploration of the posterior on the sample space $\mathbb{R}^{n_p}$ is out of question. Instead, it is necessary to draw conclusions on the probabilistic characteristics of the parameters $\boldsymbol{\theta}$ in a more efficient way. Depending on the field of application, this task is approached by a great variety of different formulations and numerical techniques. The presentation in this chapter aims at a concise but consistent introduction that motivates the algorithmic choices made with respect to large-scale patient-specific modeling, see chapter 7. The applied formulations can be introduced following a decision theoretic approach [19].

In this context, sought descriptions of characteristic properties are referred to as *estimators* $\hat{\boldsymbol{\theta}}(\mathbf{Z})$. Through the conditioning of the posterior on the data $\mathbf{Z}$, these estimators are also functions of the data. And due to the randomness in the data, an estimator is itself a random variable. A particular realization of an estimator is called an *estimate* $\hat{\boldsymbol{\theta}}$. The definition of an estimator through single valued 'optimal' estimates leads to the concept of *point estimators.* Thereby, different definitions of optimality result in point estimators with different properties. However, due to their single valued character, these estimators fail to reveal the inherent variability of the parameters. This information is provided by *interval estimators.* The term 'interval' is thereby also used to refer to regions in a multidimensional setting.

In the setting of model calibration, the identified parameters are only a means to exploit a calibrated computational model to generate further information. A description of the parameters in terms of estimators is therefore not the primary interest. Given that the information provided by the calibrated model is represented by a function $f(\boldsymbol{\theta})$, the interest actually lies in inference based on the distribution $p_{f(\boldsymbol{\theta})}(f(\boldsymbol{\theta}))$. Since the point-wise exploration of this distribution suffers from the same constraints as the exploration of the posterior, summarizing statistics such as the expected value $\mathbb{E}[f(\boldsymbol{\theta})]$ or the variance $\mathbb{V}[f(\boldsymbol{\theta})]$ are desired. In the context of this thesis, the function $f$ will eventually represent a prediction in AAA growth that is made based on an identified set of parameters $\boldsymbol{\theta}$. More specific, $f$ could represent quantities like maximal growth of an aneurysm or changes in maximum diameter. Since the maximum diameter is still the prevalent surrogate in clinical decision making, the variability in such a prediction is clearly of high importance to assess the reliability of conclusions drawn from these predictions.

The multitude of information that can be deduced from the identification problem is also reflected in the number of different numerical algorithms necessary to extract the desired information from the available data. Depending on the estimator, the application of algorithms ranges from classical optimization techniques [see e.g. 169] to Monte Carlo based techniques for sampling and integration [see e.g. 185].

The remainder of this chapter is concerned with the introduction of the specific estimators in the chapters 4.1, 4.2 and 4.3. The algorithmic techniques utilized to arrive at specific estimates based on these estimators are described in chapter 4.4. Due to the setup of the identification problem in the Bayesian setting, the presentation is only concerned with Bayesian estimation techniques. For a comprehensive introduction to statistical inference and estimation theory, the reader is referred to classical literature on estimation theory such as Lehmann and Casella [135], Koch [119] or Olive [171].

## 4.1. Bayesian point estimators

The intention of a point estimator is to represent the information provided by the posterior in an 'optimal' way. The particular definition of optimality thereby leads to a variety of different estimators. In Bayesian analysis, a notion of optimality can be derived from decision theoretic arguments [19]. To this end, a *loss-function* $L : \mathbb{R}^{n_p} \times \mathbb{R}^{n_p} \to \mathbb{R}^+$ is introduced. This loss function allows for the qualification of an estimator with respect to the true unknown parameters via $L(\hat{\boldsymbol{\theta}}(\mathbf{Z}), \boldsymbol{\theta})$. Since this functional depends on the unknown parameters $\boldsymbol{\theta}$, it is more reasonable to analyze the *posterior expected loss* defined by

$$\mathbb{E}_{\boldsymbol{\theta}|\mathbf{Z}}[L(\hat{\boldsymbol{\theta}}(\mathbf{Z}), \boldsymbol{\theta})] = \int_{\mathbb{R}^{n_p}} L(\hat{\boldsymbol{\theta}}(\mathbf{Z}), \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{Z}) \mathrm{d}\boldsymbol{\theta}. \tag{4.1}$$

**Definition 4.1.1** (Bayes estimator)**.** *A Bayes estimator is an estimator $\hat{\boldsymbol{\theta}}(\mathbf{Z})$ that minimizes the posterior expected loss defined in (4.1).*

Particular definitions of the loss-function $L$ give rise to specific estimators described in the following, see figure 4.1.

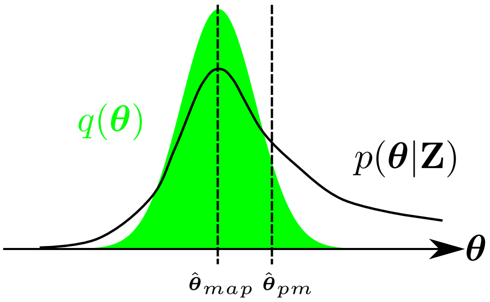

Figure 4.1.: Illustration of different estimators to be identified from the posterior. The maximum a posteriori estimate $\hat{\boldsymbol{\theta}}_{map}$ and the posterior mean estimate $\hat{\boldsymbol{\theta}}_{pm}$ represent very popular instances of *point-estimators*. Another class of estimators centers around inference on $\boldsymbol{\theta}$ based on an approximate but explicitly given classes of probabilities $q(\boldsymbol{\theta})$.

### 4.1.1. Posterior mean

A very popular choice for the loss $L$ is the squared error loss

$$L(\hat{\boldsymbol{\theta}}(\mathbf{Z}), \boldsymbol{\theta}) = \|\hat{\boldsymbol{\theta}}(\mathbf{Z}) - \boldsymbol{\theta}\|^2_{\mathbb{R}^{n_p}}. \tag{4.2}$$

Under this definition, a minimization of (4.1) with respect to $\hat{\boldsymbol{\theta}}$ results in the posterior mean (PM)

$$\hat{\boldsymbol{\theta}}_{pm}(\mathbf{Z}) = \int_{\mathbb{R}^{n_p}} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{Z}) \mathrm{d}\boldsymbol{\theta} = \mathbb{E}_{\boldsymbol{\theta}|\mathbf{Z}}[\boldsymbol{\theta}] \tag{4.3}$$

as the 'optimal' estimator [see e.g. 94]. The popularity of this estimator stems mainly from two facts: on the one hand, estimators derived from the squared error loss (4.2) are *unbiased* estimators in the sense

$$\mathbb{E}_{\mathbf{Z}|\boldsymbol{\theta}}[\hat{\boldsymbol{\theta}}(\mathbf{Z})] = \boldsymbol{\theta} \quad \forall\, \boldsymbol{\theta}; \tag{4.4}$$

on the other hand, a direct measure of the estimator's accuracy in terms of the mean squared error is given by the posterior (co)-variance

$$\mathbb{E}_{\boldsymbol{\theta}|\mathbf{Z}}[(\hat{\boldsymbol{\theta}}(\mathbf{Z}) - \boldsymbol{\theta})^2] = \mathbb{V}_{\boldsymbol{\theta}|\mathbf{Z}}[\boldsymbol{\theta}] \tag{4.5}$$

[see 19, p. 136]. Furthermore, the computation of the posterior variance usually does not require additional computational effort given a method to compute the PM.

### 4.1.2. Maximum a posteriori estimation

Another possible choice for the loss function is the so called *'0-1'* loss given by

$$L(\hat{\boldsymbol{\theta}}(\mathbf{Z}), \boldsymbol{\theta}) = \begin{cases} 0 & \|\hat{\boldsymbol{\theta}}(\mathbf{Z}) - \boldsymbol{\theta}\| \leq \epsilon \\ 1 & \|\hat{\boldsymbol{\theta}}(\mathbf{Z}) - \boldsymbol{\theta}\| > \epsilon. \end{cases} \tag{4.6}$$

In the limit $\epsilon \to 0$, this results in the maximum posterior mode being the 'optimal' estimator given by

$$\hat{\boldsymbol{\theta}}_{map}(\mathbf{Z}) = \operatorname*{argmin}_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{Z}) \tag{4.7}$$

[see e.g. 94]. Since the location of the posterior maximum is independent of the model-evidence $p(\mathbf{Z})$, it is equivalently characterized by

$$\hat{\boldsymbol{\theta}}_{map}(\mathbf{Z}) = \operatorname*{argmin}_{\boldsymbol{\theta}} p(\mathbf{Z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{4.8}$$

A clear advantage of the MAP estimator is the significant simplification in computational effort. Whereas the PM estimator requires the application of numerical integration techniques, the MAP estimator reduces to an optimization problem, which is generally easier to compute [19].

## 4.2. Estimation of credible intervals

If used in a self-contained manner, the point-estimators introduced in chapter 4.1 are only partially Bayesian. Although exploiting the full posterior information, they fail at providing insight into the variability of the conditional distribution of the parameters $\boldsymbol{\theta}|\mathbf{Z}$. Within a Bayesian setting, point estimates are usually augmented with interval estimates that are referred to as credible interval (CI). In principle, the definition of a *credible region/set* as an extension of a CI to multiple dimensions is straightforward. In practice however, CIs are mainly referred to as sections from the marginal probability densities. This is mainly due to limited possibilities of presenting multidimensional data on the one hand, but primarily due to complications in the numerical computation for high dimensional problems. Splitting the random parameters $\boldsymbol{\theta}$ into the set $(\theta_i, \boldsymbol{\theta}_{\setminus i})$, i.e., one scalar parameter of interest $\theta_i$ and the vectorial remainder $\boldsymbol{\theta}_{\setminus i}$ of the vector $\boldsymbol{\theta}$, the marginal density $p(\theta_i)$ is given by

$$p(\theta_i) = \int_{\mathbb{R}^{n_p - 1}} p(\boldsymbol{\theta}|\mathbf{Z}) \mathrm{d}\boldsymbol{\theta}_{\setminus i} = \int_{\mathbb{R}^{n_p - 1}} p(\theta_i, \boldsymbol{\theta}_{\setminus i}|\mathbf{Z}) \mathrm{d}\boldsymbol{\theta}_{\setminus i}. \tag{4.9}$$

A credible interval for the parameter $\theta_i$ can then be defined [see e.g. 94]:

**Definition 4.2.1** (Credible interval). *Given a marginal density $p(\theta_i)$ as defined in (4.9), a $\alpha \cdot 100\%$-credible interval is given by the set $(\theta_l, \theta_u)$ such that*

$$\int_{\theta_l}^{\theta_u} p(\theta_i) \mathrm{d}\theta_i = \alpha \tag{4.10}$$

*whereby $\alpha \in [0, 1]$ is a fixed number.*

This definition enables to fix the probability of the parameter $\theta_i|\mathbf{Z}$ to be contained in the interval $[\theta_l, \theta_u]$ with a probability $\alpha$. But without further specification, there are infinitely many such intervals in the domain of the marginal probability density $p(\theta_i)$. One possibility is to constrain the interval to be *equi-tailed*. Such a constraint necessitates the computation of the $(1-\alpha)/2$- and $(1+\alpha)/2$-quantiles. Another possibility to arrive at unique intervals under certain regularity conditions is the use of *highest posterior density (HPD) intervals.*

**Definition 4.2.2** (Highest posterior density interval). *For a certain level $\alpha$, an HPD interval of the marginal distribution $p(\theta_i)$ is the credible interval $I$ with*

$$p(\theta_i) \geq p(\tilde{\theta}_i) \quad \forall \quad \theta_i \in I, \tilde{\theta}_i \notin I. \tag{4.11}$$

The existence of a HPD depends on the properties of the posterior density. For instance, in the case of multimodal densities, the HPD interval does not necessarily exist.

The analytic evaluation of CIs or HPD intervals is generally only possible in certain simple cases due to the involved integrations. In practice, they are estimated by means of Monte Carlo (MC) techniques, see chapter 4.4.2. In the work presented in this thesis, the posterior variability is always reported in terms of the posterior variance and the reader is referred to Chen et al. [40, chapter 7] for a detailed description of the approximation of Bayesian intervals by means of MC methods.

## 4.3. Approximate inference

Anticipating one of the shortcomings of the numerical techniques presented in chapter 4.4, it has to be mentioned that some techniques can pose almost insuperable computational barriers. In particular, these are the techniques based on numerical integration and sampling. To overcome these shortcomings, huge effort is made most notably in the field of *machine learning* [see e.g 25] to reduce the computational cost by developing methods for approximate inference.

*Variational Bayes (VB) techniques* (see chapter 4.3.2) constitute one such class of methods. By computing optimal approximations $q(\boldsymbol{\theta})$ of the posterior $p(\boldsymbol{\theta}|\mathbf{Z})$, VB methods try to render the evaluation of integrals such as in (4.3) or in the evaluation of the evidence in (3.32) feasible. With the intention of identifying parameters for complex computational models, these methods have also been extended to be applicable in combination with models incorporating nonlinear 'input-output'-relations (see chapter 4.3.3). These extensions can also be seen as generalizations to the well-known *Laplace approximation* described in the following.

### 4.3.1. Laplace approximation

In the context of statistical inference, the *Laplace approximation* usually refers to approximations of the posterior by a normal distribution [209]. This is equivalent to a quadratic approximation to the log-posterior at the MAP point [145]. To this end, the posterior is represented as

$$p(\boldsymbol{\theta}|\mathbf{Z}) = \frac{1}{Z} q(\boldsymbol{\theta}), \tag{4.12}$$

and $\log q(\boldsymbol{\theta})$ is approximated in terms of the second-order Taylor expansion at the MAP point

$$\begin{aligned} \log q(\boldsymbol{\theta}) \approx \log q(\hat{\boldsymbol{\theta}}_{map}) &+ \cancelto{0}{\left.\frac{\mathrm{d}\log q}{\mathrm{d}\boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}_{map}}} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{map}) \\ &+ \frac{1}{2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{map}) \cdot \left.\frac{\mathrm{d}^2 \log q}{\mathrm{d}\boldsymbol{\theta}^2}\right|_{\hat{\boldsymbol{\theta}}_{map}} \cdot (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{map}) \end{aligned} \tag{4.13}$$

[see e.g. 25, 145]. Due to the definition of the MAP point as an extremal point of the posterior, the first-order term in (4.13) vanishes. The matrix

$$\mathbf{H} := -\frac{\mathrm{d}^2 \log p(\boldsymbol{\theta}|\mathbf{Z})}{\mathrm{d}\boldsymbol{\theta}^2} \tag{4.14}$$

is usually called *hessian matrix.* Provided that $\mathbf{H}$ is positive definite, $q(\boldsymbol{\theta})$ can be represented by the Gaussian distribution

$$q(\boldsymbol{\theta}) = \mathcal{N}(\hat{\boldsymbol{\theta}}_{map}, \mathbf{H}^{-1}). \tag{4.15}$$

This distribution is also shown in figure 4.1. The normalizing constant

$$Z = \int p(\mathbf{Z}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} = p(\mathbf{Z}) \tag{4.16}$$

can then be approximately computed by

$$Z \approx q(\hat{\boldsymbol{\theta}}_{map}) \frac{(2\pi)^{n_p/2}}{|\mathbf{H}|^{1/2}}. \tag{4.17}$$

This approximation of the evidence $p(\mathbf{Z})$ makes Laplace's method a popular tool for model evaluation, e.g., by means of the *Bayes factor* [111]. The use of the Gaussian-approximation at the MAP estimate can be justified from an asymptotic perspective where a large amount of data is available. Nevertheless, there is no guarantee on the quality of the approximation and the Gaussian distribution might not represent the posterior well in a particular situation [25].

### 4.3.2. Variational Bayesian approach

VB methods also aim at statistical inference based on an approximation $q(\boldsymbol{\theta})$ of the posterior $p(\boldsymbol{\theta}|\mathbf{Z})$. In contrast to the Laplace approximation – where the approximation $q(\boldsymbol{\theta})$ is *a posteriori* fitted in terms of a fixed Gaussian distribution – optimality of the approximations in the VB approach is obtained from a variational argument *a priori*. To this end, similarity between probability distributions $p_1$ and $p_2$ is measured by means of the *Kullback-Leibler divergence* [129]:

$$D_{KL}(p_1 \| p_2) = \int p_1(x) \log \frac{p_1(x)}{p_2(x)} \mathrm{d}x. \tag{4.18}$$

The evaluation of a minimum of such a functional with respect to an argument $p_1$ or $p_2$ lends itself towards variational calculus motivating the name variational Bayes. Although not being a metric since $D_{KL}(p_1 \| p_2) \neq D_{KL}(p_2 \| p_1)$, the application of (4.18) reveals very useful properties. This can be seen by a rearrangement of the Kullback-Leibler divergence of some approximation $q(\boldsymbol{\theta})$ from the posterior $p(\boldsymbol{\theta}|\mathbf{Z})$:

$$\begin{aligned} D_{KL}(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta}|\mathbf{Z})) &= \int q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{p(\boldsymbol{\theta}|\mathbf{Z})} \mathrm{d}\boldsymbol{\theta} && (4.19) \\ &= \int q(\boldsymbol{\theta}) \log \frac{q(\boldsymbol{\theta})}{p(\mathbf{Z}, \boldsymbol{\theta})} \mathrm{d}\boldsymbol{\theta} + \log p(\mathbf{Z}) && (4.20) \\ &=: -\mathcal{G}(q) + \log p(\mathbf{Z}). && (4.21) \\ \rightarrow \log p(\mathbf{Z}) &= \mathcal{G}(q) + D_{KL}(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta}|\mathbf{Z})). && (4.22) \end{aligned}$$

Since the model-evidence $p(\mathbf{Z})$ is fixed with respect to $q$, the minimization of (4.19) is equivalent to a maximization of the functional $\mathcal{G}(q)$. Thus, $\mathcal{G}(q)$ provides a *lower bound* for $p(\mathbf{Z})$ [see e.g. 25].

In order to arrive at tractable relations by performing the minimization of (4.19), the VB method usually assumes that the posterior can be factorized over partitions $\boldsymbol{\theta}_i$ of the parameters $\boldsymbol{\theta}$ according to

$$q(\boldsymbol{\theta}) = \prod_i^N q_i(\boldsymbol{\theta}_i). \tag{4.23}$$

This approximation, known as *mean field approximation*, has its foundation in the *mean field theory* of statistical physics [173]. The minimization of (4.19) is then performed with respect to the single components $q_i$ according to the classical variational argument

$$\operatorname*{argmin}_{q_i} D_{KL}(q(\boldsymbol{\theta})\|p(\boldsymbol{\theta}|\mathbf{Z})) \iff \delta D_{KL}(q\|p)[\delta q_i] \equiv 0. \tag{4.24}$$

Thereby, no *a priori* assumptions on the single components $q_i$ are made, which is an often advocated advantage of the VB approach referred to as *free-form optimization* [9]. The solution of the minimization problem (4.24) results in

$$\log q_i(\boldsymbol{\theta}_i) = \mathbb{E}_{q_{j\neq i}}[\log p(\mathbf{Z}, \boldsymbol{\theta})] + const \tag{4.25}$$

[see e.g. 25]. Thereby, $\mathbb{E}_{q_{j\neq i}}$ refers to the expectation computed with respect to the probability density $q_{j\neq i} = \prod_{j,j\neq i} q_j(\boldsymbol{\theta}_j)$. With a clever choice of the partitions $\boldsymbol{\theta}_i$ and/or with the choice of *conjugate priors* for the likelihood $p(\mathbf{Z}|\boldsymbol{\theta})$, the integral $\mathbb{E}_{q_{j\neq i}}[p(\mathbf{Z}, \boldsymbol{\theta})]$ becomes analytically tractable and the components $q_i(\boldsymbol{\theta}_i)$ can often be identified with some specific distribution [180].

Starting from the mean field approximation, the original identification problem is split into $N$ conceptually simple optimization problems (4.25). The cyclic dependency between the sub-problem suggests an iterative update procedure: starting from an initial guess $\boldsymbol{\theta}_0$, the partitions are updated sequentially, and new parameters $\boldsymbol{\theta}_i$ are used in the optimization for the next partition. This procedure is then repeated until convergence. This approach is very similar to the *expectation-maximization* scheme known from optimization of latent variable models [see e.g. 25]. In fact, it can be shown that the VB approach represent a generalization of expectation-maximization [17].

The VB approach is a relatively recent development in the field of statistical inference and it has been proven particularly useful for parameter identification. Often, partitioning the parameters into model- and noise-parameters already results in feasible computations. Also for Bayesian graphical networks or hierarchical Bayesian models, the hierarchy provides a natural way of partitioning [106]. Furthermore, the good convergence properties being inherited from the expectation-maximization scheme make the use of VB methods appealing. The general advantage of the VB approach over the conceptually easier Laplace approximation is twofold. On the one hand, VB provides a much richer class of approximations, whereas the Laplace approximation is restricted to optimal solutions being represented as Gaussian distributions. On the other hand, VB gives a lower bound for the evidence, whereas the approximation quality of the Laplace approximation can generally not be assessed. For a detailed introduction and illustrations of the approach, the reader is referred to Bishop [25], MacKay [145] and Beal [17].

### 4.3.3. Extension to nonlinear forward models

The advantage of the VB approach is mainly connected to the availability of good partitions $\boldsymbol{\theta}_i$ such that the optimal distribution $q_i$ can be identified from (4.25). Under certain circumstances, this identification might not be possible. A particular case is given by likelihoods given in terms of complex nonlinear forward models. In order to also handle nonlinear models within the VB framework, it can be combined with the Laplace approximation [67]. One possible approach suggested by Chappell et al. [39] is to use the computational model in a linearized version

$$F(\boldsymbol{\theta}) \approx F(\boldsymbol{\theta}_m) + \mathbf{J}(\boldsymbol{\theta} - \boldsymbol{\theta}_m), \tag{4.26}$$

with the matrix $[\mathbf{J}]_{ij} = \frac{\mathrm{d}F_i(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}_j}\Big|_{\boldsymbol{\theta}_m}$. Considering solely the model-parameters $\boldsymbol{\theta}$ as single partition, the minimization problem (4.25) reduces to

$$\log q(\boldsymbol{\theta}) = \log p(\boldsymbol{\theta}|\mathbf{Z}) + \log p(\boldsymbol{\theta}) + const. \tag{4.27}$$

Using the generic definition of the likelihood (3.27) and the assumption of a Gaussian prior $p(\boldsymbol{\theta}) \sim \mathcal{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$ over the parameters, the log-posterior can be expanded to

$$\log q(\boldsymbol{\theta}) = -\frac{1}{2\sigma^2}(\mathbf{Z} - F(\boldsymbol{\theta})) \cdot (\mathbf{Z} - F(\boldsymbol{\theta})) - \frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0) \cdot \boldsymbol{\Sigma}_0 \cdot (\boldsymbol{\theta} - \boldsymbol{\theta}_0) + const. \tag{4.28}$$

Inserting the linearization (4.26) and accumulating all terms that are constant with respect to $\boldsymbol{\theta}$ in the constant remainder, one obtains

$$\begin{aligned} \log q^*(\boldsymbol{\theta}) = & -\frac{1}{2}\boldsymbol{\theta} \cdot (\frac{1}{\sigma^2}\mathbf{J}^\top\mathbf{J} + \boldsymbol{\Sigma}_0) \cdot \boldsymbol{\theta} \\ & +\boldsymbol{\theta} \cdot (\frac{1}{\sigma^2}\mathbf{J}^\top \cdot ((\mathbf{Z} - F(\boldsymbol{\theta}_m)) + \mathbf{J} \cdot \boldsymbol{\theta}_m) + \boldsymbol{\Sigma}_0 \cdot \boldsymbol{\theta}_0) + const. \end{aligned} \tag{4.29}$$

In order to identify the ride hand side as the logarithm of some specific probability density function, the logarithm of a Gaussian density $\log \mathcal{N}(\mathbf{x}|\mathbf{m}, \boldsymbol{\Sigma})$ is expanded to

$$\log \mathcal{N}(\mathbf{x}|\mathbf{m}, \boldsymbol{\Sigma}) = -\frac{1}{2}\mathbf{x} \cdot \boldsymbol{\Sigma} \cdot \mathbf{x} + \mathbf{x} \cdot \boldsymbol{\Sigma} \cdot \mathbf{m} + const. \tag{4.30}$$

By comparison to (4.29) it is possible to identify

$$\boldsymbol{\Sigma} = \frac{1}{\sigma^2}\mathbf{J}^\top\mathbf{J} + \boldsymbol{\Sigma}_0, \tag{4.31}$$

$$\boldsymbol{\Sigma}\mathbf{m} = \frac{1}{\sigma^2}\mathbf{J}^\top \cdot ((\mathbf{Z} - f(\boldsymbol{\theta}_m)) + \mathbf{J} \cdot \boldsymbol{\theta}_m) + \boldsymbol{\Sigma}_0 \cdot \boldsymbol{\theta}_0. \tag{4.32}$$

This again suggests an iterative procedure: starting from an initial guess of $\boldsymbol{\theta}_m^0$, $\mathbf{m}$ is iteratively obtained as the solution of the linear system (4.32), and the next linearization point is updated as $\boldsymbol{\theta}_m^{n+1} \leftarrow \mathbf{m}$. This iteration can be readily identified as a *Gauss-Newton* algorithm for the solution of a regularized least-squares formulation [see e.g. 146]. Except for the Gauss-Newton approximation, this iteration will converge to the Laplace-approximation (4.15) with the covariance $\mathbf{H}$ approximated by

$$\mathbf{H} \approx \frac{1}{\sigma^2}\mathbf{J}^\top\mathbf{J} + \boldsymbol{\Sigma}_0. \tag{4.33}$$

This specific interpretation of the nonlinear-VB approach arises as the limiting case of a fixed a priori known noise variance $\sigma^2$ and a Gaussian prior on $\boldsymbol{\theta}$. Whereas this limiting case is of course not representative for the nonlinear VB approach, it reveals its two main disadvantages. On the one hand, the lower bound on the evidence $p(\mathbf{Z})$ is weakened due to the linear model approximation. On the other hand, the nonlinear VB is no longer an expectation maximization scheme and thus suffers from the same disadvantages as the Gauss-Newton algorithm. I.e., it is stable only close enough to the solution and generally requires extensions such as the *Levenberg-Marquard* algorithm [see e.g. 151].

Furthermore, the computation of the linearization **J** used in (4.26) poses a significant computational effort. Especially for high dimensional complex nonlinear models $F(\boldsymbol{\theta})$ in combination with high-dimensional data **Z**, the linearization **J** is exactly that quantity whose computation is avoided by state-of-the art Quasi-Newton methods, see chapter 4.4.1. The resulting Gauss-Newton approximation as a replacement for the exact quadratic approximation of the log-likelihood is also know to be inaccurate for large residuals $(\mathbf{Z} - F(\boldsymbol{\theta}))$. This fact can be directly observed from the exact second derivative

$$\frac{\mathrm{d}^2 \log p(\mathbf{Z}|\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}^2} = \frac{1}{\sigma^2}(\mathbf{J}^\top \mathbf{J} - (\mathbf{Z} - f(\boldsymbol{\theta}))\frac{\mathrm{d}^2 f}{\mathrm{d}\boldsymbol{\theta}^2}). \tag{4.34}$$

As a remedy, the exact hessian to the log-likelihood **H**, which can be obtained by adjoint approaches either exactly or approximately, can be used instead [66, 67, 175].

## 4.4. Numerical computation of estimates

This chapter introduces the numerical techniques and algorithmic implementations used in this thesis to compute estimates from the estimators introduced in the preceding chapter. Depending on the estimator, conceptually different techniques are necessary. Whereas the computation of a MAP estimate requires methods for optimization, a PM estimate requires methods for integration and sampling. Both approaches are affected by the potentially high dimension of the parameters $\boldsymbol{\theta}$ and the high computation cost in terms of evaluation time associated to the solution of the computational model $F(\boldsymbol{\theta})$.

The following presentation is not intended as a comprehensive introduction to numerical optimization and integration in general but as the description of the state-of-the-art techniques that are used in the large-scale regime and applied in this thesis. To compute the MAP estimate, the (limited-memory) Broyden-Fletcher-Goldfarb-Shanno method is introduced in chapter 4.4.1 as state-of-the-art optimization algorithm for large-scale nonlinear problems. The computation of the PM estimate will be performed by means of Monte Carlo techniques presented in the chapters 4.4.2 - 4.4.4.

### 4.4.1. Limited-memory BFGS

The Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [31, 62, 86, 200] is a popular method for the solution of nonlinear equations as arising from optimization problems [153]. One instance of such a problem is given by the necessary conditions to be fulfilled by the MAP estimate $\hat{\boldsymbol{\theta}}_{map}$ from (4.8):

$$\frac{\mathrm{d}(-\log p(\mathbf{Z}, \boldsymbol{\theta}))}{\mathrm{d}\boldsymbol{\theta}} = 0. \tag{4.35}$$

In order to keep the notation simple and in line with available literature, this system of equations in $\boldsymbol{\theta}$ will be represented by the generic system

$$\nabla h(\boldsymbol{x}) = 0 \tag{4.36}$$

whereby the function $h : \mathbb{R}^{n_p} \to \mathbb{R}$ is used to represent $-\log p(\mathbf{Z}, \boldsymbol{\theta})$ as a function of the parameters $\boldsymbol{\theta}$. The BFGS method is a particular member of a class of methods named *Quasi-Newton* methods. A Quasi-Newton method is characterized by the iterative update rule

$$\boldsymbol{x}_{n+1} = \boldsymbol{x}_n - \tau_n (\boldsymbol{B}_n)^{-1} \cdot \nabla h(\boldsymbol{x}_n). \tag{4.37}$$

The matrix $\boldsymbol{B}_n$ is an approximation to $\nabla^2 h(\boldsymbol{x}_n)$ and $\tau_n$ is a line search parameter. A Quasi-Newton scheme creates a sequence $\{\boldsymbol{B}_n\}$ of SPD matrices following a general line search paradigm:

$$(1) \quad \boldsymbol{d} \leftarrow -(\boldsymbol{B}_n)^{-1} \cdot \nabla h(\boldsymbol{x}_n) \tag{4.38}$$

$$(2) \quad \boldsymbol{x}_{n+1} \leftarrow \boldsymbol{x}_n + \tau_n \boldsymbol{d} \tag{4.39}$$

$$(3) \quad \boldsymbol{B}_{n+1} \leftarrow \boldsymbol{x}_{n+1}, \boldsymbol{B}_n \tag{4.40}$$

Thereby, the line search parameter $\tau$ is subject to certain conditions such that the update on $\boldsymbol{B}_n$ in step (3) generates a contraction for (4.37) [115]. These conditions depend on the update strategy for $\boldsymbol{B}_n$. In the class of *secant methods*, this update strategy is generated from the *secant equation*

$$\boldsymbol{B}_{n+1} \cdot \boldsymbol{s}_n = \boldsymbol{y}_n \tag{4.41}$$

with $\boldsymbol{s}_n = \boldsymbol{x}_{n+1} - \boldsymbol{x}_n$ and $\boldsymbol{y}_n = \nabla h(\boldsymbol{x}_{n+1}) - \nabla h(\boldsymbol{x}_n)$. To determine a unique $\boldsymbol{B}_{n+1}$ from (4.41), it is necessary to impose additional conditions. These can be imposed in terms of a similarity condition between the iterates $\boldsymbol{B}_n$ and $\boldsymbol{B}_{n+1}$. To this end, the solution to (4.41) is formulated as the optimization problem

$$\min \|\boldsymbol{B}_{n+1} - \boldsymbol{B}_n\|_W \quad s.t. \quad \boldsymbol{B}_{n+1}^\top = \boldsymbol{B}_{n+1},\, \boldsymbol{B}_{n+1} \cdot \boldsymbol{s}_n = \boldsymbol{y}_n. \tag{4.42}$$

The choice of the norm $\|\cdot\|_W$ then leads to different Quasi-Newton methods. In the BFGS method, this norm is chosen as a weighted Frobenius norm $\|\boldsymbol{A}\|_W := \left\|W^{1/2} \boldsymbol{A} W^{1/2}\right\|_F$ and instead of imposing conditions on the approximate hessian $\boldsymbol{B}_n$, the secant equation is reformulated in term of the inverse hessian $\boldsymbol{B}_n^{-1}$ resulting in the minimization problem

$$\min \left\|\boldsymbol{B}_{n+1}^{-1} - \boldsymbol{B}_n^{-1}\right\|_W \quad s.t. \quad \boldsymbol{B}_{n+1}^{-\top} = \boldsymbol{B}_{n+1}^{-1},\, \boldsymbol{B}_{n+1}^{-1} \cdot \boldsymbol{y}_n = \boldsymbol{s}_n. \tag{4.43}$$

Under the condition $\boldsymbol{y}_n \cdot \boldsymbol{s}_n > 0$, this leads to the *rank-two* update formula for the inverse hessian

$$\boldsymbol{B}_{n+1}^{-1} = (\mathbf{I} + \frac{\boldsymbol{s}_n \boldsymbol{y}_n}{\boldsymbol{y}_n \cdot \boldsymbol{s}_n}) \boldsymbol{B}_n^{-1} (\mathbf{I} + \frac{\boldsymbol{y}_n \boldsymbol{s}_n}{\boldsymbol{y}_n \cdot \boldsymbol{s}_n}) + \frac{\boldsymbol{s}_n \boldsymbol{s}_n}{\boldsymbol{s}_n \cdot \boldsymbol{s}_n} \tag{4.44}$$

[see e.g. 115]. To guarantee that $\boldsymbol{B}_{n+1}^{-1}$ is SPD, a line search strategy is necessary to ensure $\boldsymbol{y}_n \cdot \boldsymbol{s}_n > 0$. This can be achieved by a variety of different approaches. A line search algorithm in the context of Quasi-Newton methods generally tries to solve the sub-problem

$$\min_\tau \chi(\tau) := \min_\tau h(\boldsymbol{x}_n + \tau \boldsymbol{d}) \tag{4.45}$$

such that updates like (4.44) are valid. Throughout the work presented in this thesis, a relatively simple approach based on polynomial models $\chi_\kappa$ (with order $\kappa$) of $\chi(\tau)$ has proven extremely robust [115]. This algorithm is based on the *sufficient decrease* [7] condition

$$h(\boldsymbol{x}_n + \tau \boldsymbol{d}) - h(\boldsymbol{x}_n) < c_1 \tau \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d} \tag{4.46}$$

with the small constant $c_1$ typically chosen to $c_1 = 1.0\text{e}{-4}$ [see e.g. 169]. In order to solve the minimization problem (4.45), a sequence of cubic polynomial models is constructed. Based on the information $\chi(0) = h(\boldsymbol{x}_n)$, $\chi'(0) = \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d}$ and $\chi(\tau_0) = h(\boldsymbol{x}_n + \tau_0 \boldsymbol{d})$, this sequence is initialized by the quadratic model

$$\chi_2(\tau) = \tau h(\boldsymbol{x}_n) + \tau \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d} + \frac{h(\boldsymbol{x}_n + \tau_0 \boldsymbol{d}) - h(\boldsymbol{x}_n) - \tau_0 \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d}}{\tau_0^2} \tau^2. \tag{4.47}$$

The minimizer of this quadratic model is given by

$$\tau_+ = \frac{\tau_0^2 \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d}}{h(\boldsymbol{x}_n + \tau_0 \boldsymbol{d}) - h(\boldsymbol{x}_n) - \tau_0 \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d}}. \tag{4.48}$$

It is subsequently used for the next iteration $\tau_i \leftarrow \tau_+$ to provide the additional information $\chi(\tau_i) = h(\boldsymbol{x}_n + \tau_i \boldsymbol{d})$. Based on this additional information, the cubic model

$$\chi_3(\tau) = \tau h(\boldsymbol{x}_n) + \tau \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d} + \tau^2 c_2 + \tau^3 c_3 \tag{4.49}$$

is constructed. The constants $c_2$ and $c_3$ can be identified from the solution of the linear system

$$\begin{bmatrix} \tau_i^2 & \tau_i^3 \\ \tau_{i-1}^2 & \tau_{i-1}^3 \end{bmatrix} \begin{bmatrix} c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} h(\boldsymbol{x}_n + \tau_i \boldsymbol{d}) - h(\boldsymbol{x}_n) - \tau_i \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d} \\ h(\boldsymbol{x}_n + \tau_{i-1} \boldsymbol{d}) - h(\boldsymbol{x}_n) - \tau_{i-1} \nabla h(\boldsymbol{x}_n) \cdot \boldsymbol{d} \end{bmatrix}. \tag{4.50}$$

The minimizer $\tau_+$ is then readily obtained as

$$\tau_+ = \frac{-c_2 + \sqrt{c_2^2 - 3 c_3 \chi'(0)}}{3 c_3} \tag{4.51}$$

and can be used as a next iterate. The outline of the resulting algorithm is depicted in algorithm 1. Such an algorithm benefits from local *super-linear* convergence [115]. Reinitializing the algorithm in case $\boldsymbol{y}_n \cdot \boldsymbol{s}_n < 0$ may lead to repeated *steepest gradient descent* steps in non strictly convex regions. Theoretically this can be considered inefficient and there are more elaborate solutions available [see e.g. 169]. In practice though, algorithm 1 has proven to be very robust due to its simplicity, which is a clear advantage over more efficient line search strategies that come at the cost of increased algorithmic complexity.

The use of Quasi-Newton methods in nonlinear optimization is very appealing since they don't need second order derivatives. Instead, hessian approximations are constructed solely based on gradient information. This represents a particular advantage for large-scale optimization. In comparison to other methods used in this field, like the *nonlinear conjugate gradient* method [63], the benefit of Quasi-Newton methods is mainly given by the super-linear convergence properties. A drawback of the BFGS update (4.44) is the high storage cost of the generally non-sparse inverse hessian $\boldsymbol{B}^{-1}$. Although it is possible to use its symmetry, the storage of approximately $n_p^2/2$ entries of $\boldsymbol{B}^{-1}$ can become infeasible for large parametric dimension $n_p$. The *limited-memory BFGS (LBFGS)* method addresses this issue efficiently by the incorporation of two important features:

- the approximation $\boldsymbol{B}_n^{-1}$ is not constructed from all increments $\boldsymbol{y}_n$ and $\boldsymbol{s}_n$ but only from the $m_s$ most recent.

```
$n = 0$; $\boldsymbol{B}_n \leftarrow \mathbf{I}$
$\boldsymbol{d} \leftarrow -\boldsymbol{B}_n^{-1}\nabla h(\mathbf{x}_n)$
while $|\nabla h(\boldsymbol{x})| > tol$ do
    $i \leftarrow 0$; $\tau_i \leftarrow 1.0$
    while not sufficient decrease (4.46) do
        if $i == 0$ then
            $\tau_+ \leftarrow \min_\tau \chi_2(\tau)$
        else
            $\tau_+ \leftarrow \min_\tau \chi_3(\tau)$
        end if
        $\tau_{i+1} \leftarrow \min(\max(\beta_l\tau_i, \tau_+), \beta_h\tau_i)$
        $i \leftarrow i + 1$
    end while
    $\boldsymbol{x}_{n+1} \leftarrow \boldsymbol{x}_n + \tau_i\boldsymbol{d}$
    if $\boldsymbol{y}_n \cdot \boldsymbol{s}_n > 0$ then
        Update $\boldsymbol{B}_{n+1}^{-1}$ according to (4.44)
        $\boldsymbol{d} \leftarrow -\boldsymbol{B}_{n+1}^{-1}\nabla h(\mathbf{x}_{n+1})$
    else
        $\boldsymbol{B}_{n+1} \leftarrow \mathbf{I}$
    end if
    $n \leftarrow n + 1$
end while
```

Algorithm 1: Line search BFGS algorithm. The factors $\beta_l$ and $\beta_h$ with $0 < \beta_l < \beta_h < 1$, e.g., $\beta_l = 0.1, \beta_h = 0.5$, are *safeguarding* against step sizes becoming too low or too high due to inefficient proposals by the polynomial models.

- the approximation $\boldsymbol{B}_n^{-1}$ is never explicitly constructed, but its action $\boldsymbol{B}_n^{-1} \cdot \mathbf{v}$ on some vector $\mathbf{v}$ is directly computed from the increments $\boldsymbol{y}_n$ and $\boldsymbol{s}_n$.

These features can be implemented in the so called *two-loop recursion* formula [168], see algorithm 2. In contrast to the standard BFGS update (4.44), the two-loop recursion allows the initial scaling $\boldsymbol{B}_n^0$ to vary between iterations. A method that has proven to be efficient is given by the choice $\boldsymbol{B}_n^0 = \gamma_n\mathbf{I}$ with

$$\gamma_n = \frac{\boldsymbol{y}_n \cdot \boldsymbol{s}_n}{\boldsymbol{y}_n \cdot \boldsymbol{y}_n}. \tag{4.52}$$

This choice tries to estimate the scaling of the true hessian along the latest search direction such that the step length will be $\tau_{n+1} \approx 1$ [140]. The amount of storage for the two-loop recursion is given by the $m_s$ most recent increments $\boldsymbol{y}_n$ and $\boldsymbol{s}_n$. Although loosing super-linear properties, the convergence of the LBFGS method is reportedly superior to that of nonlinear CG already for small $m_s$ ($m_s = 3, 5, 7, 10$). This makes LBFGS the method of choice for large-scale nonlinear problems [169].

```
$\boldsymbol{d} \leftarrow \nabla h(\boldsymbol{x}_k)$
for $i = k, \dots, k - m_s$ do
    $a_i \leftarrow (\boldsymbol{s}_i \cdot \boldsymbol{d})/(\boldsymbol{y}_i \cdot \boldsymbol{s}_i)$
    $\boldsymbol{d} \leftarrow \boldsymbol{d} - a_i \boldsymbol{y}_i$
end for
$\boldsymbol{d} \leftarrow \boldsymbol{B}_n^0 \cdot \boldsymbol{d}$
for $i = k - m_s, \dots, k$ do
    $b \leftarrow (\boldsymbol{y}_i \cdot \boldsymbol{d})/(\boldsymbol{y}_i \cdot \boldsymbol{s}_i)$
    $\boldsymbol{d} \leftarrow \boldsymbol{d} + \boldsymbol{s}_i(a_i - b)$
end for
$\textit{Result} : \boldsymbol{d} \equiv \boldsymbol{B}_k^{-1} \cdot \nabla h(\boldsymbol{x}_k)$
```

Algorithm 2: Two-loop recursion algorithm to compute a direction $\boldsymbol{d}$ at an iteration $k$ of algorithm 1.

### 4.4.2. Monte Carlo methods

*MC* methods cover a wide spectrum of different applications. The most prominent ones are optimization and integration. Often, the MC method [156] is referred to as a method for the evaluation of general intervals given by

$$\mathbb{E}_x[f(\mathsf{x})] = \int_{\mathbb{R}^n} f(x)p(x)\mathrm{d}x. \tag{4.53}$$

Depending on the definition of the function $f(x)$, this integral can represent the PM (4.3) or the posterior variance (4.5). In this case $f(x)$ would be given by $f(x) = x$ or $f(x) = (x - \mathbb{E}[\mathsf{x}])^2$. Moreover, $f(x)$ can represent any output quantity of the computational model such that (4.53) can be used to compute statistics over the model output as outlined in the introduction to chapter 4.

Given a sample $(\mathsf{x}_1, \dots, \mathsf{x}_N) \sim p$, the integral (4.53) is approximated by the empirical average

$$\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(\mathsf{x}_i = x_i). \tag{4.54}$$

Since for large $N$, the error $\bar{f} - \mathbb{E}[f(\mathsf{x})]$ is $\mathcal{O}(1/\sqrt{N})$, convergence of this approximation can be observed by the variance estimate

$$\mathbb{V}[\bar{f}] = \mathbb{V}_x[f(\mathsf{x})] \tag{4.55}$$

[185], which is a consequence of the *central limit theorem* [see e.g. 59]. The general independence of the convergence of MC methods on the dimension of $x$ constitutes the biggest advantage over classical integration techniques such as *Simpson's rule* or *numerical quadrature* [see e.g. 178]. However, the applicability of the plain MC method (4.54) is restricted to the availability of a sample $(\mathsf{x}_1, \dots, \mathsf{x}_N) \sim p$. In the situation presented in this thesis, the generation of samples from the posterior $p(\boldsymbol{\theta}|\mathbf{Z})$ is not possible by standard random variate generation techniques [see e.g. 51]. To account for this, (4.53)

is more generally given by

$$\mathbb{E}_x[f(\mathsf{x})] = \int_{\mathbb{R}^{n_p}} f(x)\frac{p(x)}{\eta(x)}\eta(x)\mathrm{d}x. \tag{4.56}$$

As a consequence, choosing an *importance distribution* $\eta(x)$ that allows for the generation of a sample $(\mathsf{x}_1^\eta, \dots, \mathsf{x}_N^\eta) \sim \eta$, the MC estimate $\bar{f}$ of $\mathbb{E}[f(\mathsf{x})]$ can be computed by the weighted average

$$\bar{f} = \frac{1}{N}\sum_{i=1}^{N} w_i f(\mathsf{x}_i^\eta = x_i) \tag{4.57}$$

with the weight-function $w(x_i) = p(x_i)/\eta(x_i)$ and $w_i := w(x_i)$. Given that $supp(p) \subset supp(\eta)$, the estimate (4.57) theoretically converges to (4.53) for every importance distribution $\eta(x)$ [185]. In practice however, the amount of samples needed is highly sensitive to the choice of importance functions. Furthermore, although convergence in $\bar{f}$ is established, the variance of the estimate $\bar{f}$ can be infinite in situations where the ratio $p(x)/\eta(x)$ is not bounded [81]. To this end, a popular approach for variance reduction is to replace the estimate (4.57) by

$$\bar{f}_w = \frac{\sum_{i=1}^{N} w_i f(\mathsf{x}_i^\eta = x_i)}{\sum_{i=1}^{N} w_i}, \tag{4.58}$$

which is justified since $1/N\sum_{i=1}^{N} w_i \to 1$ for $N \to \infty$ [185]. Still, choosing a proper importance density $\eta$ is crucial to the applicability of (4.58). This eventually renders importance sampling inapplicable in situation where this choice cannot be feasibly performed. Since plain importance sampling is not considered to be used throughout this thesis due to these difficulties, the reader is referred to Robert and Casella [185] or Rubinstein and Kroese [189] for a detailed discussion on the selection of importance functions. In this thesis, sequential Monte Carlo techniques are used to generate an importance distribution sequentially. Since these techniques build on Markov chain properties, the Markov chain Monte Carlo method is briefly described in the following.

### 4.4.3. Markov chain Monte Carlo

*Markov chain Monte Carlo (MCMC)* techniques, similar to importance sampling, also rely on auxiliary probability distributions. Although, in contrast to importance sampling, the exact choice of this auxiliary distribution is less important. MCMC methods draw on the stationary properties acquired by certain classes of *Markov chains* whereby the auxiliary distributions play the role of *transition probabilities* between successive element of the chain. Formally, given a probability space $(\mathcal{S}, \mathcal{F}, P)$, a (discrete time) Markov chain $\{\Phi_n\}_{n=1}^{\infty}$ is a sequence of random variables $\Phi_n \in \mathcal{S}$ for which the *Markov property* holds, i.e.,

$$P(\Phi_{n+1} \in \mathrm{A}|\Phi_n = x_n, \Phi_{n-1} = x_{n-1}, \dots, \Phi_0 = x_0) = P(\Phi_{n+1} \in \mathrm{A}|\Phi_n = x_n). \tag{4.59}$$

A similar property was already introduced as the defining property of a MRF being defined on a graph structure, cf. chapter 3.4.3. In the context of Markov chains, this property is understood in the sense of discrete sequential random events $\Phi_n$. Despite being seemingly oversimplifying, such chains have proven extremely successful in modeling a great variety of different state-space systems, see Meyn and Tweedie [157] for an overview. The exceedingly successful application of Markov chains in MC techniques as a means of sampling and integrating is raised by *ergodicity* and *stationarity* of certain Markov chains. These properties establish the convergence of the time average

$$\frac{1}{N}\sum_{i=1}^{N} f(\Phi_i) \underset{N\to\infty}{\to} \mathbb{E}_\pi[f(\Phi)], \tag{4.60}$$

whereby $\pi$ is the so called *stationary distribution* of the Markov chain. The goal in using MCMC techniques in the Bayesian setting introduced in chapter 3.1 is to construct a Markov chain with stationary distribution $\pi(\boldsymbol{\theta}) \equiv p(\boldsymbol{\theta}|\mathbf{Z})$, and then apply (4.60) to evaluate quantities like the PM (4.3) or the posterior variance (4.5). In the following, the basic concepts that allow for this approximation are briefly introduced and one popular algorithmic implementation (Metropolis-Hastings) is described. For an extensive introduction into the context of MCMC techniques, the reader is referred to Robert and Casella [185]. An overview with respect to application in a Bayesian setting is also given in Robert [183]. Information on the general properties of Markov chains can be found in Jerrum [107] or Meyn and Tweedie [157].

A Markov chain is generally characterized by the transition between successive elements. This transition is defined by the *transition kernel*.

**Definition 4.4.1** (Transition kernel)**.** *On a measurable space $\mathcal{S}$ with $\sigma$-algebra $\mathcal{F}$, a transition kernel $\mathfrak{K} : \mathcal{S} \times \mathcal{F} \to \mathbb{R}$ is given by the following properties:*

*(i)* $\forall x \in \mathcal{S}$, $\mathfrak{K}(x, \cdot) : \mathcal{F} \to \mathbb{R}$ *is a probability measure.*

*(ii)* $\forall \mathrm{A} \in \mathcal{F}$, $\mathfrak{K}(\cdot, \mathrm{A}) : \mathcal{S} \to \mathbb{R}$ *is measurable.*

For a continuous state space $\mathcal{S}$, the kernel represents a conditional density such that the Markov property (4.59) can be expressed by

$$P(\Phi_{n+1} \in \mathrm{A}|\Phi_n = x_n) = \int_{\mathrm{A}} \mathfrak{K}(x_n, \mathrm{d}x). \tag{4.61}$$

A Markov chain is called *homogeneous* if the transition from $n \to n+1$ is given by the same kernel for all $n$. Starting from an initial $x_0$ and using the abbreviation $x_1, \ldots, x_n =: x_{1:n}$, the joint probability density $p_{x_{1:n}|x_0}$ is given by

$$p_{x_{1:n}|x_0}(x_{1:n}|x_0) = \prod_{i=1}^{N} \mathfrak{K}(x_{i-1}, x_i). \tag{4.62}$$

Setting $\mathfrak{K}^{(1)}(x_0, \mathrm{A}) = \mathfrak{K}(x_0, \mathrm{A})$, a transition kernel $\mathfrak{K}^{(n)}$ for $n$ transitions at once is obtained by marginalizing over the states of the $n-1$ intermediate steps resulting in the recursion formula

$$\mathfrak{K}^{(n)}(x_0, \mathrm{A}) = \int_{\mathcal{S}} \mathfrak{K}^{(n-1)}(y, \mathrm{A})\mathfrak{K}(x_0, dy). \tag{4.63}$$

To allow for the convergence result (4.60), these multi-step transition kernel must belong to Markov chains with certain properties. One possible starting point to define these properties is given by the *stopping time*

$$\mathfrak{t}_{\mathrm{A}} := \inf\{n \geq 1 : \Phi_n \in \mathrm{A}\} \tag{4.64}$$

and the number of visits to the set A

$$\mathfrak{i}_A := \sum_{i=1}^{\infty} \mathbb{I}_{\mathrm{A}}(\Phi_n). \tag{4.65}$$

The probabilities $P_x(\mathfrak{t}_{\mathrm{A}} < \infty)$ and $P_x(\mathfrak{i}_{\mathrm{A}} = \infty)$ describe the probabilities of finite stopping times and infinite number of visits to a set $\mathrm{A} \in \mathcal{F}$ given a certain initial state $x \in \mathcal{S}$. These quantities are used to define *irreducibility* [see 185, chapter 6.3] and *recurrence* [see 185, chapter 6.4] of a Markov chain. Simply stated, these properties establish that a chain can and will reach every 'reasonably sized' set A from every initial state $x \in \mathcal{S}$. For a detailed definition of these properties, the reader is referred to the respective literature. The result of the technical elaborations used to establish the necessary ergodic properties is given by the following definition [185, proposition 6.33 and proof]:

**Definition 4.4.2** (Harris recurrent chain)**.** *If for every* $\mathrm{A} \in \mathcal{F}$, $P_x(\mathfrak{t}_{\mathrm{A}} < \infty) = 1$ *for all* $x \in \mathcal{S}$*, then* $P_x(\mathfrak{i}_{\mathrm{A}} = \infty) = 1$ *for all* $x \in \mathcal{S}$*, and the chain* $\{\Phi_n\}$ *is called Harris recurrent.*

The second important concept is that of a stationary distribution of a Markov chain. Stationarity is established by the existence of an invariant probability measure $\pi$ such that

$$\pi(\mathrm{A}) = \int_{\mathcal{S}} \mathfrak{K}(x, \mathrm{A})\pi(\mathrm{d}x) \quad \forall \quad \mathrm{A} \in \mathcal{F}. \tag{4.66}$$

If the distribution $\pi$ exists, it is referred to as *stationary* distribution of the chain. To check in practice whether a distribution $\pi$ is stationary for a particular transition kernel, the *detailed balance condition*

$$\mathfrak{K}(x, y)\pi(x) = \mathfrak{K}(y, x)\pi(y) \tag{4.67}$$

provides a sufficient condition [185, theorem 6.46]. Given a Harris recurrent aperiodic (i.e., $P_x(\mathfrak{t}_{\mathrm{A}} = 1) > 0$) chain, it can be shown that the chain converges according to

$$\lim_{n\to\infty} \| \int_{\mathcal{S}} \mathfrak{K}^{(n)}(x, \cdot)\mu(\mathrm{d}x) - \pi(\cdot)\| = 0, \tag{4.68}$$

for every initial distribution $\mu$ [185, theorem 6.51]. Thereby $\pi$ is a unique stationary distribution of the chain, and $\|\cdot\|$ is understood as the TV norm on measures [24]. It can be shown that (4.68) directly establishes the desired property (4.60).

### Metropolis-Hastings algorithm

The *Metropolis-Hastings algorithm* [93] is one instance of the class of MCMC-based algorithms. In order to construct a chain with a desired stationary distribution $\pi(x)$, it uses a proposal distribution $q_{pp}(x|y)$. The main requirement on this proposal density is

```
Require: $x = \Phi_n$
  Generate $y \sim q_{pp}(y|x)$
  $\varrho(x,y) = \min\{\frac{\pi(y)}{\pi(x)} \frac{q_{pp}(x|y)}{q_{pp}(y|x)}, 1\}$
  if $\rho(x,y) \geq 1$ then
      $\Phi_{t+1} = y$
  else
      Generate $a \sim \mathcal{U}(0,1)$
      if $\rho(x,y) \geq a$ then
          $\Phi_{n+1} = y$
      else
          $\Phi_{n+1} = x$
      end if
  end if
```

Algorithm 3: Metropolis-Hastings kernel.

that it must be possible to easily generate samples from it. The algorithmic outline of the transition kernel to generate the next sample $\Phi_{n+1}$ from the current sample $\Phi_n$ in the chain is given in algorithm 3.

Formally, the transition kernel corresponding to the Metropolis-Hastings algorithm is given by

$$\mathfrak{K}_{mh}(x,y) = \varrho(x,y)q_{pp}(y|x) + (1 - r(x))\delta_x(y), \tag{4.69}$$

with $r(x) = \int \varrho(x,y)q_{pp}(y|x)\mathrm{d}y$ and the *acceptance probability*

$$\varrho(x,y) = \min\{\frac{\pi(y)}{\pi(x)} \frac{q_{pp}(x|y)}{q_{pp}(y|x)}, 1\}. \tag{4.70}$$

Whereas the transition probabilities represented by this kernel cannot be evaluated in practice due to the intractable integrals, it can easily be checked that the detailed balance condition (4.67) is fulfilled. Thus, by construction, $\pi$ is the stationary distribution corresponding to the kernel (4.69). With the minimal necessary condition

$$\mathit{supp}(\pi) \subset \bigcup_{x \in \mathit{supp}(\pi)} \mathit{supp}(q_{pp}(\cdot|x)) \tag{4.71}$$

such that it is generally possible for the chain to visit every possible state $x \in \mathit{supp}(\pi)$, this allows for a variety of different proposal densities. These are usually selected from a family of distributions with certain tunable location and scale parameters [41]. The efficiency of the chain, i.e., the speed of convergence, is then influenced by a proper choice of a family of proposal densities and an appropriate scaling. A popular subclass of MCMC methods are so called *random walks*. A random walk is characterized by proposals such as

$$q_{pp}(y|x) = \tilde{q}_{pp}(y - x) = \tilde{q}_{pp}(z) \tag{4.72}$$

with the update $y = x + z$. If the proposal density $\tilde{q}_{pp}$ is chosen to be Gaussian, the particular Markov chain is called *Gaussian random walk*. For certain classes of target densities $\pi$, optimal convergence of the chain can be controlled by means of the *acceptance* *rate*, i.e., the ratio of accepted versus dismissed samples. By using a Gaussian random walk, the scaling parameter of the Gaussian distribution can be used to control the acceptance ratio whereby it has been found that an acceptance rate of $\approx 1/4$ is optimal for a multidimensional target [186].

MCMC methods for integration and sampling arose with the emergence of computational power in the last century and have since then been established as the state-of-the-art tool in the high dimensional regime. In the area of Bayesian computation, their flexibility allowed for development of models with ever increasing complexity. Whereas MCMC methods themselves are in principle not affected by the 'curse of dimensionality', the choice of good proposal densities becomes often more complicated with increasing dimension. This can severely affect the convergence of the chain. Furthermore, Markov chains are built up sequentially. The application of sophisticated and complex models such as described in chapter 2 can impose high computational efforts in terms of overall computational time. As a consequence, MCMC methods can become impractical. This issue is addressed by *sequential Monte Carlo (SMC)* methods.

### 4.4.4. Sequential Monte Carlo

*Sequential Monte Carlo* tries to combine both, the flexibility of MCMC methods with the *embarrassing* parallelity of importance sampling. To this end, instead of targeting a single posterior distribution, SMC targets a sequence of distributions $\{\pi_n\}_{n=1}^t$. Such sequences are either given naturally by sequential filtering problems [see e.g. 49, 54], or they can be created artificially as, e.g., in global optimization by simulated annealing [see e.g 165]. Throughout this thesis, the sequence is constructed as a transition from an initial distribution $\pi_0$ to a target distribution $\pi$ via the intermediate distributions

$$\pi_n(x) = \pi_0(x)^{1-\lambda_n}\pi(x)^{\lambda_n}, \tag{4.73}$$

with $0 < \lambda_1 < \cdots < \lambda_t = 1$ [see 50]. Such a sequence enables the following rationale: given the initial distribution $\pi_0$ can be readily approximated via (importance) sampling, the sequence (4.73) is such that a sequence of importance distributions $\{\eta_n\}_{n=1}^t$ can be created according to

$$\eta_n(x) = \int_{\mathcal{S}} \eta_{n-1}(y)\mathfrak{K}_n(y,x)\mathrm{d}y. \tag{4.74}$$

The subscript $n$, in contrast to (4.63), indicates that $\mathfrak{K}_n$ is an indexed transition kernel for step $n$. I.e., the transition kernel is allowed to change for every transition $n \to n+1$. Starting from a set of particles $\{\mathsf{x}_0^i\}_{i=1}^N \sim \eta_0(x)$ (in the simplest case $\eta_0 = \pi_0$) and evolving each particle $\mathsf{x}_0^i \to \mathsf{X}_1^i$ using the kernel $\mathfrak{K}_{n=1}$, the resulting set of particles is distributed according to $\{\mathsf{x}_1^i\}_{i=1}^N \sim \eta_1(x)$. In the next step this set of particles is then evolved using $\mathfrak{K}_{n=2}$ and so on. This approach is also known as *sequential importance sampling.* To apply it in practice, it is necessary to evaluate the importance distribution $\eta_n(x_n)$ given by

$$\eta_n(x_n) = \int_{\mathcal{S}^n} \eta_0(x_0)\eta_{1:n}(x_{1:n})\mathrm{d}x_{0:n-1} = \int_{\mathcal{S}^n} \eta_0(x_0)\prod_{i=1}^{n}\mathfrak{K}_i(x_{i-1},x_i)\mathrm{d}x_{0:n-1} \tag{4.75}$$

for each particle $\mathsf{x}_n^i$. This however is only possible for very simple transitions. For instance, a Metropolis-Hastings kernel such as (4.69) does not allow for this evaluation.

At this point, SMC techniques [50] can be used to compensate for this deficiency in the construction of plain sequential importance sampling.

In the following, the sequence of target distributions is represented via

$$\pi_n(x) = \frac{\varsigma_n(x)}{Z_n}. \tag{4.76}$$

To circumvent the computation of (4.75), SMC sampling uses an artificial joint target distribution $\tilde{\pi}_n(x_{1:n}) = \tilde{\varsigma}_n(x_{1:n})/Z_{1:n}$ with

$$\tilde{\varsigma}(x_{0:n}) = \varsigma_n(x_n) \prod_{i=0}^{n-1} \mathfrak{L}_i(x_{i+1}, x_i). \tag{4.77}$$

Thereby, the *backward* Markov kernel $\mathfrak{L}_i : \mathcal{S} \times \mathcal{F} \to \mathbb{R}$ describes a backward transition $i+1 \to i$ equivalent to the (forward) Markov kernel from definition 4.4.1. This representation enables to use importance sampling for the joint distribution $\tilde{\pi}(x_{0:n})$ by means of the joint importance distribution $\eta_n(x_{0:n}) = \eta_0(x_0)\eta_{1:n}(x_{1:n})$. This feature of the SMC approach avoids the explicit marginalization in (4.75). Since $\tilde{\pi}(x_{0:n})$ admits $\pi_n(x_n)$ as a marginal by construction, its importance approximation also provides estimates for the approximation of $\pi_n(x_n)$ [78].

For an initial set of weighted particles $\{\mathsf{w}_0^i, \mathsf{x}_0^i\}_{i=1}^N$ with $\mathsf{x}_0^i \sim \eta_0(x)$ (whereby in the simplest case $\eta_0 = \pi_0$, the initial weights simplify to $\mathsf{w}_0^i = 1$), a path $\mathsf{x}_{0:t}^i$ for each particle through the $dim((\mathcal{S})^t)$-dimensional space is created by sequential evolution of each particle using the forward kernel $\mathfrak{K}_n$. Accordingly, this evolution generates a sequence of weight-functions $\{w_n\}_{n=1}^t$ with elements

$$\begin{aligned} w_n(x_{0:n}) &= \frac{\tilde{\varsigma}_n(x_{0:n})}{\eta_n(x_{0:n})} && (4.78) \\ &= \frac{\varsigma_n(x_n) \prod_{i=0}^{n-1} \mathfrak{L}_i(x_{i+1}, x_i)}{\eta_0(x_0) \prod_{i=1}^{n} \mathfrak{K}_n(x_{i-1}, x_i)} && (4.79) \\ &= \underbrace{\frac{\varsigma_n(x_n)\mathfrak{L}_{n-1}(x_n, x_{n-1})}{\varsigma_{n-1}(x_{n-1})\mathfrak{K}_n(x_{n-1}, x_n)}}_{\tilde{w}_n(x_n, x_{n-1})} \underbrace{\frac{\varsigma_{n-1}(x_{n-1}) \prod_{i=0}^{n-2} \mathfrak{L}_i(x_{i+1}, x_i)}{\eta_0(x_0) \prod_{i=1}^{n-1} \mathfrak{K}_n(x_{i-1}, x_i)}}_{w_{n-1}(x_{0:n-1})}. && (4.80) \end{aligned}$$

Thus, the current weights $w_n(x_{0:n})$ can be obtained by the incremental update

$$w_n(x_{0:n}) = \tilde{w}_n(x_n, x_{n-1}) w_{n-1}(x_{0:n-1}) \tag{4.81}$$

whereby the incremental weights depend on the backward kernel $\mathfrak{L}_i$. The optimal choice of these kernels with respect to the variance of the weights $w_n$ is again infeasible. But it is shown by Del Moral et al. [50] that a good approximation is given by

$$\mathfrak{L}_{n-1}(x_n, x_{n-1}) = \frac{\pi_n(x_{n-1})\mathfrak{K}_n(x_{n-1}, x_n)}{\pi_n(x_n)} \tag{4.82}$$

if $\mathfrak{K}_n$ is a MCMC kernel as introduced in sec. 4.4.3 and $\pi_{n-1} \approx \pi_n$. This choice results in the incremental weights

$$\tilde{w}_n(x_n, x_{n-1}) = \frac{\varsigma_n(x_{n-1})}{\varsigma_{n-1}(x_{n-1})}. \tag{4.83}$$

As a consequence of $\pi_n(x_n)$ being the marginal distribution of $\tilde{\pi}_{0:n}(x_{0:n})$, a particle approximation $\pi_n^N$ is given by

$$\pi_n(x) \approx \pi_n^N(x) = \sum_{i=1}^{N} \mathsf{w}_n^i \delta_{\mathsf{x}_n}(x) \tag{4.84}$$

with the normalized weights $\mathsf{w}_n^i = \mathsf{w}_{n-1}^i \tilde{w}_n(\mathsf{x}_n^i, \mathsf{x}_{n-1}^i) / \sum_{j=1}^N \mathsf{w}_{n-1}^j \tilde{w}_n(\mathsf{x}_n^j, \mathsf{x}_{n-1}^j)$. As a result, approximations of expected values such as $\mathbb{E}_{\pi_n}[f(\mathsf{x})]$ are equivalently obtained.

With the potential decrease in discrepancy between the importance distribution $\eta_n$ and the target distribution $\pi_n$ with increasing time $n$, also the variance of the estimates based on the particle approximation $\pi_n^N$ increases. To account for this, the particle basis is *resampled* whenever the discrepancy between $\eta_n$ and $\pi_n$, commonly measured by the *effective sample size (ESS)* [141], exceeds some threshold $\mathrm{T}_{red}$. The ESS ($0 < ESS < N$) as a measure of degeneracy of the particle approximation is given by

$$ESS_n = \left( \sum_{i=1}^{N} (W_n^i)^2 \right)^{-1} . \tag{4.85}$$

A common choice for the threshold is $\mathrm{T}_{red} = N/2$ [see e.g. 123]. A very simple resampling strategy, the *multinomial scheme*, is to choose the new particles according to the particle approximation $\{X_{n-1}^i\}_{i=1}^N \sim \pi_n^N$.

To preserve the approximation quality of the backward kernel (4.82), it is crucial to find a step size $\Delta\lambda_n = \lambda_{n+1} - \lambda_n$ such that $\pi_{n+1} \approx \pi$. On the one hand, the approximation will be better for smaller step sizes. On the other hand, in order to bound the computational effort, the step size cannot be arbitrarily small. Furthermore, it is to be expected that choosing a single constant step size appropriate for the whole sequence results in an inefficient scheme due to too small steps in slowly evolving regions of the sequence. It was therefore proposed by Koutsourelakis [123] to use an adaptive step size control. Thereby, the sequence of steps is determined by an acceptable reduction in ESS such that

$$ESS_n = \xi ESS_{n-1} \tag{4.86}$$

with the user defined reduction factor $\xi$. Throughout the applications presented in this thesis, $\xi = 0.95$ is chosen according to Koutsourelakis [123]. A summary of the final algorithm is given in algorithm 4.

SMC methods provide a very general and powerful framework for a broad range of statistical inference problems. The application to the Bayesian parameter identification problem enables a fully probabilistic analysis of very complex and sophisticated nonlinear models without the a priori restrictions of the approximate analysis presented in chapter 4.3. Although this benefit is accompanied by an increased computational cost, this cost can be handled by the inherent parallelity of the algorithm. This parallelity is given by the possibility to evaluate each particle independently during the rejuvenation step. Assuming that the rejuvenation process represents the majority of the computational cost within one time step, which is justified for complex nonlinear models, the overall computational time scales almost perfectly with the parallel distribution of particles.

By construction, the initial distribution $\pi_0$ is less important to the overall efficiency of the algorithm than the importance distribution $\eta$ is to plain importance sampling.

```
$n \leftarrow 0$, $\lambda_n \leftarrow 0$;
$\{\mathsf{x}_0^i\}_{i=1}^N \sim \pi_0$, $\{\mathsf{w}_0^i = 1/N\}$;
while $\lambda_n < 1$ do
    $n \leftarrow n + 1$;
    Find $\lambda_n$, s.t. $ESS_n \approx \xi ESS_{n-1}$ AND $\lambda_n \leq 1$;
    if $ESS_n < \mathrm{T}_{red}$ then
        Resample: $\{\mathsf{x}_{n-1}^i\}_{i=1}^N \sim \pi_n^N$;
    end if
    Rejuvenate: $\{\mathsf{x}_n^i\}_{i=1}^N \sim \mathfrak{K}_n(x_{n-1}, \cdot)$;
end while
```

Algorithm 4: Sequential Monte Carlo algorithm according to Del Moral et al. [50](Remark 1). Common choices for the $ESS$ reduction factor $\xi$ and the resampling threshold $\mathrm{T}_{red}$ are $\xi = 0.95$ and $\mathrm{T}_{red} = N/2$.

However, in practice, providing a good initial distribution significantly reduces the number of steps in the sequence of distributions (4.73) and accordingly also the variance in potential estimates. Additionally, with increasing dimension of the problem, this effect becomes more pronounced due to the ‘curse of dimensionality’. The same arguments apply to the choice of forward kernels $\mathfrak{K}_n$. For complex models, these choices are crucial for the successful application of the SMC algorithm. A possible approach that makes use of dimensionality reduction is presented in chapter 5.

**Computational aspects** Using the advantage of the inherent parallelity provided by the SMC algorithm requires an efficient software implementation. This is of particular importance with increasing model complexity which is often reflected in increased computational time. To this end, the implementation developed within the scope of this thesis uses the hierarchical parallel layout shown in figure 4.2. Depending on the available hardware architecture, the particles representing states in the sample space are distributed equally among different *parallel groups*. Within each of the parallel groups, the particles are evaluated sequentially. Thus, for a fixed number of particles, the computational time is inversely proportional to the number of parallel groups. This proportionality is only slightly deteriorated by the increase in communication overhead. The communication between the groups results from the resampling step, see algorithm 4, which necessitates a redistribution of particles among the parallel groups whenever the ESS falls below the resampling threshold $\mathrm{T}_{red}$.

If required by the evaluation of the surface current similarity measure, see chapter 3.3.2 page 51, the implementation allows for a hybrid OpenMP/message passing interface (MPI)-parallel layout where each MPI rank is pinned to a certain number of threads.

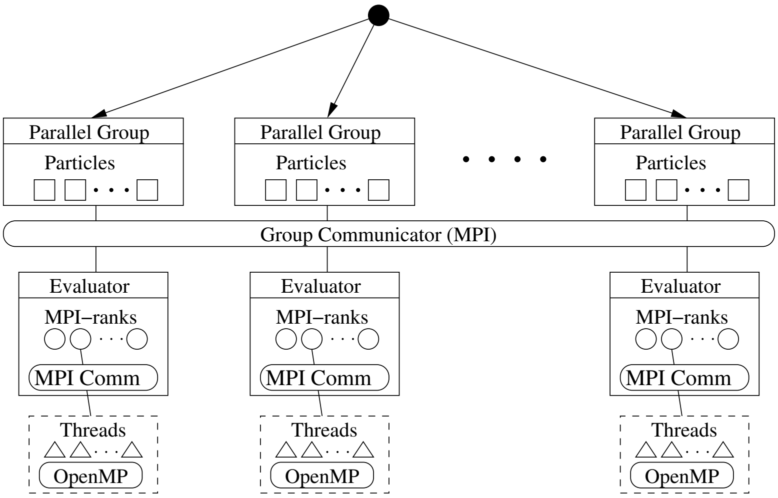

Figure 4.2.: Parallel layout for the efficient implementation of the SMCalgorithm. Particles (□) are assigned to parallel groups. Within each group, particles are evaluated on an MPI-parallel layout. Thereby, each MPI rank (◯) can be pinned to a certain number of threads (△). Communication among the different parallel groups and within a parallel group is managed by MPI communication. The thread parallel level can, e.g., be implemented with OpenMP.

# 5. A two-stage approach towards predictive modeling of AAA growth

The intention of this chapter is the description of the concepts for numerical solution introduced in chapter 4 in view of a parameter identification framework for AAA growth as introduced in the chapters 2 and 3. The aspect of a specific choice of a similarity measure is thereby left to be discussed in chapter 6.

A crucial prerequisite for the predictive modeling of AAA growth is the mathematical description of a priori knowledge in the prior-distribution. Whereas there are attempts to create such a knowledge base for population parameters [199], existing approaches for the inverse identification of spatial distributions of patient-specific growth parameters rely on prior distributions implicitly defined by sparse spatial representations of parameters [232]. Studies regarding the patient-specific inverse modeling of AAA are scarce and there is no common agreement on a population wide definition of priors for growth parameters. As a consequence, at this stage any choice of prior represents a personal belief in the statistical population of growth parameters.

In light of the prior models presented in chapter 3.4, this belief is expressed in terms of certain smoothness assumptions on the spatial distribution of parameters. The following reasoning thereby adapts to the qualitative specification of priors (3.71). In this regard, both, the Gaussian process prior with squared exponential kernel and the smoothness prior represent functions with regular spatial variation. Functions with certain specific but localized features are thus rendered unlikely. However, a 'classical' AAA is not unlikely to be a localized phenomenon with respect to the entire aorta. Whereas this does not imply the local character to be necessarily reflected in a spatial distribution of parameters describing such a phenomenon, it is clearly undesirable to define this possibility to be unlikely a priori. Hence, the choice of the more flexible TV prior represents a reasonable a priori model in the sense of (3.71) since it allows for a local character without enforcing it.

In order for the graph-based approximation of the TV norm (3.88) to accurately represent functions with bounded variation from the continuous setting (3.85) and its smoothed version (3.86) respectively, the graph structuring the underlying domain has to be accurate. I.e., its vertices have to cover the domain of the parameters $\theta$ (given by the reference configuration $\Omega_0$) accurately and uniformly. This requirement can conveniently be achieved by fixing the graph structure to be defined by the elements of the FE mesh and their connectivity, see appendix F. Hence, the dimension $n_p$ of the inverse problem is given by the discretization size $n_{ele}$.

In principle, the resulting $n_p$-dimensional problem could be tackled by any of the numerical techniques presented in chapter 4. In practice, their plain application is often rendered infeasible. Whereas the approximate methods suffer from unquantifiable bounds on the quality of the approximation in the case of nonlinear models, MC methods are implicitly affected by the ‘curse of dimensionality’. For importance sampling based methods, this is reflected in the difficulty of finding appropriate importance distributions. For SMC methods, this problem is mirrored in the definition of initial distributions. Similar difficulties arise in the definition of proposal densities for the application of MCMC methods. Furthermore, standard MCMC methods are inherently sequential methods. Thus for computational models with an evaluation time in the order of minutes to hours, a sequential evaluation in the order of 10.000 - 100.000 evaluations as required by standard MCMC methods is not an option.

In this light, applications of the inverse modeling in the context of large-scale nonlinear computational models can be approached differently. On the one hand, with the increasing influence of methods from machine learning to the field of model-based inverse problems, approximate inference has seen rapid improvements. In particular, VB approaches have been shown to be able to result in good approximations also for nonlinear problems [39, 66]. On the other hand, methods for dimensionality reduction are used to render sampling based approaches feasible. These methods are based on the assumption that low dimensional representations of the input parameter space can be found that reflect the major contributions to the posterior density. The resulting reduced dimensional problem - whereby the aspect of a reduced representation solely refers to the dimension of the input parameters - allows for the solution of large-scale, nonlinear inverse problems by means of SMC techniques. Such an approach is presented in the following.

## 5.1. Existing approaches for dimensionality reduction in cardiovascular modeling

To the author’s best knowledge, there are no existing approaches to the inverse modeling of AAA growth based on longitudinal data that explicitly account for the probabilistic nature of the problem. Although there are attempts towards the inverse modeling of patient-specific AAA growth based on the constrained mixture theory [230, 231, 232], these approaches are neither accounting for the inherent probabilistic nature of the inverse problem nor based on longitudinal data. A non-probabilistic approach using longitudinal data was presented by Tinkl [218].

However, inverse modeling and uncertainty quantification in the general field of biomechanical modeling and especially in the field of cardiovascular modeling experienced significant advances in recent years. Thereby a variety of different problems were approached. These approaches differ further by the different mathematical formulations and the according numerical techniques. Nevertheless, the applied approaches to cope with the problem dimension can be roughly grouped into a few main categories:

- Identification problems formulated as classical optimization problems in line with (3.6) are not in need of dimensionality reduction if used in combination with efficient adjoint approaches as described in chapter 3.5 [see e.g. 170]. These formulations are usually not intended to be used in a probabilistic sense.

- *Rule-based* approaches enforce some particular *rule* as a mapping from a high-dimensional to a low-dimensional parameters space a priori. As, e.g., in the case of cardiac models, these rules can be physiologically motivated [37, 164] whereby the rule is motivated from clinical research [36]. Furthermore, rules can be expressed in terms of certain sparsity assumptions which are enforced as part of the identification problem by means of primary model output. An example of such rules is, e.g., given by clustering approaches based on computed strain maps [14, 118]. Also, the a priori assignment of patches to represent the solution [20, 21, 161, 218] or the enforcement of certain polynomial representations of the solution [232] can be seen as a rule-based dimensionality reduction.

- If the definition of the prior admits an orthogonal decomposition, a *truncated* representation of the prior represents a natural choice for a reduced dimensional basis in the sense of a principle component analysis (PCA). Such approaches are naturally possible for Gaussian priors by a factorization of the covariance matrix. In the setting of Gaussian process priors, sparse representations are formalized by the *representer theorem* [see e.g. 196]. In the context of inverse problems, such formulations have been used by Koutsourelakis [123].
- High problem dimension is also an issue in the approximate inference methods, e.g., due to high storage requirements of $n_p \times n_p$ hessian matrices. As a remedy, sparse representations of the Laplace approximation can, e.g., be tackled by *matrix-free* representations of the covariance matrix [32]. One instance of such a sparse representation is, e.g., given by the two-loop recursion formula used in the LBFGS algorithm, see chapter 4.4.1. Very recently, VB approaches with built-in identification of reduced dimensional representations have also been proposed [66, 124].

The implementation of one of the existing strategies to reduce the stochastic dimension for the identification of AAA growth is not straightforward, which is reflected in the following aspects:

- In contrast to cardiac mechanics, a rule-based model for parameter reduction is not available and the most simple rule in the sense of an a priori assignment of patches suffers from severe disadvantages. On the one hand, given that a spatial scale of variation of the true parameters is know providing an estimate of a patch size, it is still unclear whether a particular patch-wise approximation is consistent with the true solution in terms of the location of the patches. On the other hand, it is unclear whether a patch-wise basis is able to cover the main covariance structure of the true solution. Since this covariance structure will be reflected in predictions of AAA growth, this is an important aspect with respect to predictive modeling.
- The TV prior does not allow for an a priori decomposition due to its nonlinearity. Thus, in contrast to Gaussian priors, it does not admit a reduced basis in terms of an a priori truncated representation.
- Although nonlinear VB approaches have shown that the resulting approximations are in good agreement to the true posterior for certain scenarios [39, 66], their performance highly depends on the nonlinearity of the model.

In conclusion, existing methods for dimensionality reduction cannot be applied directly in the inverse modeling of AAA growth. However, although not a priori accessible, some sparsity in the input parameters is expected due to the assumed boundedness in variation that is expressed by the choice of the TV prior. In the following, an approach is proposed that tries to exploit this sparsity. To evaluate its performance, it is contrasted with a patch-wise approximation of the solution.

## 5.2. A posteriori dimensionality reduction under TV prior assumptions

In this chapter, a novel PCA-like approach [see e.g. 108] is suggested. It uses a decomposition

$$\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \mathbf{P} \cdot \boldsymbol{\psi} \tag{5.1}$$

with the *reduced dimensional* parameters $\boldsymbol{\psi} \in \mathbb{R}^{n_r}$ and the *dictionary* $\mathbf{P} \in \mathbb{R}^{n_p} \times \mathbb{R}^{n_r}$. The approach is based on the assumption that a parametrization (5.1) with $n_r << n_p$ can be found. The dictionary $\mathbf{P}$ is comprised of pair-wise orthonormal columns representing a basis for the reduced dimensional space $\mathbb{R}^{n_r}$ such that $\mathbf{P}^\top \mathbf{P} = \mathbf{I}$. This setup defines a linear monotonic mapping $h_{\mathbf{P}} : \boldsymbol{\psi} \mapsto \boldsymbol{\theta}$. From a probabilistic point of view, it is most important that this mapping captures the main covariance structure in the high-dimensional layout. Assuming that this covariance structure is sufficiently represented by the Laplace approximation, a factorization of the associated covariance matrix would yield the desired basis, as discussed by Bui-Thanh et al. [32]. However, the resulting dense $n_p \times n_p$ covariance matrices cannot be handled in the large-scale regime and methods to approximate the action of the covariance matrix are usually applied. Hence, the quality of the basis would be directly coupled to the quality of the approximation of the covariance matrix. For instance, the requirements on the covariance approximation are relatively low to still obtain good convergence in the LBFGS scheme that is applied throughout this thesis. As a consequence, a factorization of the resulting approximate covariance might not be optimal.

In order to decouple the numerical solution and the dimension reduction, a different approach is proposed here. Motivated by the work of Babacan et al. [10], a local quadratic approximation of the TV functional (3.88) is utilized. Using the inequality[1]

$$\sqrt{v} \leq \frac{v+z}{2\sqrt{z}} \quad \forall z \tag{5.2}$$

with $\sqrt{v} = (v+z)/(2\sqrt{z}) \iff v = z$ and setting $v = \sum_j \mathbf{W}^G_{i,j}(x_j - x_i)^2 + \epsilon^2$, it can be seen that the TV norm is bounded by

$$TV_w(\boldsymbol{x}) \leq \sum_i \frac{\left(\sum_j \mathbf{W}^G_{i,j}(x_j - x_i)^2 + \epsilon^2\right) + z_i}{\sqrt{z_i}}. \tag{5.3}$$

---

[1] This inequality is easily verified from linearization of $\sqrt{v}$ around $z$.

This bound is quadratic in $\boldsymbol{x}$ and as such allows for a representation in terms of a symmetric operator $\mathbf{L}_{tv}(z_i)$ such that

$$TV_w(\boldsymbol{x}) \leq \boldsymbol{x} \cdot \mathbf{L}_{tv}(z_i) \cdot \boldsymbol{x} + \sum_i \frac{\epsilon^2 + z_i}{\sqrt{z_i}} + const \tag{5.4}$$

Given the eigendecompostion

$$\mathbf{L}_{tv}(z_i) = \mathbf{Y}\boldsymbol{\Lambda}\mathbf{Y}^\top \tag{5.5}$$

with the matrix of eigenvectors $\mathbf{Y}$ and the matrix of eigenvalues $\boldsymbol{\Lambda}$, a reduced basis $\mathbf{P}$ can be constructed by

$$\mathbf{P} = \mathbf{Y}_{n_r}, \tag{5.6}$$

where $\mathbf{Y}_{n_r}$ represents the columns of $\mathbf{Y}$ corresponding to the $n_r$ smallest eigenvalues. Given that the TV prior represents a class of desirable solutions, the rationale behind this choice is that the quadratic approximation (5.4) is still capable of representing the characteristics of these desirable solutions. The dominant properties should then be captured by the eigenvectors corresponding to the small eigenvalues since these represent modes of least action for the operator $\mathbf{L}_{tv}(z_i)$. See figure 5.6 for an illustration of this idea in practice. Since the operator (5.5) depends on the point of linearization, the factorization (5.5) is unique only up to the choice of $z_i$. A natural choice is to set $z_i = v$ and to evaluate $v$ at the MAP estimate. This choice then suggests a 2-stage approach:

1. Find the MAP estimate $\hat{\boldsymbol{\theta}}_{map}$ and the associated Laplace approximation (4.15). By using a Quasi-Newton method such as the LBFGS method, the covariance $\mathbf{H}^{-1}$ will only be available in terms of its approximate action on some vector $\mathbf{v}$. This aspect is notationally accounted for by using the symbol $\tilde{\mathbf{H}}^{-1}\mathbf{v}$ in the following. Anyhow, this product is never computed in terms of a matrix-vector product but in terms of the two-loop-recursion algorithm 2, see chapter 4.4.1. Set $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_{map}$. Find a reduced dimensional basis given by (5.4)-(5.6). Initialize $\boldsymbol{\psi} = \mathbf{0}$.

2. Find the PM and the posterior variance, or any other posterior statistic, by MC techniques such as SMC with a parametrization given by (5.1). Due to the monotonic behavior of the linear one-to-one mapping $h_{\mathbf{P}}$ defined by (5.1), the posterior $p(\boldsymbol{\psi}|\mathbf{Z})$ is given by the change of variables for probability densities

   $$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}|\mathbf{Z}) = |\mathbf{P}|\, p_{\boldsymbol{\psi}}(\boldsymbol{\psi}|\mathbf{Z}) \propto p(\mathbf{Z}|h_{\mathbf{P}}(\boldsymbol{\psi}))p(h_{\mathbf{P}}(\boldsymbol{\psi})). \tag{5.7}$$

   Again, since the change in functional dependency is expected to be clear from the argument, the labeling indices are omitted resulting in $p(\boldsymbol{\theta}|\mathbf{Z}) \propto p(\boldsymbol{\psi}|\mathbf{Z})$. The two important ingredients for a successful application of the SMC algorithm are the initial distribution $\pi_0$ and an appropriate choice of transition kernels $\mathfrak{K}_n$. For the initial distribution the already available Laplace approximation projected to the reduced dimensional parameters is a natural candidate. With the linear transformation (5.1), this projection is given by

   $$\pi_0(\boldsymbol{\psi}) = \mathcal{N}(\mathbf{0}, \mathbf{P}^\top\tilde{\mathbf{H}}^{-1}\mathbf{P}). \tag{5.8}$$

   Again, the product $\mathbf{H}_{\mathbf{P}}^{-1} := \mathbf{P}^\top\tilde{\mathbf{H}}^{-1}\mathbf{P}$ is not computed as a dense matrix-matrix product but column-wise, which is feasible since $n_r << n_p$. Additionally, assuming that the Laplace approximation reasonably captures the dimensions of the true posterior, it can be made use of as a proposal density in the sense of (4.72) according to

$$q(x|y) = \mathcal{N}(y, \sigma_P \mathbf{H}_{\mathbf{P}}^{-1}). \tag{5.9}$$

Thereby $\sigma_P$ is an adjustable scaling parameter such that an optimal acceptance ratio of the transition kernel is obtained.

The proposed approach incorporates several beneficial properties:

- Step 1 of the approach can be achieved by state-of-the-art large-scale optimization. Quasi-Newton methods, routinely resulting in approximate Laplace approximations, can thereby additionally be used to initialize step 2. This approximation will have a direct effect on the efficiency of step 2. Anyhow, the dimensionality reduction relies solely on the existence of the MAP estimate which can be found by a great variety of different algorithmic setups.
- As can be seen from (5.3), the sparsity pattern of the linear operator $\mathbf{L}_{tv}$ is entirely defined by the spatial connectivity of the parameters $\boldsymbol{\theta}$. E.g., for an element-wise layout, the sparsity pattern is defined by the connectivity of the elements of the FE mesh. Depending on the specific definition of this connectivity, see appendix F, this results in extremely sparse graph structures. Thus, the eigendecomposition can be performed by standard large-scale sparse eigenanalysis [see e.g. 6].
- The Laplace approximation, which cannot be considered a good approximation in general [25, 144], is only used in the sense of preconditioning for the second step. So even in situations where the Laplace approximation is not suitable to be used as direct importance distribution for the posterior, *globalization strategies* like SMC can make use of it in a meaningful way.

### 5.2.1. Patch-wise approximations

In the setting of TV prior assumptions, the representation of a point estimate such as the MAP solution in terms of a priori patches seems to be justified at first sight. Such a patch-wise solution can also formally be represented by (5.1). Let a *patch* be defined by a connected set of elements $\mathcal{K}_i = \{\mathcal{E}_k\}_{k=1}^{n_{ele}^{\mathcal{K}_i}}$ of the mesh $\mathcal{K}$, see chapter 2.3, such that

- $\mathcal{K}_i \subset \mathcal{K}$,
- $\mathcal{K} = \bigcup_{i=1}^{N_p} \mathcal{K}_i$,
- $\mathcal{K}_i \cap \mathcal{K}_j = \varnothing \quad \forall i \neq j$,

with the number of patches $N_p$ and the number of elements $n_{ele}^{\mathcal{K}_i}$ per patch $\mathcal{K}_i$. In this case, the columns of the dictionary $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_{n_r}]$ are comprised of the vectors

$$\mathbf{p}_i = \{\boldsymbol{v} \in \mathbb{R}^{n_p} : v_k = 1/\sqrt{n_{ele}^{\mathcal{K}_i}} \iff \mathcal{E}_k \in \mathcal{K}_i, v_k = 0 \iff \mathcal{E}_k \notin \mathcal{K}_i\}. \tag{5.10}$$

Obviously, an a priori definition of such a patch-wise basis is only possible if the patch size is chosen small enough so that, justified by prior knowledge, the spatial variations of the solution can be represented. If not justified by expert knowledge, this might not be possible in general. In the sense of the a posteriori dimensionality reduction introduced above, it is also possible to fit a patch-wise basis to the MAP estimate. This then results in an ‘optimal’ distribution of patches. Here, it is therefore suggested to construct the a posterior patches based on a histogram of the MAP estimate. Let a histogram of the MAP estimate with bins $\mathfrak{m}_i$ be given such that

$$n_{ele} = \sum_{i=1}^{n_{bin}} \mathfrak{m}_i \tag{5.11}$$

with $n_{bin}$ being the number of bins and the set $\mathfrak{M} := \{\mathfrak{m}_1, \dots, \mathfrak{m}_{n_{bin}}\}$ being the set of bins with decreasing magnitude, i.e., $\mathfrak{m}_1 > \mathfrak{m}_2, \dots > \mathfrak{m}_{n_{bin}} \geq 0$. Patches can then be created by the association of element-wise values $\hat{\theta}_{map,\mathcal{E}}$ to the ‘closest’ bin $\mathfrak{m}_j$ and an inter-connectivity check within each patch. The creation of the patch-wise basis is outlined in algorithm 5.

```
Given a set of ordered bins $\mathfrak{M} = \{\mathfrak{m}_1, \dots, \mathfrak{m}_{n_{bin}}\}$;
$i = 1$; Choose a tolerance $tol_p$;
repeat
    Find element-wise association to the bins:
        $\forall \mathcal{E} \in \mathcal{K},\ d_{\mathcal{E}} \leftarrow \underset{j}{\operatorname{argmin}}(\left| m_j - \hat{\theta}_{map,\mathcal{E}} \right|), s.t.\, j \leq i$;
    Create sets $\mathcal{K}_j = \{\mathcal{E} \in \mathcal{K} : d_{\mathcal{E}} = j\}$;
    Split each set $\mathcal{K}_j$ into subsets of interconnected elements;
    Create basis according to (5.10);
    $\boldsymbol{\psi} \leftarrow \mathbf{P}^\top \hat{\boldsymbol{\theta}}_{map}$;
    $c \leftarrow \|\hat{\boldsymbol{\theta}}_{map} - \mathbf{P}\boldsymbol{\psi}\| / \|\hat{\boldsymbol{\theta}}_{map}\|$;
    $i \leftarrow i + 1$;
until $c > tol_p$ OR $i \leq n_{bin}$
```

Algorithm 5: Patch creation algorithm.

Therein, the check for inter-connectivity between all the elements within one patch is performed by standard search algorithms such as the *Breadth-first-search* algorithm [see e.g. 191]. By choosing a low tolerance $\xi$, the approximation quality with respect to the MAP estimate can be controlled. However, the requirement on the resulting basis to represent the main posterior covariance structure is not necessarily guaranteed and can lead to wrong probabilistic interpretations. This will be shown by the following example.

## 5.3. Proof of concept

The intention of the following proof of concept is twofold: First, it is used to generate an intuition of the functionality of the SMC algorithm 4 with respect to the choice of different initial distributions. Second, it is shown that the proposed dimensionality reduction approach results in feasible reduced dimensional approximations that can be efficiently used for predictive modeling.

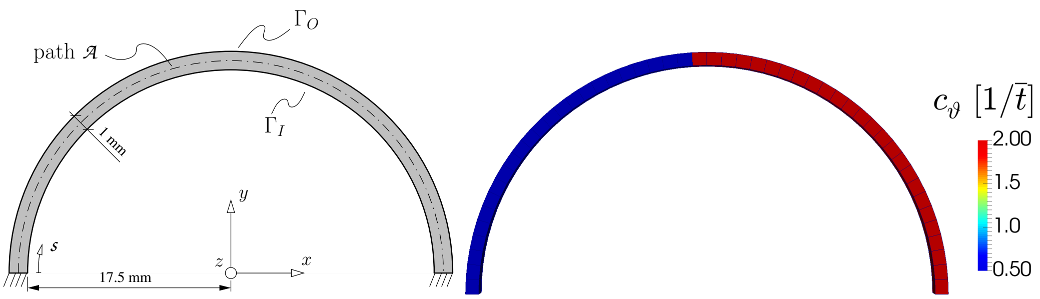

Figure 5.1.: Geometrically simplified model geometry mimicking a segment of the aorta (left). Ground truth/reference solution of the growth parameter $c_\vartheta \in \{0.5, 2.0\}$ $[1/\bar{t}]$ for the creation of synthetic measurement data (right). Growth is considered over a generic time period $\bar{t}$.

To this end, a geometrically simplified model is employed, see figure 5.1. In combination with the continuum mechanical model described in chapter 2 and with the growth model (2.65) in particular, this model mimics the setup in a patient-specific simulation of arterial growth. In this context, the geometry shown in figure 5.1 can be interpreted as a segment from the aorta. Along this segment, the model is discretized with 50 hexahedral elements. To model the constitutive properties of the arterial wall, a simple but widely used model proposed by Raghavan and Vorp [179] is applied. It is given in a nearly incompressible formulation by

$$\Psi_{\text{AAA}}(\hat{\mathbf{C}}_e) = \alpha(\hat{\mathrm{I}}_{1,e} - 3) + \beta(\hat{\mathrm{I}}_{1,e} - 3)^2 + \Psi_{vol}(\mathbf{C}_e). \tag{5.12}$$

Therein, $\hat{\mathrm{I}}_{1,e}$ represents the first invariant of the distortional component of the elastic (i.e., growth-free) right Cauchy-Green tensor $\hat{\mathbf{C}}_e = J_e^{-1/3}\mathbf{C}_e = (J_e^{-1/3}/\vartheta^2)\mathbf{C}$. The volumetric penalty is chosen as

$$\Psi_{\text{vol}}(\mathbf{C}_e) = \frac{\kappa}{4}(2\ln J_e + J_e^2 - 1). \tag{5.13}$$

The parameters for these constitutive relations are taken from Maier et al. [148] resulting in $\alpha = 0.174$ MPa, $\beta = 1.881$ MPa and $\kappa = 104.4$ MPa. For the simulation of growth, a patch-wise spatial distribution $c_{\vartheta,P}$ of the growth parameter $c_\vartheta$ is prescribed as shown in figure 5.1. Growth is simulated over a generic time period $\bar{t}$. Prior to the growth simulation, the model is prestressed by the application of an orthonormal pressure on the inner surface $\Gamma_I$ of the arch in the initial configuration. By applying a pressure of $\mathrm{p} = 1.066\mathrm{e}{-2}$ MPa $\approx 80$ mmHg, diastolic conditions are sought. Thereby, the MULF approach [77] is applied to project the pressurized state to the initial configuration, see chapter 2.5. The overall nonlinear solution is controlled by the load-control scheme given in figure 5.2.

To define a synthetic inverse problem, the solution of the forward problem $\mathbf{A}(\boldsymbol{\theta} = c_{\vartheta,P})$ is used to construct a measurement in the space of currents via the measured surface $\mathbf{Z} = \varphi^{\mathbf{A}(c_{\vartheta,P})}(\Gamma_O)$ as the push-forward of the model surface $\Gamma_O$ given the parameters $c_{\vartheta,P}$. To avoid the most obvious version of the *inverse crime* [110], this measurement is blurred with centered Gaussian noise ($\sigma_N = 0.1$ mm) in the space of currents, see chapter 3.3.2. The spatial scale of the kernel defining the space of test functions $\mathcal{W}$ is

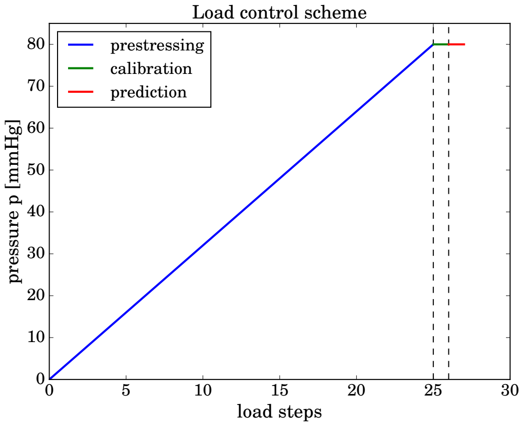

Figure 5.2.: Load control scheme. In the first 25 steps, the load is incrementally increased to a pressure of p = 80 mmHg whereby MULF is applied. Growth is then initiated from this prestressed state. Step 26 represent a state corresponding to measured data (i.e. a state from which a measurement is synthetically created). Step 27 computes a prediction of further growth.

chosen to $\sigma_W = 0.5$ mm. Assuming the noise variance $\sigma_N$ to be known, the likelihood is then given in terms of (3.69).

### 5.3.1. 1 dimensional example

In a first step, the above model is used to analyze the functionality of the SMC algorithm 4. To this end, the parametrization of the growth parameter along the segment of the aortic arch is a priori set to be uniformly constant and thus given by one single parameter. This allows to explore the posterior in a structured way. Thus, expectations and variances computed from integration rules can serve as exact reference solutions. Without any restrictions to the analysis, the posterior is assumed to be entirely defined by the likelihood at this point. The structured evaluation of the posterior on the interval $c_\vartheta \in [1.2, 1.35]$ in 300 steps, see figure 5.3, allows to compute the reference solution

$$\mathbb{E}[c_\vartheta] \approx 1.266\ 1/\bar{t}, \quad \mathbb{SD}[c_\vartheta] \approx 1.002\text{e}{-2}\ 1/\bar{t}. \tag{5.14}$$

The results of the application of the SMC algorithm 4 are given in table 5.1 for different initial distributions $\pi_0$, see figure 5.3. On the one hand, a very ‘bad’ initial guess with respect to mean and variance is chosen as $\pi_0 = \mathcal{N}(1.0, 1.0)$. On the other hand, a ‘good’ initial guess is provided by the Laplace approximation $\pi_0 = \mathcal{N}(\hat{\boldsymbol{\theta}}_{map} \approx 1.266, \tilde{\mathbf{H}}^{-1} \approx 3.769\text{e}{-3})$. The Laplace approximation is computed by the BFGS algorithm 1 (convergence tolerance $tol = 1.0\text{e}{-6}$, storage size $m_s = 10$; convergence was achieved after 5 iterations). The deviation between the approximate Laplace approximation and the almost Gaussian posterior density is due to the very few Quasi-Newton iterations and the corresponding rough hessian approximation. To reduce the variance in the statistical estimates, multiple runs of the SMC algorithm ($N = 5$), each with 96 particles, are performed simultaneously and the resulting estimates are averaged over the number of

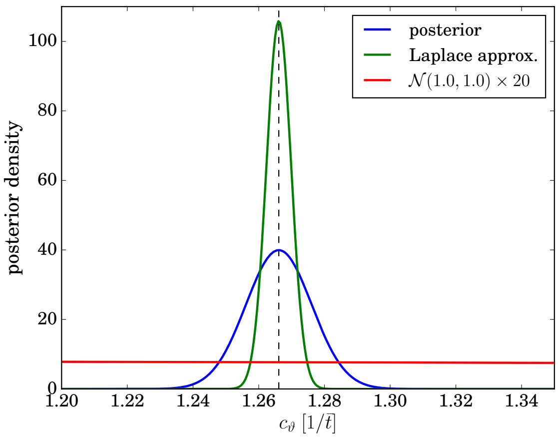

Figure 5.3.: Sampled posterior density (blue) and initial distributions for the SMC algorithm. A very uninformative Gaussian distribution $\mathcal{N}(1.0, 1.0)$ (red, scaled by a factor of 20 for visualization) and an approximation of the Laplace approximation given by $\mathcal{N}(1.266, 3.769\text{e}{-3})$ (green).

runs [see e.g. 50]. Based on the final particle approximation, statistical estimates of the primal solution of the forward problem – i.e. the deformation due to growth in terms of displacements $\mathbf{u}$ – are additionally reported in table 5.1. Therefore, the displacement $u_y^{max}$ in $y$-direction at $x, y = (0.0 \text{ mm}, 18.5 \text{ mm})$ (see figure 5.1) is monitored at the end of the calibration phase.

| initial distribution $\pi_0$ | $\mathbb{E}[\boldsymbol{\theta}](\pm\mathbb{SD}[\boldsymbol{\theta}])$ $[1/\bar{t}]$ | $\mathbb{E}[u_y^{max}](\pm\mathbb{SD}[u_y^{max}])$ $[1/\bar{t}]$ | numiter |
|---|---|---|---|
| $\mathcal{N}(1.0, 1.0)$ | 1.266(±1.119e-02) | 1.167(±9.985e-03) | 42.8 |
| $\mathcal{N}(1.266, 3.769 \cdot 10^{-3})$ | 1.267(±8.668e-03) | 1.168(±7.738e-03) | 17.0 |

Table 5.1.: Statistical estimates obtained by the application of the SMC algorithm with different initial distributions that highly influence the average number of iterations and consequently the computational cost.

As expected, the final approximations are of equal quality for both initial guesses. Compared to the bad initial guess, the Laplace approximation as initial guess allows for a significant reduction in the average number of iterations per SMC run. In this particular example, the Laplace approximation represents a relatively good approximation to the posterior. But even samples from the bad initial guess cover the domain of the posterior reasonably well enough to allow the SMC algorithm to converge quickly. In multiple dimensions though, this discrepancy is expected to worsen rendering the application of the SMC algorithm with uninformed initial distributions infeasible.

### 5.3.2. 50 dimensional example

In a second step, the predictive capabilities of the proposed dimensionality reduction approach are analyzed. Therefore, a spatially varying solution of the growth parameter $c_\vartheta$ is allowed by using the element-wise basis (2.105). The expected blocky structure of the solution is modeled by the TV prior (3.93) ($\alpha_{tv} = 1.0$, $\epsilon = 1.0\text{e}{-2}$). With a problem dimension of $n_p = 50$, this setup does not constitute a large-scale inverse problem. However, the dimension is already high enough to render the application of plain MCMC highly inefficient. Furthermore, the dimension is high enough to allow for meaningful reduced dimensional approximations. At the same time, the dimension is low enough to allow for the plain application of the SMC algorithm without any dimensionality reduction providing a reference solution. The predictive quality of this *full-space* solution can then be compared against a reduced-dimensional solution. The predictive quality is thereby measured in the sense of the predictive capabilities of the calibrated model to provide an *informed extrapolation* of future growth. This prediction is monitored as the growth/displacement $u_y(s)$ in $y$-direction along path $\mathscr{A}$ (see figure 5.1) after the prediction step (see figure 5.2). The displacement $u_y(c_\vartheta)$ is a function of the parameters and as such its probabilistic interpretation is clearly of interest. Furthermore, since $u_y(s)$ is a function of the parameters only through the solution of the forward problem, it captures the influence of the covariance structure of the parameters. Thus, it provides a natural means to qualify reduced dimensional approximations. The solution of the identification problem is reported in terms of posterior-mean $\mathbb{E}_I[c_\vartheta]$ and posterior standard deviation $\mathbb{SD}_I[c_\vartheta]$. The predictive qualities are assessed in term of expected value $\mathbb{E}_I[u_y]$ and standard deviation $\mathbb{SD}_I[u_y]$ of the predicted growth $u_y$. The index $I \in (\boldsymbol{\theta}, q(\boldsymbol{\theta}), \boldsymbol{\psi}_p, \boldsymbol{\psi}_{tv})$ thereby refers to the evaluation of the statistics with respect to the full-dimensional parameter layout indicated by $I = \boldsymbol{\theta}$, the Laplace approximation indicated by $I = q(\boldsymbol{\theta})$ and the reduced dimensional layouts indicated by $I = \boldsymbol{\psi}_p$ for the patch-wise approximation and $I = \boldsymbol{\psi}_{tv}$ for the approximation based on the TV approximation.

**Reference solution** The identified parameters of the full-dimensional solution and the associated prediction are shown in figure 5.4. The Laplace approximation is computed with the BFGS algorithm 1 (convergence tolerance $tol = 1.0\text{e}{-8}$; storage size $m_s = 200$; although 52 iterations were sufficient to achieve convergence). The SMC algorithm was then started from this Laplace approximation as initial distribution with 5 simultaneous runs each comprised of 3024 particles. Such a particle approximation, with $\approx 60$ particles per parametric dimension, showed excellent agreement in terms of mean and variance with a MCMC solution.

It can be seen from figure 5.4 that the exact solution is reasonably well covered by both, the mean of the Laplace approximation and the PM computed with SMC. The relative good alignment of the PM solution and the MAP solution suggest that the posterior density is more or less symmetric. The prediction $u_y$ has a minimum of sensitivity, measured in terms of variance, at the center of the arch. Given that the surface current similarity measure can very well detect mismatch in out-of-surface direction but that it is insensitive with respect to a mismatch in in-surface direction, this reflects the expected behavior. The often mentioned capability of a reduced dimensional approximation to ‘capture the main posterior covariances’ can subsequently be measured in terms of the capability to reproduce the predictive solution $u_y$ with respect to mean and variance. In the following, the two parameter reduction techniques proposed in chapter 5.2 are analyzed in this regard.

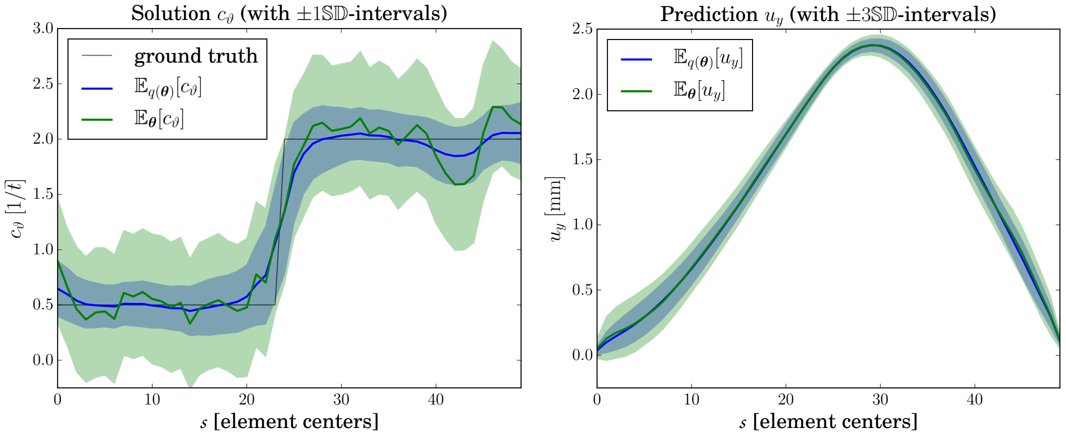

Figure 5.4.: Parameters $c_\vartheta$ (left) and prediction $u_y$ (right) using an element-wise basis for the spatial representation of the parameters. The results from the particle approximation of the posterior obtained via the SMC algorithm are given in green. The Laplace approximation is given for comparison in blue. Intervals are given as $\pm 1\sigma$-regions and $\pm 3\sigma$-regions around the mean, respectively.

**Patch-wise dimensionality reduction** First, the 'optimal' patch-wise approximation is applied. Using a histogram of the MAP solution with $n_{bin} = 10$ bins, algorithm 5 stopped after 2 iterations with an approximation tolerance

$$tol_p = \|\hat{\boldsymbol{\theta}}_{map} - \mathbf{P}\boldsymbol{\psi}\| / \|\hat{\boldsymbol{\theta}}_{map}\| = 10\%. \tag{5.15}$$

The two resulting patches match the patches of the reference solution exactly. For the application of the SMC algorithm, 120 particles were used per run resulting again in 60 particles per parametric dimension. The initial distribution was set to the projected Laplace approximation according to (5.8).

It can be seen from figure 5.5 (left) that this patch-wise approximation represents a reasonable approximation of the PM. However, the standard deviation does not reflect the variability of the full-space solution any longer. In itself, this does not necessarily disqualify the reduced basis, given it could still represent the main posterior covariance structure. However, from the inspection of the prediction $u_y$, see figure 5.5 (right), it becomes clear that the patch-wise basis fails in this respect. Whereas the prediction mean $\mathbb{E}_{\boldsymbol{\psi}_p}[u_y]$ is still very close to the full-space solution, the prediction standard deviation $\mathbb{SD}_{\boldsymbol{\psi}_p}[u_y]$ is even worse than the standard deviation approximated by the Laplace approximation $\mathbb{SD}_{q(\boldsymbol{\theta})}[u_y]$. It is clear that such an approximation is not suited to draw probabilistic conclusions. This holds especially in situations where the definition of patches is made a priori.

**TV-based dimensionality reduction** The improvements offered by the dimensionality reduction based on the decomposition of the linearized TV functional are able to overcome the deficiencies of the patch-wise approximation. A motivation for the approach based on the TV decomposition is given by the inspection of the eigenspectrum of the linear operator (5.5) evalutated at the MAP point, see figure 5.6. The eigenspectrum

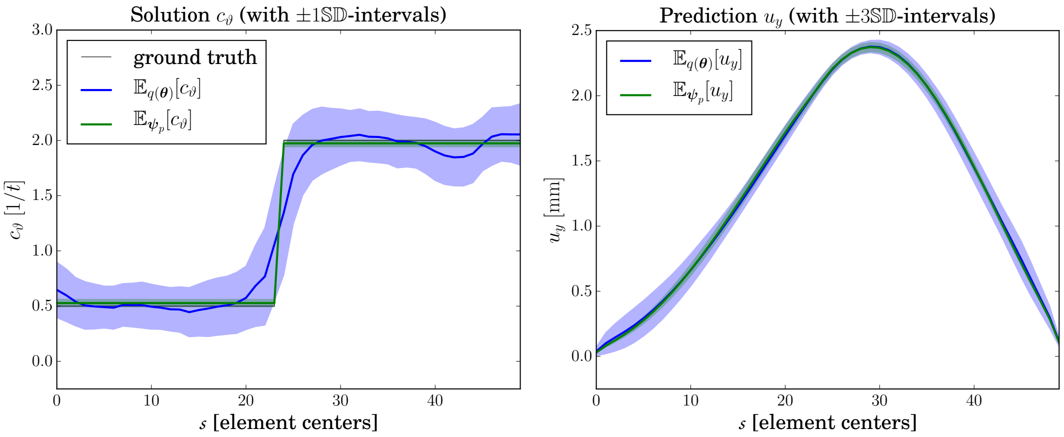

Figure 5.5.: Parameters $c_\vartheta$ (left) and prediction $u_y$ (right) using a patch-wise basis with 2 patches. The results from the particle approximation of the posterior obtained via the SMC algorithm are given in green. The Laplace approximation in terms of the element-wise basis is given for comparison in blue. Intervals are given as $\pm 1\sigma$-regions and $\pm 3\sigma$-regions around the mean, respectively. The variability of the patch-wise approximation is drastically reduced compared to the reference solution shown in figure 5.4.

stagnates in terms of eigenvalues at the end of the spectrum. This indicates that there is a set of eigenvectors to which the operator applies a similar action. Furthermore, since these are the eigenvectors with low eigenvalue, they are associated to higher prior densities, i.e., likely solutions from the a priori point of view. This suggests that the operator's action on desirable solutions can be well approximated by the reduced set of eigenvectors corresponding to these distinct eigenvalues. Furthermore, it implies that solutions favored by the linearized TV prior can be reasonably represented by these eigenvectors. Additionally, it can be seen from the shape of these eigenvectors that they indicate the 'singularity'/jump in the MAP solution. From a practical point of view, these facts make this choice of basis promising.

The results obtained with a reduced basis constructed from the linearized TV prior are given in figure 5.7. The smallest 10 eigenvectors are used to construct the reduced approximation according to (5.6). The SMC-algorithm was started from the projected Laplace approximation with 756 particles per run, which again results in $\approx 60$ particles per parametric dimension. The PM $\mathbb{E}_{\boldsymbol{\psi}_{tv}}[c_\vartheta]$ shows similar characteristics as the full-space PM $\mathbb{E}_{\boldsymbol{\theta}}[c_\vartheta]$. The posterior standard deviation $\mathbb{SD}_{\boldsymbol{\psi}_{tv}}[c_\vartheta]$ is in the range of the standard deviation $\mathbb{SD}_{q(\boldsymbol{\theta})}[c_\vartheta]$ from the Laplace approximation. However, the quality of the prediction $u_y$ is very similar to the full-space solution in terms of mean and variance. This aspect is shown in more detail in figure 5.8.

These results suggest that the information provided by the linearized TV prior is well suited to represent the main posterior covariance structure. Furthermore, the sparsity of solutions favored by the TV prior is reflected in the linearization (5.4) and can be used to construct a reasonably reduced dimensional approximation.

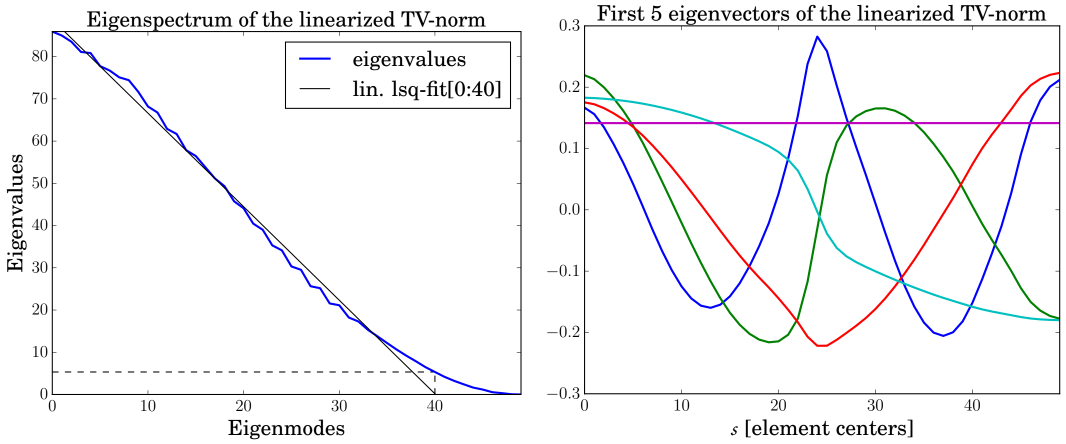

Figure 5.6.: Eigenanalysis of the linearized TV norm. The smallest 5 eigenvectors corresponding to the smallest eigenvalues are shown on the right. Due to the rank-1 deficiency of the operator, the eigenvectors are given with respect to the arbitrary level 0.0. The null-space is represented by the constant mode.

## 5.4. Conclusion

In this chapter, a practical approach to dimensionality reduction under TV prior assumptions was proposed. The reduction of the dimensionality is thereby enabled in terms of the eigendecomposition of the linear operator associated to the quadratic approximation of the TV norm at the MAP estimate. The applicability of this approach is not constrained by the dimensionality of the problem but by the spatial variation of the MAP solution. However, this variation is a priori expected to be bounded as a consequence of the choice of the TV prior. Thus, given efficient algorithms for the computation of the MAP solution, the proposed approach can be readily applied in the large-scale regime. The obtained reduced dimensional basis then allows the use of MC techniques such as SMC which can be efficiently applied in the regime of sophisticated and computationally expensive models. This application is shown in chapter 7.

It was demonstrated on a simple example that the efficiency of the SMC algorithm in the context of an artificially created sequence of distributions depends on the quality of the initial distributions. Throughout the work presented in this thesis, the Laplace approximation, or a Quasi-Newton approximation thereof, is used. However, in situations where the Laplace approximation is not suitable, i.e., for highly unsymmetric posterior densities or multimodal posterior densities, the initial distribution can be obtained from more sophisticated posterior approximations [see e.g. 65, 66].

Furthermore, the proposed dimensionality reduction approach was compared against the widely used patch-wise approximation of the spatial variation of parameters. It was shown that the patch-wise approximation can result in misleading probabilistic conclusions. In this regard, the proposed approach is able to sustain the predictive capabilities of the model under a reduced dimensional approximation. Thus, a practical alternative was presented that is able to overcome the deficiencies of a patch-wise approximation.

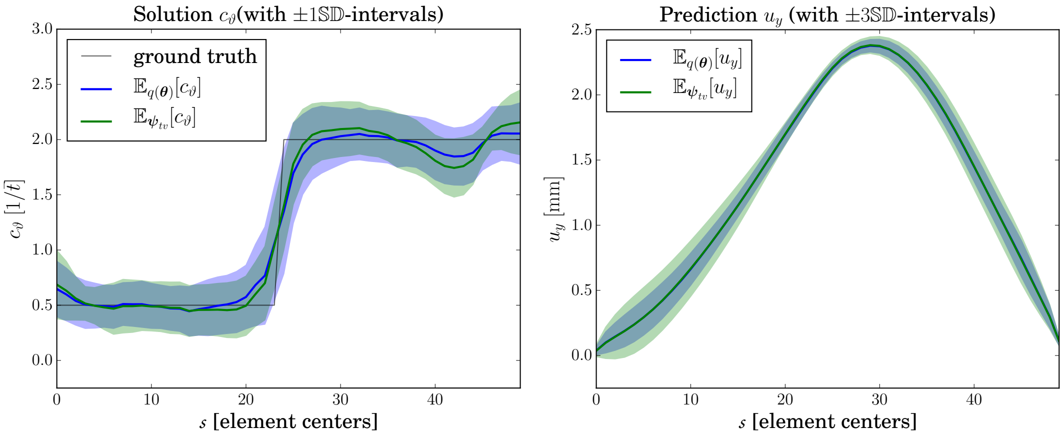

Figure 5.7.: Parameters $c_\vartheta$ (left) and prediction $u_y$ (right) using a reduced basis constructed from the linearized TV norm. The results from the particle approximation of the posterior obtained via the SMC algorithm are given in green. The Laplace approximation in terms of the element-wise basis is given for comparison in blue. Intervals are given as $\pm 1\sigma$-regions and $\pm 3\sigma$-regions around the mean, respectively.

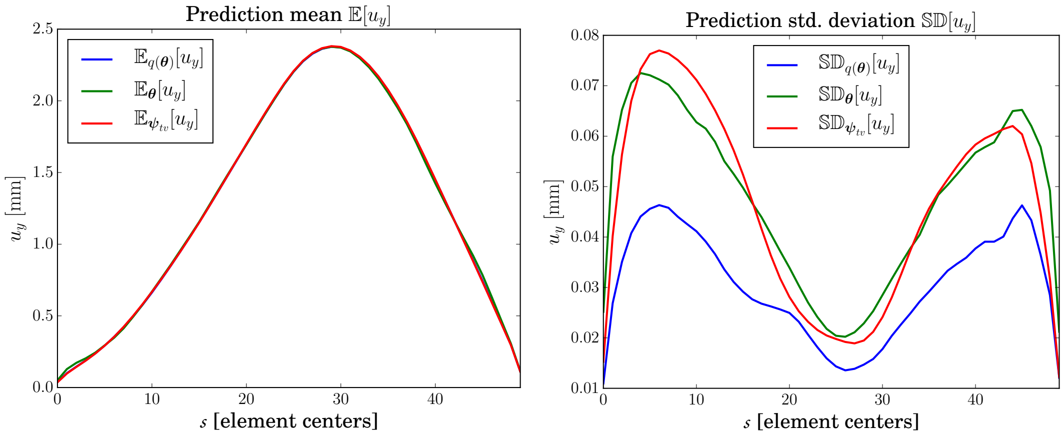

Figure 5.8.: Comparison of the predictive capabilities obtained under the full-space element-wise approximation and the reduced basis given by the TV linearization. It can be seen, that the reduced dimensional approach covers the true solution very well in terms of mean (left) and in terms of variance (right). The prediction from the Laplace approximation is given for comparison.

# 6. Analysis of the choice of similarity measure using synthetic data

As indicated in the introduction, one major aspect for UQ in the setting of model calibration is introduced through the observation error. By using the relation (3.26), this observation error is entirely modeled by the generic error term $\boldsymbol{\zeta}$. Its physical interpretation depends on the definition of the space of measurements $\mathcal{Z}$. The difficulties in this definition were highlighted in chapter 3.2. In particular, it was shown that the use of image registration can introduce a significant systematic error that is not modeled by the Gaussian assumption (3.24). This fact will lead to a bias in the calibrated parameters and accordingly in the predictive capabilities of the calibrated model. Being constrained to the definition of measurements given by point-wise displacement data, this bias could theoretically be accounted for by incorporating a model for image registration in the computational model $C \circ \mathbf{A}(\boldsymbol{\theta})$. However, models for image registration are in the same order of complexity than the computational model for AAA growth itself. Thus, such an approach is clearly undesirable and adds unnecessary and probably infeasible complexity to the calibration problem. Rather, it was proposed that the systematic error associated to image registration can be avoided by defining measurements in terms of surface data.

To highlight this effect, this chapter provides an analysis based on synthetic data. To this end, measured data is generated from a computational model and subjected to both, random and systematic errors. The propagation of these errors to the identified parameters under different definitions of the space of measurements is then analyzed. More specific, the computational model is intended to mimic growth of a synthetic AAA. The data generation is then given by the growth of this model under a certain set of parameters as the deformation of the model subject to random noise. The systematic error is introduced by generating the final measured deformed model from a registration step that maps the undeformed model to the noise corrupted deformed model. This measured deformation is then used within the parameter identification in two different ways. On the one hand, it is applied in terms of point-wise displacements at the discrete degrees of freedom of the FE model boundary. On the other hand, it is applied in the sense of discrete surface currents (3.59) as the deformed model boundary itself. Since model errors are excluded, this setup allows for a relative quantification of the quality of the results obtained under these two definitions of the space of measurements.

The presentation is thereby not intended towards a quantification of model inadequacy in a probabilistic sense. Rather, it concentrates on the avoidance of the propagation of artificially introduced systematic errors through image registration. For an introduction towards the incorporation of model inadequacy, the reader is referred to Kennedy and O'Hagan [116]. The work presented in this chapter is partly based on a similar analysis shown previously by Kehl and Gee [112].

## 6.1. Data generation

The synthetic AAA model is shown in figure 6.1. Its geometry incorporates an asymmetric distribution of intraluminal thrombus (ILT) and a thin layer of 1 mm of aneurysmatic arterial wall. Its diameter in the descending aorta is 22 mm from where it enlarges to mimic a rather large aneurysm with a maximum diameter of 72 mm. For the computational model, symmetry in longitudinal and lateral direction is used and, despite not being completely physiological, Dirichlet conditions according to (2.39) are prescribed at the proximal inlet of the aneurysm. For later reference, the luminal (inner) model surface is referred to as $\Gamma_I$ and the outer model surface is referred to as $\Gamma_O$.

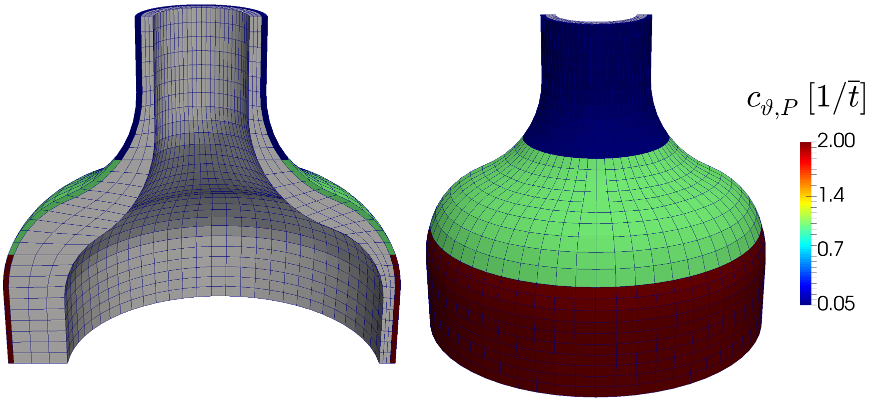

Figure 6.1.: Discretization of the synthetic AAA model which utilizes symmetry in lateral and longitudinal direction. The model incorporates an asymmetric distribution of ILT and a patch-wise distribution of the growth parameter $c_\vartheta$ ($c_{\vartheta,P} \in \{0.05, 1.0, 2.0\}$) in the arterial wall.

The ILT is modeled as a coupled compressible neo-Hookean material [99] whose strain energy function is given in terms of the first invariant $\mathrm{I}_{1,e}$ of the elastic component $\mathbf{C}_e$ of right Cauchy-Green tensor according to

$$\Psi_{ILT} = c_1(\mathrm{I}_{1,e} - 3) + \frac{c_1}{\beta_1}(J_e{-2\beta_1} - 1) \tag{6.1}$$

with $c_1 = 18.0$ kPa and $\beta_1 = 4.5$ taken from Maier et al. [148]. The arterial wall is modeled as the nearly-incompressible neo-Hookean type material given in (5.12) and (5.13).

For the simulation of growth according to (2.65), a patch-wise spatial distribution $c_{\vartheta,P}$ of the growth parameter $c_\vartheta$ is prescribed as shown in figure 6.1. Growth is simulated over a generic time period $\bar{t}$. Prior to the growth simulation, the model is prestressed by the application of an orthonormal pressure on the inner surface $\Gamma_I$ of the synthetic AAA in the initial configuration. By applying a pressure of $\mathrm{p} = 1.0666\mathrm{e}{-2}$ MPa $\approx 80$ mmHg, diastolic conditions are sought. Thereby, the MULF approach is applied to project the pressurized state to the initial configuration [77]. The overall nonlinear solution is controlled by the load-control scheme given in figure 5.2 whereby the prediction-phase is not of interest for this example. Except for the simplification in geometry, this setup very closely resembles the approach pursued in state-of-the-art patient-specific computational experiments.

A synthetic measurement is given by the solution of the forward problem $\mathbf{A}(\boldsymbol{\theta} = c_{\vartheta,P})$ as

$$\mathbf{D}_{\Gamma_O} := C(\mathbf{A}(c_{\vartheta,P})) + \boldsymbol{\zeta}, \tag{6.2}$$

whereby $\boldsymbol{\zeta}$ is a vector of white Gaussian noise ($\sigma = 0.1$, which corresponds to a signal-to-noise ratio (SNR) $\approx 30\%$) and $C$ is the restriction to the degrees of freedom on $\Gamma_O$. The associated push-forward of the model surface $\Gamma_O$ is given by

$$\Gamma_O^{\mathbf{A}(c_{\vartheta,P})} := \varphi^{\mathbf{A}(c_{\vartheta,P})+\boldsymbol{\zeta}}(\Gamma_O), \tag{6.3}$$

i.e., the outer model surface deformed via the displacements $\mathbf{D}_{\Gamma_O}$. The current representation of this surface is given according to (3.50).

### 6.1.1. Data registration via surface matching

To introduce a systematic error into the data generation, a registration step is applied. Thereby, the practical situation where the measurements $\mathbf{D}_{\Gamma_O}$ are extracted from longitudinal image data is emulated. Since the synthetic model is not based on image data, this registration step cannot be performed in the sense of image registration as introduced in chapter 3.2. Instead, a registration algorithm is applied that operates on surface data directly. To this end, the surface matching algorithm introduced in Vaillant and Glaunès [221] is used here. It applies the LDDMM framework [see e.g. 57]. Since the specific properties of such an algorithm are not of interest for the effect to be shown, the reader is referred to appendix E for an introduction of the surface matching framework. In summary, for two surfaces $T_1 \subset \Omega_0^{\mathrm{I}}$ and $T_2 \subset \Omega_t^{\mathrm{I}}$ embedded in the ambient space $\hat{\Omega} = \Omega_0^{\mathrm{I}} \cup \Omega_t^{\mathrm{I}}$, the surface matching tries to find a diffeomorphism $\varphi_R(\hat{\mathbf{x}})$ with $\varphi_R(\hat{\mathbf{x}}) = \hat{\mathbf{x}} + \mathbf{V}_R$ such that an energy functional based on the surface current distance

$$\|T_2 - \varphi_R(T_1)\|_{\mathcal{W}^*} \tag{6.4}$$

is minimized, see (E.3) for details. Consequently, when applied to the model surfaces $\Gamma_O$ and the deformed model surface $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$, the matching results in an optimal mapping $\Gamma_O \mapsto \Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$ with respect to the norm $\|\cdot\|_{\mathcal{W}^*}$. The quality of this mapping is depicted in figure 6.2. It can be seen that the bulk of point-wise distances between $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$ and $\varphi_R(\Gamma_O)$ is $< 0.1$ mm. This indicates that the surface matching accurately performs the mapping $\Gamma_O \mapsto \Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$. However, the in-plane accuracy of the registration, which cannot be detected by the surface current norm $\|\cdot\|_{\mathcal{W}^*}$, might be severely corrupted. This setup closely resembles the application of image registration as outlined in chapter 3.2 where uniqueness of deformations is enforced by modeling assumptions not necessarily in line with the underlying physical processes.

The deformation resulting from the application of the surface matching algorithm allows for two different definitions of measurements. On the one hand, the point-wise measurement

$$\mathbf{V}_{R,\Gamma_O} := C_{R,\Gamma_O}(\mathbf{V}_R) \tag{6.5}$$

is generated by restricting the registered deformation to the associated $n_{dof}^{\Gamma}$ degrees of freedom of the boundary $\Gamma_O$. This defines a likelihood in terms of the similarity measure

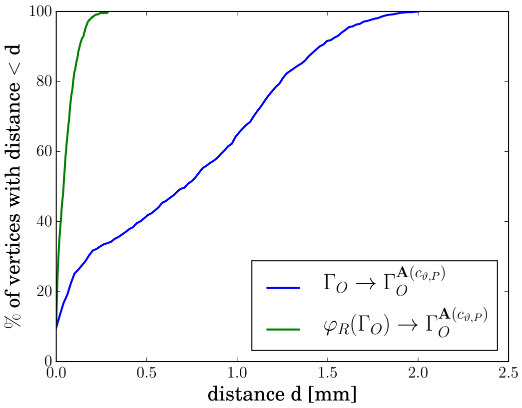

Figure 6.2.: Result of the surface matching algorithm in terms of the projected distance (3.44). The initial distances between $\Gamma_O$ and $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$ is shown in blue. The result of the registration in terms of the distance between the target $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$ and the registered surface $\varphi_R(\Gamma_O)$ is shown in green.

(3.38) via

$$-\log p(\boldsymbol{\theta}|\mathbf{Z}) = -\log p_D(c_\vartheta|\mathbf{V}_{R,\Gamma_O}) = \frac{1}{2\sigma_D^2}\|C(\mathbf{A}(c_\vartheta)) - \mathbf{V}_{R,\Gamma_O}\|^2_{\mathbb{R}^{n^\Gamma_{dof}}}. \tag{6.6}$$

On the other hand, the registered surface

$$\Gamma_O^R := \varphi_R(\Gamma_O) \tag{6.7}$$

can be used as a measurement in the space of currents. According to (3.69), a negative log-likelihood is given by

$$-\log p(\boldsymbol{\theta}|\mathbf{Z}) = -\log p_W(c_\vartheta|\Gamma_0^R) = \frac{1}{2\sigma_N^2}\|\varphi^{\mathbf{A}(c_\vartheta)}(\Gamma_O) - \Gamma_O^R\|^2_{\mathcal{W}^*}. \tag{6.8}$$

Both likelihood functions could have been equivalently defined if the measurements $\mathbf{V}_{R,\Gamma_O}$ had been obtained from image registration. This highlights the use of the surface matching algorithm as an emulation of systematic errors introduced through the application of image registration.

### 6.1.2. Inverse problem setup

The result of the registration step is the creation of a registered displacement measurement $\mathbf{V}_{R,\Gamma_O}$ that closely resembles the result of a measurement process in applications involving image registration. Furthermore, by just changing the definition of the space of measurements from point-wise displacement measurements to surface measurements, the same measurement is obtained as a measured surface $\Gamma_O^R$. Since both measurements are subject to the exact same systematic error, the resulting parameter estimates can be compared. A prerequisite for this comparison is the consistent determination of the noise parameters $\sigma_D$ and $\sigma_N$. Obviously, $\sigma_D$ is best chosen to be the standard deviation of the synthetically applied measurement noise $\boldsymbol{\zeta}$. The consistent choice of $\sigma_N$, which describes the measurement noise transformed into the space of currents, is not straightforward due to the complicated relation of the noisy push-forward relation of the associated current $\varphi^{\mathbf{A}(c_{\vartheta,P})+\boldsymbol{\zeta}}(\Gamma_O)(\omega)$. Furthermore, normally distributed noise in the space of discrete currents is defined by a covariance structure on the normals defining the surface, see chapter 3.2. It is therefore unlikely that white Gaussian noise on the measurements $\mathbf{D}_{\Gamma_O}$ transfers to the correct noise fitting into the assumptions defining the current likelihood (6.8). Hence, a comparison of the two different parameter estimation problems in terms of an absolute quantification of the deviation of the estimates from the ground truth is not possible.

However, the purpose of the comparison was to quantify the susceptibility of the applied measures of similarity with respect to the systematic error contained in the measurements. This can still be achieved since there is a systematic error free measurement $\mathbf{D}_{\Gamma_O}$ and $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$ for each of the measurements $\mathbf{V}_{R,\Gamma_O}$ and $\Gamma_O^R$ respectively. By simply replacing the measurements in (6.6) and (6.8) with their systematic error free counterparts, the resulting estimation problems, or the corresponding estimates, constitute a reference solution for each of the two problems (6.6) and (6.8). The susceptibility of each of the problems can now be measured in terms of the deviation of the solution from the respective reference solution. Due to the relative character of this measure of susceptibility, the comparison is independent of the choice of the noise parameters $\sigma_D$ and $\sigma_N$. In the following, these are set to $\sigma_D = 0.18$ mm and $\sigma_N = 1.0$ mm resulting in similar values for the log-likelihood functions (6.6) and (6.8) such that the effect of the prior assumptions is similar for each of the problems.

To complete the setup of the parameter identification problem, the TV prior (3.93) ($\alpha_{tv} = 0.1$, $\epsilon = 1.0\mathrm{e}{-2}$) is assumed over the parameters. Details on the graph structure defining the adjacency matrix are given in appendix F. In combination with the two different likelihood functions given by (6.6) and (6.8), this constitutes the two different parameter estimation problems to be compared.

## 6.2. Comparison of parameter estimates

The following analysis is carried out by means of the MAP estimates which are computed using the LBFGS algorithm 1 and 2 (convergence tolerance $tol = 1.0\mathrm{e}{-6}$, storage size $m_s = 200$), see chapter 4.4.1. The solutions are parametrized using the element-wise basis (2.105).

### 6.2.1. Point-wise measurements

The MAP estimate corresponding to (6.6) is denoted by $c_{\vartheta,D}$. The associated reference solution, which is obtained by replacing $\mathbf{V}_{R,\Gamma_O}$ with $\mathbf{D}_{\Gamma_O}$, is denoted by $\tilde{c}_{\vartheta,D}$. Both solutions are depicted in figure 6.3.

It can be seen that the reference solution $\tilde{c}_{\vartheta,D}$ clearly shows the 3 different patches from the ground truth, see also figure 6.4 (left). Due to the inverse-crime nature of this reference solution, the deviation from the ground truth is entirely defined by the synthetic noise vector $\boldsymbol{\zeta}$. Since the measurements $\mathbf{V}_{R,\Gamma_O}$ and $\mathbf{D}_{\Gamma_O}$ differ solely by the

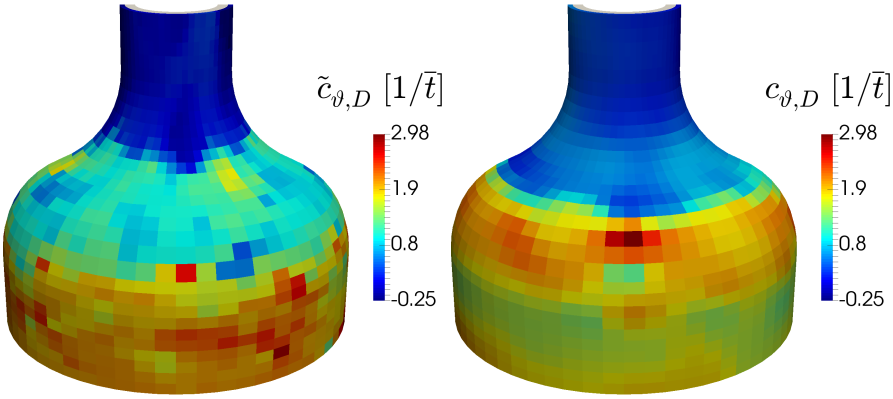

Figure 6.3.: Reference solution $\tilde{c}_{\vartheta,D}$ (left) and solution $c_{\vartheta,D}$ (right). The clear difference between the two is entirely caused by the propagation of the systematic error contained in the measurements $\mathbf{V}_{R,\Gamma_O}$.

systematic error component, the deviation shown by $c_{\vartheta,D}$ is caused exclusively by this systematic error. As can also be seen in figure 6.4 (right), this leads to a drastic deviation from the ground truth. To quantify the influence of the systematic error component, a relative error $\Delta_D$ is computed as

$$\Delta_D = \frac{\|c_{\vartheta,D} - \tilde{c}_{\vartheta,D}\|}{\|\tilde{c}_{\vartheta,D}\|} \approx 0.41. \tag{6.9}$$

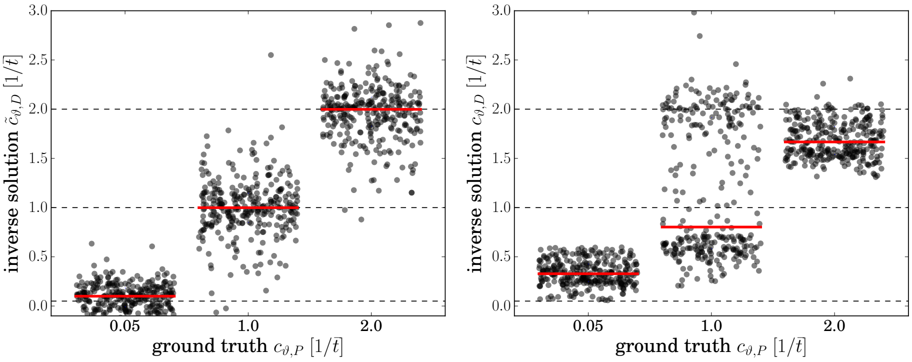

Figure 6.4.: Reference solution $\tilde{c}_{\vartheta,D}$ (left) and solution $c_{\vartheta,D}$ (right) in comparison to the ground truth $c_{\vartheta,P}$. The estimated solutions are grouped according to the ground truth. The group-wise mean is indicated by the red lines. For visualization, random horizontal jitter is applied within each group.

## 6.2.2. Surface measurements

The MAP estimate corresponding to (6.8) is denoted by $c_{\vartheta,W}$. The associated reference solution, which is obtained by replacing $\Gamma_O^R$ with $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$, is denoted by $\tilde{c}_{\vartheta,W}$. Both solutions are depicted in figure 6.5.

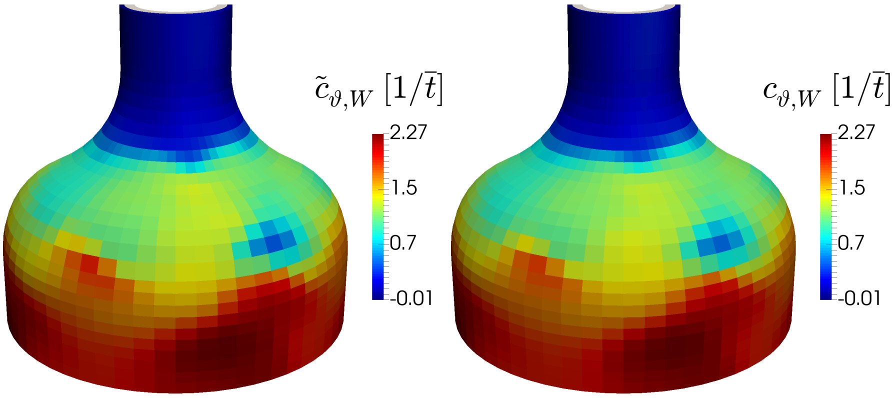

Figure 6.5.: Reference solution $\tilde{c}_{\vartheta,W}$ (left) and solution $c_{\vartheta,W}$ (right). Although visually almost indistinguishable, the difference between the two is entirely caused by the propagation of the systematic error contained in the measurements $\Gamma_O^R$.

Both solutions, which are visually almost indistinguishable, show the patch-wise character of the ground truth, see also figure 6.6. As for the displacement-based solutions, the difference, although small, is exclusively caused by the systematic error in which $\Gamma_O^R$ differs from $\Gamma_O^{\mathbf{A}(c_{\vartheta,P})}$. The propagation of this difference is measured in terms of the relative error $\Delta_W$ as

$$\Delta_W = \frac{\|c_{\vartheta,W} - \tilde{c}_{\vartheta,W}\|}{\|\tilde{c}_{\vartheta,W}\|} \approx 0.02. \tag{6.10}$$

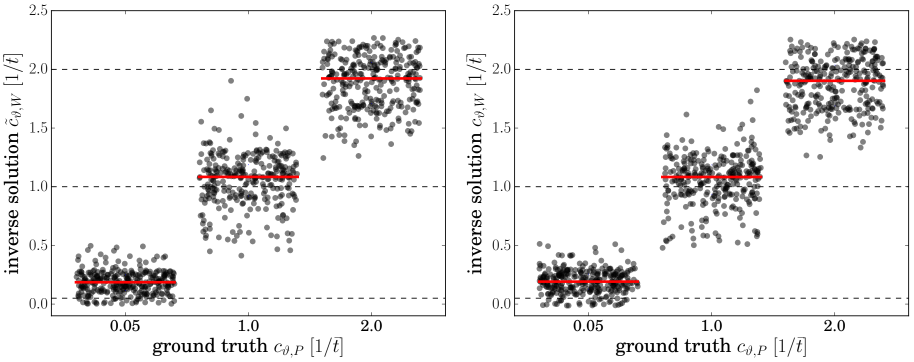

Figure 6.6.: Reference solution $\tilde{c}_{\vartheta,W}$ (left) and solution $c_{\vartheta,W}$ (right) in comparison to the ground truth $c_{\vartheta,P}$. The estimated solutions are grouped according to the ground truth. The group-wise mean is indicated by the red lines. For visualization, random horizontal jitter is applied within each group. Both solutions clearly represent the patch-wise character of the ground-truth.

## 6.3. Discussion

The results presented in chapter 6.2.1 and chapter 6.2.2 indicate that the same systematic error contained in both measurements $\mathbf{V}_{R,\Gamma_O}$ and $\Gamma_O^R$ is propagated differently by the different measures of similarity in (6.6) and (6.8). By measuring this propagation with respect to a similarity-measure-specific reference solution, obtained from measurements free of systematic error, an objective comparison between the susceptibility of the respective measures of similarity is possible. Given the relative errors $\Delta_D$ and $\Delta_W$, a quantitative measure of the error propagation is presented. It is revealed from (6.9) and (6.10) that the error propagation through the displacement-based similarity measure is an order of magnitude larger than the error propagation through the surface current based similarity measure.

On the one hand, this is due to the hard point-wise correspondences in terms of the component-wise associations $\{\mathbf{V}_{R,\Gamma_O}\}_i \leftrightarrow \{\mathbf{D}_{\Gamma_O}\}_i$. This correspondence was already identified as a source of ill-posedness for the classical optimization problem (3.5). The effect shown here is just a manifestation of this ill-posed character, which is reflected in a high sensitivity of the MAP solution with respect to systematic errors. On the other hand, the surface current norm $\|\cdot\|_{\mathcal{W}^*}$ is specifically tailored to be insensitive with respect to in-surface errors by relaxing these point-wise correspondences. Since image registration is well-known to be more accurate in the direction of steep image gradients [see e.g. 162] - eventually segmented as surfaces - the use of the surface current based likelihood (6.8) is advocated in scenarios involving registration.

With respect to patient-specific data and the intention of a probabilistic analysis, the derivation of the likelihood-function (6.8) from the assumption of normally distributed random currents is an issue to be addressed. Computational methods, like the FEM, make heavy demands on the smoothness of model boundaries: to allow for the mesh generation on the one hand and to fit into the modeling assumptions imposed by the computational methods on the other hand. As a consequence, the resulting smooth representations of the model boundaries don't reflect the variability in the available image data. Rather, the whole process of mesh generation is prone to the incorporation of systematic errors. This issue will be discussed in more detail in chapter 7.

# 7. Application to patient-specific data

The enhancement of clinical decision making by computational methods depends on the personalization of these methods on a patient-specific basis. With respect to the prediction of AAA growth by means of a continuum mechanical model as presented in chapter 2, this personalization spans a great variety of different aspects. The most obvious aspect is the use of patient-specific geometries, which are usually directly obtained from medical imaging. Further patient-specific model characteristics are, e.g., given by the surrounding tissue, the loading, stress-free reference configurations with respect to prestress and residual stresses, spatial distributions of material parameters or homeostatic conditions. Whereas all these aspects can play a significant role in the accurate mechanical modeling of AAA growth, it is impossible to a priori assess their influence on the predictive capabilities of a computational model. But *predictive value* is not a quality inherent to a specific model. Rather, it is defined by its purpose. Thus, a model with obvious deficiencies in terms of patient-specific modeling can still provide relevant results. For instance, it has already been shown that merely the use of patient-specific geometries in combination with population averaged material properties can statistically contribute to the prediction of AAA rupture [72, 148]. These studies are based on well-established constitutive equations with available populations parameters [73, 74, 179].

Among others, one reason for the failure of the application of such approaches to AAA growth is the lack of a common mathematical formulation of growth. As a consequence, population averaged parameters for particular growth models have not been established. Such attempts are further complicated by intricate mathematical formulations of some models and the difficulty to measure the involved parameters for a sufficiently large population. Thus, predictive capabilities of theoretical models of arterial growth have not been assessed so far. Nevertheless, it was already shown that they are in principle applicable to patient-specific geometries [128, 230]. In such a setting, also the inference on parameters of growth models for healthy aortas was already attempted. To this end, for a given imaged state subject to prestressed conditions, inverse approaches have been applied to identify parameters of constrained mixture models subject to the condition that the imaged state reflects homeostasis [232]. Exemplary growth of these models, induced by arbitrary manual destruction of the homeostatic balance, has then been shown. Given that in the pathology of AAAs, the imaged state will not picture homeostasis but already the homeostatic breakdown, it is unclear to the author how such approaches can be extended to practical cases.

For the first time, the aspect of a patient-specific prediction of AAA growth was addressed by Tinkl [218]. Based on longitudinal data of a AAA, an inverse approach was pursued with the intention of prediction of future growth. However, the lack of knowledge of the homeostatic conditions revealed issues with respect to the identifiability of parameters of the constrained mixture model. Whereas this issue could be ameliorated by the introduction of assumptions on the transient behavior of the turn-over of material components, it generally renders the state-of-the-art constrained mixture theory a great challenge for inverse approaches.

Furthermore, none of the available approaches towards patient-specific AAA growth accounts for the inherent probabilistic nature of the identification problem. However, for a reliable assessment of the predictive value, the quantification of uncertainty contained in computed predictions can be important. Hence, this chapter describes the application of the approach introduced in chapter 5 to patient-specific data.

As in the previous chapters, thereby the simple growth law (2.65) is applied. With respect to approaches based on the constrained mixture theory, the gain in identifiability of the parameters due to the independence on an explicit homeostatic reference configuration compensates for the loss in accurate physiological modeling. Hence, this approach creates patient-specific calibrated models that can eventually be tested for their predictive capabilities.

## 7.1. Preparation of the data

The data is available in terms of CT data of a 68 year old male patient with a relatively small aneurysm having a maximum diameter of $\approx 3$ cm, see figure 7.1. The resolution of the CT data is 0.79 mm $\times$ 0.79 mm in the transversal plane and 5 mm in out-of-plane direction. In contrast to the impression given in figure 7.1 of a homogeneous resolution, the resolution in the out-of-plane direction is reduced by a factor of $\approx 6.3$. This represents a typical situation in clinical CT imaging of AAAs, where the accuracy in imaging has to be balanced with the duration of the examination and the exposure to radiation [87]. Longitudinal data for this patient is given with a time lag $\bar{t} = 2$ years. During this period, the change in maximum diameter was approximately 0.5 mm, see figure 7.8. Compared to trial data of small AAAs ($3 - 3.5$ cm) with a mean growth rate of $1.65 \pm 2.41$ mm during the first year of a 5 year follow-up period [45, Table 2], the present case can be considered typical, although below average.

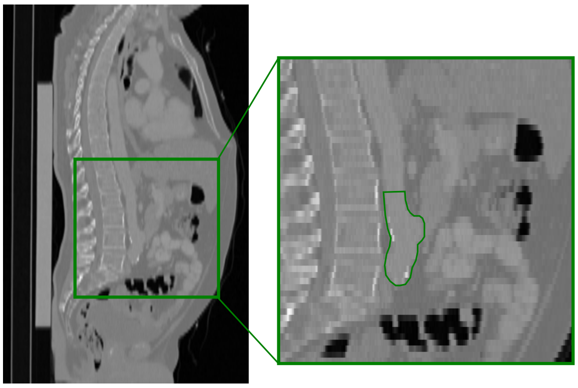

Figure 7.1.: Sagittal view of the CT data[1] of the presented case.

---

[1] Medical images are courtesy of the Institut für diagnostische und interventionelle Radiologie, Klinikum rechts der Isar, Technische Universität München. The use of clinically induced CT-imaging was approved by the ethics commission of the TUM School of medicine; 'M-AAA-rker Studie' (Nr. 1897/07, 27.9.2007, Amendment 19.3.2012).

Based on a segmentation of the image data resulting in the luminal volume and in a thin layer of ILT, smoothed geometric reconstructions are obtained. For a detailed description of this process, the reader is referred to Maier [147]. Subsequently, the resulting ILT volume is meshed with an unstructured mesh of hexahedral elements and a mesh size of ≈ 1mm. The arterial wall, which cannot be detected from the CT data, is then extruded to this mesh. Thereby, a uniform wall thickness of $1mm$ is assumed according to Maier et al. [148], see figure 7.2 (left). The resulting mesh contains 14,856 elements distributed equally between the ILT and the arterial wall. The same meshing procedure is pursued on the longitudinal data so that the respective surfaces dividing ILT and arterial wall can be used in the similarity measure for the parameter identification, see figure 7.2 (right). To account for the misalignment of the patient in the CT scanner between the two snapshots, the segmented surfaces have been mutually aligned with respect to the rigid body modes prior to meshing, see appendix E.

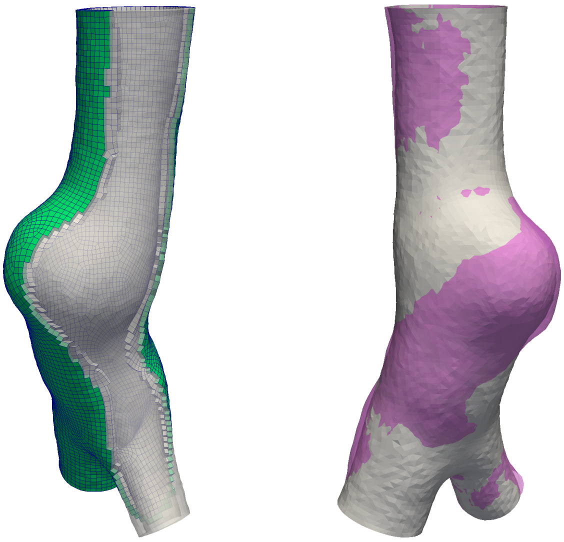

Figure 7.2.: Left: Patient-specific geometry as a mesh with arterial wall (green) and ILT (gray). Right: Surfaces dividing ILT and arterial wall as the entities to be compared in the parameter identification. Model surface Γ (gray) and the longitudinal measurement **Z** (pink)

The constitutive modeling of the elastic behavior of the wall and the ILT is equivalent to that of the synthetic AAA model in the chapters 5 and 6. For the wall, modeled by (5.12) and (5.13), this choice is motivated by the evidence that the used model is well suited to represent diseased arterial tissue [179]. As discussed comprehensively by Maier [147], accurate modeling of the ILT is a topic of ongoing research. It's development via accumulation and deposition of coagulated blood is often modeled by constitutive laws which explicitly account for a layered structure [73]. Since the accurate modeling of ILT is not at the core focus of this thesis and the ILT in the present patient is only existent as a very thin layer, the simple model (6.1) is used here. This model has previously been used in the context of AAA modeling in Maier et al. [148].

The definition of boundary conditions depends on the partial character of the model of the AAA as a segment of the aorta. Thus, the determination of the behavior of the model at the segment interfaces in advance is a problem of circular reasoning. Rather, the deformation at the proximal and distal model surfaces should be inferred from the data as well. In a first attempt towards a predictive simulation setup, this possibility is not exploited and the model is fixed at the inlet and outlet surfaces. Further simplifications are applied to the model by ignoring the effects of surrounding tissue. Although it was shown by Moireau et al. [161] that this information can be obtained for a patient-specific aorta model from motion-MRI data, such data is not available in the present case. A diastolic pressure of 80 mmHg is assumed to act on the luminal surface. As introduced in chapter 2.5, this is motivated by the assumption that the cyclic effects of blood flow play a minor role for the loading conditions with respect to the fixed diastolic point in time that is monitored over the growth period. With the intention of mimicking the diastolic conditions during imaging, the load control scheme shown in figure 5.2 is used.

## 7.2. Parameter estimation

Due to the beneficial properties of the treatment of surface data in the sense of currents described in chapter 6, the parameter identification problem is set up in the context of measurements provided as surfaces.

### 7.2.1. Inverse problem specification

To this end, the model surface $\Gamma$ and the measured surface $\mathbf{Z}$, depicted in figure 7.2, are used in the definition of the likelihood via (3.69) and (3.70) ($\sigma_N^2 = 2.0\ \text{mm}^2$, $\sigma_W^2 = 0.25\ \text{mm}^2$). The kernel scale $\sigma_W$ reflects a user defined assumption over the spatial features to be detected by the norm (3.70). As described in chapter 3.3.2, page 49, this choice is also related to the discretization size of the triangulation of the surfaces and cannot be chosen arbitrarily large due to degeneration of the likelihood function (3.69).

The noise variance $\sigma_N^2$ is chosen according to an expected variation of the surface segmentation. The discretization size of the model surfaces is implied by the accurate resolution of the computational model and the eventual discrete model is the result of a model generation process with the intention of smooth surface representations. This incorporates a great deal of smoothing operations [147, appendix A]. The resulting variation in the normals of the surfaces of the computational model can be very roughly estimated from planar sections of the model to $\sigma_N^2 \approx 0.002\ \text{mm}^2$. However, as already indicated in chapter 3.3.2, page 50, this variation does not reflect the uncertainty incorporated in the image data. Given the low spatial resolution in $z$-direction of 5 mm, a triangulation that accurately represents the *available* data is expected to have a discretization size $> 5$ mm. An accurate determination of the variability of such a triangulation is not straightforward since it has to account for the propagation of the error in the imaging machinery to the segmented surface. With the image resolution between 0.79 mm and 5 mm, the noise level is roughly estimated to lie in the same order of magnitude as the image resolution and is set to $\sigma_N^2 = 2.0\ \text{mm}^2$.

As argued in the introduction to this chapter, the phenomenological growth law (2.65) is used to model growth. Following the argumentation brought forward in chapter 5, the TV prior (3.93) ($\alpha_{tv} = 0.5$, $\epsilon^2 = 1.0\text{e}{-2}$) is assumed over the parameters $c_\vartheta$. Details on the graph structure defining the adjacency matrix are given in appendix F.

### 7.2.2. Numerical solution

To solve the specified inverse problem, the two-stage approach proposed in chapter 5 is applied. It starts from the general representation of growth parameters in terms of the element-wise basis (2.105). Thus, the initial dimension of the inverse problem is $n_p = 7428$ since growth is only considered in the arterial wall.

The computation of the first stage is performed by the LBFGS algorithm described in chapter 4.4.1 (convergence tolerance $tol = 1.0\text{e}{-8}$, storage $m_s = 300$, convergence achieved after 279 iterations). In contrast to a standard application of the LBFGS algorithm, a large storage capacity is chosen to enable the approximation of the Laplace approximation as accurate as possible. The resulting MAP estimate $\hat{c}_{\vartheta,map}$ and the corresponding standard deviation $\mathbb{SD}_q[c_\vartheta]$ from the Laplace approximation are shown in figure 7.3.

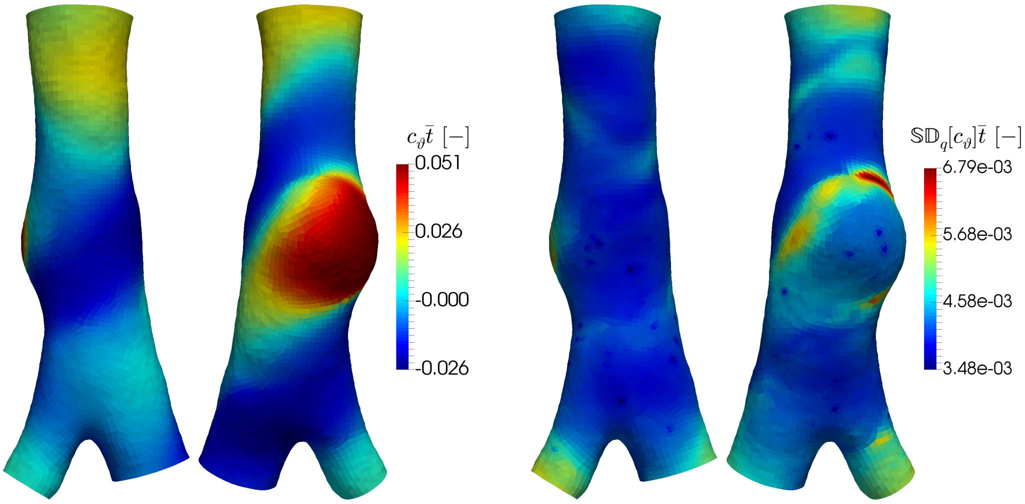

Figure 7.3.: MAP estimate (left) and standard deviation computed from the Laplace approximation (right). The results are visualized as the volumetric growth $\vartheta = c_\vartheta \bar{t}$ for the time period $\bar{t} = 2$ years.

The MAP solution clearly shows the characteristic features of functions of bounded variation, which is a prerequisite for the application of the dimensionality reduction approach presented in chapter 5.2. The core feature of this approach is the determination of a reduced basis from the *significant* eigenvectors of the linearized TV operator (5.5). In the synthetic data example presented in chapter 5.3, a significant set of eigenvectors could be naturally detected from a subset of rapidly decaying eigenvalues, see figure 5.6. For the inverse problem on real patient-specific data presented in this chapter, the eigenspectrum of the first 50 eigenmodes of the linearized TV operator is shown in figure 7.4 (left). This spectrum does not show a distinct set of eigenvalues that could be utilized for the definition of a basis. To define a set of eigenvectors, it is therefore necessary to quantify the approximation quality of the reduced basis. In a predictive model, this approximation quality can be *measured* in terms of the computed prediction. To this end, the standard deviation of the primal solution $\mathbf{u}$ – i.e., the displacements due to growth – under a reduced dimensional representation of the Laplace approximation according to

(5.8) is monitored over the size of the basis. This is motivated by the assumption that the variability in the prediction is essentially influenced by the main covariance structure in the parameters. In fact, it can be seen that the standard deviation of the prediction (approximated with $N = 1680$ particles) converges quickly, see figure 7.4 (right). Based on this convergence, the size of the reduced basis was chosen to $n_r = 30$.

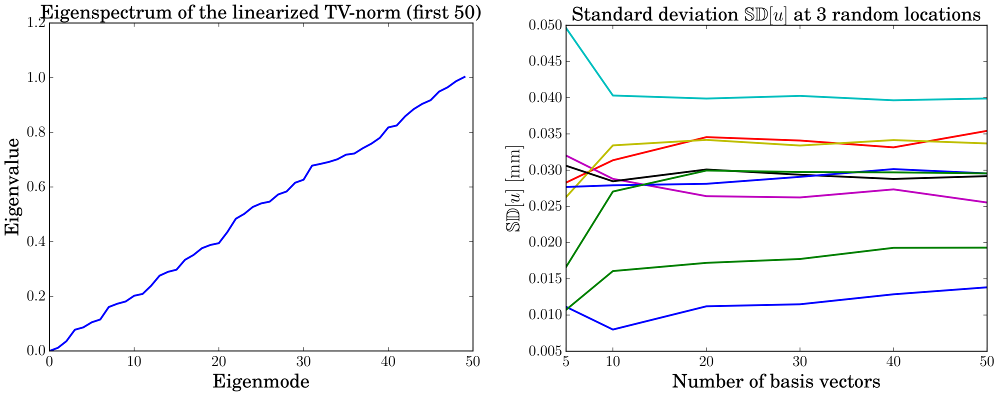

Figure 7.4.: Left: Eigenspectrum of the first 50 eigenvectors corresponding to the smallest eigenvalues of the linearized TV operator. Right: Convergence of the standard deviation of the growth displacements $\mathbf{u}(c_\vartheta)$ under the projected Laplace approximation (5.8). For the purpose of visualization, displacements in $x, y, z$-direction at 3 randomly chosen locations of the model are monitored. $N = 1680$ particles are used for the approximation of the standard deviation.

Given this reasonably designed reduced basis, particle approximations of the projected posterior (5.7) can be computed by means of the SMC algorithm described in chapter 4.4.4. Thereby, the approximation is based on a set of $N = 1680$ particles. This choice results in $\approx 60$ particles per parametric dimension, which showed excellent performance for the examples shown in chapter 5.3. The exact number of particles also has to be chosen with respect to the available parallel layout, see chapter 7.2.3. The projected Laplace approximation is used as initial distribution, which proved to be a feasible approach for the synthetic example in chapter 5.3. The resulting particle approximations of the PM and the posterior standard deviation are shown in figure 7.5. The PM solution is qualitatively similar to the MAP solution. But a change of 15% in terms of the maximum value indicates a significant quantitative difference between the MAP solution and the PM solution. However, the impact of this difference on the *quality of the fit* at the MAP or the PM estimate is negligible. For the MAP solution, the quality of the fit is reported in terms of the projected distance (3.45), see figure 7.6. It can be seen that the fit is accurate in the area of the aneurysm sac, whereas there are a few distinct sites where the initial distance seems to be not sufficiently reduced. The posterior standard deviation shows a characteristic as already indicated by the Laplace approximation, cf. figure 7.3. Whereas the absolute values differ by a factor of 2, it is expected that the underlying particle approximation captures the main covariance structure represented by the posterior as shown in the analysis provided in chapter 5.

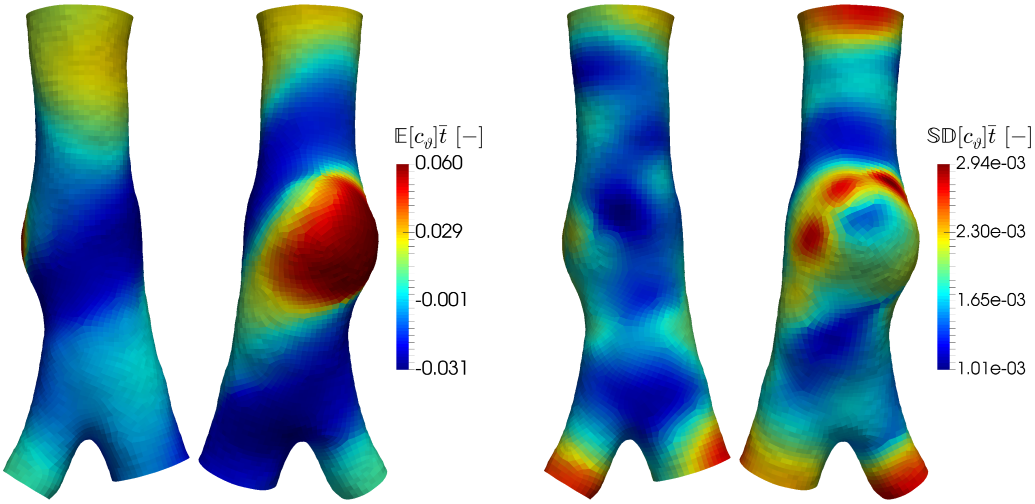

Figure 7.5.: Expected value $\mathbb{E}[c_\vartheta]$ and standard deviation $\mathbb{SD}[c_\vartheta]$ from the reduced dimensional approximation. The results are visualized as the volumetric growth $\vartheta = c_\vartheta \bar{t}$ for the time period $\bar{t} = 2$ years.

### 7.2.3. Computational aspects

The evaluation time of the computational model described in chapter 7.1 is in the order of minutes. As can be seen from the load control scheme shown in figure 5.2, thereby the majority of the computational effort is invested into the prestressing computation. But the modeling of growth does not affect the prestressed state and vice versa. So for repeated evaluations of the model for different sets of growth parameters, the nonlinear solution process can be restarted from the prestressed state. This very simple form of a parameter continuation scheme effectively reduces the computational burden. E.g., on 2 cores of a *Intel® Xeon® Processor E5-2697 v3* processor on the 'Phase 2' partition of SuperMUC[2], the evaluation time takes approximately 30 s. 2 threads per core are assigned for the evaluation of the measure (3.70) using OpenMP through Trilinos' Kokkos package.

Despite this reduction in evaluation time, the sequential evaluation of such a model using MCMC techniques, see chapter 4.4.3, can easily become infeasible. E.g., the overall computation time for 20,000 - 50,000 MCMC steps is in the range of 7 - 17 days. Such time ranges are impractical towards the application in the clinical management routine. The application of the SMC algorithm in combination with the parallel layout shown in figure 4.2 can improve this issue. Therefore, the 1680 particles used for the particle approximation are distributed into 840 parallel groups. These groups are distributed on 60 nodes on the 'Phase 2' partition of SuperMUC. This allows for the evaluation of the particles within a group as described above. The application of the SMC algorithm 4 (see chapter 4.4.4) with the adaptive step-size control (4.86) results in 192 time steps and an overall computation time of $\approx$ 15 h for 322,560 model evaluations. With respect to a sequential evaluation of the same number of model evaluations, this represents a speed-up factor of $\approx$ 180.

---

[2] `https://www.lrz.de/services/compute/supermuc/systemdescription/`

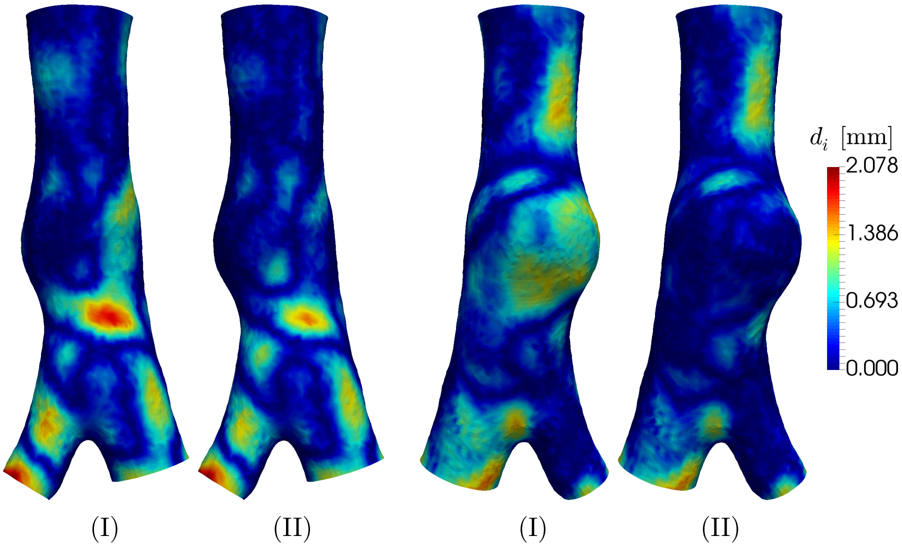

Figure 7.6.: Projected distances $d_i$ according to (3.44) of the initial undeformed computational model surface $\Gamma$ to the measured data $\mathbf{Z}$ (I) and of the fitted deformed model $\varphi(\Gamma)$ to the measured data $\mathbf{Z}$ (II). (Left: posterior perspective; Right: anterior perspective)

## 7.3. Towards prediction of future growth

The main focus of the presented framework is to enable a predictive use of the parametrized computational model of AAA growth. Despite the immense computational effort for the evaluation of descriptive statistics of the posterior distribution over the parameters such as the PM or the posterior standard deviation, these quantities are of minor interest in itself. In particular, this holds for parameters of phenomenological descriptions that cannot be interpreted in terms of explicit physical parameters. Nevertheless, the calibrated model presented in chapter 7.2 allows for an informed extrapolation of future growth. And given a particle approximation of the posterior distribution over the parameters, statistics of this extrapolation as a function $f(c_\vartheta)$ can be readily evaluated by means of (4.54) or (4.58).

In the simplest case, the function $f(c_\vartheta)$ can be chosen to represent the primal variables $\mathbf{u}(c_\vartheta)$ of the computational model after the prediction step. The expected value and standard deviation of a prediction of growth in terms of the deformation $\mathbf{u}$ are shown in figure 7.7 for a prediction period of further 2 years.

In a clinical context, the information content of predictive models and of available data is often condensed to a drastically reduced set of variables. With respect to AAA growth, the prevailing descriptive variable is given by the maximum diameter [28, 207]. In order for computational models to be eventually applied in the clinical management process, it will be important to show their predictive capabilities with respect to the maximum diameter and to validate the results on a sufficiently large cohort of patients. As a step towards this goal, the changes in maximum diameter for the present patient are presented in figure 7.8. The determination of the variability of the predicted maximum diameters is straightforward, and can be evaluated by directly defining the maximum

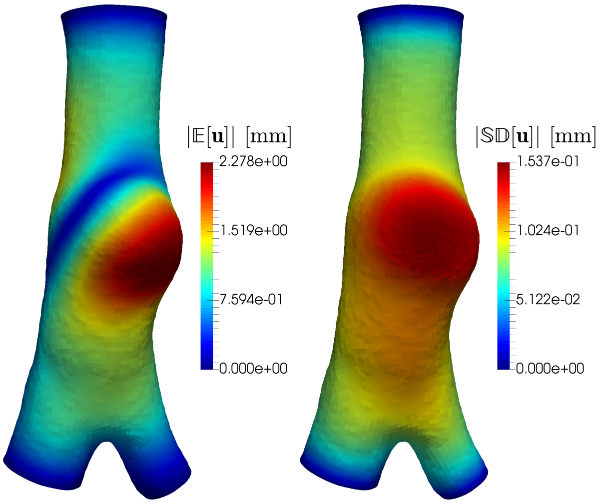

Figure 7.7.: Prediction of growth in terms of model displacements $\mathbf{u}$ after $+2$ years from the time point of the follow-up imaging.

diameter as the function $f(c_\vartheta)$.

## 7.4. Discussion

Given the probabilistic nature of the problem, a general validation of the presented approach cannot be based on the application to a single patient but must be performed with respect to a statistically significant cohort of patients. However, the results shown in the previous chapters already reveal a few characteristics that are to be discussed since they might have to be taken into consideration in a statistical validation of the model.

The MAP solution as well as the PM solution, see figures 7.3 and 7.5, show the characteristic properties of functions of bounded variation. In this sense, they are characterized by localized features such as a clearly distinct site of growth in the area of the aneurysm sac. Whereas this does of course not justify the correctness of the prior assumptions, it is an indication towards their validity. Furthermore, it enables the dimensionality reduction approach presented in chapter 5, and consequently it allows for a probabilistic treatment of the identification problem by means of particle approximations. These approximations enable the computation of statistics of output quantities of the model and accordingly also the predictive use of the model.

Beside the predictive use of the calibrated model, the particle approximation of the posterior can also be used in the context of model selection [121] since it enables the computation of the evidence $p(\mathbf{Z}|\mathbb{M})$ as introduced in (3.32). Using an approach based on Bayes factor [111], the model evidence could be used to formally compare the applied volumetric growth formulation against the constrained mixture formulation. In this view, an indication of the quality of the applied model was already shown in figure 7.6. But this quality of the fit cannot be assumed to be entirely induced by the mathematical formulation of growth. On the one hand, it might result from model deficiencies in terms of surrounding tissue, loading, Dirichlet boundary conditions at the inlet and

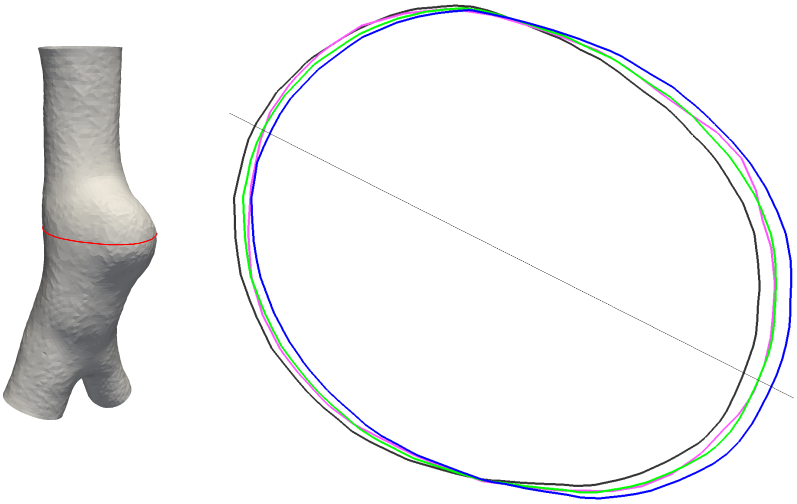

Figure 7.8.: Maximum diameters during growth, located at the aneurysm sac and measured along the indicated direction. Undeformed model (black): 31.9 mm, measurement after 2 years: 32.6 mm (pink), fitted model (green): 32.6 mm and +2 years-prediction (blue): 33.2 mm.

outlets or the lack of patient-specific elastic properties of the tissue. On the other hand, the fine discretized model represents a level of detail not justified by the available image data. It is therefore likely that the local deviations seen in figure 7.6 are artificially created due to a potentially wrong and over-complete representation of the data. Within the current formalism, the discrepancy between the scale of accuracy of the image data and the apparent scale of accuracy of the smooth computational model can in theory be conveniently accounted for by the spatial scale $\sigma_W$ of the the kernel (3.68). But in practice, the choice of $\sigma_W$ is bounded by the size of the surface discretization due to the increasing ill-posedness of the likelihood (3.69) with increasing $\sigma_W$. In this regard, the choice $\sigma_W^2 = 0.25$ mm$^2$ is rather made to account for the well-posedness of the problem than to reflect the expected spatial correlation structure. Thus, it is desirable to decouple the surface representations for the evaluation of the likelihood in terms of currents from the smooth surface representation needed for the computational model. One possible approach to be implemented in future work is to define the surfaces representing the data from a coarsening strategy applied to the surfaces of the computational model. E.g., a coarse triangulation can be created by the assignment of triangles from the fine triangulation to volume cells of a regular grid over a bounding volume. Within a sub volume, a new coarse triangle is then created by averaging all triangle centers and normals. This results in an easy and deterministic functional relation between the coarse and the fine triangulation. The computation of the tangent currents needed for a Lagrangian formulation of the identification problem, see chapter 3.5 and appendix C.1, is therefore not significantly complicated. The loss of connectivity of such a triangulation is not an issue for the current formalism. On the contrary, connectivity of the surface of the AAA model is a projection of a priori knowledge to the representation of the model. This assumption is not necessarily reflected in the raw pixel-wise information provided by the images. Thus, a disconnected triangulation represents an advantage with respect to the representation of variability given by the image data and it fits well into the assumption of Gaussian noise in the space of currents, cf. chapter 3.3.2, page 50.

In the present case, the difference in position of the patient in the CT scanner between the imaging of the initial data and the imaging of follow-up data was taken into account by an alignment of the follow-up model to the computational model with respect to the rigid body modes. As can be seen from figure 7.2, this results in a locally perfect alignment of the two surfaces in certain regions. In particular, the intersection of model and follow-up data lateral to the aneurysm sac is clearly propagated to the prediction of growth, see figure 7.7. The intensity of this effect can be influenced by the spatial scale of the kernel and is therefore also connected to the size of the surface discretization. Thus, the use of coarser surface representations as advocated above will also reduce the influence of a potential misalignment. However, this should not mask the fact that the rigid alignment of the follow-up model is not as deterministic as assumed for the application in the exemplary case studied in this chapter. Rather, it is subject to uncertainties that should be incorporated into the parameter identification problem. Given that the rigid body modes can be parametrized by 6 degrees of freedom (3 translations, 3 rotations), the increase in complexity of the identification problem is only moderate.

As can be seen from figure 7.8, simply measuring the changes in diameter can result in a misleading assessment of growth. In the present patient, the translation of the aneurysm sac is in the same order of magnitude as the enlargement in diameter. Thus, the effects of growth are to a similar extent expressed in a change of turtuosity of the AAA. However, the quantification of this effect should not be attempted without taking the uncertainties of the rigid alignment of the measurement data to the model data into consideration

### 7.4.1. Concluding remarks

This chapter showed the feasibility of the application of state-of-the-art numerical techniques in combination with a newly proposed approach for parameter reduction to models with applicability in a clinical management context. Thereby major components towards model selection and model validation in a clinical trial situation were presented. This incorporates the probabilistic approach to the personalization of patient-specific models and the propagation of the probabilistic nature to important measures used in a clinical situation, like the maximum diameter. This measure defines the current standard for the assessment of AAA growth [28]. Further, it is subject to modeling attempts by means of statistical mixed effect models [207]. Thus, the proposed computational framework has to be validated also with respect to its capabilities regarding the prediction in maximum diameter. But beyond such a contraction of information, the proposed approach presents a much more versatile source of information on the patient-specific progression of AAA growth. The benefit of this information in the clinical management of AAA has to be shown in future work.

# 8. Summary and outlook

The work presented in this thesis addressed the development of computational models of arterial growth towards clinical application. In particular, the modeling of the growth of AAAs was motivated by the need for policies of surveillance intervals in the clinical management of asymptomatic patients. The need for such policies is raised by the onset of screening programs that suggested a beneficial effect on overall mortality rates. However, common policies for surveillance intervals have not been established so far. Due to the possibility of tailoring a computational model to patient-specific scenarios by means of calibration, it is expected that such models can significantly contribute to a patient-specific management in the clinical routine. In this context, the focus of the thesis was directed towards the development of methods and approaches for the calibration and the quantification of the predictive capabilities of computational models based on existing continuum mechanical formulations of arterial growth. The application of a Bayesian point of view enabled the propagation of uncertain input parameters to a probabilistic quantification of the model output. In this setting, the work presented in this thesis was concerned with two major aspects: the effective and sensible treatment of measurement data that is necessary for the calibration process; and an approach for dimensionality reduction of the input space of the model to enable an efficient probabilistic assessment of the predictive qualities of the model.

The need for a specific treatment of the measurement data was motivated by a detailed analysis of the raw measured data. In the context of AAA growth, such data is typically available as longitudinal image data and thus the mapping of the model output to the image data, and vice versa, was identified as a non-trivial operation. It was shown how the formalism of surface currents is able to overcome the involved difficulties in an elegant and efficient way. These properties were demonstrated on a model using synthetic data. Thereby, it could be shown that the use of surface currents in a calibration framework avoids the effects of systematic errors introduced through image registration applied to the raw measured images. Moreover, it was shown how the formalism of surface currents can be integrated seamlessly in a Bayesian parameter identification framework by offering a perspective on surfaces as random variables. Thus, beside its elegant treatment of the relevant data provided by the measured images, the surface current formalism also provides a possibility to characterize the uncertainty in the data. This perspective is of particular importance for the application of the framework to patient-specific data. In principle, the vector space structure of the space of surface currents as a RKHS decoupled the representation of a discrete current from the size of the triangulation of the associated surface. It was shown that a random current is not defined by the size of the triangulation but the spatial scale $\sigma_W$ of the kernel associated to the RKHS. In practice however, the choice of this scale is implicitly bounded by the size of the triangulation due to an increase in ill-posedness of the definition of a probability density on currents with increasing scale $\sigma_W$ (for a fixed size of the triangulation). In chapter 7, it was therefore proposed to decouple the representation of the boundary of the computational model from the representation of the data. An implementation of this approach however is left to future work. In this context, it will also be necessary to investigate the Gaussian assumption imposed on the randomness of the image data represented in terms of surfaces.

For the solution of the Bayesian calibration problem, techniques of advanced sampling for sophisticated nonlinear computational models were described in detail. However, due to the need for accurate spatial representations of the input parameters and the implied high dimension of the associated input space, the plain application of these sampling techniques was rendered highly inefficient. Since this inefficiency is a direct consequence of the curse of dimensionality, a novel approach for dimensionality reduction was proposed that is able to maintain the main covariance structure associated to the high dimensional input space. Given that this covariance structure is reflected in the model output, the proposed approach enabled the predictive use of the model by providing a fully probabilistic quantification of the model output. The quality of the proposed approach was demonstrated numerically on a small-scale example.

Finally, the application of the presented framework to patient-specific data demonstrated the feasibility of the approach in the large-scale regime. With regard to the patient-specific simulation of AAA growth [218, 230, 231, 232], the presented application is the first to demonstrate a fully probabilistic use of a model that is capable to predict AAA growth based on patient-specific longitudinal data. Thereby, as a first step, a simple and purely kinematic formulation for the modeling of arterial growth was applied. Whereas this application can clearly be questioned, the probabilistic formulation of the calibration problem allows for a statistical testing of the applied model. Thus, the quantitative evaluation of the model output in terms of a prediction of growth can be performed. Guided by such an assessment of the predictive quality, also future improvements and developments of the model can be based on a sound argumentation. In this sense, the proposed approach provides a starting point for application-oriented model development and can serve as a reference for more elaborate models of arterial growth. More elaborate models can thereby also encompass the incorporation of additional patient-specific properties of the model (e.g., with respect to boundary conditions) or the use of element technology.

## Perspectives for future work

The work presented in this thesis introduced major steps towards the applicability of models of AAA growth in the clinical practice. Beside the showcase application provided in chapter 7 and an evaluation of the model on a patient-specific basis, the application of the model in the clinical practice requires the evaluation of its predictive capabilities in a statistical sense. I.e., a hypothesis such as ‘the model is capable to predict AAA growth’ must be statistically tested. Moreover, to be applied to statistical testing, a mathematically clear definition of such an hypothesis has to be defined. With the probabilistic character of the model output given in terms of particle approximations, such a definition can, e.g., be defined by means of CIs over the model output. In the context of clinical management of AAAs with the classical measure of growth given by changes in the maximum diameter, the distribution of the model output can also be defined in terms of the maximum diameter. However, as observed by Martufi et al. [154], this measure only provides limited information about AAA expansion. In this light, the formalism of currents can again be utilized to define a measure which quantifies growth of an entire AAA. Moreover, the interpretation of a current as a random variable readily allows for the computation of the distribution over the model output given by a surface current. This distribution can then be used to define a hypothesis which is tested on a significant cohort of patients.

To enable such a statistical testing, data for at least 3 points in time is necessary for a large amount of patients. Whereas the presentation in chapter 7 relied on CT data, the presented approach works equivalently with US that is also a common technology in the screening of AAA. However, the reduced quality of US data will necessarily be reflected in the quality of the prediction of growth. It is also possible that the initial model is given by CT data and longitudinal data is provided by means of US. Generally, the main requirement is that the data allows for a segmentation of the AAA geometry.

Although, a statistical evaluation requires a significant amount of data preparation and the intensive use of computational resources, it will eventually also provide a justification for or against the use of the kinematical description of growth. Furthermore, it creates the already mentioned reference for future model development.

Beside the informed extrapolation of AAA growth, the calibrated model provides a model-based extraction of information from the image data. The information contained in this data is thereby reflected in the calibrated growth parameters and propagated to the prediction of future growth. Compared to the common lossy compression of image data in measurements of changes in maximum diameter, the information content provided by the predictive model is expected to be richer. With respect to the classification of stable and rupture prone aneurysms, it can therefore be expected that such information can provide a more sensible classification than the maximum diameter alone. One possible aspect to investigate in this regard in future work is the information provided by the deformation of the calibrated model. Whereas the feature space constructed by the deformation of the model is clearly of too high dimension, low-dimensional representations of this feature space might capture significant information for a classification of AAAs. Such low-dimensional representations can be conveniently extracted by means of the PCA.

The work presented in this thesis concentrated on the uncertainty associated to the calibration of parameters of models for growth introduced through the observation error. Other sources of parametric uncertainty have been left out of consideration at this stage. However, the modeling of AAA is associated to multiple source of parametric uncertainty: uncertainty in parameters of the applied constitutive relations, uncertainty in the initial model geometry, uncertainty in the wall thickness or uncertainty in the boundary conditions. The comprehensive consideration of all these sources of uncertainty goes beyond the approach presented in this thesis and requires multi-fidelity techniques such as proposed by Biehler et al. [22].

# A. Balance of angular momentum

Using the skew-symmetric matrix $\mathbf{M}_\mathbf{x}$ to represent the cross product $\mathbf{x} \times \bullet$ given by

$$\mathbf{M}_\mathbf{x} := \begin{bmatrix} 0 & -\mathrm{x}_3 & \mathrm{x}_2 \\ \mathrm{x}_3 & 0 & -\mathrm{x}_1 \\ -\mathrm{x}_2 & \mathrm{x}_1 & 0 \end{bmatrix}, \tag{A.1}$$

the rotational equilibrium (2.25) can be written as

$$\int_{\partial\mathcal{I}} \mathbf{M}_\mathbf{x} \cdot (\boldsymbol{\sigma}^\top \cdot \mathbf{n}) + \int_\mathcal{I} (\mathbf{x} \times \mathbf{b}) \, \mathrm{dV} = 0. \tag{A.2}$$

The application of the divergence theorem to each of the 3 components of this balance equation gives

$$\int_\mathcal{I} \mathrm{div}((\mathbf{M}_\mathbf{x}\boldsymbol{\sigma}^\top)^\top) + (\mathbf{x} \times \mathbf{b}) \, \mathrm{dV}. \tag{A.3}$$

By definition of the vector

$$\mathbf{v}_{\boldsymbol{\sigma}} := [\sigma_{23} - \sigma_{32}, \sigma_{31} - \sigma_{13}, \sigma_{12} - \sigma_{21}], \tag{A.4}$$

the expansion of the first term in (A.3) can be written as

$$\begin{aligned} \int_\mathcal{I} \mathrm{div}((\mathbf{M}_\mathbf{x}\boldsymbol{\sigma}^\top)^\top) \, \mathrm{dV} &= \int_\mathcal{I} \mathbf{v}_{\boldsymbol{\sigma}} + \mathbf{M}_\mathbf{x} \cdot \mathrm{div}\boldsymbol{\sigma} \, \mathrm{dV} \\ &= \int_\mathcal{I} \mathbf{v}_{\boldsymbol{\sigma}} + \mathbf{x} \times \mathrm{div}\boldsymbol{\sigma} \, \mathrm{dV}. \end{aligned} \tag{A.5}$$

Thus, the rotational equilibrium is given by

$$\int_\mathcal{I} \mathbf{v}_{\boldsymbol{\sigma}} \, \mathrm{dV} + \int_\mathcal{I} \mathbf{x} \times (\mathrm{div}\boldsymbol{\sigma} + \mathbf{b}) \, \mathrm{dV} = 0. \tag{A.6}$$

With the equilibrium of forces given by (2.24), rotational equilibrium is enforced by

$$\mathbf{v}_{\boldsymbol{\sigma}} = \mathbf{0} \iff \boldsymbol{\sigma} = \boldsymbol{\sigma}^\top. \tag{A.7}$$

# B. Push-forward of a surface current

Given a invertible matrix $\mathbf{M}$, and two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ the following identity holds:

$$(\mathbf{M}\mathbf{a}_1) \times (\mathbf{M}\mathbf{a}_2) = \det(\mathbf{M})\mathbf{M}^{-\top}(\mathbf{a}_1 \times \mathbf{a}_2). \tag{B.1}$$

Given two surfaces $S \in \Omega_0$ and $T \in \Omega_t$ with parametrizations $\phi_1 : U \to S$ and $\phi_2 : U \to T$ according to definition (3.3.1), the integration of a function $f : \mathbb{R}^3 \to \mathbb{R}$ along $T$ is given by

$$\int_T f(\mathbf{x})\,\mathrm{d}\Gamma = \iint_U f(\mathbf{x}(s,t))\|\frac{\partial\phi_2}{\partial s} \times \frac{\partial\phi_2}{\partial t}\|\mathrm{d}s\mathrm{d}t. \tag{B.2}$$

In the case of a surface current, the function to be integrated is given by

$$f(\mathbf{x}) := \mathbf{w}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x}) = \mathbf{w}(\mathbf{x}) \cdot \frac{(\frac{\partial\phi_2}{\partial s} \times \frac{\partial\phi_2}{\partial t})}{\|\frac{\partial\phi_2}{\partial s} \times \frac{\partial\phi_2}{\partial t}\|}. \tag{B.3}$$

Inserting this definition into the integral equation (B.2) results in the current $T(\omega)$ associated to the surface $T$ given by

$$T(\omega) = \int_T \mathbf{w}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})\,\mathrm{d}\Gamma = \iint_U \mathbf{w}(\mathbf{x}(s,t)) \cdot (\frac{\partial\phi_2}{\partial s} \times \frac{\partial\phi_2}{\partial t})\mathrm{d}s\mathrm{d}t. \tag{B.4}$$

Given a diffeomorphism $\varphi : S \mapsto T$, the corresponding push-forward current $\varphi(S)(\omega)$ is given by

$$\begin{aligned}
\varphi(S)(\omega) &:= \iint_U \mathbf{w}(\mathbf{x}(s,t)) \cdot (\frac{\partial\phi_2}{\partial s} \times \frac{\partial\phi_2}{\partial t})\mathrm{d}s\mathrm{d}t && \text{(B.5)}\\
&= \iint_U \mathbf{w}(\mathbf{x}(s,t)) \cdot (\frac{\partial\phi_2}{\partial\phi_1}\frac{\partial\phi_1}{\partial s} \times \frac{\partial\phi_2}{\partial\phi_1}\frac{\partial\phi_1}{\partial t})\mathrm{d}s\mathrm{d}t && \text{(B.6)}\\
&= \iint_U \mathbf{w}(\mathbf{x}(s,t)) \cdot \det(\frac{\partial\phi_2}{\partial\phi_1})\frac{\partial\phi_2}{\partial\phi_1}^{-\top}(\frac{\partial\phi_1}{\partial s} \times \frac{\partial\phi_1}{\partial t})\mathrm{d}s\mathrm{d}t && \text{(B.7)}\\
&= \int_S \mathbf{w}(\varphi(\mathbf{X})) \cdot J\mathbf{F}^{-\top}\mathbf{N}(\mathbf{X})\,\mathrm{d}\Gamma, && \text{(B.8)}
\end{aligned}$$

whereby $\frac{\partial\phi_2}{\partial\phi_1} = \mathbf{F}$ and $\det(\mathbf{F}) = J$ can be identified and in the 2nd step (B.1) was used. By noticing that the integrand $\varphi_*^{-1}(\omega) := \mathbf{w}(\varphi(\mathbf{X})) \cdot J\mathbf{F}^{-\top}\mathbf{N}(\mathbf{X})\,\mathrm{d}\Gamma$ is the pull back of $\omega = \mathbf{w}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})\,\mathrm{d}\Gamma$, the formal relationship

$$\varphi(S)(\omega) = S(\varphi_*^{-1}(\omega)) \tag{B.9}$$

is established.

# C. Linearizations

## C.1. Linearization of the surface current similarity

The formal definition of the Gateaux derivative $\frac{\delta D}{\delta \mathbf{U}}$ via the variation $\delta D(\mathbf{Z}, C(\mathbf{U}))[\delta \mathbf{U}]$ by

$$\langle \frac{\delta D}{\delta \mathbf{U}}, \delta \mathbf{U} \rangle = -\langle \mathbf{Z} - C(\mathbf{U}), \delta C(\mathbf{U})[\delta \mathbf{U}] \rangle_{\mathcal{Z}} \tag{C.1}$$

given in (3.108) introduces the question of how to compute the variational observation $\delta C(\mathbf{U})[\delta \mathbf{U}]$. Since the similarity measure is not an explicit function of the parameters $\theta$, the notation was simplified $\delta_\theta \mathbf{U} \to \delta \mathbf{U}$. In the case of surface currents, i.e., $\mathcal{Z} = \mathcal{W}^*$, the observation operator represents the projection from the primal solution $\mathbf{U}$ to the space of currents. Thus, for a particular model surface $\Gamma$, the observation is formally defined by (3.50) resulting in

$$C(\mathbf{U}) := \varphi(\Gamma)(\omega). \tag{C.2}$$

The functional dependency of this push-forward current on the deformation $\varphi(\mathbf{X}) = \mathbf{X} + \mathbf{U}(\mathbf{X})$ is given by (B.8). However, the computation of the variation $\delta C(\mathbf{U})[\delta \mathbf{U}]$ then involves the variation of the determinant $\delta J(U)[\delta \mathbf{U}]$ as well as the variation of the inverse transpose deformation gradient $\delta \mathbf{F}^{-\top}(\mathbf{U})[\delta \mathbf{U}]$, which results in intricate computations. To circumvent these computations, the dependency of the deformation is already contained in the differential structure defining the current of the deformed surface $\varphi(\Gamma)(\omega)$ as given by (B.5). In particular, this dependency is given by the tangent vectors $\frac{\partial \phi_2(\mathbf{U})}{\partial t}$ and $\frac{\partial \phi_2(\mathbf{U})}{\partial s}$ via $\phi_2 = \varphi(\phi_1) = \phi_1 + \mathbf{U}$. This results in the variation of the unnormalized normal according to

$$\begin{aligned} \delta \tilde{\mathbf{n}}(\mathbf{U})[\delta \mathbf{U}] &= \frac{\mathrm{d}}{\mathrm{d}\epsilon} \left( \frac{\partial \phi_1 + \mathbf{U} + \epsilon \delta \mathbf{U}}{\partial s} \times \frac{\partial \phi_1 + \mathbf{U} + \epsilon \delta \mathbf{U}}{\partial t} \right) \bigg|_{\epsilon = 0} \\ &= \left( \frac{\partial \delta \mathbf{U}}{\partial s} \times \frac{\partial \phi_2}{\partial t} \right) + \left( \frac{\partial \phi_2}{\partial s} \times \frac{\partial \delta \mathbf{U}}{\partial t} \right) =: \frac{\partial \tilde{\mathbf{n}}}{\partial \mathbf{U}} \delta \mathbf{U}, \end{aligned} \tag{C.3}$$

and the variation of the test functions $\mathbf{w}(\mathbf{x}) = \mathbf{w}(\varphi(\mathbf{X})) = \mathbf{w}(\mathbf{X} + \mathbf{U})$ according to

$$\delta \mathbf{w}(\mathbf{U})[\delta \mathbf{U}] = \nabla_\varphi \mathbf{w}(\mathbf{x}) \delta \mathbf{U}. \tag{C.4}$$

Thereby $\nabla_\varphi = \frac{\partial}{\partial \mathbf{x}}$ is used to highlight the differentiation with respect to the deformed coordinates $\mathbf{x} = \varphi(\mathbf{X})$. Although unnecessary for the application to the test functions $\mathbf{w}$, it is important for more complex dependencies on the deformation. To allow for the variation of the push-forward current $C(\mathbf{U})$, it is seen in the sense of the Dirac currents $\delta_x^{\mathbf{n}}$, cf. (3.59). This enables the formal application of the directional derivative

$$\Delta \delta_x^{\mathbf{n}}(\mathbf{U})[\delta \mathbf{U}] = \nabla_\varphi \mathbf{w}(x) \delta \mathbf{U} \cdot \tilde{\mathbf{n}}(x) + \mathbf{w}(x) \cdot \frac{\partial \tilde{\mathbf{n}}(x)}{\partial \mathbf{U}} \delta \mathbf{U} \tag{C.5}$$

whereby again the symbol $\Delta$ was used to avoid the notation $\delta\delta_x^{\mathbf{n}}$ for the variation of the Dirac current. Denoting the dual representation of the differential current $\delta C(\mathbf{U})[\delta\mathbf{U}] = \sum_x \Delta\delta_x^{\mathbf{n}}$ with $\boldsymbol{K}_{\mathbf{n}}^{\delta C}$, this allows for the definition of the norm

$$\begin{aligned}\|\delta C(\mathbf{U})[\delta\mathbf{U}]\|^2_{\mathcal{W}^*} &= \langle \boldsymbol{K}_{\mathbf{n}}^{\delta C}, \boldsymbol{K}_{\mathbf{n}}^{\delta C}\rangle_{\mathcal{W}} \\ &= \sum_{x\in C(\mathbf{U})} \nabla_\varphi \boldsymbol{K}_{\mathbf{n}}^{\delta C}(x)\delta\mathbf{U}\cdot\tilde{\mathbf{n}}(x) + \boldsymbol{K}_{\mathbf{n}}^{\delta C}(x)\cdot\frac{\partial\tilde{\mathbf{n}}(x)}{\partial\mathbf{U}}\delta\mathbf{U}.\end{aligned} \tag{C.6}$$

To identify the dual representation $\boldsymbol{K}_{\mathbf{n}}^{\delta C}$, the reproducing property (3.54) is applied to the last term in (C.6) resulting in

$$\begin{aligned}\boldsymbol{K}_{\mathbf{n}}^{\delta C}(x)\cdot\frac{\partial\tilde{\mathbf{n}}}{\partial\mathbf{U}}\delta\mathbf{U} &= \langle \boldsymbol{K}_{\mathbf{n}}^{\delta C}(x), K_x\frac{\partial\tilde{\mathbf{n}}}{\partial\mathbf{U}}\delta\mathbf{U}\rangle_{\mathcal{W}} \\ &= \sum_x \nabla_\varphi(K_x\frac{\partial\tilde{\mathbf{n}}}{\partial\mathbf{U}}\delta\mathbf{U})\delta\mathbf{U}\cdot\mathbf{n}(x) + K_x\frac{\partial\tilde{\mathbf{n}}}{\partial\mathbf{U}}\delta\mathbf{U}\cdot\frac{\partial\tilde{\mathbf{n}}(x)}{\partial\mathbf{U}}\delta\mathbf{U}. \\ &= \left(\sum_x \nabla_\varphi K_x\delta\mathbf{U}\cdot\tilde{\mathbf{n}}(x) + K_x\frac{\partial\tilde{\mathbf{n}}(x)}{\partial\mathbf{U}}\delta\mathbf{U}\right)\cdot\frac{\partial\tilde{\mathbf{n}}}{\partial\mathbf{U}}\delta\mathbf{U}.\end{aligned} \tag{C.7}$$

$K_x$ is thereby used to indicate the matrix kernel $K_x\mathbf{I}$. The dual representation $\boldsymbol{K}_{\mathbf{n}}^{\delta C}$ can thus be identified with

$$\boldsymbol{K}_{\mathbf{n}}^{\delta C} = \sum_x \nabla_\varphi K_x\delta\mathbf{U}\cdot\tilde{\mathbf{n}}(x) + K_x\frac{\partial\tilde{\mathbf{n}}(x)}{\partial\mathbf{U}}\delta\mathbf{U}. \tag{C.8}$$

Due to the definition of the representer $K_x(y) = k(x,y)$, see Theorem (3.3.1), the application $\nabla_\varphi K_x$ is not necessarily the plain spatial gradient of $K_x(y)$ motivating the notation $\nabla_\varphi$. Denoting the dual representation of the current $\mathbf{Z}(\omega)$ by $\boldsymbol{K}_{\mathbf{n}}^{\mathbf{z}}$ and the dual representation of the current $C(\mathbf{U})$ by $\boldsymbol{K}_{\mathbf{n}}^{C}$, (C.1) is given by

$$\langle\frac{\delta D}{\delta\mathbf{U}},\delta\mathbf{U}\rangle = -\langle\boldsymbol{K}_{\mathbf{n}}^{\mathbf{z}},\boldsymbol{K}_{\mathbf{n}}^{\delta C}\rangle_{\mathcal{W}} + \langle\boldsymbol{K}_{\mathbf{n}}^{C},\boldsymbol{K}_{\mathbf{n}}^{\delta C}\rangle_{\mathcal{W}}. \tag{C.9}$$

Inserting $\boldsymbol{K}_{\mathbf{n}}^{\delta C}$ into the definition (3.48) results in

$$\begin{aligned}\langle\frac{\delta D}{\delta\mathbf{U}},\delta\mathbf{U}\rangle = &-\sum_{y\in\mathbf{Z}}\tilde{\mathbf{n}}^{\mathbf{z}}(y)\cdot\sum_{x\in C(\mathbf{U})}\nabla_\varphi K_x(y)\delta\mathbf{U}(x)\cdot\tilde{\mathbf{n}}^C(x) \\ &-\sum_{y\in\mathbf{Z}}\tilde{\mathbf{n}}^{\mathbf{z}}(y)\cdot\sum_{x\in C(\mathbf{U})}K_x(y)\frac{\partial\tilde{\mathbf{n}}^C(x)}{\partial\mathbf{U}}\delta\mathbf{U}(x) \\ &+\sum_{y\in C(\mathbf{U})}\tilde{\mathbf{n}}^C(y)\cdot\sum_{x\in C(\mathbf{U})}\nabla_\varphi K_x(y)\delta\mathbf{U}(x)\cdot\tilde{\mathbf{n}}^C(x) \\ &+\sum_{y\in C(\mathbf{U})}\tilde{\mathbf{n}}^C(y)\cdot\sum_{x\in C(\mathbf{U})}K_x(y)\frac{\partial\tilde{\mathbf{n}}^C(x)}{\partial\mathbf{U}}\delta\mathbf{U}(x).\end{aligned} \tag{C.10}$$

For this variation to produce the exact gradient with respect to the discretized norm (3.70), the same triangulations should be used for the computation of the norm as well as for computation of the gradient. Furthermore, since the adjoint equation (3.107) is solved based on the discretization of the nonlinear problem (2.101), the right hand side for the adjoint equation is given by the gradient $\frac{\delta D}{\delta \mathbf{D}}$. Using the approximations introduced with (3.70) in combination with the approximation $\mathbf{U} \approx \mathbf{U}_h$ according to (2.74) results in the approximation

$$\begin{aligned}
\langle \frac{\delta D}{\delta \mathbf{U}}, \delta \mathbf{U} \rangle \approx{}& \langle \frac{\delta D}{\delta \mathbf{D}}, \delta \mathbf{D} \rangle \\
\approx{}& \sum_{\boldsymbol{\tau}_i^Z} \sum_{\boldsymbol{\tau}_j^\Gamma} \tilde{\mathbf{n}}_i^Z \cdot \frac{\mathrm{d}k(\mathbf{c}_i^Z, \mathbf{c}_j^\Gamma)}{\mathrm{d}\mathbf{c}_j^\Gamma} \frac{\mathrm{d}\mathbf{c}_j^\Gamma}{\mathrm{d}\mathbf{D}} \delta \mathbf{D}\, \tilde{\mathbf{n}}_j^\Gamma \\
&+ \sum_{\boldsymbol{\tau}_i^Z} \sum_{\boldsymbol{\tau}_j^\Gamma} \tilde{\mathbf{n}}_i^Z \cdot k(\mathbf{c}_i^Z, \mathbf{c}_j^\Gamma) \frac{\mathrm{d}\tilde{\mathbf{n}}_j^\Gamma}{\mathrm{d}\mathbf{D}} \delta \mathbf{D} \\
&- \sum_{\boldsymbol{\tau}_i^\Gamma} \sum_{\boldsymbol{\tau}_j^\Gamma} \tilde{\mathbf{n}}_i^\Gamma \cdot \frac{\mathrm{d}k(\mathbf{c}_i^\Gamma, \mathbf{c}_j^\Gamma)}{\mathrm{d}\mathbf{c}_j^\Gamma} \frac{\mathrm{d}\mathbf{c}_j^\Gamma}{\mathrm{d}\mathbf{D}} \delta \mathbf{D}\, \tilde{\mathbf{n}}_j^\Gamma \\
&- \sum_{\boldsymbol{\tau}_i^\Gamma} \sum_{\boldsymbol{\tau}_j^\Gamma} \tilde{\mathbf{n}}_i^\Gamma \cdot \frac{\mathrm{d}k(\mathbf{c}_i^\Gamma, \mathbf{c}_j^\Gamma)}{\mathrm{d}\mathbf{c}_i^\Gamma} \frac{\mathrm{d}\mathbf{c}_i^\Gamma}{\mathrm{d}\mathbf{D}} \delta \mathbf{D}\, \tilde{\mathbf{n}}_j^\Gamma \\
&- \sum_{\boldsymbol{\tau}_i^\Gamma} \sum_{\boldsymbol{\tau}_j^\Gamma} \tilde{\mathbf{n}}_i^\Gamma \cdot k(\mathbf{c}_i^\Gamma, \mathbf{c}_j^\Gamma) \frac{\mathrm{d}\tilde{\mathbf{n}}_j^\Gamma}{\mathrm{d}\mathbf{D}} \delta \mathbf{D}.
\end{aligned} \tag{C.11}$$

The variation $\delta \mathbf{D}$ is a constant vector and can thus be pulled out of the summations within every of the five terms. The gradient $\frac{\mathrm{d}D}{\mathrm{d}\mathbf{D}} = \frac{\delta D}{\delta \mathbf{D}}$ is therefore directly identified from the definition $\langle \frac{\delta D}{\delta \mathbf{D}}, \delta \mathbf{D} \rangle$. The exactness of this gradient with respect to the norm (3.70) can be checked by direct differentiation of (3.70).

## C.2. Explicit differentiation of the weak form

### C.2.1. Partial differentiation with respect to the displacements

The variation of the weak from (2.100) with respect to the displacement $\mathbf{U}$ is given by (2.91), (2.92) and (2.93). In the following these terms are specified for a parametrization given by the growth stretch $\vartheta$. For a growth law such as (2.62), it is important to highlight that the dependency on the displacement in the parametrized weak form is actually given by

$$\delta W(\mathbf{U}, \delta \mathbf{U}, \vartheta) = \delta W(\mathbf{U}, \delta \mathbf{U}, \vartheta(\mathbf{U})). \tag{C.12}$$

The dependency $\vartheta(\mathbf{U})$ is thereby made explicit through a time discretization of the growth law given by

$$\vartheta_n = \vartheta_{n-1} + \Delta t f(\vartheta, \mathbf{C}_e, \mathbf{S}_e). \tag{C.13}$$

Thereby, the choice for an explicit or an implicit scheme is intentionally left open by not specifying time indices for the arguments to the function $f$. Given that $\mathbf{S}_e = 2\frac{\partial \Psi_e}{\partial \mathbf{C}_e}$, the growth stretch $\vartheta$ can be entirely defined as a function $f_\vartheta(\mathbf{C}_e)$ that represents the solution of (C.13):

$$\vartheta = f_\vartheta(\mathbf{C}_e). \tag{C.14}$$

The variation of the variational Green-Lagrange strain $\Delta\delta\mathbf{E}[\delta\mathbf{U}]$ is already entirely specified by (2.91) and the variation of the traction $\mathbf{t}_0$ has no dependency on $\vartheta(\mathbf{U})$. However, the variation of the stresses $\delta\mathbf{S}[\Delta\mathbf{U}]$ is affected by the growth and therefore its computation deviates from a classical hyperelastic formulation. This is seen from the expansion

$$\mathbf{S}(\mathbf{E}(\vartheta),\vartheta) = \frac{1}{\vartheta^2}\mathbf{S}_e(\mathbf{E}_e(\vartheta)), \tag{C.15}$$

whereby $\vartheta$ is itself still a function of $\mathbf{E}$ through (C.14). From this follows the gradient

$$\frac{\partial\mathbf{S}}{\partial\mathbf{E}} = \left(-\frac{2}{\vartheta^3}\mathbf{S}_e + \frac{1}{\vartheta^2}\frac{\partial\mathbf{S}_e}{\partial\mathbf{E}_e} : \frac{\partial\mathbf{E}_e}{\partial\vartheta}\right)\frac{\mathrm{d}\vartheta}{\mathrm{d}\mathbf{E}} + \frac{1}{\vartheta^2}\frac{\partial\mathbf{S}_e}{\partial\mathbf{E}_e} : \frac{\partial\mathbf{E}_e}{\partial\mathbf{E}}. \tag{C.16}$$

With the elastic Green-Lagrange strain given by

$$\mathbf{E}_e(\vartheta) = \frac{1}{2}(\frac{1}{\vartheta^2}\mathbf{C} - \mathbf{I}), \tag{C.17}$$

the partial derivative $\frac{\partial\mathbf{E}_e}{\partial\vartheta}$ is given as $\frac{\partial\mathbf{E}_e}{\partial\vartheta} = -\frac{1}{\vartheta^3}\mathbf{C}$ and $\frac{\partial\mathbf{E}_e}{\partial\mathbf{E}} = \frac{1}{\vartheta^2}\mathbf{I}$. This finally results in the expression

$$\frac{\partial\mathbf{S}}{\partial\mathbf{E}} = -\frac{1}{\vartheta}\left(2\mathbf{S} + \frac{1}{\vartheta^4}\mathcal{C}_e : \mathbf{C}\right)\frac{\mathrm{d}\vartheta}{\mathrm{d}\mathbf{E}} + \frac{1}{\vartheta^4}\mathcal{C}_e. \tag{C.18}$$

The specific implementation of a growth law is thereby only reflected through the term $\frac{\mathrm{d}\vartheta}{\mathrm{d}\mathbf{E}} = 2\frac{\mathrm{d}\vartheta}{\mathrm{d}\mathbf{C}}$. For the simple growth law (2.65), the first term in (C.18) vanishes since $\frac{\mathrm{d}\vartheta}{\mathrm{d}\mathbf{E}} = 0$. As a result, since $\frac{1}{\vartheta^4}\mathcal{C}_e = \mathcal{C}$, the standard term from hyperelasticity is obtained.

### C.2.2. Partial differentiation with respect to parameters of volumetric growth

The partial variation of the weak form (2.100) with respect to the parametrization $\theta$ is given by (3.112) and (3.113). In the following these terms are specified for a parametrization given by the growth stretch $\vartheta$. The displacements $\mathbf{U}$ are not explicit functions of the growth stretch $\vartheta$ since the relation $\mathbf{U}(\theta)$ is only given implicitly through the solution of the nonlinear problem. Thus, the derived kinematical quantities are constant under a partial variation $\delta\vartheta$, $\mathbf{F} = \mathit{const}$, $\mathbf{C} = \mathit{const}$ and $\mathbf{E} = \mathit{const}$. Consequently, it follows directly from (2.18) that

$$\Delta_\theta\delta\mathbf{E}[\delta\theta] = 0. \tag{C.19}$$

However, the corresponding elastic kinematic quantities exhibit an explicit dependency on the growth stretch through the definition of isotropic volumetric growth (2.55) resulting in the elastic Green-Lagrange strains

$$\mathbf{E}_e(\vartheta) = \frac{1}{2}(\frac{1}{\vartheta^2}\mathbf{C} - \mathbf{I}). \tag{C.20}$$

The stress $\mathbf{S}$ as a function of the growth stretch $\vartheta$ is therefore given by

$$\mathbf{S}(\mathbf{E}) = \frac{1}{\vartheta^2}\mathbf{S}_e(\mathbf{E}_e) = \frac{1}{\vartheta^2}\mathbf{S}_e(\mathbf{E}_e(\vartheta)) \tag{C.21}$$

and the variation $\delta_\vartheta \mathbf{S}[\delta\vartheta]$ is computed from the derivative

$$\begin{aligned} \frac{\partial \mathbf{S}}{\partial \vartheta} &= -\frac{2}{\vartheta^3}\mathbf{S}_e(\mathbf{E}_e) + \frac{1}{\vartheta^2}\frac{\partial \mathbf{S}_e}{\partial \mathbf{E}_e} : \frac{\partial \mathbf{E}_e}{\partial \vartheta} = \\ &= -\frac{1}{\vartheta}\left(2\mathbf{S} + \frac{1}{\vartheta^4}\mathcal{C}_e : \mathbf{C}\right). \end{aligned} \tag{C.22}$$

The differentiation with respect to parameters of some particular growth law is then simply given by the application of the chain rule.

# D. Pseudo inverse and singular value decomposition

This chapter summarizes the necessary results from matrix algebra and properties of the singular value decomposition (SVD) that allow to identify the smoothness prior (3.83) as a Gaussian probability density. For references with respect to matrix algebra, the reader is referred to classical literature such as Golub and Van Loan [88] or Gentle [79].

A *unitary* matrix is a square matrix whose transpose is also its inverse. It is defined by

$$\mathbf{U}\mathbf{U}^\top = \mathbf{I}. \tag{D.1}$$

For a $n \times m$-matrix $\mathbf{M}$, the SVD is given by

$$\mathbf{M} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{V}^\top \tag{D.2}$$

with the unitary $n \times n$-matrix $\mathbf{U}$, the unitary $m \times m$-matrix $\mathbf{V}$ and the diagonal $n \times m$-matrix $\boldsymbol{\Lambda}$. The diagonal entries of $\boldsymbol{\Lambda}$ are the so called *singular values* of $\mathbf{M}$. Given this, it follows directly that

$$\mathbf{M}\mathbf{M}^\top = \mathbf{U}\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top\mathbf{U}^\top, \tag{D.3}$$

$$\mathbf{M}^\top\mathbf{M} = \mathbf{V}\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda}\mathbf{V}^\top. \tag{D.4}$$

The columns of $\mathbf{U}$, the so called *left-singular* vectors, form an orthonormal set of the eigenvectors of $\mathbf{M}\mathbf{M}^\top$ and the columns of $\mathbf{V}$, the so called *right-singular* vectors, form an orthonormal set of the eigenvectors of $\mathbf{M}^\top\mathbf{M}$. The diagonal elements of $\boldsymbol{\Lambda}$ are thus identified with the square roots of the eigenvalues of $\mathbf{M}\mathbf{M}^\top$ and $\mathbf{M}^\top\mathbf{M}$, respectively. If $\operatorname{rank}(\mathbf{M}) = p < \min(m, n)$ (i.e., there a diagonal elements in $\boldsymbol{\Lambda}$ that are zero) the SVD can be written more compactly as

$$\mathbf{M} = \mathbf{U}_r\boldsymbol{\Lambda}_r\mathbf{V}_r^\top \tag{D.5}$$

with the $n \times p$-matrix $\mathbf{U}_r$ and the $m \times p$-matrix $\mathbf{V}_r$ being comprised of the singular vectors associated to non-zero singular values. The $p \times p$-matrix $\boldsymbol{\Lambda}_r$ is comprised of the diagonal of non-zero singular values.

The so called *pseudo-inverse* $\mathbf{M}^+$ is given in terms of the SVD according to

$$\mathbf{M}^+ = \mathbf{V}\boldsymbol{\Lambda}^+\mathbf{U}^\top \tag{D.6}$$

whereby the pseudo-inverse $\boldsymbol{\Lambda}^+$ is obtained by inverting the diagonal elements and taking the transpose. The pseudo-inverse of a product $(\mathbf{A}\mathbf{B})^+$, with $\mathbf{B} = \mathbf{A}^\top$, is given as

$$(\mathbf{A}\mathbf{B})^+ = \mathbf{B}^+\mathbf{A}^+. \tag{D.7}$$

And the pseudo inverse of an invertible matrix $\mathbf{A}$ is identical to the inverse

$$\mathbf{A}^+ = \mathbf{A}^{-1}. \tag{D.8}$$

In case $\mathbf{M}$ is a symmetric $n \times n$ matrix given by $\mathbf{M} = \mathbf{Q}\mathbf{Q}^\top$ and $\mathbf{Q}$ is a $n \times k$-matrix with $\mathbf{Q} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$, the pseudo inverse $\mathbf{M}^+$ is computed via

$$\begin{aligned}\mathbf{M}^+ = \left(\mathbf{Q}\mathbf{Q}^\top\right)^+ &= \left(\mathbf{V}\mathbf{S}^\top\mathbf{U}^\top\right)^+ \mathbf{V}\mathbf{S}\mathbf{U}^\top \\ &= \mathbf{U}\left(\mathbf{S}\mathbf{S}^\top\right)^+ \mathbf{U}^\top. \end{aligned} \tag{D.9}$$

In the special case where the columns of $\mathbf{Q}$ are given by a set of $k$ orthonormal basis vectors $\mathbf{q}_i$, the pseudo inverse $\mathbf{M}^+$ reduces to

$$\mathbf{M}^+ = \mathbf{Q}\mathbf{Q}^\top. \tag{D.10}$$

## D.1. Singular value decomposition of $\mathrm{M}^\top\mathrm{M} + \mathrm{Q}\mathrm{Q}^\top$

Let $\mathbf{M}$ be a $n \times n$-matrix and let $\mathbf{Q}$ be a full-rank $n \times k$-matrix. And let the respective SVDs be given as

$$\mathbf{M} = \mathbf{U}_M\mathbf{S}_M\mathbf{V}_M^\top \tag{D.11}$$

$$\mathbf{Q} = \mathbf{U}_Q\mathbf{S}_Q\mathbf{V}_Q^\top. \tag{D.12}$$

Then the sum $\mathbf{M}^\top\mathbf{M} + \mathbf{Q}\mathbf{Q}^\top$ can be written in a block-matrix version as

$$\mathbf{A} := \mathbf{M}^\top\mathbf{M} + \mathbf{Q}\mathbf{Q}^\top = \left[\mathbf{V}_M, \mathbf{U}_Q\right] \begin{bmatrix} \mathbf{S}_M^\top\mathbf{S}_M & 0 \\ 0 & \mathbf{S}_Q\mathbf{S}_Q^\top \end{bmatrix} \left[\mathbf{V}_M, \mathbf{U}_Q\right]^\top \tag{D.13}$$

$$= \left[\mathbf{V}_M, \mathbf{U}_Q\right] \begin{bmatrix} \tilde{\mathbf{S}}_M & 0 \\ 0 & \tilde{\mathbf{S}}_Q \end{bmatrix} \left[\mathbf{V}_M, \mathbf{U}_Q\right]^\top. \tag{D.14}$$

This is generally **not** a SVD since the matrix $[\mathbf{V}_M, \mathbf{U}_Q]$ is not unitary. In the special case $k < n$, $\operatorname{rank}(\mathbf{M}) = n - k$ and with the constraint

$$\mathbf{V}_{M,r} \perp \mathbf{U}_{Q,r}, \tag{D.15}$$

the matrices $\mathbf{S}_M$, $\mathbf{S}_Q$, $\mathbf{V}_M$ and $\mathbf{U}_Q$ can be replaced by their compact representations $\mathbf{S}_{M,r}$, $\mathbf{S}_{Q,r}$, $\mathbf{U}_{Q,r}$ and $\mathbf{V}_{M,r}$, according to (D.5). The resulting matrix $[\mathbf{V}_{M,r}, \mathbf{U}_{Q,r}]$ is then square. Let $[\mathbf{V}_{M,r}, \mathbf{U}_{Q,r}]$ be given in terms of its columns as

$$[\mathbf{V}_{M,r}, \mathbf{U}_{Q,r}] = [\mathbf{v}_1, ..., \mathbf{v}_{n-k}, \mathbf{v}_{n-k+1}, ..., \mathbf{v}_n], \tag{D.16}$$

it can be seen that $[\mathbf{V}_{M,r}, \mathbf{U}_{Q,r}]$ is also unitary since

$$\mathbf{v}_i^\top\mathbf{v}_j = \begin{cases} 1 & \text{if}\, i = j, \\ 0 & \text{if}\, i \neq j. \end{cases} \tag{D.17}$$

Thus, $\mathbf{A}$ admits an inverse $\mathbf{A}^{-1}$ given in terms of the eigendecomposition

$$\mathbf{A}^{-1} = \tilde{\mathbf{U}}\tilde{\mathbf{\Lambda}}^{-1}\tilde{\mathbf{U}}^{-1}, \tag{D.18}$$

with the identifications

$$\tilde{\mathbf{U}} = [\mathbf{V}_{M,r}, \mathbf{U}_{Q,r}] \quad \text{and} \quad \tilde{\mathbf{\Lambda}} = \begin{bmatrix} \tilde{\mathbf{S}}_{M,r} & 0 \\ 0 & \tilde{\mathbf{S}}_{Q,r} \end{bmatrix}. \tag{D.19}$$

Unrolling the block-matrices appearing in (D.18), the inverse $\mathbf{A}^{-1}$ is given by

$$\mathbf{A}^{-1} = \mathbf{M}^{+}\left(\mathbf{M}^{+}\right)^\top + \left(\mathbf{Q}\mathbf{Q}^\top\right)^{+}. \tag{D.20}$$

Under the conditions resulting in (D.10) the last term can be replaced accordingly.

# E. Surface matching

The examples shown in chapter 6 and in chapter 7 make use of the surface matching algorithm introduced by Vaillant and Glaunès [221]. For the sake of completeness of the presentation, it is summarized here. For a detailed description, the reader is referred to Vaillant and Glaunès [221] and the references therein.

The problem of finding a diffeomorphism that uniquely maps between surfaces is conceptually very similar to the image registration problem described in chapter 3.2. This similarity stems from the fact that the information provided by the images, or the surfaces, does usually not allow for the determination of a unique mapping. This issue is accounted for by modeling assumptions and regularization. To this end, the surface matching introduced by Vaillant and Glaunès [221] is embedded in the LDDMM framework [see e.g. 57, 159].

**Large deformation framework** Using the notation introduced in chapter 2.1, the surface matching is considered on the ambient space $\hat{\Omega} := \Omega_0^{\mathrm{I}} \cup \Omega_t^{\mathrm{I}}$. For coordinates $\hat{\mathbf{x}} \in \hat{\Omega}$, the LDDMM framework parametrizes a time dependent deformation $\varphi_R : t \times \hat{\Omega} \to \hat{\Omega}$ via a *velocity* field $\boldsymbol{v}_t$ according to the evolution equation

$$\frac{\partial \varphi_R(t, \hat{\mathbf{x}})}{\partial t} = \boldsymbol{v}_t(\varphi_R(t, \hat{\mathbf{x}})) \tag{E.1}$$

with $\varphi_R(t = 0, \hat{\mathbf{x}}) = \hat{\mathbf{x}}$. Given a Hilbert space with inner product $\langle \cdot, \cdot \rangle_{V_R}$ and a smooth velocity field $\boldsymbol{v} \in V_R$ such that

$$\int_0^1 \|\boldsymbol{v}_t\|_{V_R} \, dt < \infty, \tag{E.2}$$

then $\varphi_{R,1} := \varphi_R(t = 1, \hat{\mathbf{x}})$ is a diffeomorphism uniquely characterized by the associated velocity field $\boldsymbol{v}_t$ [159, 221].

**Surface matching problem** Given a metric between two surfaces $S$ and $T$ defined by the formalism of surface currents described in chapter 3.3.2, an optimal deformation $\varphi_R^*$ is defined as the minimizer of the functional

$$\int_0^1 \|\boldsymbol{v}_t\|_{V_R}^2 \, dt + \frac{1}{\sigma_R^2} \|\varphi_{R,1}(S) - T\|_{\mathcal{W}^*}^2 \, . \tag{E.3}$$

If the space of velocities is specified as reproducing kernel Hilbert space, uniquely defined by its kernel $k_{V_R}(\hat{\mathbf{x}}_j, \hat{\mathbf{x}}_i)$ (see chapter 3.3.2), minimizer of (E.3) have a representation given by

$$\boldsymbol{v}_t^*(\hat{\mathbf{x}}) = \sum_i k_{V_R}(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}) \boldsymbol{\alpha}_{t,i} \tag{E.4}$$

with the *momentum vectors* $\boldsymbol{\alpha}_{t,i} \in \mathbb{R}^3$ [see e.g. 109]. E.g., for the $\hat{\mathbf{x}}_i$ defined as the points of a triangulation $\mathcal{T}_S$ of the surface $S$, cf. definition 3.3.2, the minimization of (E.3) is performed with respect to the $N_S^p \times N_t$ momentum vectors $\boldsymbol{\alpha}$. $N_t$ is the number of discrete time steps used to solve (E.1).

**Rigid matching** If a rigid deformation between surfaces $S$ and $T$ is sought, the parametrization (E.1) can be replaced by the rigid body deformation

$$\varphi_{rig}(\hat{\mathbf{x}}) = \mathbf{R} \cdot \hat{\mathbf{x}} + \hat{\mathbf{x}}_m \tag{E.5}$$

with the $3 \times 3$ rotation matrix $\mathbf{R}$ and the translation vector $\hat{\mathbf{x}}_m$. The minimization with respect to the 6 degrees of freedom of rotation and translation can usually be computed unregularized by the direct minimization of

$$\|\varphi_{rig}(S) - T\|^2_{\mathcal{W}^*} . \tag{E.6}$$

# F. Graph based total variation on non-uniform meshes

The definition of the graph based version of the TV functional

$$TV_w(\mathbf{b}) = \sum_i \left( \sum_j w^G(i,j)(b_j - b_i)^2 + \epsilon^2 \right)^{1/2} , \qquad \text{(F.1)}$$

cf. (3.88), allows for a very flexible representation of functions with bounded variation on complex domains. The spatial 'structure' of the domain has to be encoded in the adjacency matrix $[\mathbf{W}^G]_{ij} = w^G(i,j)$. Thereby, the term 'structure' refers to the spatial connectivity as well as the spatial scale of this connectivity. The former is represented by the connectivity pattern of the adjacency matrix $\mathbf{W}^G$, the latter by the values of its entries, the so called *edge-weight*. For signals being defined on a mesh, this structure is naturally given by the mesh-connectivity, see figure F.1. A representation of signals in terms of the element-wise basis (2.105) lends itself towards the definition of the vertices of the graph underlying the adjacency matrix $\mathbf{W}^G$ to be the elements of the mesh. The most straightforward definition of the connectivity among these vertices, i.e., the edges of the graph, is then given by the faces dividing one element from another. At the same time, a face's area can directly be used as a measure of the scale of connectivity in the definition of the edge-weights $w^G$, see figure F.1 (left). Another possibility of connectivity is given by the definition of edges given by any lower dimensional entity (faces, lines, nodes). However, the natural definition of the spatial scale via a unique entity defining edges is lost. Instead, the edge-weight can be defined via the spatial distance between element centers in this case, see figure F.1 (right). Throughout the work presented in this thesis, the connectivity of the adjacency matrix is defined solely via element faces

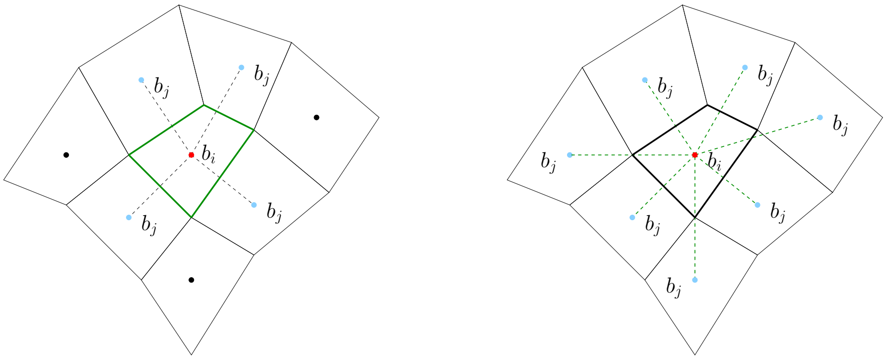

Figure F.1.: Different connectivity setups for an element-wise parameter layout.

since this choice generates graphs that are sparser compared to the graphs defined from connectivity given by faces, lines and nodes.

By denoting the area $a_{ij}$ as the area of the face dividing element $\mathcal{E}_i$ from element $\mathcal{E}_j$, the edge-weight $w^G(i,j)$ can be defined by

$$w^G(i,j) = \begin{cases} \frac{\hat{a}}{a_{ij}} & \iff a_{ij} \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{F.2}$$

whereby the average area $\hat{a} = \sum_{i=1}^{n_f} a_i$ is computed as the average of all $n_f$ internal faces $a_i$. This definition is motivated by the regularizing effect of the norm (3.88) on a function defined as element-wise constant function on a mesh. For the convenience of easy meshing or in the case of mesh size adaptivity, meshes are not necessarily homogeneous. But this inhomogeneity should not affect the regularizing properties of the TV norm. In fact, in the sense of the inverse problem set up in chapter 3, the resulting solution should be penalized in a spatially uniform manner by the TV regularization. This is achieved by (F.2), which is shown in the following using synthetic data.

To this end, the synthetic model introduced in chapter 6 is used, see figure 6.1. This model incorporates an inhomogeneous distribution of the mesh size, see figure F.2, due to the structured meshing with hexahedral elements.

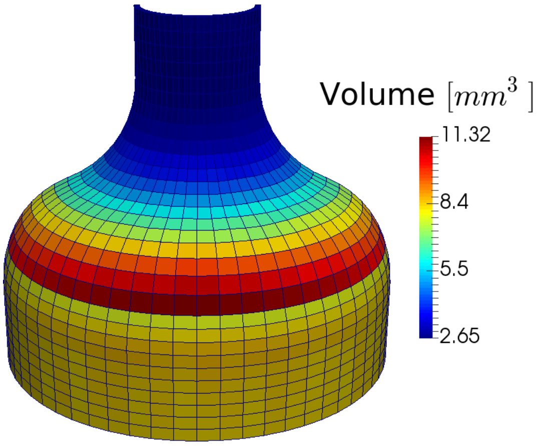

Figure F.2.: Local mesh size in terms of element volume.

The synthetic inverse problem is defined in terms of measurements given as surface data and the likelihood given by (6.8). This measured surface is constructed by the solution of the forward problem $\mathbf{D} = \mathbf{A}(\boldsymbol{\theta} = c_{\vartheta,P})$ and its respective push-forward action on the outer model surface $\varphi^{\mathbf{A}(c_{\vartheta,P})}(\Gamma_O)$. In contrast to the problem setup in chapter 6, synthetic noise is applied on this surface in the sense of centered Gaussian noise in the space of currents ($\sigma_N^2 = 0.001$ mm$^2$, $\sigma_W^2 = 1.0$ mm$^2$), see chapter 3.2. The spatial scale of this noise is not defined by the discretization of the surface but by the spatial scale $\sigma_W$ of the kernel defining the space $\mathcal{W}$, see figure 3.7. Despite the change in the size of the triangulation of the model surfaces, the action of the synthetic noise is propagated in a spatially uniform manner via the application of the surface current perspective. Thus, beside the jumps at the patch boundaries, the inverse solution is expected to show a constant spatial variability, given that the prior penalizes uniformly. This property is verified for the MAP solution of the inverse problem defined by the likelihood in terms of currents and the TV prior (3.93) ($\alpha_{tv} = 1.0, \epsilon = 1.0\text{e}{-2}$). For its definition, the use of (F.2) is contrasted with a uniform weight function given by

$$w^G(i,j) = \begin{cases} 1 & \iff a_{ij} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{F.3}$$

The corresponding MAP solutions are labeled $c_{\vartheta,U}$ for the uniform edge-weights according to (F.3) and $c_{\vartheta,A}$ for the area weighted edge-weights according to (F.2), see figure F.3. It can be seen that both solutions reflect the patch-wise character of the ground truth $c_{\vartheta,P}$. Visually, a difference in the spatial variability within each of the patches is difficult to detect due to the small local variability compared to the overall range of values covered by the solution. However, this overall range cancels out by analyzing the spatial gradient[1] $|\nabla c_\vartheta|$, see figure F.4. It can be seen that the spatial gradient, except for the transition between the patches, shows the uniform distribution for the solution obtained with area weighted edge-weights. The solution obtained with uniform edge-weights shows a much higher gradient in regions with a significantly reduced element size.

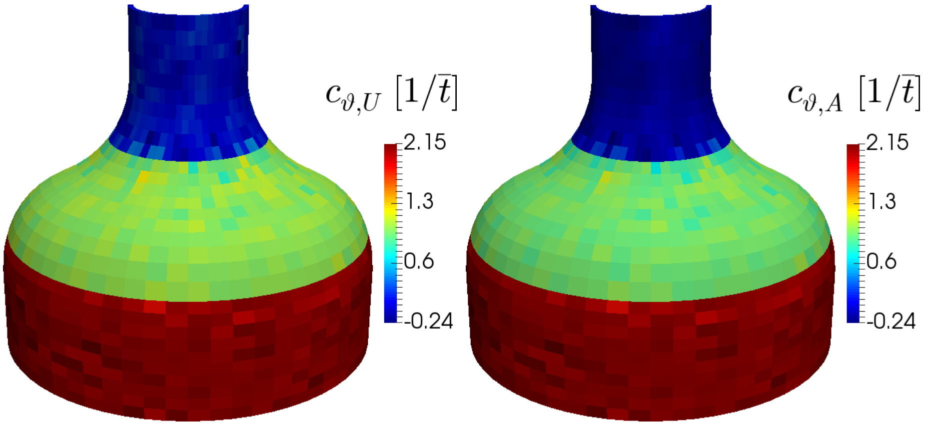

Figure F.3.: MAP solutions $c_{\vartheta,U}$ corresponding to the uniform edge-weights (left) and $c_{\vartheta,A}$ corresponding to the area weighted edge-weights (right). Both solutions reflect the patch-wise character from the reference solution.

This example demonstrated that the use of the area weighted edge-weight (F.2) avoids the effects of a heterogeneous mesh size to be reflected in the modeling of prior information. In contrast, the uniform weight function (F.3) does not account for the spatial variability of smoothness implied by a variation in mesh size. Therefore, definition (F.2) is applied throughout this thesis.

---

[1] The gradient is approximated by projecting the element-wise constant cell data to the point data and followingly using the elemental shape functions to interpolate the gradient at the cell center. This gradient is computed by Paraview's GradientOfUnstructuredDataSet function using vtkCellderivatives.

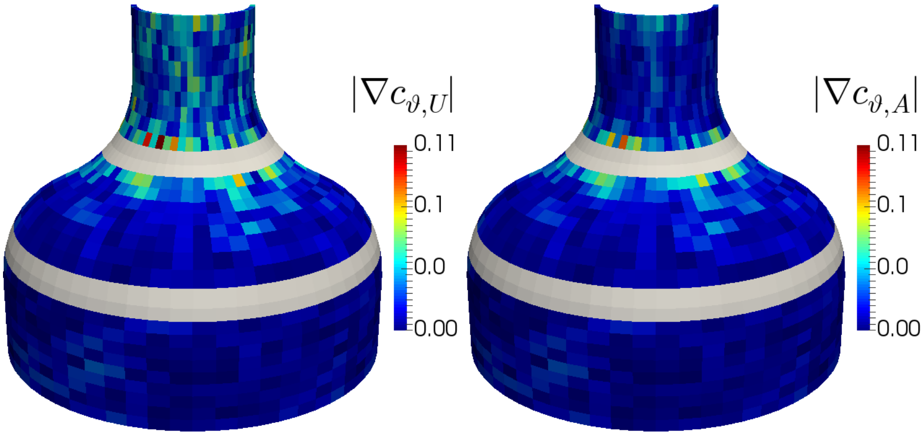

Figure F.4.: Spatial gradient norm $|\nabla c_{\vartheta,U}|$ corresponding to the uniform edge-weight (left) and spatial gradient norm $|\nabla c_{\vartheta,A}|$ corresponding to the area weighted edge-weights (right). The gradient at the transition between the patches is not shown due to the large magnitude which is outside the chosen color range.

# Abbreviations

| | |
|---|---|
| **AAA** | Abdominal aortic aneurysm |
| **BFGS** | Broyden-Fletcher-Goldfarb-Shanno |
| **CI** | Credible interval |
| **CT** | Computed tomography |
| **ESS** | Effective sample size |
| **FD** | Finite difference |
| **FE** | Finite element |
| **FEM** | Finite element method |
| **FOM** | Full order model |
| **FSI** | Fluid structure interaction |
| **G&R** | Growth and remodeling |
| **HPD** | Highest posterior density |
| **ILT** | Intraluminal thrombus |
| **LBFGS** | Limited memory BFGS |
| **LDDMM** | Large deformation diffeomorphic metric mapping |
| **MAP** | Maximum a posteriori |
| **MC** | Monte Carlo |
| **MCMC** | Markov chain Monte Carlo |
| **MPI** | Message passing interface |
| **MRF** | Markov random field |
| **MRI** | Magnetic resonance imaging |
| **MULF** | Modified updated Lagrangian formulation |
| **OpenMP** | Open Multi-Processing |
| **PCA** | Principle component analysis |
| **PDE** | Partial differential equation |
| **PM** | Posterior mean |
| **POD** | Proper orthogonal decomposition |
| **PTC** | Pseudotransient continuation |
| **RKHS** | Reproducing kernel Hilbert space |
| **RPI** | Rupture potential index |
| **SEF** | Strain energy function |
| **SMC** | Sequential Monte Carlo |
| **SNR** | Signal-to-noise ratio |
| **SPD** | Symmetric positive definite |
| **SVD** | Singular value decomposition |

| | |
|---|---|
| **TV** | Total Variation |
| **UQ** | Uncertainty quantification |
| **US** | Ultrasound |
| **VB** | Variational Bayes |

# Nomenclature

## General

| | |
|---|---|
| $(\cdot)^\top$ | Transpose of $(\cdot)$ |
| $\delta A[\delta B]$ | Variation of $A$ under admissible variation $\delta B$ |
| $\Delta A[\delta B]$ | Variation of $A$ under admissible variation $\delta B$, used if $A = \delta C$ |

## Nonlinear structural elasticity

### Domains and boundaries

| | |
|---|---|
| $\Omega_0$ | Reference configuration |
| $\Omega_t$ | Current configuration |
| $\Omega_0^\mathrm{I}$ | Ambient space/image domain ($\Omega_0 \subset \Omega_0^\mathrm{I}$) |
| $\Omega_t^\mathrm{I}$ | Ambient space/image domain ($\Omega_t \subset \Omega_t^\mathrm{I}$) |
| $\Omega_g$ | Intermediate configuration resulting from growth |
| $\Omega_z$ | Stress-free configuration |
| $\partial\Omega_i$ | Boundary of a configuration $\Omega_i, i \in (0, t, g)$ |
| $T_x\Omega_i$ | Tangent space of $\Omega_i$ at $x$, $x \in (\mathbf{X}, \mathbf{x})$, $i \in (0, t)$ |
| $T_x^*\Omega_i$ | Dual to the tangent space $T_x\Omega_i$ |
| $\mathrm{dV}_0, \mathrm{dV}, \mathrm{dV}_g$ | Volume measure for $\Omega_0$, $\Omega_t$, $\Omega_g$ |
| $\mathrm{d}\Gamma_0, \mathrm{d}\Gamma$ | Volume measure for $\partial\Omega_0$, $\partial\Omega_t$ |
| $\mathbf{N}$ | Normal vectors on $\partial\Omega_0$ |
| $\mathbf{n}$ | Normal vectors on $\partial\Omega_t$ |
| $\mathbf{X}$ | Material coordinates $\in \Omega_0$ |
| $\mathbf{x}$ | Material coordinates $\in \Omega_t$ |
| $\mathbf{X}_z$ | Material coordinates $\in \Omega_z$ |

### Kinematic description

| | |
|---|---|
| $\varphi$ | Diffeomorphic deformation $\Omega_0 \mapsto \Omega_t$ |
| $\varphi_z$ | Diffeomorphic deformation $\Omega_z \mapsto \Omega_0$ |
| $\mathfrak{c}$ | Constraint on $\varphi$ |
| $\varphi_*(\bullet)$ | Push-forward action associated to $\varphi$ |
| $\varphi_*^{-1}(\bullet)$ | Pull-back action associated to $\varphi$ |
| $\mathbf{F}$ | Deformation gradient in $\Omega_0$ |
| $\hat{\mathbf{F}}$ | Deformation gradient in $\Omega_t$ |
| $\mathbf{F}_z$ | Deformation gradient associated to prestress |

| | |
|---|---|
| $\tilde{\mathbf{F}}_z$ | Approximation of $\mathbf{F}_z$ |
| $J$ | Determinant of $\mathbf{F}$ |
| $\mathbf{U}$ | Displacement field in $\Omega_0$ |
| $\mathbf{u}$ | Displacement field in $\Omega_t$ |
| $\mathbf{C}$ | Right Cauchy-Green tensor (in $\Omega_0$) |
| $\mathbf{E}$ | Green-Lagrange strain tensor (in $\Omega_0$) |
| $\boldsymbol{\ell}$ | Left Cauchy-Green tensor (in $\Omega_t$) |
| $\mathbf{e}$ | Euler-Almansi strain tensor (in $\Omega_t$) |
| $\mathbf{e}^*$ | Euler-Almansi strain tensor in $\Omega_0$ |
| $\delta\mathbf{U}, \boldsymbol{\lambda}, \Delta\mathbf{U}$ | Admissible variation of $\mathbf{U}(\mathbf{X})$ |
| $\delta\mathbf{u}$ | Admissible variation of $\mathbf{u}(\mathbf{x})$ |

## Kinetic description and virtual work

| | |
|---|---|
| $\mathcal{I}$ | Arbitrary subset $\mathcal{I} \subset \Omega_t$ |
| $\mathbf{t}$ | Traction in $\Omega_t$ |
| $\hat{\mathbf{t}}$ | Prescribed traction at $\partial\Omega_t$ |
| $\mathbf{t}_0$ | Prescribed traction at $\partial\Omega_0$ |
| $\boldsymbol{\sigma}$ | Cauchy stress in $\Omega_t$ |
| $\boldsymbol{\sigma}^*$ | Cauchy stress in $\Omega_0$ |
| $\mathbf{S}$ | Second Piola-Kirchhoff stress in $\Omega_0$ |
| $\mathbf{b}$ | Body force (force/current volume) in $\Omega_t$ |
| $\mathbf{b}^*$ | Body force (force/current volume) in $\Omega_0$ |
| $\mathbf{b}_0$ | Body force (force/reference volume) in $\Omega_0$ |
| $\mathrm{p}$ | Pressure |
| $\mathrm{p}_{\mathrm{dia}}$ | Diastolic pressure ($\approx 80$ mmHg) |
| $\delta W$ | Virtual work |

## Hyperelasticity

| | |
|---|---|
| $\Psi, \tilde{\Psi}, \Psi_{\mathrm{I}_1,\mathrm{I}_2,\mathrm{I}_3}$ | Strain energy (density) function |
| $\hat{\Psi}$ | Distortional strain energy function |
| $\Psi_{vol}$ | Volumetric strain energy function |
| $\mathrm{I}_i$ | Invariants ($i \in (1,2,3)$) of $\mathbf{C}$ |
| $\hat{\mathbf{C}}, \hat{\mathbf{C}}_e$ | Distortional component of $\mathbf{C}$, $\mathbf{C}_e$ |
| $\hat{\mathrm{I}}_i, \hat{\mathrm{I}}_{i,e}$ | Invariants ($i \in (1,2,3)$) of $\hat{\mathbf{C}}$, $\hat{\mathbf{C}}_e$ |
| $\mathcal{C}$ | Elasticity tensor in $\Omega_0$ |
| $\mathcal{C}_e$ | Elastic component of $\mathcal{C}$ in $\Omega_0$ |

## Isovolumetric growth

| | |
|---|---|
| $\varphi_g$ | Non-diffeomorphic deformation $\Omega_0 \mapsto \Omega_g$ due to growth |
| $\mathbf{F}_g$ | Growth component of $\mathbf{F}$ |
| $\mathbf{F}_e$ | Elastic component of $\mathbf{F}$ |
| $J_g$ | Determinant of $\mathbf{F}_g$ |
| $J_e$ | Determinant of $\mathbf{F}_e$ |
| $\vartheta$ | Growth stretch in $\Omega_0$ |
| $\mathbf{C}_e$ | Elastic component of $\mathbf{C}$ |
| $\Psi_e$ | Elastic strain energy function |
| $\mathbf{S}_e$ | Second Piola-Kirchhoff stress due to elastic deformation in $\Omega_0$ |
| $m$ | Mass |
| $\rho_t$ | Density (mass/current volume) in $\Omega_t$ |
| $\rho_t^{**}$ | Density (mass/current volume) in $\Omega_g$ |
| $\rho_t^{*}$ | Density (mass/current volume) in $\Omega_0$ |
| $\rho_0$ | Density (mass/reference volume) in $\Omega_0$ |
| $k_\vartheta^+, k_\vartheta^-$ | Parameters of specific growth evolution functions |
| $\vartheta^+, \vartheta^-$ | Parameters of specific growth evolution functions |
| $m_\vartheta^+, m_\vartheta^-$ | Parameters of specific growth evolution functions |
| $c_\vartheta$ | Parameter of specific growth evolution functions |

## Finite element method

| | |
|---|---|
| $C^2$ | Differentiability class of order 2 |
| $\alpha$ | Multi index |
| $D^\alpha(\cdot)$ | Partial derivatives |
| $\mathrm{L}^p$ | Lebesgue spaces |
| $W_k^p(\Omega_0)$ | Sobolev spaces |
| $\mathrm{H}^k(\Omega_0)$ | Hilbert spaces $\mathcal{W}_k^2(\Omega_0)$ |
| $\gamma$ | Trace operator $\mathrm{H}^1(\Omega_0) \rightarrow \mathrm{L}^2(\partial\Omega_0)$ |
| $\mathrm{H}_0^1(\Omega_0)$ | Hilbert space $\mathrm{H}^1(\Omega_0)$ with evaluation at the boundary |
| $\mathcal{V}$ | Extension of $\mathrm{H}_0^1$ to $d$ dimensions |
| $\mathcal{K}$ | Mesh discretizing $\Omega_0$ |
| $\mathcal{E}$ | Element of $\mathcal{K}$ |
| $n_{ele}$ | Number of elements in $\mathcal{K}$ |
| $\mathcal{P}_\kappa(\mathcal{E})$ | Class of polynomials in $\mathcal{E}$ with order $\kappa$ |
| $\mathcal{V}_h$ | Finite dimensional approximation of $\mathcal{V}$ |
| $\mathbf{U}_h$ | Finite dimensional approximation of $\mathbf{U}$ |
| $\delta\mathbf{U}_h$ | Finite dimensional approximation of $\delta\mathbf{U}$ |
| $N_i$ | Shape function associated to node $i$ |
| $n_{nod}^{\mathcal{E}}$ | Number of nodes per element $\mathcal{E}$ |
| $\mathbf{d}_i$ | Coefficients for shape function $N_i$ associated to $\mathbf{U}_h$ |
| $\delta\mathbf{d}_i$ | Coefficients for shape function $N_i$ associated to $\delta\mathbf{U}_h$ |
| $\mathcal{E}_g$ | Generic reference element |
| $\boldsymbol{\xi}$ | Coordinates describing $\mathcal{E}_g$ |
| $\mathrm{dV}_{\mathcal{E}_g}$ | Volume measure of the generic element $\mathcal{E}_g$ |

| | |
|---|---|
| $\mathbf{D}$ | Assembled vector of nodal coefficients $\mathbf{d}_i$ |
| $\delta\mathbf{D}$ | Assembled vector of nodal coefficients $\delta\mathbf{d}_i$ |
| $n_{dof}$ | Number of degrees of freedom, i.e. $\dim(\mathbf{D})$ |
| $n_{dof}^{\gamma}$ | Number of degrees of freedom associated to $\partial\Omega_0$ |
| $\mathbf{f}$ | Nonlinear system resulting from discretization of $\delta W$ |

### Nonlinear solution

| | |
|---|---|
| $\mathbf{D}^k$ | Solution at nonlinear iteration $k$ |
| $\Delta\mathbf{D}^k$ | Increment at nonlinear iteration $k$ |
| $\mathbf{f}^{\alpha}$ | Instance of sequence of nonlinear problems $\mathbf{f}$ |
| $\hat{\mathbf{t}}^{\alpha}$ | Instance of a sequence of traction loads $\hat{\mathbf{t}}$ |
| $\tau_k$ | Artificial time / line search parameter |
| $\mathsf{K}$ | Symmetric linear operator representing $\Delta\delta W(\mathbf{U}, \delta\mathbf{U})[\Delta\mathbf{U}]$ |
| $\boldsymbol{\mathsf{K}}$ | Discretization of $\mathsf{K}$ (stiffness matrix) |

### Parametrization

| | |
|---|---|
| $\theta$ | Parametrization of the weak form |
| $\Psi_{\theta}$ | Parametrized strain energy function |
| $\delta W_{\theta}$ | Parametrized virtual work |
| $\mathcal{X}$ | Parameter space |
| $\theta_h$ | Discrete parametrization |
| $\mathcal{X}_h$ | Discretization of $\mathcal{X}$ |
| $g_{\mathcal{E}}$ | Basis functions for $\mathcal{X}_h$ |
| $\mathbf{G}^{\top}$ | Vector of basis functions $g_{\mathcal{E}}$ |
| $\boldsymbol{\theta}$ | Vector of basis coefficients |
| $A$ | Nonlinear operator $\mathcal{X} \to \mathcal{V}$ |
| $A_h$ | Nonlinear operator $\mathcal{X} \to \mathcal{V}_h$ |
| $\mathbf{A}$ | Discrete nonlinear operator $\mathbb{R}^{n_{ele}} \to \mathcal{V}_h$ |

## Inverse problem

### General

| | |
|---|---|
| $\mathcal{Z}$ | Measurement space |
| $\mathbf{Z}$ | Generic measurement $\in \mathcal{Z}$ |
| $C$ | Generic observation operator $\mathcal{V} \to \mathcal{Z}$ |
| $\hat{\mathbf{U}}$ | Real state of the system modeled by $\mathbf{U}$ |
| $R$ | Regularization |

## Statistical quantities

| | |
|---|---|
| $(\mathcal{S}, \mathcal{F}, P)$ | Probability space |
| $\mathsf{X}, \mathsf{Y}, \mathsf{\Phi}$ | Random variables |
| $\mathrm{A}, \mathrm{B}$ | Sets $\in \mathcal{F}$ |
| $\mathcal{B}, \mathcal{D}$ | Borel sets on $\mathbb{R}^n$ |
| $p_{\mathsf{X}}$ | Probability density function of a random variable $\mathsf{X}$ |
| $\mathbb{E}_{\mathsf{X}}[\cdot], \mathbb{E}_x[\cdot]$ | Expected value of $(\cdot)$ wrt. density $p_{\mathsf{X}}(x)$ |
| $\mathbb{V}_{\mathsf{X}}[\cdot], \mathbb{V}_x[\cdot]$ | Variance of $(\cdot)$ wrt. density $p_{\mathsf{X}}(x)$ |
| $\mathbb{SD}_{\mathsf{X}}[\cdot]$ | Standard deviation of $(\cdot)$ wrt. density $p_{\mathsf{X}}$ |
| $\mathcal{N}(\boldsymbol{x}\|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ | Normal distribution with mean $\boldsymbol{\mu}$, and covariance $\boldsymbol{\Sigma}$ |
| $\mathcal{U}(a, b)$ | Uniform distribution on the interval $[a, b]$ |

## Statistical identification problem

| | |
|---|---|
| $n_p$ | Dimension of the parameter identification problem |
| $p(\mathbf{Z}\|\boldsymbol{\theta})$ | Likelihood |
| $p(\boldsymbol{\theta})$ | Prior |
| $p(\mathbf{Z})$ | Evidence |
| $p(\boldsymbol{\theta}\|\mathbf{Z})$ | Posterior |
| $\boldsymbol{\zeta}$ | Measurement noise |
| $\boldsymbol{\Sigma}_{\mathbf{Z}}$ | Measurement covariance matrix |
| $\sigma$ | Measurement noise standard deviation |
| $F$ | Computational model $\mathbb{R}^{n_p} \to \mathcal{Z}$ |
| $D$ | Similarity measure |
| $\boldsymbol{\Theta}$ | Stacked vector of model parameters $\boldsymbol{\theta}$ and noise parameters $\sigma$ |
| $\mathbb{M}$ | Class of computational models |

## Similarity measures on images

| | |
|---|---|
| $\varphi_{\mathrm{I}}$ | Diffeomorphic deformation $\Omega_0^{\mathrm{I}} \mapsto \Omega_t^{\mathrm{I}}$ |
| $\mathbf{V}$ | Smooth displacement field associated with $\varphi_{\mathrm{I}}$ |
| $\mathcal{V}_{\mathrm{I}}$ | Space of displacements $\mathbf{V}$ |
| $\mathrm{I}_0$ | Image in $\Omega_0^{\mathrm{I}}$ |
| $\mathrm{I}_t$ | Image in $\Omega_t^{\mathrm{I}}$ |
| $\hat{\varphi}_{\mathrm{I}}$ | Estimation of $\varphi_{\mathrm{I}}$ from image registration |
| $\hat{\mathbf{V}}$ | Smooth displacement field associated with $\hat{\varphi}_{\mathrm{I}}$ |
| $C_{\mathrm{I}}$ | Observation operator $\mathcal{V}_{\mathrm{I}} \to \mathbb{R}^{n_{dof}}$ |
| $C_{\mathrm{I},\gamma}$ | Observation operator $\mathcal{V}_{\mathrm{I}} \to \mathbb{R}^{n_{dof}^{\gamma}}$ |

## Similarity measures on surfaces

| | |
|---|---|
| $U$ | Open interval $U \subset \mathbb{R}^2$ |
| $s, t$ | Coordinates of $U$ |
| $T_q U$ | Tangent space to $U$ at $q \in U$ |

| | |
|---|---|
| $\phi_\alpha, \phi_\beta$ | Surface parametrization |
| $\boldsymbol{\nu}, \boldsymbol{\eta}$ | Tangent vectors |
| $\mathbf{g}$ | Metric tensor |
| $\omega$ | Differential 2-form |
| $\mathbf{w}$ | Vector field associated to $\omega$ |
| $\mathbf{n}^S$ | Field of normals of a surface $S$ |
| $\tilde{\mathbf{n}}^S$ | Field of unnormalized normals $\mathbf{n}\,\mathrm{d}\Gamma$ of a surface $S$ |

| | |
|---|---|
| $\mathcal{T}_S$ | Triangulation of a surface $S$ |
| $\boldsymbol{\varkappa}_i^S$ | Point of the triangulation $\mathcal{T}_S$ |
| $\boldsymbol{\tau}_i^S$ | Triangle $i$ of the triangulation $\mathcal{T}_S$ |
| $\mathbf{c}_i^S$ | Center of triangle $\boldsymbol{\tau}_i^S$ |
| $\tilde{\mathbf{n}}_i^S$ | Normal of triangle $\boldsymbol{\tau}_i^S$ |
| $N_S$ | Number of triangles in $\mathcal{T}_S$ |
| $N_S^p$ | Number of vertices in $\mathcal{T}_S$ |
| $d_i$ | Closet point projection |

| | |
|---|---|
| $S(\omega)$ | Surface current associated to the surface $S$ |
| $\mathcal{W}^*$ | Space of surface currents |
| $\mathcal{W}$ | Dual space to $\mathcal{W}^*$ |
| $\boldsymbol{K}_{\mathbf{n}}^S$ | Dual representation of a surface current $S(\omega)$ |
| $\delta_x^{\mathbf{n}}$ | Dirac current |

| | |
|---|---|
| $\mathcal{H}_k$ | Reproducing kernel Hilbert space |
| $\delta_x$ | Linear evaluation functional $\mathcal{H}_k \to \mathbb{R}$ |
| $K_x$ | Generic representer $\in \mathcal{H}_k$ of $\delta_x$ |
| $k$ | Kernel defining $\mathcal{H}_k$ |
| $\sigma_W$ | Spatial scale of the kernel $k$ |
| $\sigma_N$ | Variance of a random current |

| | |
|---|---|
| $\underline{K}$ | Kernel matrix |
| $\underline{\tilde{\mathbf{n}}}_i$ | Stacked vector of the $i$-th spatial component of all $\tilde{\mathbf{n}}$ |
| $\boldsymbol{\mathsf{n}}_i$ | Random variable $\underline{\tilde{\mathbf{n}}}_i$ |

## Prior assumptions

| | |
|---|---|
| $\sigma_g$ | Spatial scale of the Gaussian prior |
| $k_\Sigma$ | Covariance function |
| $\ell$ | Spatial scale of $k_\Sigma$ |
| $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{y}_0$ | Vectors $\in \mathbb{R}^n$ |
| $\boldsymbol{\mathsf{x}}, \boldsymbol{\mathsf{y}}, \boldsymbol{\mathsf{z}}$ | Random vectors |
| $\mathbf{L}$ | Structural information encoding discrete linear operator |
| $\sigma_s$ | Spatial scale of the smoothness prior |

| | |
|---|---|
| $\mathbf{Q}$ | Basis for $\ker(\mathbf{L})$ |
| $\mathbf{q}_i$ | Columns of $\mathbf{Q}$ |
| $\mathsf{h}_i$ | Random coefficients for $\mathbf{Q}$ |
| $\mathbf{h}$ | Vector of random coefficients $\mathsf{h}_i$ |
| $\mathbb{R}^{n_p}_{\setminus\mathbf{Q}}$ | Factor space $\mathbb{R}^{n_p} \setminus \ker(\mathbf{L})$ |

| | |
|---|---|
| $TV$ | Total variation |
| $TV_\epsilon$ | Smoothed total variation |
| $TV_w$ | Graph based total variation |
| $\epsilon$ | Smoothing factor |
| $\mathrm{J}$ | Index set |
| $G$ | Graph |
| $V$ | Vertices of the graph $G$ |
| $E$ | Edges of the graph $G$ |
| $w^G$ | Weight function associated to the edges of the graph $G$ |
| $\mathbf{W}^G$ | Matrix of edge-weights |
| $\boldsymbol{\iota}_i$ | Maximal clique of the graph $G$ |
| $\psi$ | Clique potential |
| $H$ | Clique energy |
| $\alpha_{tv}$ | Scaling for the total variation prior |

### Lagrangian formulation

| | |
|---|---|
| $\tilde{F}$ | Computational model $\mathcal{X} \to \mathcal{Z}$ |
| $\mathcal{J}$ | negative log-posterior |
| $\tilde{\mathcal{L}}$ | Lagrangian |
| $\mathcal{L}$ | Augmented Lagrangian |

## Numerical solution

### Estimators

| | |
|---|---|
| $\hat{\boldsymbol{\theta}}(\mathbf{Z})$ | Estimator |
| $\hat{\boldsymbol{\theta}}$ | Estimate |
| $L$ | Loss-function |
| $\hat{\boldsymbol{\theta}}_{pm}(\mathbf{Z})$ | Posterior mean estimator |
| $\hat{\boldsymbol{\theta}}_{map}(\mathbf{Z})$ | Maximum a posteriori estimator |
| $\theta_i$ | $i$-th entry of $\boldsymbol{\theta}$ |
| $\boldsymbol{\theta}_{\setminus i}$ | Remainder vector $\boldsymbol{\theta} \setminus \theta_i$ |

### Approximate inference

| | |
|---|---|
| $q$ | Approximate posterior |
| $\mathbf{H}$ | Hessian matrix |

| | |
|---|---|
| $D_{KL}$ | Kullback-Leibler divergence |
| $\mathcal{G}$ | Lower bound on $p(\mathbf{Z})$ |
| $\mathbf{J}$ | Jacobian matrix |

## LBFGS

| | |
|---|---|
| $h$ | Representation of $-\log p(\boldsymbol{\theta}\vert\mathbf{Z})$ |
| $\boldsymbol{d}$ | Search direction |
| $\boldsymbol{B}$ | Approximate hessian |
| $\boldsymbol{s}$ | Solution increment |
| $\boldsymbol{y}$ | Gradient increment |
| $\chi$ | Line search problem |
| $c_1$ | Sufficient decrease parameter |
| $\chi_\kappa$ | Polynomial models of $\chi$ with order $\kappa$ |
| $c_2, c_3$ | Coefficients for $\chi_3$ |
| $\tau_+$ | Optimal line search parameter |
| $\beta_l, \beta_h$ | Safeguard parameters |
| $m_s$ | Size of vector storage |
| $\gamma_n$ | Initial hessian scaling |

## Monte Carlo methods

| | |
|---|---|
| $\eta$ | Importance densities |
| $w$ | Weight function |
| $\pi$ | Stationary distribution, target distribution |
| $\mathfrak{K}$ | Markov kernel |
| $\mathfrak{K}_{mh}$ | Metropolis-Hasting transition kernel |
| $\mathfrak{t}_{\mathrm{A}}$ | Stopping time for the set A |
| $\mathfrak{i}_{\mathrm{A}}$ | Number of visits to the set A |
| $\mathbb{I}_{\mathrm{A}}$ | Indicator function of the set A |
| $q_{pp}, \tilde{q}_{pp}$ | Proposal densities |
| $\varrho$ | Metropolis-Hastings acceptance probability |

| | |
|---|---|
| $\pi_0$ | Initial distribution |
| $\lambda_n$ | Sequence parameters |
| $\varsigma_n$ | Unnormalized density |
| $\tilde{\pi}_n$ | Artificial probability distributions |
| $\tilde{\varsigma}_n$ | Unnormalized $\tilde{\pi}_n$ |
| $\mathfrak{L}$ | Backward Markov kernel |
| $\tilde{w}$ | Incremental weight function |
| $ESS$ | Effective sample size |
| $\xi$ | $ESS$ reduction factor |
| $\mathrm{T}_{red}$ | Resampling threshold |
| $\mathsf{w}$ | Normalized weights |
| $\pi_n^N$ | Particle approximation of $\pi_n$ |

## Predictive approach

| | |
|---|---|
| $\mathbf{P}$ | Dictionary of basis vectors |
| $\mathbf{p}_i$ | Column $i$ of $\mathbf{P}$ |
| $\boldsymbol{\psi}$ | Reduced dimensional parameters |
| $\mathbf{L}_{tv}$ | Linear operator representing the total variation bound |
| $\mathbf{Y}$ | Matrix of eigenvectors |
| $\boldsymbol{\Lambda}$ | Matrix of eigenvalues |
| $\mathbf{Y}_{n_r}$ | Matrix of significant eigenvectors |
| $\tilde{\mathbf{H}}^{-1}$ | Approximate action of the hessian $\mathbf{H}$ |
| $\mathbf{H}_{\mathbf{P}}^{-1}$ | Projected inverse hessian |
| $\sigma_P$ | Scaling factor for $\mathbf{H}_{\mathbf{P}}^{-1}$ |

| | |
|---|---|
| $\mathcal{K}_i$ | Patch of connected elements from the mesh $\mathcal{K}$ |
| $N_p$ | Number of patches |
| $n_{ele}^{\mathcal{K}_i}$ | Number of elements per patch $\mathcal{K}_i$ |
| $\mathfrak{M}$ | Set of bins $\mathfrak{m}_i$ of a histogram |
| $n_{bin}$ | Number of bins |
| $\hat{\theta}_{map,\mathcal{E}}$ | Entry of $\hat{\boldsymbol{\theta}}_{map}$ associated to element $\mathcal{E}$ |
| $tol_p$ | Tolerance for patch creation |
| $d_{\mathcal{E}}$ | Element to patch association |
| $\boldsymbol{\psi}_p$ | Reduced parameters associated to patch-wise basis |
| $\boldsymbol{\psi}_{tv}$ | Reduced parameters associated to TV decomposition |

## Synthetic data model

| | |
|---|---|
| $\Psi_{AAA}$ | Strain energy function for aneurysmatic wall |
| $\Psi_{ILT}$ | Strain energy function for intraluminal thrombus |
| $\alpha, \beta$ | Stiffness parameters for $\Psi_{AAA}$ |
| $c_1, \beta_1$ | Stiffness parameters for $\Psi_{ILT}$ |
| $\kappa$ | Bulk modulus |
| $\Gamma_O, \Gamma_I$ | Model surfaces |
| $c_{\vartheta,P}$ | Patch-wise distribution of growth stretch |
| $\bar{t}$ | Generic time period of growth |
| $\mathscr{A}$ | Path along aortic arch model |
| $s$ | Coordinate of path $\mathscr{A}$ |
| $u_y(s)$ | Displacement along path |
| $u_y^{max}$ | Maximal displacement in $y$-direction |

### Surface matching

| | |
|---|---|
| $\hat{\Omega}$ | Ambient space $\Omega_0^{\mathrm{I}} \cup \Omega_t^{\mathrm{I}}$ |
| $\hat{\mathbf{x}}$ | Coordinate in the ambient space $\hat{\Omega}$ |

| | |
|---|---|
| $\varphi_R$ | Diffeomorphic deformation from surface matching |
| $\boldsymbol{v}$ | Velocity field |
| $V_R$ | Kernel space of velocity fields |
| $k_{V_R}$ | Kernel associated to $V_R$ |
| $\boldsymbol{\alpha}$ | Field of moments parametrizing $\boldsymbol{v}$ |
| $\varphi_{rig}$ | Rigid deformation |
| $\mathbf{R}$ | Rotation matrix |
| $\hat{\mathbf{x}}_m$ | Translation vector |

| | |
|---|---|
| $\mathbf{D}_{\Gamma_O}$ | Dofs associated to the surface $\Gamma_O$ |
| $n_{dof}^{\Gamma}$ | Number of dofs $\mathbf{D}_{\Gamma_O}$ |
| $\mathbf{V}_R$ | Displacement field associated to $\varphi_R$ |
| $\mathbf{V}_{R,\Gamma_O}$ | $\mathbf{V}_R$ evaluated at the dofs $\mathbf{D}_{\Gamma_O}$ |
| $C_{R,\Gamma_O}$ | Observation operator $\mathbf{V}_R \mapsto \mathbf{V}_{R,\Gamma_O}$ |
| $p_D$ | Likelihood in terms of displacement data |
| $p_W$ | Likelihood in terms of surface data |
| $c_{\vartheta,D}, c_{\vartheta,W}$ | Inverse solutions corresponding to $p_D$, $p_W$ |
| $\tilde{c}_{\vartheta,D}, \tilde{c}_{\vartheta,W}$ | Reference solutions corresponding to $p_D$, $p_W$ |
| $\Delta_D$ | Relative error of $c_{\vartheta,D}$ |
| $\Delta_W$ | Relative error of $c_{\vartheta,W}$ |

# List of Figures

Figures have been generated using Paraview [1], Matplotlib [103], Xfig [213] and Inkscape [13].

# Bibliography

[1] Ahrens, J., Geveci, B., and Law, C. (2005). *ParaView: An End-User Tool for Large Data Visualization.* Visualization Handbook, Elsevier.

[2] Allgower, E. L. and Georg, K. (1990). *Numerical Continuation Methods: An Introduction.* Springer, Berlin, Heidelberg.

[3] Alt, H. W. (1999). *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung.* Springer, Berlin, Heidelberg.

[4] Álvarez, M. A., Rosasco, L., and Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. *Foundations and Trends in Machine Learning*, 4(3):195–266.

[5] Ambrosi, D., Ateshian, G., Arruda, E., Cowin, S., Dumais, J., Goriely, A., Holzapfel, G., Humphrey, J., Kemkemer, R., Kuhl, E., Olberding, J., Taber, L., and Garikipati, K. (2011). Perspectives on biological growth and remodeling. *Journal of the Mechanics and Physics of Solids*, 59(4):863–883.

[6] Arbenz, P., Hetmaniuk, U. L., Lehoucq, R. B., and Tuminaro, R. S. (2005). A comparison of eigensolvers for large-scale 3d modal analysis using amg-preconditioned iterative methods. *International Journal for Numerical Methods in Engineering*, 64(2):204–236.

[7] Armijo, L. (1966). Minimization of functions having lipschitz continuous first partial derivatives. *Pacific Journal of Mathematics*, 16(1):1–3.

[8] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, 1:337–404.

[9] Attias, H. (2000). A variational bayesian framework for graphical models. In *Advances in Neural Information Processing Systems 12*, pages 209–215. MIT Press.

[10] Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2008). Parameter estimation in tv image restoration using variational distribution approximation. *IEEE Transactions on Image Processing*, 17(3):326–339.

[11] Bachman, D. (2012). *A geometric approach to differential forms.* Springer, New York.

[12] Baek, S., Rajagopal, K. R., and Humphrey, J. D. (2005). A theoretical model of enlarging intracranial fusiform aneurysms. *Journal of Biomechanical Engineering*, 128(1):142–149.

[13] Bah, T. (2017). *Inkscape: Guide to a vector drawing program*, 5th edition.

[14] Balocco, S., Camara, O., Vivas, E., Sola, T., Guimaraens, L., van Andel, H. A. G., Majoie, C. B., Pozo, J. M., Bijnens, B. H., and Frangi, A. F. (2010). Feasibility of estimating regional mechanical properties of cerebral aneurysms in vivo. *Medical physics*, 37(4):1689–1706.

[15] Banks, H. T. and Kunisch, K. (1989). *Estimation Techniques for Distributed Parameter Systems.* Birkhäuser, Boston.

[16] Barber, D. C., Valverde, I., Shi, Y., Brown, A., Beerbaum, P., and Rodney Hose, D. (2014). Derivation of aortic distensibility and pulse wave velocity by image registration with a physics-based regularisation term. *International Journal for Numerical Methods in Biomedical Engineering*, 30(1):55–68.

[17] Beal, M. J. (2003). *Variational Algorithms for Approximate Bayesian Inference.* PhD thesis, Gatsby Computational Neuroscience Unit, University College London.

[18] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396.

[19] Berger, J. O. (1985). *Statistical Decision Theory and Bayesian Analysis.* Springer, New York.

[20] Bertoglio, C., Barber, D., Gaddum, N., Valverde, I., Rutten, M., Beerbaum, P., Moireau, P., Hose, R., and Gerbeau, J.-F. (2014). Identification of artery wall stiffness: In vitro validation and in vivo results of a data assimilation procedure applied to a 3d fluid-structure interaction model. *Journal of Biomechanics*, 47(5):1027–1034.

[21] Bertoglio, C., Moireau, P., and Gerbeau, J.-F. (2012). Sequential parameter estimation for fluid-structure problems: Application to hemodynamics. *International Journal for Numerical Methods in Biomedical Engineering*, 28(4):434–455.

[22] Biehler, J., Gee, M. W., and Wall, W. A. (2015). Towards efficient uncertainty quantification in complex and large-scale biomechanical problems based on a bayesian multi-fidelity scheme. *Biomechanics and Modeling in Mechanobiology*, 14(3):489–513.

[23] Bilionis, I. and Zabaras, N. (2014). Solution of inverse problems with limited forward solver evaluations: a bayesian perspective. *Inverse Problems*, 30(1):015004.

[24] Billingsley, P. (1995). *Probability and Measure.* Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken.

[25] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning.* Information Science and Statistics. Springer, New York.

[26] Björck, A. (1996). *Numerical Methods for Least Squares Problems.* Society for Industrial and Applied Mathematics, Philadelphia.

[27] Bonet, J. and Wood, R. D. (1997). *Nonlinear continuum mechanics for finite element analysis.* Cambridge University press, Cambridge.

[28] Brady, A. R., Thompson, S. G., Fowkes, F. G. R., Greenhalgh, R. M., Powell, J. T., and on behalf of the UK Small Aneurysm Trial Participants (2004). Abdominal aortic aneurysm expansion: Risk factors and time intervals for surveillance. *Circulation*, 110(1):16–21.

[29] Brewster, D. C., Cronenwett, J. L., Hallett, John W., J., Johnston, K., Krupski, W. C., and Matsumura, J. S. (2003). Guidelines for the treatment of abdominal aortic aneurysms: Report of a subcommittee of the joint council of the american association for vascular surgery and society for vascular surgery. *Journal of Vascular Surgery*, 37(5):1106–1117.

[30] Brezis, H. (2011). *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York.

[31] Broyden, C. G. (1965). A class of methods for solving nonlinear simultaneous equations. *Mathematics of Computation*, 19(92):577–593.

[32] Bui-Thanh, T., Burstedde, C., Ghattas, O., Martin, J., Stadler, G., and Wilcox, L. C. (2012). Extreme-scale uq for bayesian inverse problems governed by pdes. In *International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 1–11. IEEE Computer Society Press, Los Alamos.

[33] Bui-Thanh, T., Willcox, K., and Ghattas, O. (2008). Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288.

[34] Calvetti, D. and Somersalo, E. (2008). Hypermodels in the bayesian imaging framework. *Inverse Problems*, 24(3).

[35] Carmeli, C., De Vito, E., and Toigo, A. (2006). Vector valued reproducing kernel hilbert spaces of integrable functions and mercers theorem. *Analysis and Applications*, 4(4):377–408.

[36] Cerqueira, M. D., Weissman, N. J., Dilsizian, V., Jacobs, A. K., Kaul, S., Laskey, W. K., Pennell, D. J., Rumberger, J. A., Ryan, T., and Verani, M. S. (2002). Standardized myocardial segmentation and nomenclature for tomographic imaging of the heart. *Circulation*, 105(4):539–542.

[37] Chabiniok, R., Moireau, P., Lesault, P.-F., Rahmouni, A., Deux, J.-F., and Chapelle, D. (2012). Estimation of tissue contractility from cardiac cine-mri using a biomechanical heart model. *Biomechanics and Modeling in Mechanobiology*, 11(5):609–630.

[38] Chan, T. F. and Tai, X.-C. (2003). Identification of discontinuous coefficients in elliptic problems using total variation regularization. *SIAM Journal on Scientific Computing*, 25(3):881–904.

[39] Chappell, M., Groves, A., Whitcher, B., and Woolrich, M. (2009). Variational bayesian inference for a nonlinear forward model. *IEEE Transactions on Signal Processing*, 57(1):223–236.

[40] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. (2000). *Monte Carlo Methods in Bayesian Computation*. Springer Series in Statistics. Springer, New York.

[41] Chib, S. and Greenberg, E. (1995). Understanding the metropolis-hastings algorithm. *The American Statistician*, 49(4):327–335.

[42] Chopin, N. (2004). Central limit theorem for sequential monte carlo methods and its application to bayesian inference. *The Annals of Statistics*, 32(6):2385–2411.

[43] Dacorogna, B. (2008). *Direct Methods in the Calculus of Variations*. Springer, New York.

[44] D'Agostini, G. (2003). Bayesian inference in processing experimental data: principles and basic applications. *Reports on Progress in Physics*, 66(9):1383–1419.

[45] de Céniga, M. V., Gómez, R., Estallo, L., Rodríguez, L., Baquer, M., and Barba, A. (2006). Growth rate and associated factors in small abdominal aortic aneurysms. *European Journal of Vascular and Endovascular Surgery*, 31(3):231–236.

[46] de Putter, S., Wolters, B., Rutten, M., Breeuwer, M., Gerritsen, F., and van de Vosse, F. (2007). Patient-specific initial wall stress in abdominal aortic aneurysms with a backward incremental method. *Journal of Biomechanics*, 40(5):1081–1090.

[47] de Rham, G. (1960). *Variétés différentiables: formes, courants, formes harmoniques*. Actualités scientifiques et industrielles. Éditions Hermann, Paris.

[48] Degroote, J., Couckuyt, I., Vierendeels, J., Segers, P., and Dhaene, T. (2012). Inverse modelling of an aneurysm's stiffness using surrogate-based optimization and fluid-structure interaction simulations. *Structural and Multidisciplinary Optimization*, 46(3):457–469.

[49] Del Moral, P. (2004). *Feynman-Kac Formulae: Genealogical and Interacting Particle Systems with Applications*. Springer, New York.

[50] Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.

[51] Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.

[52] do Carmo, M. P. (1994). *Differential Forms and Applications*. Springer, Berlin, Heidelberg.

[53] Donea, J., Huerta, A., Ponthot, J.-P., and Rodríguez-Ferran, A. (2004). *Arbitrary Lagrangian–Eulerian Methods*, volume 1 of *Encyclopedia of Computational Mechanics*, chapter 14, pages 413–437. John Wiley & Sons, Inc., Hoboken.

[54] Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential monte carlo sampling methods for bayesian filtering. *Statistics and Computing*, 10(3):197–208.

[55] Doyley, M. M. (2012). Model-based elastography: a survey of approaches to the inverse elasticity problem. *Physics in Medicine and Biology*, 57(3):R35–R73.

[56] Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B*, 195(2):216–222.

[57] Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of Applied Mathematics*, 56(3):587–600.

[58] Durrleman, S. (2010). *Statistical models of currents for measuring the variability of anatomical curves, surfaces and their evolution.* Phd thesis, Université Nice Sophia Antipolis.

[59] Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I., and Tutz, G. (2016). *Statistik: Der Weg zur Datenanalyse.* Springer, Berlin, Heidelberg.

[60] Filardo, G., Powell, J. T., Martinez, M. A.-M., and Ballard, D. J. (2015). Surgery for small asymptomatic abdominal aortic aneurysms. *Cochrane Database of Systematic Reviews.*

[61] Fillinger, M. F., Raghavan, M., Marra, S. P., Cronenwett, J. L., and Kennedy, F. E. (2002). In vivo analysis of mechanical wall stress and abdominal aortic aneurysm rupture risk. *Journal of Vascular Surgery*, 36(3):589–597.

[62] Fletcher, R. (1970). A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322.

[63] Fletcher, R. and Reeves, C. M. (1964). Function minimization by conjugate gradients. *The Computer Journal*, 7(2):149–154.

[64] Formaggia, L., Nobile, F., Quarteroni, A., and Veneziani, A. (1999). Multiscale modelling of the circulatory system: a preliminary analysis. *Computing and Visualization in Science*, 2(2):75–83.

[65] Franck, I. and Koutsourelakis, P.-S. (2017). Multimodal, high-dimensional, model-based, bayesian inverse problems with applications in biomechanics. *Journal of Computational Physics*, 329:91–125.

[66] Franck, I. M. and Koutsourelakis, P.-S. (2016). Sparse variational bayesian approximations for nonlinear inverse problems: Applications in nonlinear elastography. *Computer Methods in Applied Mechanics and Engineering*, 299:215–244.

[67] Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., and Penny, W. (2007). Variational free energy and the laplace approximation. *NeuroImage*, 34(1):220–234.

[68] Fung, Y. C. (1990). *Biomechanics: Motion, Flow, Stress, and Growth.* Springer, New York.

[69] Fung, Y.-C. (1993). *Biomechanics: Mechanical Properties of Living Tissues.* Springer, New York.

[70] Galbally, D., Fidkowski, K., Willcox, K., and Ghattas, O. (2010). Non-linear model reduction for uncertainty quantification in large-scale inverse problems. *International Journal for Numerical Methods in Engineering*, 81(12):1581–1608.

[71] Garikipati, K., Arruda, E., Grosh, K., Narayanan, H., and Calve, S. (2004). A continuum treatment of growth in biological tissue: the coupling of mass transport and mechanics. *Journal of the Mechanics and Physics of Solids*, 52(7):1595–1625.

[72] Gasser, T., Auer, M., Labruto, F., Swedenborg, J., and Roy, J. (2010). Biomechanical rupture risk assessment of abdominal aortic aneurysms: Model complexity versus predictability of finite element simulations. *European Journal of Vascular and Endovascular Surgery*, 40(2):176–185.

[73] Gasser, T. C., Görgül, G., Folkesson, M., and Swedenborg, J. (2008). Failure properties of intraluminal thrombus in abdominal aortic aneurysm under static and pulsating mechanical loads. *Journal of Vascular Surgery*, 48(1):179–188.

[74] Gasser, T. C., Ogden, R. W., and Holzapfel, G. A. (2006). Hyperelastic modelling of arterial layers with distributed collagen fibre orientations. *Journal of The Royal Society Interface*, 3(6):15–35.

[75] GBD 2013 Mortality and Causes of Death Collaborators (2015). Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 385(9963):117–171.

[76] Gee, M., Reeps, C., Eckstein, H., and Wall, W. (2009). Prestressing in finite deformation abdominal aortic aneurysm simulation. *Journal of Biomechanics*, 42(11):1732–1739.

[77] Gee, M. W., Förster, C., and Wall, W. A. (2010). A computational strategy for prestressing patient-specific biomechanical problems under finite deformation. *International Journal for Numerical Methods in Biomedical Engineering*, 26(1):52–72.

[78] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

[79] Gentle, J. E. (2007). *Matrix Algebra: Theory, Computations, and Applications in Statistics*. Springer, New York.

[80] Gerbeau, J.-F., Vidrascu, M., and Frey, P. (2005). Fluid–structure interaction in blood flows on geometries based on medical imaging. *Computers & Structures*, 83(2–3):155–165.

[81] Geweke, J. (1989). Bayesian inference in econometric models using monte carlo integration. *Econometrica*, 57(6):1317–1339.

[82] Girolami, M. and Calderhead, B. (2011). Riemann manifold langevin and hamiltonian monte carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(2):123–214.

[83] Giusti, E. (1984). *Minimal Surfaces and Functions of Bounded Variation*. Monographs in Mathematics. Birkhäuser, Boston.

[84] Glaunès, J., Trouvé, A., and Younes, L. (2004). Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 712–718.

[85] Glaunès, J. A. (2005). *Transport par difféomorphismes de point, de mesures et de courant pour la comparaison de formes et l'anatomie numérique*. PhD thesis, Université Paris 13.

[86] Goldfarb, D. (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26.

[87] Goldman, L. W. (2007). Principles of ct: Radiation dose and image quality. *Journal of Nuclear Medicine Technology*, 35(4):213–225.

[88] Golub, G. H. and Van Loan, C. F. (2013). *Matrix computations*. Johns Hopkins studies in the mathematical sciences. Johns Hopkins University Press, Baltimore, 4th edition.

[89] Grinfeld, E. M. (2015). *Mathematical Tools for Physicists*. Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, 2nd edition.

[90] Hall, A. J., Busse, E. F., McCarville, D. J., and Burgess, J. J. (2000). Aortic wall tension as a predictive factor for abdominal aortic aneurysm rupture: Improving the selection of patients for abdominal aortic aneurysm repair. *Annals of Vascular Surgery*, 14(2):152–157.

[91] Hammersley, J. and Clifford, P. (1971). Markov fields on finite graphs and lattices. http://www.statslab.cam.ac.uk/ grg/books/hammfest/hamm-cliff.pdf, Accessed 31 May 2017.

[92] Hartmann, S. and Neff, P. (2003). Polyconvexity of generalized polynomial-type hyperelastic strain energy functions for near-incompressibility. *International Journal of Solids and Structures*, 40(11):2767–2791.

[93] Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

[94] Held, L. and Sabanés Bové, D. (2014). *Applied statistical inference : likelihood and bayes*. Springer, Berlin,Heidelberg.

[95] Heroux, M., Bartlett, R., Hoekstra, V. H. R., Hu, J., Kolda, T., Lehoucq, R., Long, K., Pawlowski, R., Phipps, E., Salinger, A., Thornquist, H., Tuminaro, R., Willenbring, J., and Williams, A. (2003). An overview of trilinos. Technical Report SAND2003-2927, Sandia National Laboratories.

[96] Hidane, M., Lézoray, O., and Elmoataz, A. (2013). Nonlinear multilayered representation of graph-signals. *Journal of Mathematical Imaging and Vision*, 45(2):114–137.

[97] Himpel, G., Kuhl, E., Menzel, A., and Steinmann, P. (2005). Computational modelling of isotropic multiplicative growth. *Computer Modeling in Engineering & Science*, 8:119–134.

[98] Hirschvogel, M., Bassilious, M., Jagschies, L., Wildhirt, S. M., and Gee, M. W. (2017). A monolithic 3d-0d coupled closed-loop model of the heart and the vascular system: Experiment-based parameter estimation for patient-specific cardiac mechanics. *International Journal for Numerical Methods in Biomedical Engineering*, pages e2842–n/a.

[99] Holzapfel, G. A. (2000). *Nonlinear solid mechanics.* John Wiley & Sons, Inc., Hoboken.

[100] Humphrey, J. and Holzapfel, G. (2012). Mechanics, mechanobiology, and modeling of human abdominal aorta and aneurysms. *Journal of Biomechanics*, 45(5):805–814. Special Issue on Cardiovascular Solid Mechanics.

[101] Humphrey, J. D. (2002). *Cardiovascular Solid Mechanics: Cells, Tissues, and Organs.* Springer, New York.

[102] Humphrey, J. D. and Rajagopal, K. R. (2002). A constrained mixture model for growth and remodeling of soft tissue. *Mathematical Models and Methods in Applied Sciences*, 12(03):407–430.

[103] Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing In Science & Engineering*, 9(3):90–95.

[104] Imperiale, A., Routier, A., Durrleman, S., and Moireau, P. (2013). Improving efficiency of data assimilation procedure for a biomechanical heart model by representing surfaces as currents. In Ourselin, S., Rueckert, D., and Smith, N., editors, *7th International Conference on Functional Imaging and Modeling of the Heart*, volume 7945 of *Lecture Notes in Computer Science*, pages 342–351. Springer, Berlin, Heidelberg.

[105] Ismail, M., Wall, W. A., and Gee, M. W. (2013). Adjoint-based inverse analysis of windkessel parameters for patient-specific vascular models. *Journal of Computational Physics*, 244:113–130.

[106] Jaakkola, T. S. and Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10(1):25–37.

[107] Jerrum, M. (2003). *Counting, Sampling and Integrating: Algorithm and Complexity.* Lectures in Mathematics ETH Zürich. Birkhäuser, Basel.

[108] Jolliffe, I. (2002). *Principal Component Analysis.* Springer, New York.

[109] Joshi, S. and Miller, M. (2000). Landmark matching via large deformation diffeomorphisms. *IEEE Transactions on Image Processing*, 9(8):1357–1370.

[110] Kaipio, J. and Somersalo, E. (2006). *Statistical and Computational Inverse Problems.* Applied Mathematical Sciences. Springer, New York.

[111] Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795.

[112] Kehl, S. and Gee, M. W. (2017). Calibration of parameters for cardiovascular models with application to arterial growth. *International Journal for Numerical Methods in Biomedical Engineering*, 33(5):e2822–n/a.

[113] Kelley, C. (2003). *Solving Nonlinear Equations with Newton's Method.* Society for Industrial and Applied Mathematics, Philadelphia.

[114] Kelley, C. and Keyes, D. (1998). Convergence analysis of pseudo-transient continuation. *SIAM Journal on Numerical Analysis*, 35(2):508–523.

[115] Kelley, C. T. (1999). *Iterative methods for optimization.* Society for Industrial and Applied Mathematics, Philadelphia.

[116] Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464.

[117] Kent, K. C. (2014). Abdominal aortic aneurysms. *New England Journal of Medicine*, 371(22):2101–2108.

[118] Khalil, A. S., Bouma, B. E., and Kaazempur Mofrad, M. R. (2006). A combined fem/genetic algorithm for vascular soft tissue elasticity estimation. *Cardiovascular Engineering*, 6(3):93–102.

[119] Koch, K.-R. (2000). *Einführung in die Bayes-Statistik.* Springer, Berlin, Heidelberg.

[120] Kock, S. A., Nygaard, J. V., Eldrup, N., Fründ, E.-T., Klærke, A., Paaske, W. P., Falk, E., and Kim, W. Y. (2008). Mechanical stresses in carotid plaques using mri-based fluid–structure interaction models. *Journal of Biomechanics*, 41(8):1651–1658.

[121] Konishi, S. and Kitagawa, G. (2008). *Information Criteria and Statistical Modeling.* Springer, New York.

[122] Koutsourelakis, P.-S. (2009a). Accurate uncertainty quantification using inaccurate computational models. *SIAM Journal on Scientific Computing*, 31(5):3274–3300.

[123] Koutsourelakis, P.-S. (2009b). A multi-resolution, non-parametric, bayesian framework for identification of spatially-varying model parameters. *Journal of Computational Physics*, 228(17):6184–6211.

[124] Koutsourelakis, P.-S. (2016). Variational bayesian strategies for high-dimensional, stochastic design problems. *Journal of Computational Physics*, 308:124–152.

[125] Krishnamurthy, A., Villongco, C. T., Chuang, J., Frank, L. R., Nigam, V., Belezzuoli, E., Stark, P., Krummen, D. E., Narayan, S., Omens, J. H., McCulloch, A. D., and Kerckhoffs, R. C. (2013). Patient-specific models of cardiac biomechanics. *Journal of Computational Physics*, 244:4–21.

[126] Kroon, M. and Holzapfel, G. A. (2008). Estimation of the distributions of anisotropic, elastic properties and wall stresses of saccular cerebral aneurysms by inverse analysis. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 464(2092):807–825.

[127] Kuehnel, W. (2015). *Differential Geometry Curves-Surfaces-Manifolds.* Student Mathematical Library. AMS American Mathematical Society, Providence.

[128] Kuhl, E., Maas, R., Himpel, G., and Menzel, A. (2007). Computational modeling of arterial wall growth. *Biomechanics and Modeling in Mechanobiology*, 6(5):321–331.

[129] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86.

[130] Kurvers, H., Veith, F. J., Lipsitz, E. C., Ohki, T., Gargiulo, N. J., Cayne, N. S., Suggs, W. D., Timaran, C. H., Kwon, G. Y., Rhee, S. J., and Santiago, C. (2004). Discontinuous, staccato growth of abdominal aortic aneurysms. *Journal of the American College of Surgeons*, 199(5):709–715.

[131] Larson, M. and Bengzon, F. (2013). *The Finite Element Method: Theory, Implementation and Application.* Texts in Computational Science and Egineering. Springer, Berlin, Heidelberg.

[132] Lasheras, J. C. (2007). The biomechanics of arterial aneurysms. *Annual Review of Fluid Mechanics*, 39:293–319.

[133] Lederle, F. A., Wilson, S. E., Johnson, G. R., Reinke, D. B., Littooy, F. N., Acher, C. W., Ballard, D. J., Messina, L. M., Gordon, I. L., Chute, E. P., Krupski, W. C., Busuttil, S. J., Barone, G. W., Sparks, S., Graham, L. M., Rapp, J. H., Makaroun, M. S., Moneta, G. L., Cambria, R. A., Makhoul, R. G., Eton, D., Ansel, H. J., Freischlag, J. A., and Bandyk, D. (2002). Immediate repair compared with surveillance of small abdominal aortic aneurysms. *New England Journal of Medicine*, 346(19):1437–1444.

[134] Lee, E. H. (1969). Elastic-plastic deformation at finite strains. *Journal of Applied Mechanics*, 36(1):1–6.

[135] Lehmann, E. L. and Casella, G. (1998). *Theory of point estimation.* Springer texts in statistics. Springer, New York, 2nd edition.

[136] Leung, J. H., Wright, A. R., Cheshire, N., Crane, J., Thom, S. A., Hughes, A. D., and Xu, Y. (2006). Fluid structure interaction of patient specific abdominal aortic aneurysms: a comparison with solid stress models. *BioMedical Engineering OnLine*, 5(1-33).

[137] LeVeque, R. (2007). *Finite Difference Methods for Ordinary and Partial Differential Equations.* Society for Industrial and Applied Mathematics, Philadelphia.

[138] Lewicki, M. S. and Sejnowski, T. J. (2000). Learning overcomplete representations. *Neural Computation*, 12(2):337–365.

[139] Lipponen, A., Seppänen, A., and Kaipio, J. (2013). Electrical impedance tomography imaging with reduced-order model based on proper orthogonal decomposition. *Journal of Electronic Imaging*, 22(2):023008–023008.

[140] Liu, D. and Nocedal, J. (1989). On the limited memory bfgs method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528.

[141] Liu, J. S. (2004). *Monte Carlo Strategies in Scientific Computing.* Springer Series in Statistics. Springer, New York.

[142] Lu, J., Zhou, X., and Raghavan, M. L. (2007). Inverse elastostatic stress analysis in pre-deformed biological structures: Demonstration using abdominal aortic aneurysms. *Journal of Biomechanics*, 40(3):693–696.

[143] Lubarda, V. and Hoger, A. (2002). On the mechanics of solids with a growing mass. *International Journal of Solids and Structures*, 39(18):4627–4664.

[144] MacKay, D. J. (1998). Choice of basis for laplace approximation. *Machine Learning*, 33(1):77–86.

[145] MacKay, D. J. C. (2003). *Information theory, inference, and learning algorithms.* Cambridge University Press, Cambridge, 4th edition.

[146] Madsen, K., Nielsen, H. B., and Tingleff, O. (1999). Methods for non-linear least squares problems. Technical report, Technical University of Denmark.

[147] Maier, A. (2012). *Computational Modeling of Rupture Risk in Abdominal Aortic Aneurysms.* Phd thesis, Technische Universität München, München.

[148] Maier, A., Gee, M., Reeps, C., Pongratz, J., Eckstein, H.-H., and Wall, W. (2010). A comparison of diameter, wall stress, and rupture potential index for abdominal aortic aneurysm rupture risk prediction. *Annals of Biomedical Engineering*, 38(10):3124–3134.

[149] Manzoni, A., Pagani, S., and Lassila, T. (2016). Accurate solution of bayesian inverse uncertainty quantification problems combining reduced basis methods and reduction error models. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):380–412.

[150] Marchesseau, S., Delingette, H., Sermesant, M., Cabrera-Lozoya, R., Tobon-Gomez, C., Moireau, P., i Ventura, R. F., Lekadir, K., Hernandez, A., Garreau, M., Donal, E., Leclercq, C., Duckett, S., Rhode, K., Rinaldi, C., Frangi, A., Razavi, R., Chapelle, D., and Ayache, N. (2013). Personalization of a cardiac electromechanical model using reduced order unscented kalman filtering from regional volumes. *Medical Image Analysis*, 17(7):816–829.

[151] Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.

[152] Martin, J., Wilcox, L. C., Burstedde, C., and Ghattas, O. (2012). A stochastic newton mcmc method for large-scale statistical inverse problems with application to seismic inversion. *SIAM Journal on Scientific Computing*, 34(3):A1460–A1487.

[153] Martínez, J. M. (1991). Quasi-newton methods for solving underdetermined nonlinear simultaneous equations. *Journal of Computational and Applied Mathematics*, 34(2):171–190.

[154] Martufi, G., Auer, M., Roy, J., Swedenborg, J., Sakalihasan, N., Panuccio, G., and Gasser, T. C. (2013). Multidimensional growth measurements of abdominal aortic aneurysms. *Journal of Vascular Surgery*, 58(3):748–755.

[155] Matsumoto, M. and Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1):3–30.

[156] Metropolis, N. and Ulam, S. (1949). The monte carlo method. *Journal of the American Statistical Association*, 44(247):335–341.

[157] Meyn, S. P. and Tweedie, R. L. (2009). *Markov chains and stochastic stability.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2nd edition.

[158] Miga, M. I. (2003). A new approach to elastography using mutual information and finite elements. *Physics in Medicine and Biology*, 48(4):467–480.

[159] Miller, M., Trouve, A., and Younes, L. (2002). On the metrics and euler-lagrange equations of computational anatomy. *Annual Review of Biomedical Engineering*, 4(1):375–405.

[160] Modersitzki, J. (2003). *Numerical methods for image registration.* Oxford University Press, Oxford.

[161] Moireau, P., Bertoglio, C., Xiao, N., Figueroa, C. A., Taylor, C. A., Chapelle, D., and Gerbeau, J.-F. (2013). Sequential identification of boundary support parameters in a fluid-structure vascular model using patient image data. *Biomechanics and Modeling in Mechanobiology*, 12(3):475–496.

[162] Moireau, P., Chapelle, D., and Le Tallec, P. (2009). Filtering for distributed mechanical systems using position measurements: perspectives in medical imaging. *Inverse Problems*, 25(3):035010.

[163] Moireau, P., Chapelle, D., and Tallec, P. L. (2008). Joint state and parameter estimation for distributed mechanical systems. *Computer Methods in Applied Mechanics and Engineering*, 197(6–8):659–677.

[164] Nagler, A., Bertoglio, C., Stoeck, C. T., Kozerke, S., and Wall, W. A. (2017). Maximum likelihood estimation of cardiac fiber bundle orientation from arbitrarily spaced diffusion weighted images. *Medical Image Analysis*, 39:56–77.

[165] Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing*, 11(2):125–139.

[166] Neal, R. M. (2011). Mcmc using hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, Handbooks of Modern Statistical Methods. Chapman & Hall/CRC, Boca Raton.

[167] Negri, F., Rozza, G., Manzoni, A., and Quarteroni, A. (2013). Reduced basis method for parametrized elliptic optimal control problems. *SIAM Journal on Scientific Computing*, 35(5):A2316–A2340.

[168] Nocedal, J. (1980). Updating quasi-newton matrices with limited storage. *Mathematics of Computation*, 35(151):773–782.

[169] Nocedal, J. and Wright, S. J. (2006). *Numerical Optimization*. Springer, New York.

[170] Oberai, A. A., Gokhale, N. H., and Feijoo, G. R. (2003). Solution of inverse problems in elasticity imaging using the adjoint method. *Inverse Problems*, 19(2):297–313.

[171] Olive, D. J. (2014). *Statistical theory and inference*. Springer, Cham.

[172] Olufsen, M. S., Peskin, C. S., Kim, W. Y., Pedersen, E. M., Nadim, A., and Larsen, J. (2000). Numerical simulation and experimental validation of blood flow in arteries with structured-tree outflow conditions. *Annals of Biomedical Engineering*, 28(11):1281–1299.

[173] Parisi, G. (1988). *Statistical field theory*. Frontiers in physics; 66. Advanced book program. Addison-Wesley, Redwood City.

[174] Pedregal, P. (2000). *Variational Methods in Nonlinear Elasticity*. Society for Industrial and Applied Mathematics, Philadelphia.

[175] Petra, N., Zhu, H., Stadler, G., Hughes, T. J., and Ghattas, O. (2012). An inexact gauss-newton method for inversion of basal sliding and rheology parameters in a nonlinear stokes ice sheet model. *Journal of Glaciology*, 58(211).

[176] Polzer, S. and Gasser, T. C. (2015). Biomechanical rupture risk assessment of abdominal aortic aneurysms based on a novel probabilistic rupture risk index. *Journal of The Royal Society Interface*, 12(113).

[177] Powell, J. (2007). Final 12-year follow-up of surgery versus surveillance in the uk small aneurysm trial. *British Journal of Surgery*, 94(6):702–708.

[178] Quarteroni, A., Sacco, R., and Saleri, F. (2007). *Numerical Mathematics*. Springer, New York.

[179] Raghavan, M. and Vorp, D. A. (2000). Toward a biomechanical tool to evaluate rupture potential of abdominal aortic aneurysm: identification of a finite strain constitutive model and evaluation of its applicability. *Journal of Biomechanics*, 33(4):475–482.

[180] Raiffa, H. and Schlaifer, R. (1961). *Applied statistical decision theory.* Studies in managerial economics. Division of Research, Graduate School of Business Administration, Harvard University, Boston.

[181] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian Processes for Machine Learning.* The MIT Press, Cambridge, 2nd edition.

[182] Reeps, C., Gee, M., Maier, A., Gurdan, M., Eckstein, H.-H., and Wall, W. A. (2010). The impact of model assumptions on results of computational mechanics in abdominal aortic aneurysm. *Journal of Vascular Surgery*, 51(3):679–688.

[183] Robert, C. P. (2007). *The bayesian choice : from decision-theoretic foundations to computational implementation.* Springer texts in statistics. Springer, New York, 2nd edition.

[184] Robert, C. P. (2012). Bayesian computational methods. In Gentle, J. E., Härdle, W. K., and Mori, Y., editors, *Handbook of Computational Statistics: Concepts and Methods*, Springer Handbooks of Computational Statistics, pages 751–805. Springer, Berlin, Heidelberg.

[185] Robert, C. P. and Casella, G. (2004). *Monte Carlo Statistical Methods.* Springer, New York.

[186] Roberts, G. O., Gelman, A., and Gilks, W. R. (1997). Weak convergence and optimal scaling of random walk metropolis algorithms. *Annals of Applied Probability*, 7(1):110–120.

[187] Rodriguez, E. K., Hoger, A., and McCulloch, A. D. (1994). Stress-dependent finite growth in soft elastic tissues. *Journal of Biomechanics*, 27(4):455–467.

[188] Rohatgi, V. K. and Saleh, A. K. M. E. (2015). *An Introduction to Probability and Statistics.* John Wiley & Sons, Inc., Hoboken.

[189] Rubinstein, R. Y. and Kroese, D. P. (2008). *Simulation and the Monte Carlo method.* Wiley series in probability and statistics. John Wiley & Sons, Inc., Hoboken, 2nd edition.

[190] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268.

[191] Russell, S. J. and Norvig, P. (2016). *Artificial Intelligence: A modern approach.* Always learning. Pearson, Boston, 3rd edition.

[192] Saad, Y. (2003). *Iterative Methods for Sparse Linear Systems.* Society for Industrial and Applied Mathematics, 2nd edition.

[193] Sagan, H. (1992). *The Calculus of Variations.* Dover Books on Mathematics. Dover Publications, New York.

[194] Schay, G. (2016). *Introduction to Probability with Statistical Applications.* Springer, Cham.

[195] Schlösser, F. J., Tangelder, M. J., Verhagen, H. J., van der Heijden, G. J., Muhs, B. E., van der Graaf, Y., and Moll, F. L. (2008). Growth predictors and prognosis of small abdominal aortic aneurysms. *Journal of Vascular Surgery*, 47(6):1127–1133.

[196] Schölkopf, B., Herbrich, R., and Smola, A. J. (2001). A generalized representer theorem. In Helmbold, D. and Williamson, B., editors, *Computational Learning Theory COLT 2001*, volume 2111 of *Lecture Notes in Computer Science*, pages 416–426, Berlin, Heidelberg. Springer.

[197] Schölkopf, B., Smola, A., and Müller, K.-R. (1997). Kernel principal component analysis. In Gerstner, W., Germond, A., Hasler, M., and Nicoud, J.-D., editors, *Artificial Neural Networks — ICANN'97*, volume 1327 of *Lecture Notes in Computer Science*, pages 583–588. Springer, Berlin, Heidelberg.

[198] Sermesant, M., Moireau, P., Camara, O., Sainte-Marie, J., Andriantsimiavona, R., Cimrman, R., Hill, D., Chapelle, D., and Razavi, R. (2006). Cardiac function estimation from mri using a heart model and data assimilation: Advances and difficulties. *Medical Image Analysis*, 10(4):642–656. Special Issue on Functional Imaging and Modelling of the Heart (FIMH 2005).

[199] Seyedsalehi, S., Zhang, L., Choi, J., and Baek, S. (2015). Prior distributions of material parameters for bayesian calibration of growth and remodeling computational model of abdominal aortic wall. *Journal of Biomechanical Engineering*, 137(10):101001–101001–13.

[200] Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656.

[201] Shum, J., Martufi, G., Di Martino, E., Washington, C. B., Grisafi, J., Muluk, S. C., and Finol, E. A. (2011). Quantitative assessment of abdominal aortic aneurysm geometry. *Annals of Biomedical Engineering*, 39(1):277–286.

[202] Simon, D. (2006). *Optimal State Estimation.* John Wiley & Sons, Inc.

[203] Sirkes, Z. and Tziperman, E. (1997). Finite difference of adjoint or adjoint of finite difference? *Monthly Weather Review*, 125:3373–3378.

[204] Skalak, R., Dasgupta, G., Moss, M., Otten, E., Dullemeijer, P., and Vilmann, H. (1982). Analytical description of growth. *Journal of Theoretical Biology*, 94(3):555–577.

[205] Skalak, R., Zargaryan, S., Jain, R. K., Netti, P. A., and Hoger, A. (1996). Compatibility and the genesis of residual stress by volumetric growth. *Journal of Mathematical Biology*, 34(8):889–914.

[206] Sucar, L. E. (2015). Markov random fields. In *Probabilistic Graphical Models: Principles and Applications*, Advances in Computer Vision and Pattern Recognition, pages 83–99. Springer, London.

[207] Sweeting, M. J. and Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biometrical Journal*, 53(5):750–763.

[208] Taber, L. A. and Humphrey, J. D. (2001). Stress-modulated growth, residual stress, and vascular heterogeneity. *Journal of Biomechanical Engineering*, 123(6):528–535.

[209] Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions.* Springer Series in Statistics. Springer, New York, 3rd edition.

[210] The Multicentre Aneurysm Screening Study Group (2002). The multicentre aneurysm screening study (mass) into the effect of abdominal aortic aneurysm screening on mortality in men: a randomised controlled trial. *The Lancet*, 360(9345):1531–1539.

[211] The RESCAN Collaborators (2013). Surveillance intervals for small abdominal aortic aneurysms: A meta-analysis. *JAMA*, 309(8):806–813.

[212] The UK Small Aneurysm Trial Participants (1998). Mortality results for randomised controlled trial of early elective surgery or ultrasonographic surveillance for small abdominal aortic aneurysms. *The Lancet*, 352(9141):1649–1655.

[213] The Xfig developers (2002). *Xfig User Manual Version 3.2.5.*

[214] Theys, C. and Aime, C. (2016). Reconstructing images in astrophysics, an inverse problem point of view. In Rozelot, J.-P. and Neiner, C., editors, *Cartography of the Sun and the Stars*, volume 914 of *Lecture Notes in Physics*, pages 1–23. Springer, Cham.

[215] Thomaz, C., Giraldi, G., Costa, J., and Gillies, D. (2013). A priori-driven pca. In Park, J.-I. and Kim, J., editors, *Computer Vision - ACCV 2012 Workshops: ACCV 2012*, volume 7729 of *Lecture Notes in Computer Science*, pages 236–247. Springer, Berlin, Heidelberg.

[216] Thompson, D. W. (1917). *On growth and form.* Cambridge University Press, Cambridge.

[217] Tikhonov, A. N., Goncharsky, A. V., Stepanov, V. V., and Yagola, A. G. (1995). *Regularization methods.* Springer, Dordrecht.

[218] Tinkl, S. (2015). *Towards a predictive computational growth model for aneurysms.* Phd thesis, Technische Universität München.

[219] Tolk, A. (2015). Learning something right from models that are wrong: Epistemology of simulation. In Yilmaz, L., editor, *Concepts and Methodologies for Modeling and Simulation: A Tribute to Tuncer Ören*, Simulation Foundations, Methods and Applications, pages 87–106. Springer, Cham.

[220] Truesdell, C. and Noll, W. (2004). *The Non-Linear Field Theories of Mechanics.* Springer, Berlin, Heidelberg.

[221] Vaillant, M. and Glaunès, J. (2005). Surface matching via currents. In Christensen, G. E. and Sonka, M., editors, *Information Processing in Medical Imaging*, volume 3565 of *Lecture Notes in Computer Science*, pages 381–392. Springer, Berlin, Heidelberg.

[222] Vorp, D. A. (2007). Biomechanics of abdominal aortic aneurysm. *Journal of Biomechanics*, 40(9):1887–1902.

[223] Wall, W. A. and Gee, M. W. (2010). Baci: A parallel multiphysics simulation environment. Technical report, Institute for Computational Mechanics, Technische Universität München.

[224] Wang, Z., Navon, I. M.and Dimet, F. X., and Zou, X. (1992). The second order adjoint analysis: Theory and applications. *Meteorology and Atmospheric Physics*, 50(1):3–20.

[225] Weisbecker, H., Pierce, D. M., and Holzapfel, G. A. (2014). A generalized prestressing algorithm for finite element simulations of preloaded geometries with application to the aorta. *International Journal for Numerical Methods in Biomedical Engineering*, 30(9):857–872.

[226] Wells, W. M., Viola, P., Atsumi, H., Nakajima, S., and Kikinis, R. (1996). Multimodal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51.

[227] Wilmink, A., Quick, C., Hubbard, C. S., and Day, N. (2000). The association between connective tissue laxity and the risk of an abdominal aortic aneurysm. *European Journal of Vascular and Endovascular Surgery*, 20(3):290–295.

[228] Xiao, N., Humphrey, J. D., and Figueroa, C. A. (2013). Multi-scale computational model of three-dimensional hemodynamics within a deformable full-body arterial network. *Journal of Computational Physics*, 244:22–40.

[229] Yang, H. and Veneziani, A. (2015). Estimation of cardiac conductivities in ventricular tissue by a variational approach. *Inverse Problems*, 31(11):115001.

[230] Zeinali-Davarani, S. and Baek, S. (2012). Medical image-based simulation of abdominal aortic aneurysm growth. *Mechanics Research Communications*, 42:107–117.

[231] Zeinali-Davarani, S., Raguin, L., and Baek, S. (2011a). An inverse optimization approach toward testing different hypotheses of vascular homeostasis using image-based models. *The International Journal of Structural Changes in Solids*, 3(2):33–45.

[232] Zeinali-Davarani, S., Raguin, L., Vorp, D., and Baek, S. (2011b). Identification of in vivo material and geometric parameters of a human aorta: toward patient-specific modeling of abdominal aortic aneurysm. *Biomechanics and Modeling in Mechanobiology*, 10(5):689–699.

[233] Zienkiewicz, O., Taylor, R., and Zhu, J. (2013). *The Finite Element Method: Its Basis and Fundamentals.* Butterworth-Heinemann, Oxford, 7th edition.