Adaptive Algorithms in Optimization under PDE Constraints

Dissertation

zur Erlangung des Grades eines
Doktors der Naturwissenschaften (Dr. rer. nat.)

am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von

Marian P. Moldenhauer

Berlin, November 2020

Erstgutachter:
Prof. Dr. Ralf Kornhuber
Fachbereich Mathematik und Informatik
Freie Universität Berlin
Arnimalle 6
14195 Berlin

Zweitgutachter:
Prof. Dr. Anton Schiela
Mathematisches Institut
Universität Bayreuth
95440 Bayreuth

Tag der Disputation: 09. Juni 2021

Selbstständigkeitserklärung

Name: Moldenhauer
Vorname: Marian

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht.

Mit einer Prüfung meiner Arbeit durch ein Plagiatsprüfungsprogramm erkläre ich mich einverstanden.

Berlin, den

.....
Unterschrift

Contents

1 Introduction

Hip replacement surgeries are a frequent procedure in developed countries, e.g., more than 167,000 cases were reported to the EPRD (Endoprosthesis register Germany) for Germany in 2018 [73] of which 10% are replacement surgeries. Eurostat* even lists as many as 257,487 hip replacement surgeries for Germany in 2018. The trend of increasing replacement surgeries has been rising since 2006 not only in Germany, but also in the USA [20, 165]. Unfortunately, most joint reconstructions have to be replaced after 10–15 years, with at least 10% failing within the first ten years [15, 73]. The necessary secondary replacements are particularly expensive and burdensome for the patient.

Reasons for failure include wear, loosening, and migration of the implant. To some extent these factors are affected by the positioning of the implant during surgery. Current practice of surgery planning is based on planar X-ray images, with recent developments to take the 3D patient anatomy into account [43]. While kinematics, in particular range of motion and impingement, have been considered in the research literature [4, 28, 151, 152], even including hip joint implants [29, 30], the mechanical loading due to day-to-day motions notably addressing the issue of stress shielding has not been included in surgery planning so far. Yet it is the mechanical situation in the joint that drives bone remodeling and micro fractures and therefore mainly determines implant loosening and migration. A model based optimization of implant positioning, taking the mechanical loads into account, can therefore be expected to improve implant life time and to reduce the number of secondary replacements.

Here we investigate a comprehensive approach to mathematical optimization of implant position optimization. As a starting point, we set up a complete optimization problem in Sec. 2.1 that would ideally be solved, specifying design variables, objective, loads, and state equation.

Since this problem will turn out to be out of reach of reasonable computing power and availability of patient-specific data in the foreseeable future, we will consider pragmatic model simplifications and end up with a reduced optimization problem that is still demanding, but not beyond current computing capacities. An approach will be explained how to model the motion parameter domain to implicitly include the time component and which allows for patient-specific optimization. The simulation also needs to incorporate a contact condition, which will be briefly introduced.

Subsequently, we will develop efficient algorithmic approaches for solving this optimization problem. Since the main source of computational demand is the integration over a moderately high-dimensional load domain, we consider appropriate surrogate modeling with Kriging for objective and gradient evaluation in Sec. 2.2, with adaptive selection of Kriging evaluation points, treating smooth and irregular parts of the objective separately, and Monte Carlo integration on the surrogate model further explained in Sec. 3.

A gradient-based optimization algorithm exploiting the accuracy-effort trade-off provided by the surrogate model is then considered in Sec. 4, and particular attention paid to line search in the presence of grid topology-induced discontinuities of the objective.

Finally, preliminary numerical results are given in Sec. 5.

The last section will give a summary and an outlook to possible future research.

*https://ec.europa.eu/eurostat/de/web/health/data/database, table: hlth_co_proc2, codes: ICD-9-CM81.51-81.53

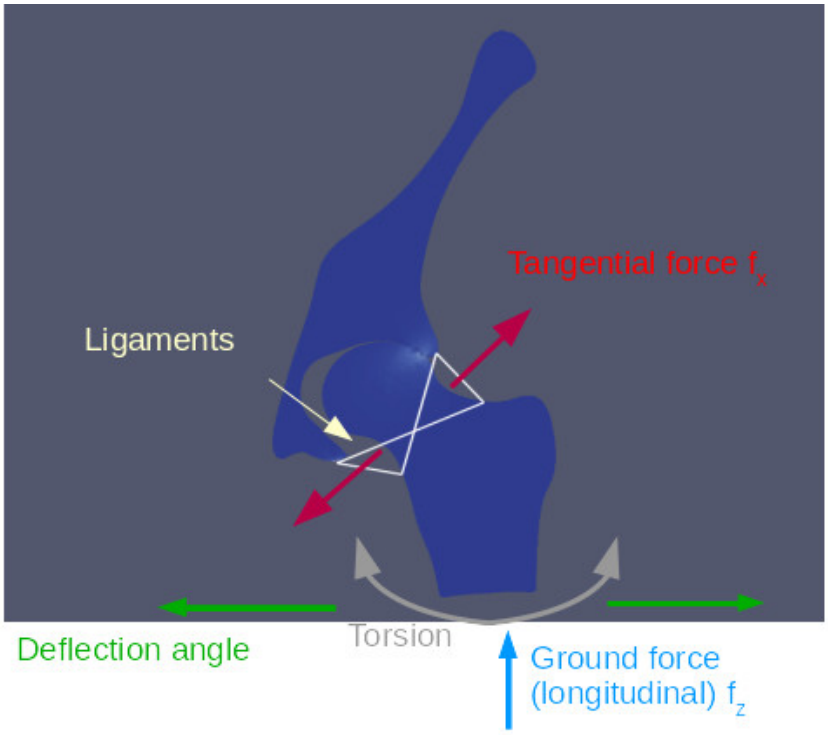

Fig. 1: *The 2D setup of the hip including ligaments next to longitudinal and tangential forces and torsion.*

1.1 Problem setting and computational challenges

Before we explain the computational and algorithmic issues and how to deal with them, we will first introduce the geometric setup and some notations. Looking at Fig. 1, we see a cross-section of a hip joint with a part of the pelvis on the top and the upper end of the femur on the bottom. In this exemplary 2D setting, the depicted loads are the longitudinal $f_z \in F_z$ and tangential forces $f_x \in F_x$ next to torsion $m_y \in M_y$. Motions $m \in \mathcal{M}$ are trajectories in the load domain $\mathcal{L}$, i.e., they map from a time interval $[0, T_m]$ to smooth curves in $\mathcal{L}$ with

$$\mathcal{L} := F_x \times F_y \times F_z \times M_x \times M_y \times M_z.$$

The intricate ligaments (cf. Fig. 2) are approximated in a rather simplistic way of four lines. The setup including the implant can be seen in Fig. 3 while Fig. 4 gives an example of a coarse FE-discretization. Here, it becomes clear that a finer discretization in the estimated contact region and possibly along the fixed contact between implant and bone is needed. For this we need to apply adaptive refinement of the discretization.

On the other hand, utilizing the measured load data from the open database OrthoLoad[†] (see, e.g., [13, 14]) we find that loads need to at least have the range (in N for the forces and Nm for the moments)

$$\begin{aligned} \mathcal{L} &:= F_x \times F_y \times F_z \times M_x \times M_y \times M_z \\ &= [-4000, 0] \times [-200, 3000] \times [0, 10{,}000] \times [-4, 5] \times [-2, 3] \times [-2, 2]. \end{aligned} \tag{1.1}$$

[†] https://orthoload.com/

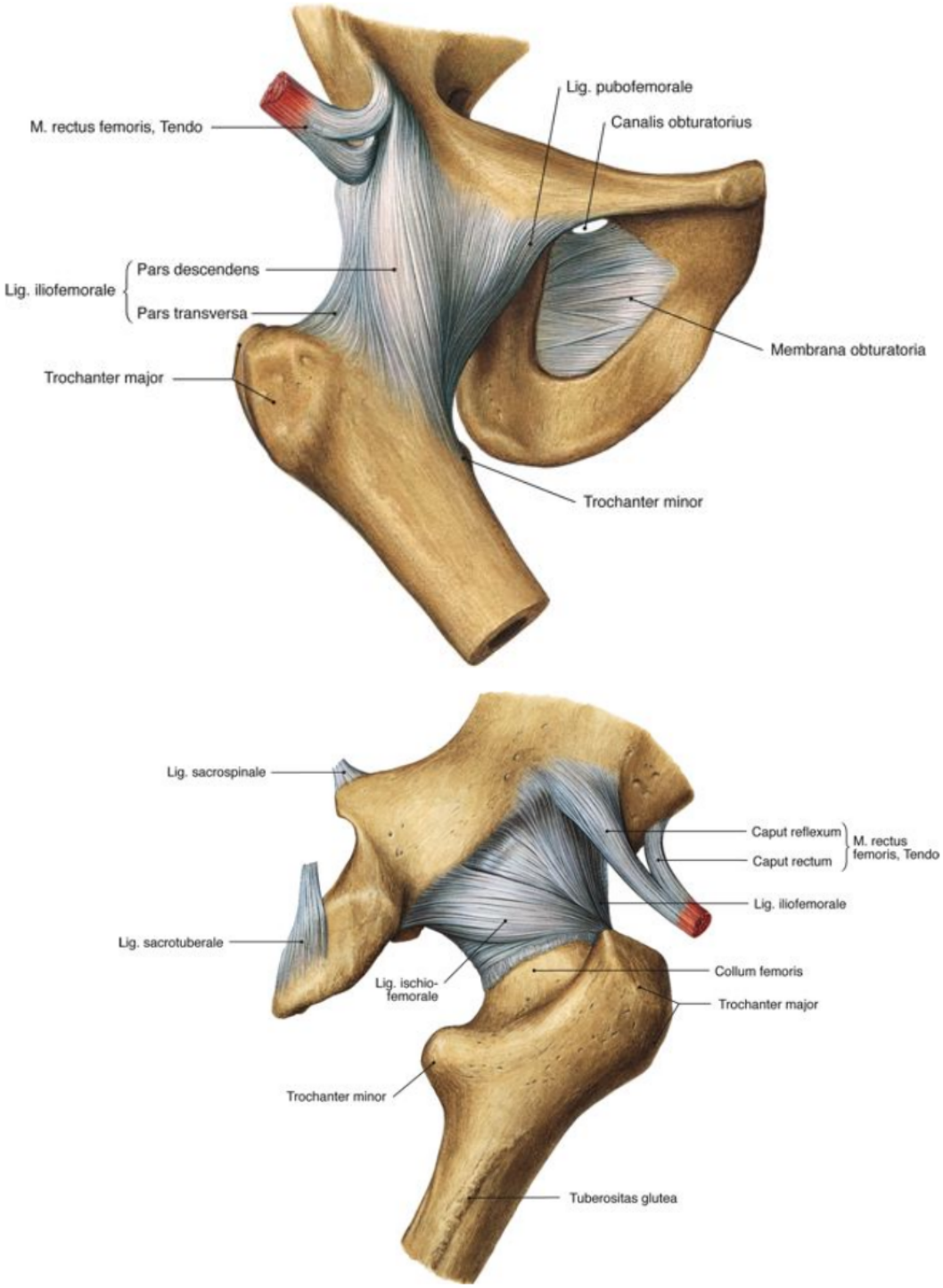

Fig. 2: *Hip anatomy: bones and ligaments [125]. Top: ventral distal; bottom: dorsal.*

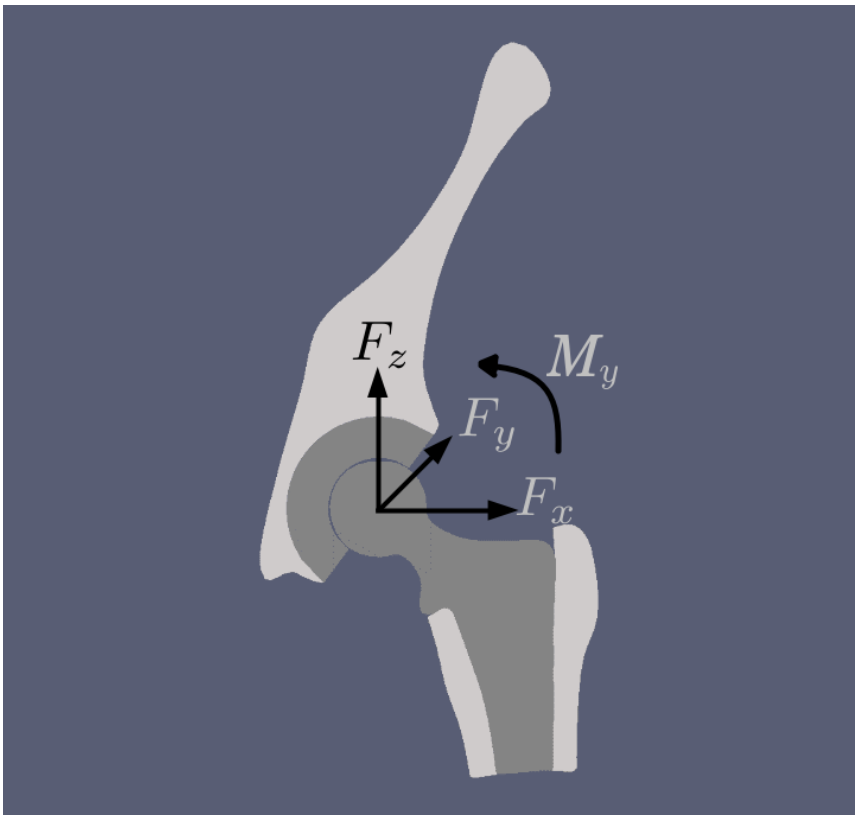

Fig. 3: *2D hip joint with an initial implant position including the used coordinate system for the applied forces F_x, F_y, F_z and exemplary moment M_y around the F_y axis.*

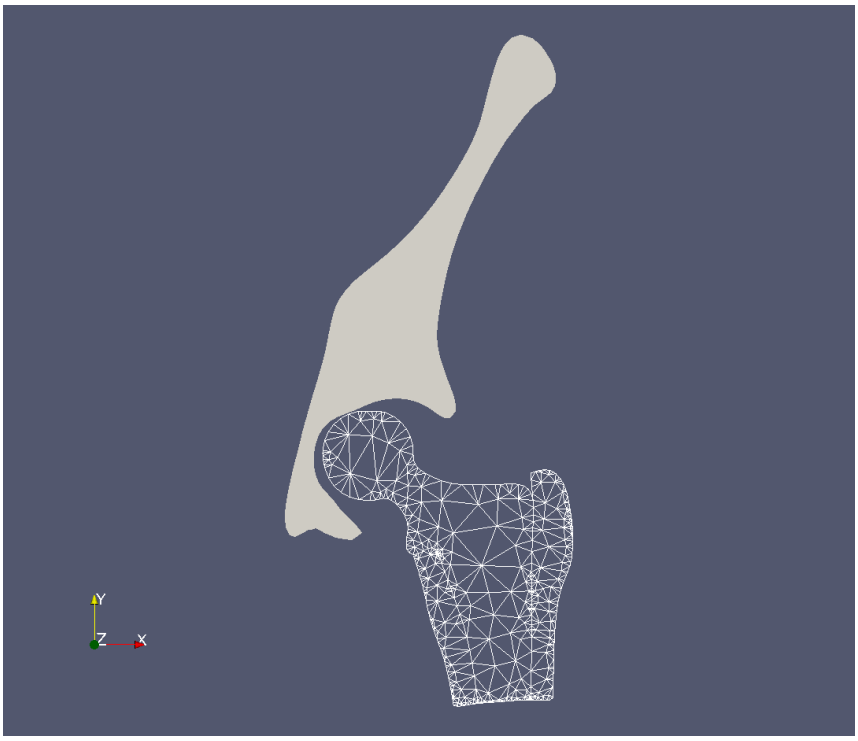

Fig. 4: *Example of a coarse triangulation.*

Computing the stress distribution in the cases with and without the implant and minimizing the difference of both for as many loads as possible needs to be accurate enough, leading to the necessity of a fine enough discretization – both the geometry as well as the load domain. Since the stress distribution is the solution to a partial differential equation (PDE) coupled with a contact problem, finding a minimum to this optimization problem is very expensive. Hence we need to find a way to reduce the computation cost, especially since the ultimate aim of the implant position's optimization is its clinical application. In order to contrast this research from the work others have done, as will be stated below, it is important to mention that we will be integrating all load tuples from the load domain $\mathcal{L}$ and not only a small number of exemplary loads.

1.2 State of the art

The approach pursued in this work will be presented in Sec. 1.3. Before doing so, other research in the area of *implant optimization* and *surrogate models* will be presented in order to better contextualize and contrast present work from other past or current studies in this field.

1.2.1 Implant optimization

Current research of implants and positioning takes a variety of approaches. One is to prevent stress shielding through shape optimization [35, 55, 160], but does not take motions into account. Lubkoll [101, 103] and Lubkoll et al. [102, 104] is also concerned with implant shape design, but the focus is on facial implants where he largely needs not concern himself with stress distribution due to external forces, but is able to concentrate on general function restoration and an unobtrusive aesthetic outcome. Still he includes elasticity and contact calculation along with an extensive mathematical background.

A different approach limits the optimization of implant positioning to very few loads from daily motion [136] with looking only at the implant without the surrounding bone [36] or using a precomputed, fixed implant-bone geometry [38]. This research, as opposed to the presented work in this thesis, is not intended to improve patient-specific surgeries and accompanying computer tools such as "Software-Guided Hip Surgery" from Brainlab[‡], but aims at better implants and their performance on a general basis.

Herrmann et al. [77] combine hardware and software in order to test total hip arthroplasty (THA) stability with this hybrid approach, yet also in the hardware setup without the bone-surrounding soft tissue, muscles, or ligaments.

For studies that deal with computational aspects of implantation, i.e., that aim to improve computer assistance tools such as the above mentioned "Software-Guided Hip Surgery", it is advisable to turn to [8, 57]. Barbič et al. [8] in particular investigate "real-time" computation of deformed objects including nonlinear stress computation combined with contact. Galloway et al. [57] on the other hand focus on the generation of a population of FE models from two statistical models – one for the shape and elastic modulus of the tibia, the other for the tibiofemoral joint loads over a gait cycle. With this, a method was developed to automatically size, position, and implant the tibial try in each tibia.

Still in the broader scope of surgery assistance tools, here for the acquisition of the patient's geometry, Lamecker [96] is discussing the setup of a pipeline for 3D reconstruction

[‡] https://www.brainlab.com/surgery-products/orthopedic-surgery-products/hip-navigation-application

for clinics. He touches on statistical shape models and deals with (automated) segmentation of medical images to increase computation speed before he combines the different techniques to present a means of 3D reconstruction from 2D X-ray images.

Improving computer-aided surgery to automate preoperative planning in THA based on a statistical implant plan atlas is the aim of Kagiyama et al. [85]. Though one should note, that the statistics are based on preoperative plans by experienced surgeons and the outcome may not fit the specific patient at all if he or she is not represented. For a further discussion of possible drawbacks and pitfalls not addressed in [85] we refer to the extensive discussion of statistical shape models in [96, 139] which pose a similar idea.

The article by Wang et al. [159] gives a good overview of the current state of hip implant research, though concentrating on topological design and additive manufacturing, i.e., 3D printing of implants. Although new implant designs need to be extensively tested before they can be used in medical practice, this issue is only briefly mentioned in the chapter "Challenges and future directions". One main focus of implant optimization is the prevention of stress shielding, which primarily occurs due to the noticable difference in the Young's moduli of bone and implant. The approach is to weaken the implant by increasing its porosity and permeability, which then allows for bone ingrowth (*osseointegration*). This contradicts the aim of a stiff implant with a good mechanical strength. The in silico testing for osseointegration calls for a dynamical testing environment with daily loading over a period of time in which the ingrowth is to happen. This setup is not mentioned as part of the past research.

Similar to 3D printing is the idea that Bram pursues in his habilitation treatise [18]. Focusing on powder metallurgy, he analyzes different materials and their usability in production of porose implants for an increase of osseointegration.

In the paper by Cilla et al. [35] the focus lies on the prevention of stress shielding through changing the implant design. The difference in approach here is to alter the mechanical parameters and study the change in stress shielding along with lightening the implant up along the coronal and sagittal section. The alteration of parameters gave the result of very little change in stress shielding. Even reducing Young's modulus and the Poisson ratio to bone properties showed the perseverance of stress shielding though highly reduced. A change (increase and decrease) in the implant's top and intermediate cross-section had very little effect on the stress shielding. Only the lightening along the coronal and sagittal section had bigger and mainly positive effects on the shielding. Cilla et al. [35] identify a number of drawbacks of their own research, being that only loads of walking are included, only one bone geometry and one implant position are considered as well as the neglect of cartilage and the simplified bone-implant interface that assumes full implant integration into the bone without friction.

A different approach still is taken by Quental et al. [126] in looking at total shoulder arthroplasty, and specifically modeling the change of bone while considering different implants and implant types. This, on the other hand, is done with a fairly simple model for bone resorption combined with a 3D geometry of the shoulder joint with implant. The model for bone remodeling was developed by Fernandes et al. [51], assuming rectangular holes in the discretized bone grid, the change of which affects the bone density and the fourth-order bone material tensor. A change over time due to repeated loading and, e.g., cellular adaption is not integrated and the applied static loads are with 12 still very limited.

Research from a more (bio-)engineer perspective was conducted by Effenberger [47]

and Oldani et al. [117]. While the former reports on different hip implant styles for stem and cup including, e.g., their stability and disadvantages of different designs, the latter cover the topic of titanium as a biomaterial for implants, its mechanical properties, but also present an FE analysis for fatigue test. The test is done for one load only and aims to compare the three alloys under regard.

1.2.2 Surrogate models

Concerning the surrogate model for the stress response surface resulting from the different loads, we will be working with Kriging coupled with an adaptive strategy that takes advantage of a trade-off between global exploration and local exploitation. The usage of Kriging in combination with deformation and contact, but excluding equations of elasticity, can be found in Forsberg et al. [53].

Iuliano [83] uses the same surrogate model as well as radial basis functions to compare with, next to a variety of criteria to adaptively enlarge the included samples and improve the surrogate. While Iuliano's aim – finding a minimum of an essentially unknown function – differs from our aim – approximating a response surface whose integral renders one point on the objective function to be minimized – the overall target in improving the accuracy of the surrogate is the same. In general terms, Iuliano describes the setup and improvement of a surrogate model in three steps: starting with the design exploration over the given domain (initial samples), the second step is that of adaptive sampling, completed by a sequential metamodel optimization. For the initialization stage, Iuliano suggests common techniques such as Latin hypercube sampling or Latinized central Voronoi tessellation techniques. In our case, the initialization samples in the force domain will be driven by the probability field and will be described later in the Sec. 2.2.2. For the second step, Iuliano describes a variety of factorized infill and expected improvement criteria. The common idea is to achieve a balance between exploration, i.e., sampling in undersampled areas, and exploitation.

For *factorized infill*, the evaluation function for new samples is split into two function parts h and g, one for the exploration, the other for the exploitation criterion, which are multiplied with each other. If *exploration* is favored, new samples will be chosen in areas that were previously unsampled – unexplored. If the concentration is on *exploitation*, then a new sample will be close to an already used sample. This is the case, if some area contains more information than already extracted. He describes the approaches of leave-one-out, weighted leave-one-out, where function values with lower value are heavier weighted, Lipschitz constant, concentrating on areas with high complexity, i.e., larger first derivative, and weighted distance criterion, which favors domain exploration.

For the *expected improvement-based infill*, Iuliano again describes two functions for exploration and exploitation. Here they are mostly coupled by addition. One term scores a new sample point higher, the closer they are to low, already sampled function values, bearing in mind that Iuliano wants to find a minimum value of the unknown function. The second term gives a higher score to samples which have a high uncertainty, i.e., a higher variance. Should the model setup not naturally provide a value for the variance, this can be approximated with the expected improvement-like criterion. The *expected improvement for global fit (EIGF)* does not focus on finding a minimum, but to produce a good (global) function fit. Lastly, in the *generalized EIGF*, the square $(h + g)^2$ is used rendering a part where the two functions interact by multiplication $2hg$ driving the selection of new samples to areas where both factors are amplified.

As for Iuliano's third step – the sequential metamodel optimization – this is useful if one wants to find an optimum on the approximated function, but is irrelevant in the present context.

Other possible surrogate models are gradient-enhanced Kriging, radial basis functions, polynomial response surfaces, support vector machines, and artificial neural networks besides others. In the following, the description of their different characteristics ensues. See also [140] for an overview of further surrogate models.

The *gradient-enhanced Kriging* uses adjoint solvers for gradient computation. The inclusion of more information leads to an improved response surface. The downside is that this method does not scale well with the number of sampling points due to the rapid growth in the size of the correlation matrix. It is unclear whether a possible gain in the response surface is achieved next to numeric errors induced through contact surface discretization and contact order[§] in contact calculation.

Radial basis functions are a class of functions whose function response depends on the distance of a predefined sample. With their fixation to a priori given locations, they are less flexible than, e.g., the samples in Kriging. This means if an area is unsupported by radial basis functions, an extra function needs to be introduced. The shape parameters can be adjusted but this infers extra work. See Driscoll et al. [45] for more information on radial basis functions, here in combination with collocation methods.

Polynomial response surfaces is a different term to polynomial regression. While this approach is – depending on the polynomial degree – generally able to interpolate smooth functions, extrapolation is not a strong feature with even the danger of oscillatory behavior close to the boundary. This technique is not fit for adaptive sample refinement or able to capture locally high nonlinear behavior, or dramatic changes in the response in simpler cases. Forsberg et al. [53] compare ordinary Kriging with a linear polynomial response surface approach and conclude that Kriging performs better. For a comparison between Kriging and a quadratic polynomial interpolation model see Giunta et al. [62].

Another possibility for a surrogate model is the *sparse grid* approach, see, e.g., [24, 59]. This method allows to overcome the curse of dimension to some extent, as the authors themselves put it. With the sparse grid approach, the number of degrees of freedom can be reduced while the accuracy stays high. The authors demonstrate the method's application in the solution of PDEs and show that the grid can be optimized to the appropriate underlying space of the target function. Besides the linear ansatz functions, locally adaptive methods, higher order polynomial or wavelet discretizations can be applied. Even for numerical integration this technique can be used.

When comparing sparse grids to Kriging it is a priori unclear which method performs better – if hierarchical basis combined with a pre-optimized (sparse) grid outperforms an adaptive Kriging scheme, that adds samples and thus complexity in areas of problem-defined interest or vice versa. Additionally, it is unclear whether a dominance of one technique over the other is true for all – accuracy, computation time, and convergence – or just a number of those criteria. The answer to this question needs more investigation.

The idea behind *support vector machines (SVM)* is that of linearly dividing two classes of objects. When applying the kernel-trick a nonlinear hyperplane can be constructed. Similarly in our case this dividing plane can be compared with the response surface. While in the present case, not two classes of objects are classified, but much rather a

[§]The *contact order* determines the number of discrete points on a contact boundary. An increasing order refines the precision of the contact solution.

hyperplane is aimed at, the extension of a SVM for regression poses a possible surrogate model candidate. With the expected nonlinearities between the responses, only support vector regression (SVR) or Bayesian SVM can be applied analogously, both including the kernel-trick and the adjustment of several parameters. When adding new samples in SVR, a regression problem needs to be solved every time where the solving of a matrix-vector problem is more straight forward in Kriging. Depending on the kernel function, e.g., (inhomogeneous) polynomial, Gaussian radial basis, or hyperbolic tangent, the computation of the interpolate is more expensive than the matrix-vector product of Kriging.

One last surrogate model to be mentioned here is *artificial neural networks (ANN)*. The network is made up of interconnected units or nodes, each representing a function (linear or nonlinear depending on the user setup). More intricate neural networks are described with *deep learning* (or *deep neural networks (DNN)*) include multiple hidden layers, i.e., intermediate nodes between input and output. The nodes can be trained, i.e., the function parameters adjusted, to fit the desired input-output. In order to train the nodes, usually a lot of training data is needed. Past research has already achieved the solution of a classification problem, e.g., identifying the depicted animal or object on pictures. Through supervised learning the ANN or DNN are trained to best fit a given input to an output, similar to a generic function $f : X \to Y$ from pairs $(x, y)_i$.

1.3 Our approach

The given problem of optimizing the hip implant position to increase the implant's longevity is indeed an interlacement of different subproblems and subalgorithms. The aim of preventing stress shielding not only for as many as possible daily motions, but for all possible motions leads to a moderately high dimensional load domain for which the equations of motion have to be solved repeatedly. That becomes increasingly computationally expensive the more degrees of freedom (DOF) are added to the geometric grid.

Here, the first two problem relaxations will be applied: first, the solving of the PDE will be done on a geometric grid with relatively little DOF that are to be extended throughout the optimization; second, the PDE is solved for only a few loads. The results will be interpolated to a response surface using Kriging interpolation.

Since high stresses are to be prevented as to sustain the bone as healthy as possible, locally high stresses will be penalized leading itself to locally high nonlinearities in the response surface that is to be integrated. These locally high stresses may stem from motions like stumbling or worse. To keep both the interpolation and the integration efficient, these two operations will be separated. Therefore, the locally high stress responses will be interpolated and the evaluation of the interpolate with the penalty function will be integrated.

As the load domain is moderately high dimensional, standard quadrature rules were to suffer the curse of dimension if applied. This curse will be avoided by utilizing Monte Carlo integration. Monte Carlo itself has a comparatively low rate of convergence, but it benefits from its independence of the space dimension d, thus needs less function evaluations to achieve a certain accuracy and thus proves more efficient than standard quadrature rules.

A possible application of *sparse grids* as in Bungartz et al. [24] was considered, but Kriging in combination with Monte Carlo is much more straight forward and offers less function evaluations.

All three, the precision of the solution of the PDE as well as the accuracy of the Kriging surface and the Monte Carlo integral will be adaptively increased depending on the

geometric discretization of bone and implant, which is to be refined throughout optimization.

For the adaptive refinement of Kriging, the approach by Iuliano [83] will be applied, making use of local exploitation and global exploration. The exploitation part will be evaluated via the cross-validation strategy, while the exploration is added through the variance term inherent to Kriging interpolation. Both are multiplied with a factor that, depending on the last added sample, favors either the one or the other, but in smooth transition from one to the other also allows for a balancing of both aspects.

Concerning the geometry discretization, it is to be said that for the bone-implant geometry, we start with a finer discretization at the contact boundary than the interior of the domain and the non-contact boundary. The pure bone geometry will be finer discretized from the start than the bone-implant geometry, since the stress computation for each applied load only needs to be computed once, as opposed to the bone-implant geometry. Here the stress response to the applied loads needs to be recomputed every time the implant position is changed.

Last but not least, implant position changes may lead to a remeshing of the bone-implant geometry. This results in discontinuities in the target function which will in optimization be dealt with special line search.

2 Mathematical Modeling

In this chapter, the complete mathematical model with its aim of optimization, the design variables, and the constraints will be introduced. This optimization problem will prove to be too difficult to solve, necessitating model simplifications that will be established in Sec. 2.2, and the subsequent use of efficient solution algorithms which will be the focus of Sec. 3.

2.1 Implant position optimization

The insertion of implants as it is the clinical practice takes place with very little automation and relies heavily on the experience of the surgeons. With X-ray or CT information about the patient's geometry, they can estimate a good position of the implant but they fail to incorporate the dynamic loads as there is no means to deduce those merely from images. One does not fail to notice that the cyclic loading of the joint and hence the implant is intrinsic and its negligence can lead to undesired results. Hence it is the aim here to approach and overcome this problem with the only admissible means – mathematical optimization. In order to do so, the translation of the given problem into a mathematical setting is mandatory.

2.1.1 Design variables

For a coordinate system fixed in the femur head with axes (x, y, z) in 3D, see Fig. 3, the initial degrees of freedom that can be used to optimize the implant's position in the patient's joint are the translatory variables $\boldsymbol{t} \coloneqq (t_x, t_y, t_z)$, transitioning the femur implant along the axes, as well as the torsion angles $\boldsymbol{\alpha} \coloneqq (\alpha_x, \alpha_y, \alpha_z)$ around three coordinate axes (e.g., α_y along the same path as moment M_y as depicted in Fig. 3). In total, they account for six degrees of freedom.

With the objective of retaining the center of motion, the design variable $\boldsymbol{t}$ is already fixed leaving the torsion angles as the only degrees of freedom.

2.1.2 Objective functional

In order to achieve an increased longevity of the implant and thus reduction of pain and complications for the patient, the issues of the retention of the center of motion, stress shielding (cf. [43]), and risk of fracture needs to be addressed. Failing to retain the center of motion would lead to a reduced range of motion for the patient and consequently to a change in the usual movements. This can lead to a misuse of the joint and hence to pain and failure of the implant.

Preventing stress shielding means, that the stress distribution Σ in the (remaining) bone pre- and post-operative is unchanged for all of the patient's movements $m \in \mathcal{M}$. Note that the motion load domain $\mathcal{M}$ represents time-dependent trajectories in the load domain as opposed to the load domain $\mathcal{L}$ which describes time-*in*dependent loads. The transition from the former to the latter will be explained in Sec. 2.2. In the context of $\mathcal{M}$, T_m is the final time of the given motion m.

A change in the stress distribution Σ in the remaining bone would otherwise lead to a restructuring of the bone, bearing the immanent risk of implant loosening or shifting the center of motion. To measure the change in stress distribution define $\bar{\Sigma}$ as the stress distribution in the patient's healthy bone before the surgery – as well as this can be

retained. With the domain Ω defining the whole bone-implant (or bone) domain, the stress difference is computed in remaining bone

$$\Omega_{\text{bone}}(\boldsymbol{\alpha}) \coloneqq \Omega \backslash \Omega_{\text{implant}}(\boldsymbol{\alpha}), \tag{2.1}$$

which depends on the implant's position. Lastly, preventing fracture will be incorporated by penalizing motion loads that induce stresses that exceed the threshold of bone stability Σ_{thres}. This is done by a function $p : \mathbb{R} \to \mathbb{R}$ that receives $\Sigma_{\max}\big(m(t); \boldsymbol{\alpha}\big)$ as an input. The value $\Sigma_{\max}$ is determined for each time $t \in [0, T_m]$ by evaluating the maximum of the Frobenius norm defined for tensors.

For all admissible implant positions $\boldsymbol{\alpha}$ optimize for those minimizing the integral over the motion load domain $\mathcal{M}$ (cf. Sec. 2.1.4)

$$\min_{\boldsymbol{\alpha}} j(\boldsymbol{\alpha}) = \int_{m \in \mathcal{M}} \left(\int_0^{T_m} \int_{x \in \Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2} \big\| \Sigma(m(t); \boldsymbol{\alpha}) - \bar{\Sigma}(m(t)) \big\|_F^2 \, \mathrm{d}x + p\Big(\Sigma_{\max}\big(m(t); \boldsymbol{\alpha}\big)\Big) \, \mathrm{d}t \right) \mathrm{d}m. \tag{2.2}$$

2.1.3 Dynamic contact equilibrium constraints

Elastodynamics. This section will give an introduction to finite strain theory. For further reading, the following literature is recommended: For a basic introduction to finite strain theory, Wriggers [172] poses a good starting point, while for in-depth theory Ciarlet [34] or Altenbach [3] are suggested, who directly include mathematical theory. For the objective (2.2) the stress distribution Σ needs to be computed.

To obtain the stress distribution, one needs to solve a PDE describing the equilibrium of forces, see, e.g., [34, Ch. 2],

$$\rho \ddot{u} - \operatorname{div}(F\Sigma) = f, \qquad \text{in } \Omega \times [0, T_m], \tag{2.3}$$
$$u = 0, \qquad \text{on } \Gamma_D \subseteq \Gamma = \partial\Omega, \tag{2.4}$$
$$F\Sigma n = g_N, \qquad \text{on } \Gamma_N \subseteq \Gamma = \partial\Omega, \tag{2.5}$$
$$u(x, 0) = u_0(x), \qquad \forall x \in \Omega, \tag{2.6}$$
$$\dot{u}(x, 0) = \dot{u}_0(x), \qquad \forall x \in \Omega. \tag{2.7}$$

The quantity $P \coloneqq F\Sigma$ defines the *first Piola-Kirchhoff stress tensor*, with $\Sigma = \Sigma(u)$ denoting the *second Piola-Kirchhoff stress tensor*, with $u \in \big(H^1(\Omega)\big)^d \; \forall t \in [0, T_m]$ being the displacement. The matrix $F \coloneqq \nabla\varphi$ denotes the gradient of deformation $\varphi(x)$ (see Eq. (2.8) below for a definition) with respect to x. Further details will be elaborated shortly. For reasons of readability when writing $\Sigma(u)$ the dependency of u on x, t, and $\boldsymbol{\alpha}$ is left out. The space $H^1(\Omega)$ is the Sobolev space $W^{1,2}(\Omega)$. The functions f and g_N are vector fields and denote a body force in Ω or a surface force on Γ_N, the Neumann boundary. For the right hand side functions square integrability in their respective domains is assumed (cf. [150]). The parameter ρ is the mass density of the regarded material, i.e., bone or implant in the present case, and $T_m > 0$ is the final time of the considered motion. With the addition of the Dirichlet boundary condition (2.4) on Γ_D mathematical well-posedness is ensured (cf. [135]).

Table 1: *Lamé constants of the bone and titanium implant.*

	λ [Pa]	μ [Pa]
Bone	$9.81{\cdot}10^9$	$6.54{\cdot}10^9$
Implant	$83.92{\cdot}10^9$	$43.23{\cdot}10^9$

When the body is transformed under the influence of body and surface forces, this transformation is given in terms of the displacement u resulting in the deformation function

$$\varphi(x) \coloneqq x + u(x). \tag{2.8}$$

Its gradient, the deformation gradient, is used describe the Green-St. Venant strain tensor E. With $F = \nabla\varphi$ as before and with I the identity the tensor E is

$$E(u) \coloneqq \frac{1}{2}(F^T F - I) = \frac{1}{2}\big(\nabla u + \nabla u^T + \nabla u^T \nabla u\big). \tag{2.9}$$

Note that in the present context, the second order term $\nabla u^T \nabla u$ is kept and no linearization is presumed. The displacement may be small, but rotations are to be expected which can only be treated correctly in the nonlinear setting.

Also note, that the *gradient* of the deformation $\nabla\varphi$ is written, but this is due to notational imprecision. As it can be found in Deuflhard et al. [42, Ch. 2.3] the geometric nonlinearity in the Green-Lagrange strain tensor in its usual form is not described by the gradient ∇u, but by the derivative u', i.e., $(u')^T u' = \nabla u \nabla u^T$. Due to its common use the usual notation will be used here, too, but the reader is advised to be aware of the difference in notation and (later) implementation.

Assuming the case of an isotropic material, the second Piola-Kirchhoff stress tensor Σ has the form

$$\Sigma(u) = \lambda\big(\operatorname{tr} E(u)\big) I + 2\mu E(u) = \mathcal{C} : E(u). \tag{2.10}$$

The parameters λ, μ are the Lamé constants. The Lamé constants for bone and titanium can be found in Tab. 1. The Hooke tensor $\mathcal{C}$ is the fourth-order elasticity tensor [17]

$$\mathcal{C}_{ijkl} = \lambda\delta_{ij}\delta_{kl} + \mu(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})$$

with δ_{ij} being the Kronecker symbols. The ':'-notation denotes the contraction of two tensors:

$$\Sigma(u)_{ij} = \mathcal{C}_{ijkl} E(u)_{kl},$$

see, e.g., [171, Ch. 3.3.2].

Materials that obey the relationship in Eq. (2.10) are called *St. Venant-Kirchhoff materials*. Let it be noted here, that St. Venant-Kirchhoff materials with Σ defined as in Eq. (2.10) are *hyperelastic* with stored energy function [34, Ch. 4.4]

$$W\big(E(u)\big) = \frac{\lambda}{2}\Big(\operatorname{tr} E(u)\Big)^2 + \mu \operatorname{tr} E(u)^2. \tag{2.11}$$

Inside the respective bodies of bone and implant homogeneity is assumed.

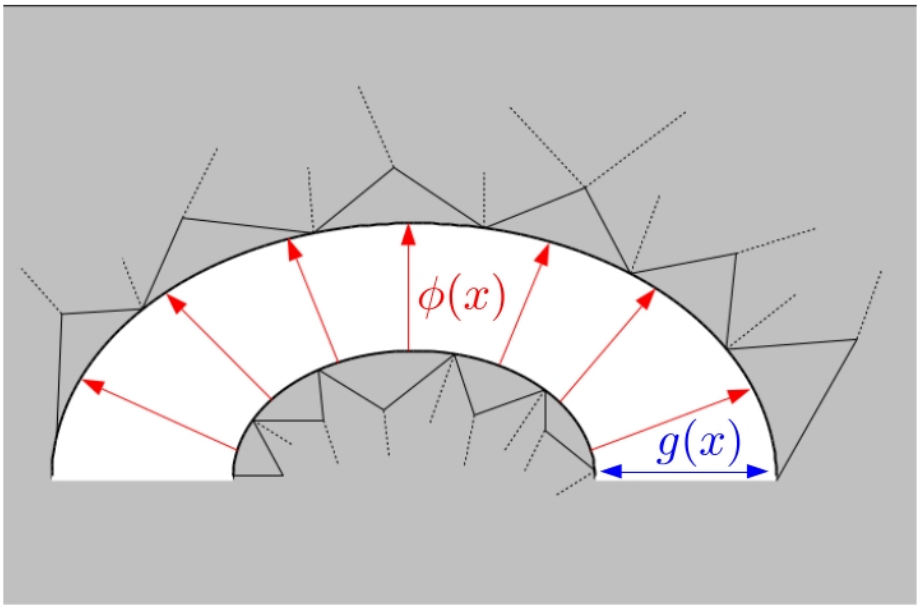

Fig. 5: *Visualization of a contact mapping between the femur and the pelvis.*

Remark. *This condition marks a huge model simplification, since bone is an inhomogeneous and anisotropic material. The implementation of the correct material model would generally be possible (inhomogeneous, anisotropic material parameters can, e.g., be found here [52, 167] for bone and here [7, 105, 115] for titanium) and lead to a better approximation of bone but it would also lead to an increase in computation time. Since the interest in the current work is in relative optimization results and not in rebuilding a rigorous bone or bone-implant model, the simplified model is used here.*

With this now solve the equation of equilibrium only in terms of the displacement u as (2.3) now reads

$$\rho\ddot{u} - \operatorname{div}\big(F\mathcal{C} : E(u)\big) = f.$$

The derivation of the *stress principle of Euler and Cauchy* is extensively studied in literature such as [17, 34, 135].

Contact problem. Naturally when applying forces that move two bodies together contact occurs. For the problem to be physically stable, non-penetration has to be enforced. Mathematically this incurs the inclusion of a contact inequality condition, as the two geometries are allowed to touch, but not to interpenetrate. Following, e.g., [135] an inequality condition is introduced with $\Gamma_C \subseteq \partial\Omega$ denoting the contact boundary. It is assumed, that $\Gamma_D \cup \Gamma_N \cup \Gamma_C = \Gamma$ and that the boundary sets are respectively disjoint. The non-penetration condition coupled with the assumption of small strain and little tangential motion at the contact boundary are known as the *Signorini conditions* for contact problems [46, 87]. Youett et al. [177] investigated the contact model with nonlinear contact constraints allowing its application in a nonlinear elasticity setting. For the non-penetration condition, a

bijective and smooth mapping ϕ_{bone} between the femur and the pelvis needs to be defined, and a mapping ϕ_{implant} with the same characteristics between the implant components. These *contact mappings* identify the two respective contact boundaries with each other as depicted in Fig. 5. With the contact mapping, one can define the *reference gap function* $g \in \left(L^2(\Gamma_C)\right)^d$ between the bodies

$$g_i : \Gamma_{\text{pel}(i),C} \to \mathbb{R}, \qquad g_i(x) \coloneqq |x - \phi_i(x)|, \qquad i \in \{\text{bone}, \text{implant}\}$$

and the relative displacement in normal direction

$$[u \cdot n]_{\phi_i} \coloneqq \Big(u_{\text{pel}(i)}(x) - u_{\text{fe}(i)}\big(\phi_i(x)\big)\Big) \cdot n(x), \qquad x \in \Gamma_{\text{pel}(i),C}.$$

Denote with $\Gamma_{\text{pel}(i),C}$ the contact boundary of the pelvis geometry and let $u_{\text{pel}(i)}, u_{\text{fe}(i)}$ describe the displacement of the pelvis and femur geometry, respectively. With this, the inequality condition denoting the contact inequality

$$[u \cdot n]_\phi \leq g \text{ on } \Gamma_C \tag{2.12}$$

is added as a further constraint to the PDE (2.3).

2.1.4 Motion loads

Motions $m \in \mathcal{M}$ are defined as trajectories in $\mathcal{L}$. That is to say, that m maps from a time interval $[0, T_m]$ to $\mathcal{L}$

$$m : [0, T_m] \to \mathcal{L}, \quad t \mapsto l = m(t),$$

with the load domain $\mathcal{L}$ as previously given in Eq. (1.1)

$$\mathcal{L} = F_x \times F_y \times F_z \times M_x \times M_y \times M_z.$$

The motion loads $m \in \mathcal{M}$ enter the PDE (2.3) through surface traction g_N. The translatory and tangential forces as well as torsion are modeled as acting on the Neumann boundary. Ideally the motion load domain $\mathcal{M}$ is made up of recorded loads from the patient's healthy state.

Here, one can already see a difficulty of the model, since in the diseased patient, there is no means of measuring the (contact) forces and moments from the healthy state. Contrariwise, in a healthy person, one would not prophylactically measure the contact forces and moments not to mention that this is not yet possible altogether. Moreover, a healthy person would not have their motion forces measured. Therefore, an approximation of the domain $\mathcal{M}$ will be needed.

2.1.5 Penalty function

The function p penalizes exceeding stresses. The quantity $\Sigma_{\max}\big(m(t); \boldsymbol{\alpha}\big)$ is defined as

$$\Sigma_{\max}\big(m(t); \boldsymbol{\alpha}\big) \coloneqq \max_{x \in \Omega_{\text{bone}}} \big\|\Sigma\big(u(x)\big)\big\|_{\text{F}} \tag{2.13}$$

again leaving out the dependency of u on t and $\boldsymbol{\alpha}$. The stress tensor Σ (2.10) is a symmetric $d \times d$ matrix, thus the Frobenius norm is well-defined. Note, that without further proof we expect no stress singularities in the bone since the implant is always stiffer and we assume the bone surface to be smooth. The maximum norm could be approximated by

the L^p-norm with sufficiently big p. This idea of approximating the maximum norm will return when computing the derivative in Sec. 4.3. The function p is thus defined as

$$p\Big(\Sigma_{\max}\big(m(t);\boldsymbol{\alpha}\big)\Big) \coloneqq b \cdot \exp\bigg(c\Big(\Sigma_{\max}\big(m(t);\boldsymbol{\alpha}\big) - \Sigma_{\text{thres}}\Big)\bigg). \tag{2.14}$$

Here, Σ_{thres} is chosen well below 200 MPa where bone is susceptible to fracture. The factor b is chosen as such the penalty term is of little influence on the objective function, if $\Sigma_{\max}$ is below Σ_{thres}. For a sharp increase of p when the value of $\Sigma_{\max}$ approaches that of Σ_{thres}, set c to a high value. The maximum $\Sigma_{\max}$ is taken for all times $t \in [0, T_m]$ and only the bone geometry, since the material of the implant is much more durable.

2.1.6 Complete optimization problem

With the given specifications above, the following optimization problem has to be solved

$$\min_{\boldsymbol{\alpha}} j(\boldsymbol{\alpha}) = \int_{m\in\mathcal{M}} \left(\int_0^{T_m} \int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2} \big\| \Sigma(m(t);\boldsymbol{\alpha}) - \bar{\Sigma}\big(m(t)\big) \big\|_F^2 \,\mathrm{d}x + p\Big(\Sigma_{\max}\big(m(t);\boldsymbol{\alpha}\big)\Big) \,\mathrm{d}t \right) \mathrm{d}m \tag{2.15}$$

s.t.

$$\rho\ddot{u} - \operatorname{div}(F\Sigma) = f, \qquad \text{in } \Omega \times [0, T_m], \tag{2.16a}$$
$$u = 0, \qquad \text{on } \Gamma_D \times [0, T_m], \tag{2.16b}$$
$$F\Sigma n = g_N, \qquad \text{on } \Gamma_N \times [0, T_m], \tag{2.16c}$$
$$u(x, 0) = u_0(x), \qquad \text{in } \Omega, \tag{2.16d}$$
$$\dot{u}(x, 0) = \dot{u}_0(x), \qquad \text{in } \Omega, \tag{2.16e}$$
$$[u \cdot n]_\phi \leq g, \qquad \text{on } \Gamma_C. \tag{2.16f}$$

This problem setting admits some difficulties and some impossibilities. One impossibility is the acquisition of the healthy patient's motion load domain $\mathcal{M}$ as mentioned before in Sec. 2.1.4.

Difficult or at least unsuitable for the clinical application is the solution of the full problem (2.15)–(2.16). Even though the equations (2.16) can be solved for a given angle $\boldsymbol{\alpha}$ and a given motion m, one solve alone is, depending on the required accuracy, computationally expensive. With the optimization problem (2.15), the PDE (2.16) has to be solved for several (possibly infinite) motions m and resolved when the implant position changes. The storage of results or parts thereof may become increasingly difficult the more positions are evaluated. At least, this issue could be countered by applying compression and storage results found in [65, 72, 164].

Summing up, model simplifications are inevitable.

2.2 Modeling simplifications

Since on the one hand solving dynamic contact problems incurs a high computational effort, and on the other hand the inertia of the bone and implant are negligible compared to the surrounding soft tissue and the limb as a whole, a quasi-static approximation of the motion trajectory is an attractive approach. This lays out the course for model simplification. This approach also fits well to the problem setting, because the displacements in the given geometry are small and plastic deformation does not take place. The transition from a dynamic to a quasi-static motion model moreover allows, as will be seen below, for a different and much simpler parameterization of the loads experienced by the joint. This is why it is omitted from Eq. (2.16a), turning a parabolic PDE into an elliptic PDE.

2.2.1 Quasi-static approximation

As inertia is relatively small, $\rho\ddot{u}$ is negligible and will be omitted from Eq. (2.16a) meaning that the time component can be completely erased including the initial conditions (2.16d) and (2.16e). Additionally, one changes over from time-dependent motion $m(t)$, i.e., load trajectories, to single loads $l \in \mathcal{L}$. Hence the notation changes from, e.g., $\bar{\Sigma}\big(m(t)\big)$ with $m \in \mathcal{M}$, $t \in [0, T_m]$ to $\bar{\Sigma}(l)$, $l \in \mathcal{L}$. This results in the use of the *spatial equilibrium equation* rather than the equation of motion

$$\min_{\boldsymbol{\alpha}} j(\boldsymbol{\alpha}) = \int_{l\in\mathcal{L}} \pi(l) \int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2}\big\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\big\|_F^2 \,\mathrm{d}x + p\big(\Sigma_{\max}(l;\boldsymbol{\alpha})\big)\,\mathrm{d}l \tag{2.17}$$

s.t.

$$-\operatorname{div}(F\Sigma) = f, \qquad \text{in } \Omega, \tag{2.18a}$$
$$u = 0, \qquad \text{on } \Gamma_D, \tag{2.18b}$$
$$F\Sigma n = g_N, \qquad \text{on } \Gamma_N, \tag{2.18c}$$
$$[u \cdot n]_\phi \leq g, \qquad \text{on } \Gamma_C, \tag{2.18d}$$

with $f \in \big(L^2(\Omega)\big)^d$, $g_N \in \big(L^2(\Gamma_N)\big)^d$, and $g \in \big(L^2(\Gamma_C)\big)^d$ as above. For the displacement u it is assumed that $u \in \big(H_0^1(\Omega)\big)^d$.

The concise change over from motions $m \in \mathcal{M}$ to $l \in \mathcal{L}$ which implicates the introduction of a probability density function π is the focus of the next section.

2.2.2 From trajectories to load densities

As already mentioned in Sec. 2.1.4, an approximation of the motion domain $\mathcal{M}$ is needed as one cannot achieve the true $\mathcal{M}$ for reasons stated before. Therefore one has to start with a discretized $\mathcal{M}_q$ which only includes the measured motions.

The forces and momenta from different daily motions, such as walking and stair climbing, acting on the hip joint were measured and recorded for different persons in the open database https://orthoload.com/.

The motion trajectories $m \in \mathcal{M}$ are intrinsically dynamic, mapping from $[0, T_m]$ to the load domain $\mathcal{L}$. Now, the point of view will be turned around and instead of m defining the loads l, determine for any load $l \in \mathcal{L}$ its probability of belonging to a given motion m,

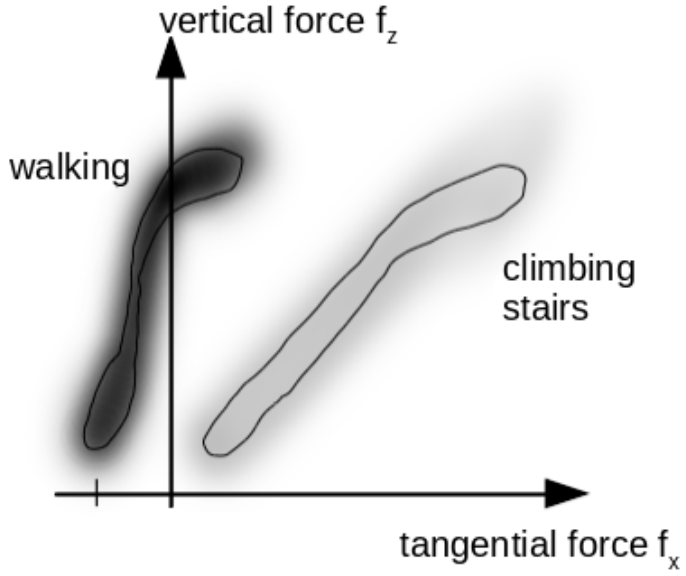

Fig. 6: *The probability density function π_m (schematic) for two exemplary motion trajectories with mean value μ_m and covariance S_m (here projected to 2D); the gray scale denotes the patient specific weight $w(m)$.*

i.e., determine $\pi_m(l)$. In order to compute this quantity, we utilize the motion's inherent density.

Realizing that motion trajectories are indeed not deterministic, each motion gait is much rather a realization of a probability distribution P_m described by a density function π_m. Naturally one could equip each load $l = m(t)$ for $t \in [0, T_m]$ with a load density $1/\left\|\dot{m}(t)\right\|_2$. However, the motion m is not measured for all $t \in [0, T_m]$, but at discrete time instances $0 = t_0 < t_1 < \cdots < t_k = T_m$, $k \in \mathbb{N}$, defining

$$l_{m_0} \coloneqq m(t_0), \ldots, l_{m_k} \coloneqq m(t_k).$$

With the same time discretization $t_0, \ldots, t_k$ for one motion, but for multiple measurements (from the same person and other people), one can compute a sample mean μ_{m_j} and covariance matrix S_{m_j} for each $j = 0, \ldots, k$.

With $\pi_{m_j} = \mathcal{N}(\mu_{m_j}, S_{m_j})$ sufficiently given, now the functions π_m and π need to be defined.

Quantifying probability densities. Considering the discretization $0 = t_0 < t_1 < \cdots < t_k = T_m$, $k \in \mathbb{N}$, of a motion's time interval $[0, T_m]$ with associate loads $l_{m_j} = m(t_j)$ and with mean load μ_{m_j} and a covariance matrix S_{m_j} for each t_j, then a discrete set of normal distributions is given

$$\Pi_m = \{\mathcal{N}(\mu_{m_0}, S_{m_0}), \ldots, \mathcal{N}(\mu_{m_k}, S_{m_k})\}. \tag{2.19}$$

Per normal distribution for $j \in \{0, \dots, k\}$, the probability density function $\pi_{m_j} \in \Pi_m$ is given by (d denoting the number of dimensions; with $l \in \mathcal{L} \subset \mathbb{R}^6$ it is $d = 6$)

$$\pi_{m_j}(l) = \frac{1}{\sqrt{(2\pi)^d \det\left(S_{m_j}\right)}} \exp\Big(-\frac{1}{2}\left(l - \mu_{m_j}\right)^T S_{m_j}^{-1}\left(l - \mu_{m_j}\right)\Big), \quad j \in \{0, \dots, k\}.$$

With this, one evaluates the probability of any load $l \in \mathcal{L} \subset \mathbb{R}^6$, with $\mathcal{L}$ as in Eq. (1.1), to be acting during a certain motion by summing

$$\pi_m(l) \coloneqq \frac{1}{k+1} \sum_{j=0}^{k} \pi_{m_j}(l). \tag{2.20}$$

Now one does not want to integrate over $\mathcal{M}_q$ but over $\mathcal{L}$. Hence one needs to generalize from a specific π_m to an m-independent π that applies for all $l \in \mathcal{L}$. In order to compute such a general probability function, one needs to take the average over the motions m. This average is necessarily a weighted one, as each motion has a different daily frequency $w(m)$. This frequency introduces, next to the patient-specific hip joint geometry, another level of patient-specificity to the implant positioning problem. Thus one computes

$$\pi(l) \coloneqq \frac{\sum_{m \in \mathcal{M}_q} w(m)\pi_m(l)}{\sum_{m \in \mathcal{M}_q} w(m)}. \tag{2.21}$$

The data from *OrthoLoad* allows to set up probability distributions π_m as described above. In combination with setting up the patient-specific $w(m)$, e.g., by filling in a questionnaire which allows an estimation on the distribution of daily motions, this allows to determine $\pi(l)$ which finalizes the shift from a dynamic to a static problem definition.

Looking at Fig. 6, two exemplary weighted density functions $w(m)\pi_m$ for the motions of walking and climbing the stairs are schematically depicted. The visible lines represent the mean value μ_m and the clouds around the line represent the covariance S_m projected to 2D. The gray scale denotes the respective weight $w(m)$. Thus one sees a representation of $\pi(l)$ limited to two recorded motions in $\mathcal{M}_q$.

2.2.3 Closing remarks

With the model simplification, suitable algorithms to solve (2.17)–(2.18) efficiently need to be applied. Though the problem to solve is a simplified one, one solve of (2.18) is still expensive. This said, it is to be expected that for optimization a solution u of PDE (2.18) needs to be computed many times which still results in an infeasible long solution time to find $\boldsymbol{\alpha}$. Similar to Erway et al. [50], a trade-off between function evaluations and accuracy of the solution is the focus of the following section.

3 Efficient Objective Evaluation

One of the main challenges in optimizing the implant position Eq. (2.17) under regard of motion loads is the long computation time for single loads combined with the need to integrate over the whole of the load domain $\mathcal{L}$:

$$\min_{\boldsymbol{\alpha}} j(\boldsymbol{\alpha}) = \int_{\mathcal{L}} \pi(l) \int_{\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2} \left\| \Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l) \right\|_F^2 \mathrm{d}x + p\big(\Sigma_{\max}(l;\boldsymbol{\alpha})\big)\, \mathrm{d}l.$$

Reducing the accuracy for the computation of the PDE solution, coarsening the discretization of the FE-grid and the load domain, or relaxing the stopping criterion for different computations, is a common approach; a balance of computation time and accuracy needs to be attained.

First and foremost, solving the contact problem is the most expensive part in the optimization problem (2.17). Thus, the main effort needs to be spent in reducing the number of PDE solves.

On the other hand, the load domain $\mathcal{L}$ is moderately high dimensional so that a reasonable discretization would result in a very high number of loads. This is why Monte Carlo (MC) integration will be applied to compute the integral circumventing the *curse of dimension*. The advantage on the one side, i.e., the error between the approximate and the real integral value diminishing with increasing number of Monte Carlo samples and the independence of the dimension, poses also the disadvantage of needing many evaluations of the function under the integral – the rate of convergence will be given in Sec. 3.2. With this the need is defined to evaluate the stress response to a given load l efficiently. The need will be satisfied by using the interpolation method of Kriging. When computing only a few solutions of the contact problem (2.18), one retains the option of keeping the accuracy of one solve high.

The penalty function introduces local high nonlinearities which is why not the function under the integral of the load domain $\mathcal{L}$, i.e.,

$$\pi(l) \int_{\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2} \left\| \Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l) \right\|_F^2 \mathrm{d}x + p\big(\Sigma_{\max}(l;\boldsymbol{\alpha})\big)$$

will be interpolated, but the difference of stress responses $\int_{\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2} \left\| \Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l) \right\|_F^2 \mathrm{d}x$ and $\Sigma_{\max}(l;\boldsymbol{\alpha})$. This is relevant in that way that the so-called *response surface*, i.e., the interpolated surface will be a smooth one that will not capture distinct features like locally high gradients. Either those local features would not be represented at all, or if one of the Kriging samples was in a spot of such a localized high nonlinearity, the smooth surface would be distorted to comply with the distinct feature of the response.

The already named interpolation method of *Kriging* has the following setup: a priori, one has to choose a drift model and a model for the local interdependence of function responses, the so called *variogram*, in more detail explained below. These model choices define parameters $\boldsymbol{\theta}$ which are fitted by solving a maximum likelihood problem to the given samples and responses. Now the interpolation value in an arbitrary location is a weighted addition of responses where the weights are estimated through few matrix-vector products. This calculation involves the computation of matrix inversions which can be precomputed once the model parameters $\boldsymbol{\theta}$ are known and which stay unchanged as long as the samples and responses are unchanged.

Some of the advantages of Kriging are, that it needs little work to set up interpolation, i.e., relatively simple parameter estimation, the computation of interpolation is realized by rather simple matrix-vector products, it can directly compute a variance which is utilized in adaptive sampling, and it allows for a sequential built of response surface. Generally, a down side of this model is the need for computing one or more inverse matrices. Compared to the effort of one PDE solve this is in the present work negligible. A variety of other possible surrogate models and a reasoning why Kriging is favored over these models has been given in Sec. 1.2.2.

Yet another possibility to increase the efficiency of the solution algorithm were to apply model reduction such as *Proper Orthogonal Decomposition (POD)* or *Discrete Empirical Interpolation Method (DEIM)* [31, 60, 79, 156]. The combination of model reduction with surrogate models and possibly Monte Carlo integration will be left for future work.

This chapter will cover existence and uniqueness results of the (linearized) Green-St. Venant strain tensor and how the contact problem will be solved. Also the design of both, Monte Carlo integration and Kriging interpolation, will be explained next to stating known convergence results. Section 3.4 will introduce error estimators that are used to adaptive refine the solving of the PDE, MC integration, or Kriging, respectively. The chapter will finish with giving the complete algorithm to compute the objective in Sec. 3.5.

3.1 Elastostatic contact solution

In this section current results for existence and uniqueness results for Eq. (2.18) are stated. For the linearized counterpart of (2.9) – with and without contact – with

$$\varepsilon(u) = \frac{1}{2}\big(\nabla u + \nabla u^T\big)$$

and with the associated linearized second Piola-Kirchhoff stress tensor σ

$$\sigma(u) = \lambda\big(\operatorname{tr}\varepsilon(u)\big)I + 2\mu\varepsilon(u) = \mathcal{C} : \varepsilon(u)$$

there exists extensive research on existence and uniqueness results for a displacement u, cf. [16, 48, 91, 134, 135]. There is further literature and results for the quasistatic contact (with friction), see for example [5, 25, 26, 123, 128, 143, 148]. For an existence and uniqueness result in linear elastodynamics see, e.g., Antonietti et al. [6]. More on contact problems and their numerical solution can yet be found in [41, 70, 89, 90, 98, 145, 146, 147, 168, 169], which partly includes error estimates and convergence results.

Turning towards existence and uniqueness results for the geometric nonlinear case, i.e., Eq. (2.18) with nonlinear tensors Σ and E, results become sparse. Ciarlet [34, Thm. 6.4-1] found, that for linear St. Venant-Kirchhoff material as given in Eq. (2.10) in case of pure displacement, i.e., homogeneous Dirichlet boundary conditions, without other boundary or contact conditions, with sufficiently smooth boundary Γ and function f, the existence of a locally unique solution can be proven.

For the case of the displacement-traction problem with a unilateral boundary condition, i.e., a problem setting with a Dirichlet boundary condition next to a Neumann boundary and a contact condition, under rigorous assumptions Ciarlet proves existence of a solution [34, Thm. 7.8-1].

As to uniqueness of a solution, results are even less satisfying. It can be found in Weiser et al. [163, Sec. 5, Ex. 1–2] even without contact the uniqueness is not given.

An existence and uniqueness result utilizing canonical duality theory and the algebraic (tensor) equation in stress-space, additionally to the requirement of the second Piola-Kirchhoff stress Σ being positive definite, can be found in Gao et al. [58].

SQP to solve the contact problem. The current setting is solved with the discretize-then-optimize approach which means that all quantities in Eq. (2.18) – spaces, functions, and the control – are discretized a priori and then the PDE is written in the discrete, numerically programmable form.

In practice, this means that the geometry Ω is discretized resulting in the grid $\mathcal{T}_h$ with grid elements $T \in \mathcal{T}_h$. The quantity h is defined as

$$h \coloneqq \max_{T \in \mathcal{T}_h} \operatorname{diam}(T).$$

According to the resulting grid $\mathcal{T}_h$, define finite-dimensional subspaces of L^2 and H_0^1, cf. [21, 33], with a piecewise linear finite element discretization. Rewrite the PDE (2.18) in a typical weak form, cf. [21, 34, 74, 150], so that $u_h \in \big(H^1(\Omega)\big)^d$ solves

$$\int_{\mathcal{T}_h} \Sigma(u_h) : E'(u_h)[v_h] \,\mathrm{d}x = \int_{\mathcal{T}_h} f_h v_h \,\mathrm{d}x + \int_{\mathcal{T}_{N,h}} g_{N,h} v_h \,\mathrm{d}s + \int_{\mathcal{T}_{D,h}} \gamma_D u_h v_h \,\mathrm{d}s \tag{3.1}$$

for all test functions $v_h \in \big(H^1(\Omega)\big)^d$, where both u_h and v_h satisfy the contact condition (2.18d), with functions f_h, $g_{N,h}$, g_h, and u_h being the finite-dimensional equivalents of f, g_N, g, and u and with discrete Neumann $\mathcal{T}_{N,h}$ and Dirichlet boundary $\mathcal{T}_{D,h}$. Note, that $E'(u_h)[v_h]$ is the directional derivative of the Green-St. Venant strain tensor $E(u_h)$ (2.9) in direction v_h

$$E'(u_h)[v_h] = \frac{1}{2}\big(\nabla v_h + \nabla v_h^T + \nabla v_h^T \nabla u_h + \nabla u_h^T \nabla v_h\big).$$

Also, the homogeneous Dirichlet boundary condition directly enters the weak form of the PDE with penalty parameter $\gamma_D \in \mathbb{R}$.

As mentioned in Sec. 2.1.3, the material is presumed to be St. Venant-Kirchhoff which is therefore hyperelastic [34, Ch. 4.4] and has a stored energy function $W\big(E(u)\big)$ as written in Eq. (2.11). It is known, that solving the equilibrium equations (2.18) is formally equivalent to finding a stationary point of the *total energy function* I, i.e., find admissible $u_h \in \big(H^1(\Omega)\big)^d$ such that $I'(u_h)v_h = 0$ for all $v_h \in \big(H^1(\Omega)\big)^d$ (cf. [34]). For this to be true, one needs that the applied forces are *conservative*, i.e., if there exist functionals $\mathcal{F}$ and $\mathcal{G}$, such that $\mathcal{F}'(u_h)v_h = \int_{\mathcal{T}_h} f_h v_h \,\mathrm{d}x$ and $\mathcal{G}'(u_h)v_h = \int_{\mathcal{T}_{N,h}} g_{N,h} v_h \,\mathrm{d}s$. This holds true in the present context with $\mathcal{F}(u_h) \coloneqq \int_{\mathcal{T}_h} f_h u_h \,\mathrm{d}x$ and $\mathcal{G}(u_h) \coloneqq \int_{\mathcal{T}_{N,h}} g_{N,h} u_h \,\mathrm{d}s$. One can further define a functional $\mathcal{D}(u_h) \coloneqq \frac{1}{2}\int_{\mathcal{T}_{D,h}} \gamma_D u_h^2 \,\mathrm{d}s$ with $\mathcal{D}'(u)v = \int_{\mathcal{T}_{D,h}} \gamma_D uv \,\mathrm{d}s$. Thus, defining the total energy function

$$I(u_h) \coloneqq \int_{\mathcal{T}_h} W\big((E(u_h)\big) \,\mathrm{d}x - \mathcal{F}(u_h) - \mathcal{G}(u_h) - \mathcal{D}(u_h) \tag{3.2}$$

solving the weak form of PDE (2.18) is equivalent to solving $I'(u_h)v_h = 0$ for all admissible $v_h \in \big(H^1(\Omega)\big)^d$.

This results in a discrete nonlinear problem (NLP) (e.g., [40, Ch. 4]) which is solved by the well-known *sequential quadratic programming (SQP)* method, cf. [116, Ch. 18] or

[150, Ch. 4.11]. The linear approximation $I'(u_h)v_h$ of $I(u_h)$ is as given by Eq. (3.1). It renders a vector that will be defined as $c \coloneqq I'(u_h)v_h$. The quadratic approximation reads as

$$I''(u_h)[v_h, w_h] = \int_{\mathcal{T}_h} \Sigma(u_h) : E''(u_h)[v_h, w_h] + E'(u_h)[v_h] : \mathcal{C} : E'(u_h)[w_h] \,\mathrm{d}x \\ - \int_{\mathcal{T}_{D,h}} \gamma_D v_h w_h \,\mathrm{d}s \tag{3.3}$$

for $v_h, w_h \in \left(H^1(\Omega)\right)^d$. The resulting matrix is defined as $A \coloneqq I''(u_h)[v_h, w_h]$. The second directional derivative $E''(u_h)[v_h, w_h]$ of the tensor $E(u_h)$ is

$$E''(u_h)[v_h, w_h] = \frac{1}{2}\left(\nabla v_h^T \nabla w_h + \nabla w_h^T \nabla v_h\right).$$

The contact inequality in discrete form reads as $Bu_h \leq b$. The SQP is formulated as the Powell-Hestenes-Rockafellar (PHR) variant of the augmented Lagrangian (cf. [129]). It solves for an update p to u_h

$$\min_p \frac{1}{2} p^T A p + c^T p + \frac{\gamma_C}{2} \left\| Bp - \left(b + \frac{\lambda}{\gamma_C}\right) \right\|_+^2 \tag{3.4}$$

where $\|x\|_+ \coloneqq \max\{x, 0\}$ and with penalty parameter $\gamma_C \in \mathbb{R}_+$. Compared to the standard augmented Lagrangian where a linear term $\lambda^T(Bp - b)$ is added, the PHR version with shifted penalties has the advantage that in the inner non-smooth, piecewise quadratic minimization problems that are to be solved the strongly active penalties remain strongly active. In the standard formulation, they become weakly active in the course of the augmented Lagrangian iteration, which can induce frequent switching of penalties between strongly and weakly active. The resulting problem (3.4) is in each SQP iteration solved using the semismooth Newton method [78, 82, 130, 153].

Convergence. As SQP is a standard method, convergence results from [116] – first a global convergence result, then results on local rate of convergence – will be informally stated.

First, recall the idea of SQP. For a given target function f and constraint functions c_i, where $i \in \mathcal{E}$ denotes the equality constraints and $i \in \mathcal{I}$ the inequality constraints, the Lagrangian $\mathcal{L}(u, \lambda)$ is defined as $\mathcal{L}(u, \lambda) \coloneqq f(u) - \lambda^T c(u)$. The SQP method sets up a problem, where the Lagrangian is approximated quadratically and the constraints linearly. In each iteration k a search direction p_k is computed. The Hessian $\nabla_{uu}\mathcal{L}$ is either computed directly or approximated by some symmetric and positive definite matrix B_k. Then, citing [116, Thm. 18.3], it is assumed that the quadratic problem is feasible, the solution p_k is bounded, and the sequences $\{u_k\}$ and $\{u_k + p_k\}$ are contained in a closed, bounded, convex region of design space in which the functions f and c_i have continuous first derivatives. Furthermore one needs, that the matrices B_k and multipliers λ_k are bounded for all k. Then all limit points of the sequence $\{u_k\}$ are KKT points of the given problem, i.e., the first-order necessary conditions are fulfilled. As Nocedal and Wright state, these conditions for the conclusion of the theorem are restrictive. They hint to more realistic conditions surveyed by Conn, Gould, and Toint [37] and their – Nocedal's and Wright's – own usage of those results in Thm. 19.2.

Depending on the computation of the second derivative in the quadratic program, different rates of convergence can be achieved. Before citing those results, three assumptions need to hold which are stated first.

Assumption 3.1. *[116, Ass. 18.2] Let u^* be a local solution of the given problem at which the following conditions hold.*

1. *The functions f and c_i are twice differentiable in a neighborhood of u^* with Lipschitz continuous second derivatives.*
2. *The active constraint gradients with respect to u are linearly independent at u^*. This condition implies that the KKT conditions (first-order necessary conditions) are satisfied for some vector of multipliers λ^*.*
3. *The second-order sufficient conditions hold at (u^*, λ^*) (i.e., the Hessian is strictly positive definite in (u^*, λ^*) for all (non-zero) vectors from the critical cone).*

In the present application, the second derivative is computed exactly. In this case and with Ass. 3.1 holding, the rate of convergence is quadratic, when the iterates (u_k, λ_k) are close enough to a local solution (u^*, λ^*).

Generally, if one did not have the exact Hessian, one could achieve only slower rates of convergence. Were one to use a quasi-Newton approximate Hessian B_k, one can show at best superlinear convergence of u_k to u^*. If the reduced-Hessian SQP was applied, the sequence $\{u_k\}_{k\in\mathbb{N}}$ method converges at most two-step superlinearly to u^* [116].

3.2 Monte Carlo objective integration

Writing (2.17) in short as $j = \int \pi(l)h(l)\,dl$ with

$$h(l) \coloneqq \int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2}\left\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\right\|_F^2 \mathrm{d}x + p\big(\Sigma_{\max}(l;\boldsymbol{\alpha})\big)$$

for a fixed $\boldsymbol{\alpha}$ and assuming the evaluation of both functions π and h is efficiently possible, Monte Carlo integration is applied to evaluate the integral.

As a first approach, one draws $N_{\text{MC}} \in \mathbb{N}$ samples $l^{(i)}$ independently and identically distributed (iid) in the integration domain $\mathcal{L}$, i.e., one computes

$$j_{N_{\text{MC}}} = \frac{|\mathcal{L}|}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} \pi\Big(l^{(i)}\Big) h\Big(l^{(i)}\Big).$$

This form of the Monte Carlo integration renders inefficient when the probability measure π is concentrated in small regions. This issue is theoretically easily overcome by drawing the samples directly from the density function π, giving

$$j_{N_{\text{MC}}} = \frac{1}{N_{\text{MC}}} \sum_{i=1}^{N_{\text{MC}}} h\Big(l^{(i)}\Big), \qquad l^{(i)} \overset{iid}{\sim} \pi. \tag{3.5}$$

Following [40, 119], $j_{N_{\text{MC}}}$ converges to the sought after quantity j for $N_{\text{MC}} \to \infty$ (see, e.g., Owen [119, Thm. 9.1]), i.e., $j_{N_{\text{MC}}} \xrightarrow[N_{\text{MC}}\to\infty]{} j$.

The drawing of samples from the inherent density function π is "theoretically" easy because the theory, on the one hand, is easy, but generally the drawing is not possible.

In this case the function needs to be approximated by a proposed density distribution q from which a drawing of samples is easy then. This variant of Monte Carlo integration is known as *importance sampling*. More about the theory and application of importance sampling can be found in [40, 49, 54, 75, 119, 149].

Luckily in the present case, the application of importance sampling is not necessary as the given probability density function π is a composite of numerous normal distributions (see Eq. (2.19) and (2.20)). This allows a drawing of samples in the following composite way:

First, for the given ordered list of motions $\mathcal{M}_q = \{m_1, m_2, \ldots\}$ the cumulative sum of frequencies $w(m_i)$ is computed. That is, for the number $M = |\mathcal{M}_q|$ of motions, the vector $\hat{W}$ stores

$$\hat{W}_k \coloneqq \sum_{i=1}^{k} w(m_i),\ m_i \in \mathcal{M}_q,\ k = 1, \ldots, M.$$

If $\hat{W}_M \neq 1$ – say not all daily motions are measured and included – each frequency $w(m_i)$ is divided by $\hat{W}_M$ thus giving the normalized cumulative sum

$$W_k \coloneqq \frac{1}{\hat{W}_M} \sum_{i=1}^{k} w(m_i),\ m_i \in \mathcal{M}_q,\ k = 1, \ldots, M.$$

With this normalized cumulative sum W one can draw a sample u_1 from a uniform distribution $\mathcal{U}(0, 1)$ and the biggest index i in W for which $u_1 \leq W_i$ defines the motion m_i.

In a similar second step, the time step t_j is sampled from m_i using a normalization of the recorded (equidistant) time steps for motion m_i for which a second sample u_2 is drawn again from a uniform distribution.

Given $l_{ij} = m_i(t_j)$, a random load l is drawn from the normal distribution $\mathcal{N}(l_{ij}, S_{l_{ij}})$, as previously elaborated in Eq. (2.19).

Convergence. The interest here lies with the rate of convergence of j_{MC} from Eq. (3.5) to the true objective value j. The convergence of Monte Carlo integration is given by the strong law of large numbers, cf. [40, Ch. 8.2]. The rate of convergence is in low dimensions comparatively slow with an error bound of $\mathcal{O}(1/\sqrt{N_{\mathrm{MC}}})$ (see, e.g., [111, 141]), but this error bound is independent of the space-dimension d which makes Monte Carlo integration the powerful tool it is. For these results to apply, the involved probability distributions need to be continuous which in the present case they are. While this error bound may not be as good as error bounds of other quadrature rules, e.g., that of the simple trapezoidal rule is $\mathcal{O}(N^{-2d})$, the number of function evaluations to achieve a certain accuracy does not scale with the dimension d [40, Ch. 9.8].

Hence, with Monte Carlo, one has a relatively low convergence rate, but prevents to suffer the curse of dimension by evaluating a multitude of function responses.

This said, it is yet unclear how to choose N_{MC}. A suitable estimator will be presented in Sec. 3.4.

3.3 Kriging interpolation of stresses

The overall aim of optimizing the implant position bears the problem of repeated solves of the contact problem (2.18). This time-consuming contribution to the optimization

algorithm is to be strongly reduced. Therefore the input of the contact problem in the objective function (2.17), namely the quantities

$$\int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2}\left\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\right\|_F^2 \,\mathrm{d}x \quad \text{and} \quad \Sigma_{\max}(l;\boldsymbol{\alpha}),$$

will be interpolated, such that after the interpolation model is set up, one function solve will be comparatively cheap as opposed to solving the PDE (2.18). Note, that the interpolation surface is smooth and cannot capture discontinuities or local high nonlinearities. This is why not the penalty function p itself is interpolated, but the penalty function input $\Sigma_{\max}$.

Since in the following Kriging interpolation will be comprehensively introduced and explained, the section will be structured in the following way: after a general summary of the idea of Kriging interpolation, different Kriging methods will be briefly presented with their interpolation aims and the involved assumptions. Depending on the interpolation approach, the kernel involved in this technique can be a correlation/covariance function C or a variogram function γ. The relationship between these functions, under what conditions they are equivalent, and especially the nature and meaning of a variogram will be discussed in a preceding subsection. Also an assortment of kernels will be given. Eventually, for the chosen Kriging method the estimation of kernel parameters $\boldsymbol{\theta}$ will be explained, which are prerequisite to computing the Kriging weights $\boldsymbol{\omega}$. The section will be finalized by stating some convergence results.

3.3.1 Kriging: a summary

Background. Kriging interpolation – or short Kriging – is also known as Gaussian process regression (GPR), see, e.g., [127, 176]. It is used to construct surrogate models for interpolation as in its present application, but also for classification, supervised and active learning. GPR constructs a statistical model of a partially observed process under the assumption that the observations are a realization of a Gaussian process (GP) (cf. [1]). A GP is uniquely described by its mean and covariance function which are generally unknown and thus have to be estimated. The general advantages of Kriging are that the prediction on the one hand interpolates the given function responses, and on the other hand the prediction is probabilistic so that one can compute empirical confidence intervals. These intervals can be used to decide if one should refit the prediction in some region of interest.

A possible disadvantage is that GPR is not sparse, i.e., it uses *all* input sample-response pairs $\big(l_i, z(l_i)\big)$ to perform the prediction, and it loses efficiency if the number of these pairs becomes too big. This is mostly due to the need of matrix inversion as will be seen in Eq. (3.18). Suitable approximation methods can be applied to counter this downside. But the issue of the inverse as well as the increasing number of floating point operations will eventually slow the method down when the number of inputs becomes high.

In the present setting this disadvantage does not take effect, since the solution of the contact problem is much more expensive compared to the applied linear algebra. Additionally, the number of sample-response pairs is to be kept as low as possible anyway, as it is the proclaimed aim to minimize solves of the contact problem (2.18) to start with. On the other hand the advantages of GPR are fit for exploitation in the given problem setting.

Kriging in a nutshell. The basic idea of Kriging is a fairly simple one: Given $N_\mathrm{K} \in \mathbb{N}$ samples l_i along with function responses $z(l_i)$, so-called sample-response pairs $\big(l_i, z(l_i)\big)$, find weights ω_i such that the linear combination of weighted responses $\sum_{i=1}^{N_\mathrm{K}} \omega_i z(l_i)$ best approximates the function response for an unsampled input l_0

$$z^*(l_0) \coloneqq \sum_{i=1}^{N_\mathrm{K}} \omega_i(l_0) z(l_i) \tag{3.6}$$

with z^* denoting the interpolation function to the sought after function z. Therefore, with the following vector notation

$$\boldsymbol{z}^T \coloneqq \big(z(l_1), \dots, z(l_{N_\mathrm{K}})\big), \qquad \boldsymbol{\omega}^T \coloneqq (\omega_1, \dots, \omega_{N_\mathrm{K}}),$$

Eq. (3.6) can also be written as $z^*(l_0) = \boldsymbol{\omega}(l_0)^T \boldsymbol{z}$.

When inventing this interpolation technique, the South African mining engineer and statistician Danie Krige [95] assumed a local dependency, i.e., the spatial correlation of the responses $z(l_i)$ depends on the samples' relative distance $h_{ij} = |l_i - l_j|$ to each other as opposed to the samples' absolute position. Simply put: sample locations that are close to one another are more likely to have a similar function response than ones that are farther away from each other irrespective of where in a region of interest the samples are placed. In this setting, the unknown function z is interpreted as a random field or a stochastic process.

One is now interested in finding the right representation of the spatial correlation and with that estimate the Kriging weights ω_i for a specific sample l_0. In other words: assuming the right correlation model is known, fit the model parameters $\boldsymbol{\theta}$ to the sample-response pairs $\big(l_i, z(l_i)\big)$. If the sample-response pairs stay unchanged, so do the model parameters $\boldsymbol{\theta}$. With the fixed correlation model and parameters the Kriging weights ω_i can now be estimated for any $l \in \mathcal{L}$.

For reasons, elaborated shortly, in some instances a correlation function does not exist. For these cases a shift in point of view occurs and the model parameters $\boldsymbol{\theta}$ are fitted for a variogram function γ.

3.3.2 Correlation, covariance, and variogram

As it was mentioned at the end of Sec. 3.3.1, one is interested in finding the right representation of the spatial correlation and fit model parameters $\boldsymbol{\theta}$ to either the covariance function cov (or C, as will shortly be introduced) or the variogram γ. In this section, the relationship between correlation, covariance, and the variogram is established. Further, the meaning of the variogram itself is explained. Additionally, the definitions of *second-order stationary* and *intrinsically stationary* will be given, which are later needed as assumptions for the different Kriging methods introduced in Sec. 3.3.3.

First note, that with a covariance cov the dimensionless correlation corr can also be defined. The two measures are related by the equality

$$\mathrm{corr}(X, Y) = \mathrm{cov}(X, Y)/(\sigma_X \sigma_Y),$$

where X and Y are random variables and σ_X, σ_Y are their respective standard deviations. In the following, only the covariance function will be used, knowing that the covariance and correlation are equivalent.

When fitting a covariance function, in the Kriging context the assumption of *second-order stationarity* has to hold.

Definition 3.1. *(Second-order stationarity) Given a spatial domain $\mathcal{D} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, a random function z is called* second-order stationary *if its expectation and covariance are translation invariant over the domain, i.e.,*

$$\begin{aligned} \mathbb{E}[z(x+h)] &= \mathbb{E}[z(x)], \\ \operatorname{cov}[z(x+h), z(x)] &=: C(h), \end{aligned}$$

for all $x, (x+h) \in \mathcal{D}$. This implicates the mean value μ to be constant: $\mathbb{E}[z(x)] = \mu \in \mathbb{R}$ for all $x \in \mathcal{D}$.

Note that cov describes a general, possibly not translation invariant covariance function that takes two arguments. Should the covariance indeed be translation invariant, it depends only on one argument, namely the separating vector h between x and $x+h$, for which the function notation C is defined. Under the assumption of second-order stationarity, the covariance function is bounded: $C(0) < \infty$ [108].

The translation invariance of the covariance means that it is the same between any two points that are the same distance h (and direction, if anisotropy applies) apart no matter where in the region those two points are, see, e.g., Wackernagel [157, Ch. 10]. Since the assumption of second-order stationarity does not always hold, one then usually fits a variogram function $\gamma(h)$.

The variogram. In some instances the covariance cannot be defined [161, Ch. 4.5] and the assumption of second-order stationarity does not hold. Matheron [107, 108] recognized this problem and introduced a new point of view: Instead of looking at the mean, i.e., the expectation, and the variance of the underlying function z, he changed the focus to the expectation and variance of differences for small distance $|h|$

$$\mathbb{E}[z(x) - z(x+h)] =: \mu(h), \qquad \operatorname{Var}[z(x) - z(x+h)] =: 2\gamma(h).$$

The first defines the mean of increments h, the second defines the *variogram function* $\gamma(h)$. For these two measures, one defines *intrinsic stationarity*:

Definition 3.2. *(Intrinsic stationarity) Given a spatial domain $\mathcal{D} \subset \mathbb{R}^d$, $d \in \mathbb{N}$, a random function z is called* intrinsic stationary of order two *if the expectation of differences h is zero, the variance of differences is finite and it only depends on the value h and not the position in the domain, i.e.,*

$$\begin{aligned} \mathbb{E}[z(x+h) - z(x)] &= \mu(h) = 0, \\ \operatorname{Var}[z(x+h), z(x)] &= 2\gamma(h) < \infty, \end{aligned}$$

for all $x, (x+h) \in \mathcal{D}$. The variance of differences $\gamma(h)$ is called variogram.

Note that if second-order stationarity holds, the variogram and the covariance function are equivalent. Then $\lim_{h\to\infty} \gamma(h)$ exists and the equivalence is given by

$$\begin{aligned} &\gamma(h) = C(0) - C(h), \\ &C(h) = \gamma(\infty) - \gamma(h) \qquad \text{with } \gamma(\infty) \coloneqq \lim_{h\to\infty} \gamma(h) < \infty, \end{aligned} \tag{3.7}$$

(cf. Wackernagel [157, Ch. 10] and Webster et al. [161, Ch. 4.5]).

Table 2: *Parametric variogram and corresponding covariance models (where applicable) with parameters $a > 0$ (the range), $b > 0$ (the sill), ν (smoothness), and gamma function Γ and modified Bessel function of second kind K_ν; $h \geq 0$ for all models.*

Bounded linear variogram

$$\gamma_{a,b}(h) = \begin{cases} b\Big(\frac{h}{a}\Big), & \text{if } 0 \leq h \leq a, \\ b & \text{otherwise,} \end{cases}$$

$$C_{a,b}(h) = \begin{cases} b\Big(1 - \frac{h}{a}\Big), & \text{if } 0 \leq h \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

Matérn variogram

$$\gamma_{a,b,\nu}(h) = b\Big(1 - \tfrac{1}{2^{\nu-1}\Gamma(v)}\big(\tfrac{h}{a}\big)^{\nu} K_\nu\big(\tfrac{h}{a}\big)\Big),$$

$$C_{a,b,\nu}(h) = b\Big(\tfrac{1}{2^{\nu-1}\Gamma(v)}\big(\tfrac{h}{a}\big)^{\nu} K_\nu\big(\tfrac{h}{a}\big)\Big),$$

Exponential variogram

$$\gamma_{a,b}(h) = b\Big(1 - \exp\big(-\tfrac{h}{a}\big)\Big),$$

$$C_{a,b}(h) = b\exp\big(-\tfrac{h}{a}\big),$$

Gaussian variogram

$$\gamma_{a,b}(h) = b\Big(1 - \exp\big(-\tfrac{h^2}{a^2}\big)\Big),$$

$$C_{a,b}(h) = b\exp\big(-\tfrac{h^2}{a^2}\big),$$

Power variogram

$$\gamma_{a,b}(h) = bh^a$$

(no corresponding covariance available).

For the choice of variogram models given in Tab. 2[¶], exempt from the power model, a covariance function C can be named for which the equivalence (3.7) holds true. Hence, in case of second-order stationarity the power model cannot be applied as a kernel function. If only the weaker assumption of intrinsic stationarity holds, all listed variograms can be used as a kernel. Only the correspondingly named translation invariant covariance functions C do not represent the correct covariance of the function z.

Where it applies (i.e., not the power variogram), the sill value b describes the supremum of the co-/variance value of differences h. The range parameter a describes in the instance of the bounded linear variogram at what distance h the sill is reached. For the Matérn, the exponential, and the Gaussian model, the sill value is only reached asymptotically for $h \to \infty$. The smoothness ν which only appears in the Matérn model, describes exactly that: the smoothness of the co-/variogram function.

3.3.3 Kriging methods

When introducing *the* Kriging interpolation, it is necessary to differentiate between different Kriging interpolation techniques each with its own set of assumptions. Next to *Kriging the Mean* and *Simple Kriging*, the most common Kriging models are *Ordinary Kriging (OK)* and *Universal Kriging (UK)*.

The first mentioned technique, Kriging the Mean, is not an interpolation per se, but way to estimate the (constant) mean of the underlying distribution.

Simple Kriging then is an interpolation method that uses the information of a known, constant mean in its setup.

[¶]Cf. [157, 161] for further models.

Table 3: *Overview of Kriging interpolation methods.*

Kriging method	*Aim*	*Assumptions*
Kriging the Mean	predict mean value $\mu \in \mathbb{R}$	mean μ unknown and constant, function z is second-order stationary
Simple Kriging	interpolate function z	mean μ known and constant, function z is second-order stationary
Ordinary Kriging	interpolate function z	mean μ unknown and constant, function z is intrinsically stationary
Universal Kriging	interpolate function z	mean μ unknown and not necessarily constant, function z can be decomposed in deterministic drift $\mu(l)$ and intrinsically stationary residual $y(l)$ that has itself zero mean

Ordinary Kriging on the other hand, sets up the interpolation model without prior knowledge of the mean. Yet the assumption is still, that the mean value is constant. As opposed to the first two methods, where second-order stationarity is assumed, here only the weaker assumption of intrinsic stationarity has to hold.

Even though OK may be the most common Kriging method, its assumption of constant mean needs not necessarily be true. This is why the method of Universal Kriging with nonstationary trend $\mu(l)$ will be presented and deployed, of which OK is a special case, since $\mu(l) \equiv \beta_0 = \mu$ is a possible submodel.

More on Kriging and its application can be found in [61, 88, 132, 133, 144] where Sakata et al. [133], using Ordinary Kriging, show that analytical integration is in general possible, making obsolete the need for a numerical integration method, e.g., in the present case that of Monte Carlo integration. Unfortunately, this cannot be applied here due to the local high nonlinearities that are introduced through the penalty function as was already mentioned in Sec. 1.3 and again in the beginning of Sec. 3. Only the smooth response surfaces $\int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2}\left\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\right\|_F^2 \mathrm{d}x$ and $\Sigma_{\max}(l;\boldsymbol{\alpha})$ (cf. Eq. (2.13)) are interpolated and the interpolation result of the latter is then inserted into the penalty function introducing the local distinct features which cannot be integrated.

3.3.4 Kernel parameter estimation for the UK method

In this section information on Universal Kriging from [44, 106, 121, 157, 161] was utilized. Recall from beginning of Sec. 3.3 that there are two functions to be interpolated

$$z_{\Delta\Sigma}(l;\boldsymbol{\alpha}) \coloneqq \int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2}\left\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\right\|_F^2 \mathrm{d}x, \qquad z_{\max}(l;\boldsymbol{\alpha}) \coloneqq \Sigma_{\max}(l;\boldsymbol{\alpha}), \tag{3.8}$$

for fixed angle $\boldsymbol{\alpha}$. For simplicity, all theory will be developed for a generic $z(l)$. The general scheme, when the kernel parameters $\boldsymbol{\theta}$ are not given – $\boldsymbol{\theta}$ representing the vector of parameters, e.g., a, b, and ν of the kernel model of choice, see Tab. 2 –, is to determine

those parameters from the sample-response pairs $\big(l_i, z(l_i)\big)$ at hand. Only then one can compute the Kriging weights $\boldsymbol{\omega}$ necessary to receive an interpolation value $z^*(l_0)$.

Universal Kriging splits the response function $z(l)$ into two parts: a deterministic trend, also called drift function $\mu(l)$, and a residual random function $y(l)$

$$z(l) = \mu(l) + y(l).$$

By definition, only the random function has an associated covariance or variogram. As the samples $z(l_i)$ contain both, the trend and the residual, one cannot directly infer the kernel parameters from the sample-response pairs, but needs a way to purge the bias of the drift from the pairs. A suitable projection matrix X will be derived below. This will then allow to derive the parameters $\boldsymbol{\theta}$ by solving a log-likelihood problem. But before solving for $\boldsymbol{\theta}$, some assumptions and notation need to be introduced.

Assumptions and notation for UK. The random function y is assumed to be second-order stationary with zero mean and variogram function γ, in the context of UK γ is the *residual variogram function*, such that

$$\mathbb{E}[z(l)] = \mu(l), \qquad \gamma(h) = \frac{1}{2}\mathbb{E}\Big[\big(y(x) - y(x+h)\big)^2\Big].$$

The drift μ is defined via the linear combination of $L+1$, $L \in \mathbb{N}_0$, deterministic functions $f_0, f_1, \ldots, f_L$ with $f_k : \mathcal{L} \to \mathbb{R}, k = 0, \ldots, L$. One needs $L+1 < N_\mathrm{K}$ for the problem to be well-defined and which hence enables one to define a projection matrix X as will be seen in the next paragraph. Conventionally, $f_0(l) = 1$. Usually it suffices to model linear or quadratic drift for which one chooses low order polynomial functions f_k correspondingly. For the combination of these functions define unknown coefficients $\beta_k \in \mathbb{R}$ such that

$$\mu(l) = \sum_{k=0}^{L} \beta_k f_k(l).$$

Note, that for the setting of Ordinary Kriging, L is set to zero, hence $\beta_k = 0$ for $k > 0$, and with that $\mu(l) \equiv \beta_0 = \mu$. For abbreviated writing define vectors

$$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_L)^T, \qquad \boldsymbol{f}(l) = \big(f_0(l), \ldots, f_L(l)\big)^T,$$

and similarly

$$\boldsymbol{z} = \big(z(l_1), \ldots, z(l_{N_\mathrm{K}})\big)^T, \qquad \boldsymbol{y} = \big(y(l_1), \ldots, y(l_{N_\mathrm{K}})\big)^T,$$

as well as the matrix

$$F = \begin{pmatrix} 1 & f_1(l_1) & \cdots & f_L(l_1) \\ 1 & f_1(l_2) & \cdots & f_L(l_2) \\ \vdots & \vdots & & \vdots \\ 1 & f_1(l_{N_\mathrm{K}}) & \cdots & f_L(l_{N_\mathrm{K}}) \end{pmatrix}. \tag{3.9}$$

With this, one can now rewrite Eq. (3.6) as

$$z^*(l_0) = \boldsymbol{\omega}^T \boldsymbol{z} = \boldsymbol{\omega}^T (F\boldsymbol{\beta} + \boldsymbol{y}). \tag{3.10}$$

Estimation of kernel parameters $\boldsymbol{\theta}$ via (REML). As the response function $z(l)$ is assumed to be the sum of the *deterministic* drift $\mu(l)$ and the random residual $y(l)$ with only y having an associated covariance or variogram function, the kernel parameters $\boldsymbol{\theta}$ cannot be directly inferred from the measured data pairs $\big(l_i, z(l_i)\big)$. If one did so, the parameters $\boldsymbol{\theta}$ would be biased by the trend $\mu(l)$ which leads to wrong estimates [121]. Thus one has to find a way to determine the best fit $\boldsymbol{\theta}$ without the bias $F\boldsymbol{\beta}$. This is realized by multiplying $\boldsymbol{z} = F\boldsymbol{\beta} + \boldsymbol{y}$ with a matrix X that filters the trend out, i.e., $XF\boldsymbol{\beta} = 0$. This idea will be further pursued below.

Then, following [44, 106, 121, 161], a maximum likelihood method is applied to predict $\boldsymbol{\theta}$. Since the response $\boldsymbol{z}$ will be restricted by matrix X, this method of parameter estimation is called *restricted maximum likelihood (REML)*. As Pardo-Igúzquiza et al. [121] explain, the assumption of second-order stationarity holds for y, since the drift function accounts for long-range variation and the residual for short range variation, i.e., finite correlation length for the covariance, hence the variogram has a sill value and is bounded. This in turn means, that the equivalence (3.7) is true and the parameters $\boldsymbol{\theta}$ can be estimated utilizing the covariance, as it is done in the aforementioned studies that apply the maximum likelihood method.

On returning to matrix X that renders $F\boldsymbol{\beta}$ zero, all references [44, 106, 121, 161] name the projection matrix

$$P = I - F(F^T F)^{-1} F^T$$

which indeed has the desired effect. As Pardo-Igúzquiza et al. state, the matrix P is indeed idempotent and has rank $N_\mathrm{K} - (L+1)$ which the next Lemma proves.

Lemma 3.1. *Assume $(L+1) < N_K$, the samples $l_1, \ldots, l_{N_K}$ are pairwise unequal, and the functions $f_0, \ldots, f_L$ are chosen in such a way, that the matrix $F \in \mathbb{R}^{N_K \times (L+1)}$ defined in Eq. (3.9) has full rank $(L+1)$. With I_{N_K} being the $N_K \times N_K$ identity matrix, define the $P \coloneqq I_{N_K} - F(F^T F)^{-1} F^T$.*

Then the matrix P is idempotent and has rank $N_K - (L+1)$.

Proof. To show the idempotence of P, calculate PP:

$$\begin{aligned} PP &= \big(I_{N_\mathrm{K}} - F(F^T F)^{-1} F^T\big)\big(I_{N_\mathrm{K}} - F(F^T F)^{-1} F^T\big) \\ &= I_{N_\mathrm{K}} - 2F(F^T F)^{-1} F^T - F(F^T F)^{-1} F^T F(F^T F)^{-1} F^T \\ &= I_{N_\mathrm{K}} - F(F^T F)^{-1} F^T = P. \end{aligned}$$

For idempotent matrices it holds true, that the rank of matrix equals its trace [122, Ch. 9.4]. This can be comprehended in two steps. The first is, that the trace of a matrix equals the sum of its eigenvalues. The second step is to see that idempotent matrices only have ones and zeros as eigenvalues. For this, assume λ is an eigenvalue of P with non-zero eigenvector x. Hence $Px = \lambda x$. Then the following is true

$$\lambda x = Px = P^2 x = P \cdot Px = P \cdot \lambda x = \lambda Px = \lambda^2 x.$$

From this follows, that $\lambda = \lambda^2$ and therefore λ is either 0 or 1. Since the number of eigenvalues with value 0 give the dimension of the kernel, the number of ones give the idempotent matrix's rang. Thus $\mathrm{rank}(P) = \mathrm{tr}(P)$.

Now three characteristics of traces will be used, namely

1. $\mathrm{tr}(A + B) = \mathrm{tr}(A) + \mathrm{tr}(B)$, for all $A, B \in \mathbb{R}^{n \times n}$,

2. $\operatorname{tr}(cA) = c\operatorname{tr}(A)$, for all $A \in \mathbb{R}^{n\times n}, c \in \mathbb{R}$,

3. $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, for all $A \in \mathbb{R}^{m\times n}, B \in \mathbb{R}^{n\times m}$.

Then one finds

$$\begin{aligned}\operatorname{rank}(P) = \operatorname{tr}(P) &= \operatorname{tr}\big(I_{N_\mathrm{K}} - F(F^TF)^{-1}F^T\big) = \operatorname{tr}\big(I_{N_\mathrm{K}}\big) - \operatorname{tr}\big(F(F^TF)^{-1}F^T\big) \\ &= N_\mathrm{K} - \operatorname{tr}\big(F^TF(F^TF)^{-1}\big) = N_\mathrm{K} - \operatorname{tr}\big(I_{(L+1)}\big) = N_\mathrm{K} - (L+1).\end{aligned}$$

□

Before proceeding, define the covariance matrix $C(\boldsymbol{\theta})$ of samples l_i given a covariance model C (see Tab. 2) and the parameters $\boldsymbol{\theta}$

$$C(\boldsymbol{\theta})_{i,j} \coloneqq C_{\boldsymbol{\theta}}\big(|l_i - l_j|\big).$$

Since one will be needing $XC(\boldsymbol{\theta})X^T$ to be nonsingular for the log-likelihood estimation it is necessary for X to have full rank. Hence $L+1$ linearly dependent rows need to be omitted from P. But while Webster et al. [161] state, that any $L+1$ rows from P can be deleted, or Pardo-Igúzquiza et al. [121] simply remove the last $L+1$ rows, Dietrich et al. [44] correctly note that arbitrary row deletion need not necessarily remove only linearly dependent rows. Hence, a different approach is required.

Following Dietrich et al. [44], $XF\boldsymbol{\beta} = 0$ needs to hold for all $\boldsymbol{\beta}$. This means that the range of F must be in the null space of X. An obvious choice for X is choosing a matrix whose rows are a basis for the complementary of the rang of F. Dietrich et al. achieve this through the enrichment of F by $N_\mathrm{K} - (L+1)$ zero columns on the right, yielding a square matrix $\tilde{F} = (F, \mathbf{0})$. Of $\tilde{F}$ the singular value decomposition is computed

$$\tilde{F} = W\Lambda V^T. \tag{3.11}$$

From W the last $N_\mathrm{K} - (L+1)$ columns are taken and defined as W_2, which indeed gives an orthonormal basis for the complementary of the range of F, cf. [63, Ch. 2.5.2]. With this, $X \coloneqq W_2^T$ is defined and has the desired properties.

Given X, one can now turn to the definition of the log-likelihood function. The responses in the vector $\boldsymbol{z}$ are assumed to be normally distributed with mean $F\boldsymbol{\beta}$ and covariance $C(\boldsymbol{\theta})$. Hence Dietrich et al. state that the maximum likelihood estimates for $\boldsymbol{\theta}$ and (still unrestricted) $\boldsymbol{\beta}$ are the minimizers of the negative log-likelihood function

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \ln\big(\det(C(\boldsymbol{\theta}))\big) + (\boldsymbol{z} - F\boldsymbol{\beta})^T C(\boldsymbol{\theta})^{-1}(\boldsymbol{z} - F\boldsymbol{\beta}).$$

In the restricted case, the transformed response vector $X\boldsymbol{z}$ is then normally distributed with mean zero and covariance $XC(\boldsymbol{\theta})X^T$. Therefore, the (REML) estimate for $\boldsymbol{\theta}$ is the solution of the minimization of

$$\min_{\boldsymbol{\theta}} rL(\boldsymbol{\theta}) \coloneqq \ln\big(\det(XC(\boldsymbol{\theta})X^T)\big) + \boldsymbol{z}^T X^T \big(XC(\boldsymbol{\theta})X^T\big)^{-1} X\boldsymbol{z}. \tag{3.12}$$

Thus, defining $\tilde{C} \coloneqq XC(\boldsymbol{\theta})X^T$ and $\tilde{C}' \coloneqq XC'(\boldsymbol{\theta})X^T$, where $C'(\boldsymbol{\theta})$ defines the derivative of $C(\boldsymbol{\theta})$ with respect to one of the parameters, e.g., a, one needs to find a such that the *necessary condition*

$$rL'(\boldsymbol{\theta}) = \operatorname{tr}(\tilde{C}^{-1}\tilde{C}') - \boldsymbol{z}^T X^T \tilde{C}^{-1}\tilde{C}'\tilde{C}^{-1} X\boldsymbol{z} = 0 \tag{3.13}$$

holds.

To see that $rL'(\boldsymbol{\theta})$ is indeed the derivative of rL from Eq. (3.13) with respect to one arbitrary constant parameter, realize first, that for arbitrary, invertible $A \in \mathbb{R}^{n\times n}$ with derivative A'

$$0 = I_n' = (A^{-1}A)' = (A^{-1})'A + A^{-1}A' \Rightarrow (A^{-1})' = -A^{-1}A'A^{-1}.$$

The relation $\ln\big(\det(A)\big)' = \operatorname{tr}(A^{-1}A')$ can be found in [122]. The function (3.13) has to be assembled respectively for all parameters in $\boldsymbol{\theta}$ yielding a system of equations to be solved.

Further, one sees easily, that the derivation rule $\operatorname{tr}(A)' = \operatorname{tr}(A')$ holds. Now defining $\tilde{C}'' := XC''(\boldsymbol{\theta})X^T$ analogously to $\tilde{C}'$, where $C''(\boldsymbol{\theta})$ is the second derivative of $C(\boldsymbol{\theta})$ with respect to one of the parameters. With this one can compute $rL''(\boldsymbol{\theta})$ which is used on the one hand for testing the *sufficient condition*

$$\begin{aligned} rL''(\boldsymbol{\theta}) = \operatorname{tr}(\tilde{C}^{-1}\tilde{C}'') - \operatorname{tr}\big((\tilde{C}^{-1}\tilde{C}')^2\big) \\ + 2\boldsymbol{z}^T X^T(\tilde{C}^{-1}\tilde{C}')^2\tilde{C}^{-1}X\boldsymbol{z} - \boldsymbol{z}^T X^T\tilde{C}^{-1}\tilde{C}''\tilde{C}^{-1}X\boldsymbol{z} > 0. \end{aligned} \tag{3.14}$$

On the other hand, the first derivate (3.13) and Hessian (3.14) of $rL(\boldsymbol{\theta})$ are used in the application of the Newton algorithm [116] to solve (3.12) with initial guess $\boldsymbol{\theta}_0$.

3.3.5 Estimation of Kriging weights for the UK method

With a known covariance or variogram function the estimation of the Kriging weights ω_i can now be targeted. The weights can be inferred by a rather simple matrix-vector product whose analytic derivation is the focus of this section. In order to do so, two more assumptions for Kriging need to be introduced and implemented.

Kriging is known as a *Best Linear Unbiased Prediction (BLUP)* of a random field z (cf. Wackernagel [157]). As Wackernagel explains the *unbiasedness* condition translates to

$$\mathbb{E}\big[z^*(l_0) - z(l_0)\big] = 0$$

with $z^*(l_0)$ as in Eq. (3.10). If $\beta_l \neq 0$ for all $l = 0, \dots, L$ the condition is fulfilled if

$$\boldsymbol{\omega}^T F = \boldsymbol{f_0}^T \tag{3.15}$$

holds. Here, $\boldsymbol{f_0} := \boldsymbol{f}(l_0)$.

An unbiased linear predictor is called *best*, if it has minimal prediction variance among all other unbiased linear predictors. With *mse* being the *mean squared error* this means

$$z^*(l_0) = \operatorname{argmin}_{\bar{z}\in \mathrm{ULP}(z)} mse\big(\bar{z}(l_0)\big) := \operatorname{Var}\big(\bar{z}(l_0) - z(l_0)\big) = \mathbb{E}\Big[\big(\bar{z}(l_0) - z(l_0)\big)^2\Big] \tag{3.16}$$

with $\mathrm{ULP}(z)$ representing the space of unbiased linear predictors of z. Together with the unbiasedness condition (3.15) this poses the following minimization problem

$$\begin{aligned} &\min_{\boldsymbol{\omega}} \boldsymbol{\omega}^T(2\boldsymbol{\gamma_0} - \tilde{\Gamma}\boldsymbol{\omega}), \\ &\text{s.t. } \boldsymbol{\omega}^T F = \boldsymbol{f_0}^T. \end{aligned} \tag{3.17}$$

Here, $\tilde{\Gamma}$ is the symmetric variogram matrix and $\boldsymbol{\gamma_0}$ being a variogram vector defined as

$$\tilde{\Gamma}_{ij} = \gamma\big(|l_i - l_j|\big), \qquad \gamma_{0,i} = \gamma\big(|l_i - l_0|\big).$$

The minimization problem (3.17) is solved with the Lagrange approach with multipliers $\boldsymbol{\lambda} \in \mathbb{R}^{L+1}$. In minimizing the Lagrange function, the interpolation function can finally be given as

$$z^*(l_0) = \Big(\boldsymbol{\gamma_0} - F \underbrace{\big(F^T\tilde{\Gamma}^{-1}F\big)^{-1}\big(F^T\tilde{\Gamma}^{-1}\boldsymbol{\gamma_0} - \boldsymbol{f_0}\big)}_{=\boldsymbol{\lambda}}\Big)^T \tilde{\Gamma}^{-1}\boldsymbol{z} = \underbrace{(\boldsymbol{\gamma_0} - F\boldsymbol{\lambda})^T\tilde{\Gamma}^{-1}}_{=\boldsymbol{\omega}^T}\boldsymbol{z}. \tag{3.18}$$

Note that the inverse of the variogram matrix $\tilde{\Gamma}^{-1}$ as well as $\big(F^T\tilde{\Gamma}^{-1}F\big)^{-1}$ only needs to be computed once as long as the sample-response pairs $\big(l_i, z(l_i)\big)$ stay unchanged. The computation of the inverse is usually done using factorization techniques, see, e.g., [40, 109, 138].

Summing up, the following algorithm is applied:

Algorithm 1 Compute Kriging Interpolation

1: **procedure** KrigingInterpolation$\big(l_0$,l_i,$z(l_i)\big)$
2: Estimate variogram parameters $\boldsymbol{\theta}$
a: Solve SVD (3.11) of $\tilde{F}$
b: Solve REML (3.12)
3: Compute the interpolate value $z^*(l_0)$ by solving (3.18)

3.3.6 Convergence

Though Kriging interpolation may not be as common as, for example, spline interpolation or radial basis functions are, results on convergence have been studied as early as 1985 by Yakowitz et al. [175]. Here, for stochastic processes with misspecification in the underlying variogram, pointwise mean square convergence is proven. This is done for scattered data as opposed to a regular grid or even optimally chosen sample points.

Wu et al. [173] took a deterministic function as interpolation aim. Also assuming scattered points, uniform convergence in the Euclidean norm was shown and a convergence rate depending on the mesh width h and the smoothness of the used interpolation functions. Using a Gaussian variogram model and interpolating functions $f : \mathbb{R}^n \to \mathbb{R}$ having generalized Fourier transform $\hat{f}(t)$ satisfying $\int_{\mathbb{R}^n} \big\|\hat{f}(t)\big\|_2^2 \exp\big(\|t\|_2^2\big)\,\mathrm{d}t < \infty$ they find that the local interpolation error – local to the samples l_i – can be arbitrarily small [173, Thm. 6.1]. The result even goes so far that the radius ρ around samples l_i where is error result applies, is not further limited, i.e., the interpolation error can be arbitrarily small in the whole domain. It is to mention here, that Wu et al. mainly focused on proving their results for radial basis functions as interpolation scheme, but utilized and proved bounds for Kriging as well.

Wang et al. [158] present the most recent research result. Next to giving an overview of past convergence results thus contextualizing their own work, they prove uniform convergence in the L_p-norm for an underlying Gaussian process with misspecification and scattered points. The rate of convergence in terms of the fill distance $h_{\boldsymbol{L}}$ as defined in Eq. (3.19) below when using the Matérn model, with ν being the smoothness factor of the Matérn kernel, is given as $h_{\boldsymbol{L}}^{\nu}\big(\log(1/h_{\boldsymbol{L}})\big)^{1/2}$. This error bound is sharp.

Not wanting to state all the technical details, an outline of the needed assumptions will be given for the rate of convergence to hold. For this, define the set of samples

$$\boldsymbol{L} \coloneqq \{l_1, \ldots, l_{N_K}\},$$

the fill distance

$$h_{\boldsymbol{L}} \coloneqq \sup_{l\in\mathcal{L}} \min_{l_j\in\boldsymbol{L}}, \|l - l_j\|_2 \tag{3.19}$$

and a *power function*

$$P_C(l) \coloneqq \left(1 - c_0^T \boldsymbol{C}^{-1} c_0\right)^{1/2}$$

with the covariance vector c_0 and the covariance matrix $\boldsymbol{C}$

$$\boldsymbol{C}_{ij} = C\big(|l_i - l_j|\big), \qquad c_{0,i} = C\big(|l_i - l_0|\big).$$

The supremum of the power function is defined as

$$P_C \coloneqq \sup_{l\in\mathcal{L}} P_C(l). \tag{3.20}$$

Now Wang et al. [158] prove convergence of z^* to z under the condition, that for true covariance function $\mathcal{C}$ of z and imposed covariance C of z^*, C may be no smoother than $\mathcal{C}$. Further, a certain moment condition on the spectral density needs to be fulfilled, which – as stated – holds for the Gaussian and Matérn kernels. Lastly, the set of samples $\boldsymbol{L}$ needs to be dense enough for P_C to exist and to be bounded. Then

$$\mathbb{E}\Big[|z(l) - z^*(l)|_\infty\Big] = \mathcal{O}\Big(P_C\big[(L+1)A + \log^{1/2}(1/P_C)\big]\Big), \tag{3.21}$$

holds, where A depends on the drift functions f_j and the matrix F from Eq. (3.9). It is furthermore noted, that in many situations $(L+1)A$ is bounded from above and thus $\mathcal{O}\big(P_C(L+1)A\big) = \mathcal{O}(P_C)$ holds.

In a last step, let the transfer from the upper bound P_C to the fill distance $h_{\boldsymbol{L}}$ be recollected. For that, regularity assumptions on the domain need to hold. In the case of the Gaussian kernel, the assumption is $\mathcal{L} = [0,1]^d$. For the Matérn model, the conditions are more general, namely $\mathcal{L}$ needs to be compact and convex with a positive Lebesgue measure. Then in both instances, the existence of constants c, h_0 depending on $\mathcal{L}$ and model parameters a, b, and ν (for the Matérn model) can be proven, such that $P_C \leq h_{\boldsymbol{L}}^{c/h_{\boldsymbol{L}}}$ (Gaussian model) or $P_C \leq c h_{\boldsymbol{L}}^{\nu}$ (Matérn model) for $h_{\boldsymbol{L}} \leq h_0$ holds.

In combination with Eq. (3.21) this leads to results on the rate of convergence of $O_{\mathbb{P}}\big(h_{\boldsymbol{L}}^{c/h_{\boldsymbol{L}}-1/2} \log^{1/2}(1/h_{\boldsymbol{L}})\big)$ for the Gaussian model and $O_{\mathbb{P}}\big(h_{\boldsymbol{L}}^{\nu} \log^{1/2}(1/h_{\boldsymbol{L}})\big)$ which concludes this section.

3.4 Adaptivity

The general strategy of adaptivity described below follows the ideas outlined in [67, 94, 154, 180].

In the beginning of Sec. 3 and in Sec. 1.3 it was stated, that for the initial setup of the solution algorithm starts with a coarse discretization and applies adaptive refinement throughout. For the adaptivity the error entering the solution algorithm due to discretization needs to be controlled. In the present case, the discretization errors that need to be controlled come from the FE-mesh on Ω, the solution of the PDE, the interpolation, and Monte Carlo integration. One can see in Fig. 7 how the different approximation errors sequentially enter the computation of the objective (2.17) and its derivative to be discussed in Sec. 4.3. The errors previous to the discrete displacement u_h and u_h itself are assumed to be accurate enough and will in the following be disregarded.

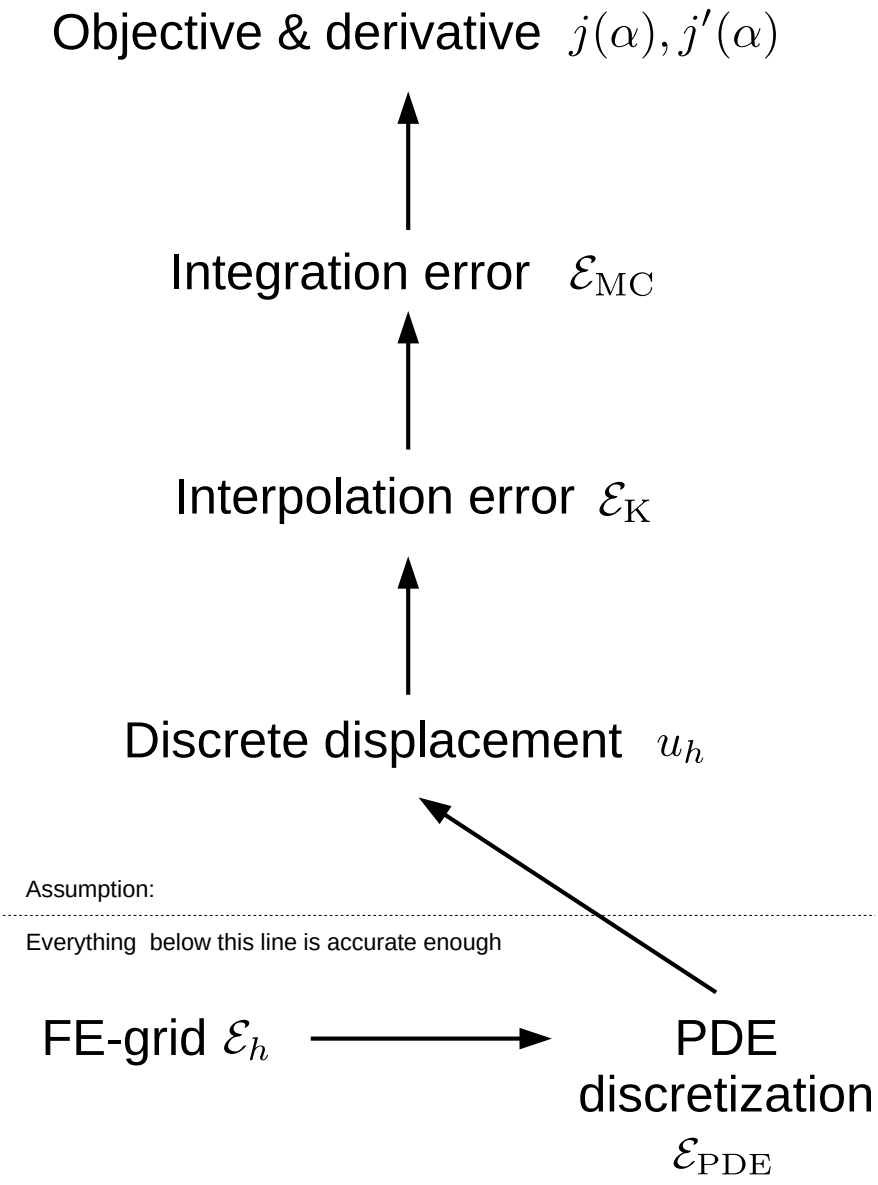

Fig. 7: *Chart of different errors in the computation of the objective* (2.17) *and its derivative; discrete displacement* u_h *and all errors before are assumed to be accurate enough.*

As grid adaptivity is well understood (see, e.g., [10, 23, 27, 97, 137] to name a few) this will not be included in the present work even though eventually this needs to be part of the final optimization algorithm.

This leaves the solution accuracy of the PDE, the accuracy of Kriging, and the accuracy of MC integration as quantities where the discretization error needs to be controlled.

Given a grid discretization $\mathcal{T}_h$ of Ω with grid elements $T \in \mathcal{T}_h$, and the quantity h

$$h = \max_{T \in \mathcal{T}_h} \operatorname{diam}(T),$$

one obtains a grid discretization error τ_h that dictates the accuracy that can at most be attained by the succeeding algorithms of PDE-solution, interpolation, and integration. Equilibrating the errors of these three methods, a classical $3 \cdot \frac{\tau_h}{3}$ argument will be applied. This means that each error estimator of the three, the PDE solution, Kriging, and Monte Carlo integration needs to indicate a performance at least as accurate as $\frac{\tau_h}{3}$. Otherwise more Newton steps in case of the PDE solution, more samples for Kriging, or more draws in the case of Monte Carlo integration are added.

The error estimator $\mathcal{E}_{\mathrm{PDE}}$ for the PDE solution is the easily computable residual of the discretized PDE (2.18) measured in the L^2-norm.

Further, the PDE discretization error needs to be included as well. A priori estimates of the full PDE error depending on the mesh discretization h and convergence results can for example be found in [9, 34, 69, 89, 90, 91, 92, 147, 168, 169]. It is assumed here, that the solution of the PDE is accurate enough and will not be further investigated.

For the Monte Carlo integration the (mean) squared error $\mathcal{E}_{\mathrm{MC}}$ will be introduced in this section, that gives a much better idea of how far the Monte Carlo-estimated value $j_{N_{\mathrm{MC}}}$ is away from the real value j.

Similarly one can estimate the root mean squared error $\mathcal{E}_{\mathrm{K}}$ in Kriging which will be used to decide whether further samples should be added to the model or whether the current model is trustworthy enough. Generally, the Kriging model is suited to sequentially add further samples which is a characteristic that fits very well to an adaptive scheme. When adding new samples, we will employ a hybrid decision tool that balances between global exploration and local exploitation. For the local exploitation we will utilize cross-validation errors, while the Kriging-intrinsic variance will be taken advantage of in global exploration.

Principally, the discretization error constitutes a major influence on the overall error and thus needs to be considered. Yet this influence is on the one hand well researched and including in depth discussion of the discretization error here would on the other hand go beyond scope of this work. Therefore the focus is – deliberately simplifying – limited to the discrete FE-model which is for this thesis regarded as the model to optimize. This is, however, not stringently true, since for bigger implant position adjustments the discretization of the geometry changes and with it so does the finite dimensional model (cf. Sec. 4.2 further below). Nevertheless the discretization error is – piecewise in the space of the control variable – disregarded. Following this line of argument, when any of the three aforementioned subalgorithm's accuracy is adaptively refined, these refinements will in subsequent calculations not be reverted. This holds until the algorithm initiates a remeshing. Then the discretization $\mathcal{T}_h$ changes to $\mathcal{T}_{h'}$ and with it the discretization error shifts to $\tau_{h'}$. Consequently, the subsequent errors change, too, such that all adaptive refinements are nullified.

For further research and inclusion of the aforementioned left out mesh adaptivity, the research by Becker et al. [10], Lamecker [97], and Schiela et al. [137] is to be mentioned.

Carstensen et al. [27] apply mesh refinement in a problem setting that includes an obstacle. Another possibility presents itself through the application of sparse grids and adaptivity in extending the sparse grid. As Bungartz et al. [23] state that this approach can lift the curse of dimension to some extent.

For adaptivity in the solving process of the PDE, one could use multigrid methods by Kornhuber et al. [90] – here in combination with contact and linear elasticity – or collocation by Kouri et al. [93].

Should one want to apply a different surrogate model, Chkifa et al. [32] present adaptivity in polynomial solution interpolation, while Driscoll et al. [45] utilize radial basis functions.

General goal-oriented error estimators and hence goal-oriented driven adaptivity is investigated by Weiser [162], by Janon et al. [84], and by Becker and Rannacher [11].

Monte Carlo sampling. A benefit of the Monte Carlo method is, that one can use the sample values to get a rough estimate σ^2_{MC} of the variance σ^2. The average squared error in Monte Carlo sampling is σ^2/n which also denotes the variance of $j_{N_{\mathrm{MC}}} - j$. Applying the central limit theorem (see Sec. 3.2), we expect the mean value of this difference to be 0. Hence the variance provides us with a confidence interval, i.e., how close $j_{N_{\mathrm{MC}}}$ to j actually is, and its square root gives us the order of error. One uses this approximation of the standard deviation σ to balance with a given tolerance $\frac{\tau_h}{3}$ with which we decide, whether the estimate $j_{N_{\mathrm{MC}}}$ is good enough or we need more sampling. This tolerance derives as stated at the beginning of Sec. 3.4 from the grid discretization $\mathcal{T}_h$ and the subsequent discretization error τ_h.

Here, following [118, Ch. 2], the root mean squared total error for N_{MC} sample points is computed as

$$\left(\mathbb{E}[(j_{N_{\mathrm{MC}}} - j)^2]\right)^{\frac{1}{2}} = \frac{\sigma_{\mathrm{MC}}}{\sqrt{N_{\mathrm{MC}}}},$$

where σ_{MC} is the standard deviation of the random variable (cf. Eq. (3.5)). The Monte Carlo error $\mathcal{E}_{\mathrm{MC}}$ will be measured with the *relative* root mean squared quantity defined as

$$\mathcal{E}_{\mathrm{MC}} \coloneqq \frac{\sigma_{\mathrm{MC}}}{j_{N_{\mathrm{MC}}}\sqrt{N_{\mathrm{MC}}}}, \tag{3.22}$$

To get a computable error estimate, σ_{MC} is approximated using the N_{MC} sample points

$$[\sigma_{\mathrm{MC}}]^2 = \left([h^2] - [h]^2\right) \tag{3.23}$$

with

$$[h^2] = \frac{1}{N_{\mathrm{MC}}}\sum_{i=1}^{N_{\mathrm{MC}}} h_i^2, \qquad [h] = \frac{1}{N_{\mathrm{MC}}}\sum_{i=1}^{N_{\mathrm{MC}}} h_i$$

and $h_i \coloneqq h(l_i)$. Owen [118, Ch. 2.3] suggests a numerically stable way to compute an approximation of σ_{MC} to avoid roundoff errors. For this one sets a quantity $S_{N_{\mathrm{MC}}}$ to $S_{N_{\mathrm{MC}}} = \sum_{i=1}^{N_{\mathrm{MC}}}(h_i - j_{N_{\mathrm{MC}}})^2$. Starting with $j_1 = h_1$ and $S_1 = 0$, the updates are

$$\begin{aligned} \delta_i &= h_i - j_{i-1}, \\ j_i &= j_{i-1} + \frac{1}{i}\delta_i, \\ S_i &= S_{i-1} + \frac{i-1}{i}\delta_i^2, \end{aligned} \tag{3.24}$$

for $i = 2, \ldots, N_{\mathrm{MC}}$ and then set $\sigma^2_{\mathrm{MC}} = S_{N_{\mathrm{MC}}} / (N_{\mathrm{MC}} - 1)$. Therefore, one arrives at

$$\mathcal{E}_{\mathrm{MC}} = \frac{1}{j_{N_{\mathrm{MC}}}} \sqrt{\frac{S_{N_{\mathrm{MC}}}}{N_{\mathrm{MC}} (N_{\mathrm{MC}} - 1)}},$$

with $j_{N_{\mathrm{MC}}}$ as in Eq. (3.5). In case of $\mathcal{E}_{\mathrm{MC}} \geq \frac{\tau_h}{3}$ and with a defined initial sample size $N_{\mathrm{MC},0}$, one adds this initial samples size to the current sample size $N_{\mathrm{MC}} = N_{\mathrm{MC}} + N_{\mathrm{MC},0}$ and reiterates Monte Carlo, drawing $N_{\mathrm{MC},0}$ more samples.

Kriging. For the adaptive enrichment of sampling points in Kriging interpolation, the exploitation of two strategies is implemented: global exploration and local exploitation, and depending on the success of the lastly added sample favor the one, the other, or balance the two. Therefore, the algorithm proposed by Liu et al. [100] is employed. To do this a short motivation and description of the procedure will be given.

Generally one seeks to minimize the overall generalization error

$$\mathcal{GE}_{\mathrm{K}} = \int_{\mathcal{L}} \mathbb{E}[\mathrm{PE}(l)] \, dl$$

where $\mathrm{PE}(l)$ is the prediction error

$$\mathrm{PE}(l) = \big(z(l) - z^*(l)\big)^2. \tag{3.25}$$

Following Liu et al. [100, Sec. 3], $\mathbb{E}\big[\mathrm{PE}(l)\big]$ can be decomposed to (with the requirement of z and z^* being stochastically independent)

$$\mathbb{E}\big[\mathrm{PE}(l)\big] = \Big(\mathbb{E}\big[z(l)\big] - \mathbb{E}\big[z^*(l)\big]\Big)^2 + \mathbb{E}\Big[\big(z^*(l) - \mathbb{E}\big[z^*(l)\big]\big)^2\Big] + \mathbb{E}\Big[\big(z(l) - \mathbb{E}\big[z(l)\big]\big)^2\Big]. \tag{3.26}$$

As Liu et al. describe, the first term on the right-hand side is the bias representing the average difference between the predicted response $z^*(l)$ and the observed response $z(l)$. The second term is the model's prediction variance which is known from Eq. (3.16). The third term represents the intrinsic noise of the data, which is not of interest at this point.

The bias accounts for local exploitation guiding the sampling towards regions with large prediction errors, while the variance ensures global exploration by avoiding the omission of undetected interesting regions.

The performance of the adaptive scheme is as Liu et al. mention greatly influenced by the trade-off between those two aspects and they propose a flexible rule which is put to practice here, too. This rule adapts to whether the bias error was under- or overestimated.

The variance can easily be computed evaluating Eq. (3.17) after the weights are computed. The bias, on the other hand, needs some elaboration.

From various options of bias computation, the cross-validation errors are evaluated here. This works as follows: One iterates over all N_{K} sampled pairs (l_i, z_i) with $z_i \coloneqq z(l_i)$. In each iteration take one sample pair out and compute how well the other samples interpolate the missing one.

Denote by $\mathcal{I} \coloneqq \{1, \ldots, N_{\mathrm{K}}\}$ the index set of the Kriging nodes corresponding to the samples, and by

$$\bar{\mathcal{I}}_j \coloneqq \{1, \ldots, j-1, j+1, \ldots, N_{\mathrm{K}}\} = \mathcal{I} \setminus \{j\} \tag{3.27}$$

the index set without index j. Then the cross-validation error at a sampled location l_j can be estimated as

$$\varepsilon_{\mathrm{K}}^{j} = \left| \sum_{i \in \bar{\mathcal{I}}_j} \bar{\omega}_i z_i - z_j \right|,$$

where $\bar{\omega}_i$ are the re-computed Kriging weights for Kriging interpolation on $\bar{\mathcal{I}}_j$. Compared to the effort of computing the sample values z_i from the PDE (2.16) the overhead for re-computing the weights for error estimation is negligible.

While the model's variance can be evaluated at any $l \in \mathcal{L}$ by computing

$$\mathrm{Var}(l) = \boldsymbol{\omega}(l)^T \big(2\boldsymbol{\gamma}(l) - \tilde{\Gamma}\boldsymbol{\omega}(l)\big) \overset{(UK)}{=} \boldsymbol{\omega}(l)^T \boldsymbol{\gamma}(l) + \boldsymbol{\lambda}(l)^T \boldsymbol{f}(l), \tag{3.28}$$

the cross-validation error can only be estimated for the sample locations $l_i, i \in \mathcal{I}$. To extend the error to the whole domain one takes the Voronoi partition of $\mathcal{L}$ according to the Kriging samples l_i and has the cells C_i defined as

$$C_i \coloneqq \{l \in \mathcal{L} : |l - l_i| \leq |l - l_j| \forall j \neq i\}, \quad 1 \leq i, j \leq N_{\mathrm{K}}.$$

With this one has

$$\varepsilon_{\mathrm{K}}(l) = \varepsilon_{\mathrm{K}}^{j}, \quad l \in C_j. \tag{3.29}$$

For the balancing of the two quantities, Liu et al. propose to introduce a parameter $\beta \in [0, 1)$ that adjusts in such a way that it accounts for the over- or underestimation of the true error $\varepsilon_{\mathrm{true}} = |z^*(l) - z(l)|$ when the last node $l_{N_{\mathrm{K}}+p-1}, p = 2, 3, \ldots$ was added

$$\beta = \begin{cases} 0.5, & p = 1, \\ 0.99 \cdot \min\left\{0.5 \frac{\varepsilon^2_{\mathrm{true}}(l_{N_{\mathrm{K}}+q-1})}{\varepsilon^2_{\mathrm{K}}(l_{N_{\mathrm{K}}+q-1})}, 1\right\}, & p > 1. \end{cases} \tag{3.30}$$

With this define the expected prediction error (EPE)

$$\mathrm{EPE}(l) = \beta\big(\varepsilon_{\mathrm{K}}(l)\big)^2 + (1 - \beta)\mathrm{Var}(l)$$

and find $l_{N_{\mathrm{K}}+p}$ by solving

$$l_{N_{\mathrm{K}}+p} = \mathrm{argmax}_{l \in \mathcal{L}} \mathrm{EPE}(l). \tag{3.31}$$

Thus for small β, the variance has a bigger impact and global exploration is favored. For β values towards 1, the bias measured with the cross-validation error and hence local exploitation are predominant in finding a new sample location.

Remark. *Recall from Sec. 3.3, that in the present context there are two function interpolations z^*, namely $z_1^* \coloneqq z_{\Delta\Sigma}^*$ and $z_2^* \coloneqq z_{\max}^*$, see also Eq. (3.8). The prediction error PE(l) (3.25) will be estimated for each function interpolation separately. Also the quantities Var(l) (3.28), $\varepsilon_K(l)$ (3.29), and ε_{true} will be computed separately and then added together, such that, e.g., $\varepsilon_K(l) \coloneqq \varepsilon_{K,1}(l) + \varepsilon_{K,2}(l)$.*

On the question of how to choose the "$l \in \mathcal{L}$" in (3.31) from which the next added sample $l_{N_{\mathrm{K}}+p}$ is estimated, there exist at least two practicable approaches. One is, to either have a prepartitioning of $\mathcal{L}$, say $\mathcal{L}_k \subset \mathcal{L}$, and evaluate the expected prediction error EPE(l) for all $l \in \mathcal{L}_k$.

The other is to sample $\mathcal{L}$ using randomly drawn samples, Latin hypercube sampling, or orthogonal sampling. The method of choice is here importance sampling, i.e., sampling $l \in \mathcal{L}$ using the probability density $\pi(l)$. The amount of drawn samples to determine the maximum expected prediction error from, will be $N_{\mathrm{MC},0}$ in each step, the initial sample size of Monte Carlo integration.

Kriging error $\mathcal{E}_{\mathrm{K}}$. In order to define the Kriging error $\mathcal{E}_{\mathrm{K}}$, which measures the precision of the interpolation, the prediction variance $\mathrm{Var}(l)$ minimized in Eq. (3.17) and again given in Eq. (3.28) will be used. If the interpolation lacks precision, it will be adaptively refined. The error will be measured relative to the objective function value j_{MC} (3.5), so that both the MC error $\mathcal{E}_{\mathrm{MC}}$ (3.22) and the Kriging error $\mathcal{E}_{\mathrm{K}}$ are relative errors.

Recall again from Eq. (3.8) that two response surfaces are interpolated. For the evaluation of the target function Eq. (2.17), the interpolation result for $\Sigma_{\max}$ is then inserted into the penalty function p from Eq. (2.14) of which the result thereof together with the interpolated value of stress difference $\Delta\Sigma$ are needed. In order to be able to compare the errors of Kriging and Monte Carlo, the relative Kriging error will be quantified in terms of its impact on the objective. Therefore, the prediction variance Eq. (3.28) will be computed and the standard deviation σ_{K} of the prediction be defined as

$$\sigma_{\mathrm{K}}(l) \coloneqq \mathrm{Var}(l)^{1/2}. \tag{3.32}$$

For the two response surfaces, there will be two standard deviations of the prediction: $\sigma_{\mathrm{K},1}$ for the response of the stress differences $\Delta\Sigma$ and $\sigma_{\mathrm{K},2}$ for the response of maximum stresses $\Sigma_{\max}$. The quantity $\sigma_{\mathrm{K},2}$ will be measured in the linear approximation of the penalty p, i.e., by $p'(\sigma_{\mathrm{K},2})\sigma_{\mathrm{K},2}$ where

$$p'(\sigma_{\mathrm{K},2}) \coloneqq \frac{\mathrm{d}}{\mathrm{d}\Sigma_{\max}(l;\boldsymbol{\alpha})} p(\sigma_{\mathrm{K},2}) = bc \cdot \exp\Big(c\Big(\sigma_{\mathrm{K},2} - \Sigma_{\mathrm{thres}}\Big)\Big). \tag{3.33}$$

Inserted in the objective function, the relative Kriging error estimate is given as

$$\mathcal{E}_{\mathrm{K}} \coloneqq \frac{1}{j_{\mathrm{MC}} \cdot N_{\mathrm{K},\mathcal{E}}} \sum_{i=1}^{N_{\mathrm{K},\mathcal{E}}} \pi(l_i)\Big[\sigma_{\mathrm{K},1}(l_i) + p'\big(\sigma_{\mathrm{K},2}(l_i)\big)\sigma_{\mathrm{K},2}(l_i)\Big],$$

for $N_{\mathrm{K},\mathcal{E}} \in \mathbb{N}$. The error can either be estimated using cross-validation with the given sample-response pairs $\big(l_j, z(l_j)\big)$, necessitating the setting up and evaluation of $2 \cdot N_{\mathrm{K}}$ Kriging interpolations, or one can utilize the probability $\pi(l)$, Eq. (2.21), and the aforementioned Monte Carlo sampling to compute

$$\mathcal{E}_{\mathrm{K}} = \frac{1}{j_{\mathrm{MC}} \cdot N_{\mathrm{K},\mathcal{E}}} \sum_{i=1}^{N_{\mathrm{K},\mathcal{E}}} \sigma_{\mathrm{K},1}\left(l^{(i)}\right) + p'\left(\sigma_{\mathrm{K},2}\left(l^{(i)}\right)\right)\sigma_{\mathrm{K},2}\left(l^{(i)}\right), \qquad l^{(i)} \overset{iid}{\sim} \pi. \tag{3.34}$$

The quantity of $\mathcal{E}_{\mathrm{K}}$ will be used to measure the interpolation precision of Kriging. If the precision check fails, more samples will be added to the Kriging model to improve the interpolation prediction.

Error equilibration. With a tolerance τ_h given by the domain discretization $\mathcal{T}_h$ one requires the error $\mathcal{E}_{\mathrm{PDE}}$ due to the PDE (2.18) as well as the errors of Monte Carlo integration $\mathcal{E}_{\mathrm{MC}}$ (3.22) and Kriging $\mathcal{E}_{\mathrm{K}}$ (3.34) to each be lower than a third of the tolerance

$$\mathcal{E}_{\mathrm{PDE}} < \frac{\tau_h}{3}, \qquad \mathcal{E}_{\mathrm{K}} < \frac{\tau_h}{3}, \qquad \mathcal{E}_{\mathrm{MC}} < \frac{\tau_h}{3}. \tag{3.35}$$

Adding these three errors up yields

$$\mathcal{E}_{\mathrm{PDE}} + \mathcal{E}_{\mathrm{K}} + \mathcal{E}_{\mathrm{MC}} < \tau_h.$$

Generally one could choose to weigh the errors differently, i.e., not uniformly with a third per contributing error. In the present context the equilibration of the errors is chosen as a first approach.

Hence for the computation of the solution of the PDE (2.18), one would require the performance of as many Newton steps as needed for the error $\mathcal{E}_{\mathrm{PDE}}$ to be below $\frac{\tau_h}{3}$. Here it is assumed though, as mentioned in the beginning of Sec. 3.4, that the solution of the PDE is accurate enough and it is not further examined.

In Kriging interpolation, the given initial sample size N_{K} is increased one sample at a time as described above until the accuracy criterion is met.

The enrichment of Monte Carlo samples with the initial sample size $N_{\mathrm{MC},0}$ each time $\mathcal{E}_{\mathrm{MC}} < \frac{\tau_h}{3}$ does not hold, was already explained at the end of the paragraph for Monte Carlo in Sec. 3.4.

The final number of Kriging samples $N_{\mathrm{K}} + p$, $p \in \mathbb{N}_0$, and MC draws $N_{\mathrm{MC},0} + oN_{\mathrm{MC},0}$, $o \in \mathbb{N}_0$, will stay unchanged at the start of the next optimization iteration, but may again be adaptively enriched. These quantities are reset when implant positioning requires a remeshing, resulting in a different mesh discretization $\mathcal{T}_{h'}$ which in consequence changes the (mesh-) discretization error to $\tau_{h'}$.

3.5 Objective computation

Summing up the chapter, this section will bring together the subroutines mentioned so far and present them in a complete algorithm.

For the evaluation of the objective function one needs to fix initial values for N_{K} sample loads, the implant position $\boldsymbol{\alpha}$ as well as the number of Monte Carlo samples N_{MC}. Then the procedure to compute target function value is given in Alg. 2.

Algorithm 2 Compute Objective Value $j(\boldsymbol{\alpha})$

1: **procedure** ObjectiveFunctionValue($\boldsymbol{\alpha}$,N_{K},N_{MC},τ_h)
2: Compute response values $\Sigma(l;\boldsymbol{\alpha})$ and $\Sigma_{\max}(l;\boldsymbol{\alpha})$ for N_{K} loads
3: Apply Kriging (Alg. 1)
a: Adaptively increase samples according to Eq. (3.31) when performance check (3.35) fails
4: Apply Monte Carlo integration using N_{MC} samples
a: Adaptively increase samples when performance check (3.35) fails

4 Optimization

For the optimization of the implant position a quasi-Newton algorithm, namely the BFGS algorithm, in combination with line search will be applied. The necessary concepts next to the analytic derivation of the gradient of $j(\boldsymbol{\alpha})$ are explained in this chapter.

Let it be said that trust-region methods pose another possible approach and alternative to line search. They can be used to handle inexactness in the objective evaluation and the objective's gradient. Both quantities are here given analytically, the latter in Sec. 4.3, and their numeric inexactness is handled with adaptivity in Kriging and Monte Carlo integration. The advantage of global convergence and the ability to minimize a non-convex function under constraints are shared with both trust-region methods and the in the following explained line search algorithm such that both concepts can be equally well applied. More on trust-region methods can be found in [50, 71, 93].

The performance and runtime of the BFGS optimization could be enhanced by combining the quasi-Newton algorithm with multilevel approaches. Multilevel, also known as multigrid or multiscale methods, implement the idea of solving on a set of grids ranging from a fine to a coarse discretization.

They were first formulated to solve PDEs as stated by Ho et al. [81], but have since been applied to the solution of optimization problems, see, e.g., [99, 113, 114]. Typically one starts with a fine grid, where the solution is expensive, and restricts the entities of importance to increasingly coarse grids until the coarsest is reached, where one solve is cheap. Then the solution is prolonged to ever finer grids until the finest level is reached again. For both, the restriction and prolongation, it is possible and common practice to perform a smoothing step. See [12, 22] and references within for the general setup.

Multilevel methods have been greatly researched for optimization, e.g., by Gratton et al. [71] in combination with trust-region schemes, by Ho et al. [80, 81] with a Galerkin model or with Newton-type optimization, or Gräser et al. [70] for a truncated nonsmooth Newton multigrid algorithm which is then again applied to solve contact problems and model, e.g., knee joint motion [135].

Weiser et al. [163] combined inexact Newton methods, that include Newton-like methods, a Newton-Truncated-CG methods and Newton-Lanczos type methods with adaptive multilevel finite element implementation for the case of nonconvex minimization.

On the other hand multilevel methods have extensively been researched in their application to (non-)linear contact- and obstacle problems [69, 90, 91, 92, 169]. Kornhuber and Krause [90] investigate the incorporation of adaptivity in combination with contact and linear elasticity.

That being said, it is left for future work to incorporate multilevel optimization in one way or the other into the implant positioning optimization.

4.1 Quasi-Newton optimization

For the optimization the BFGS algorithm (named after its inventors Broyden, Fletcher, Goldfarb, and Shanno) is used. This well-known quasi-Newton method is robust, guarantees global convergence, i.e., independent of the initial guess, for smooth functions and a superlinear convergence rate close to a (local) minimum, see, e.g., [2, 116, 166, 178] and references within. If the function's derivative is bounded in a convex domain around a minimum, then the convergence of the program is already given, even without the need of the function itself to be convex.

As for any descent algorithm, the general update rule reads as

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + s_k d_k.$$

The stepsize of the current iteration is given with s_k, while d_k represents the descent direction. The BFGS method states for d_k

$$d_k = -B_k^{-1} j'(\boldsymbol{\alpha}_k)^T, \tag{4.1}$$

where B_k is an approximation of the second derivative of function j in its current iteration $\boldsymbol{\alpha}_k$ which is the main idea of quasi-Newton: instead of computing the second derivative of the function, the Hessian is approximated through a symmetric, positive definite matrix, which is updated and adjusted when $\boldsymbol{\alpha}_k$ changes to $\boldsymbol{\alpha}_{k+1}$.

Citing Xie et al. [174], the key in the convergence analysis of quasi-Newton methods is to show that the search direction is not orthogonal to the gradient. This so-called *angle-condition* translates to

$$\cos(\phi_k) = \frac{-d_k^T j'(\boldsymbol{\alpha}_k)^T}{\|d_k\|_2 \|j'(\boldsymbol{\alpha}_k)\|_2} \neq 0. \tag{4.2}$$

The gradient is computed as described in Sec. 4.3. For an initial guess B_0, one can take the identity $B_0 = I$ or the identity multiplied with a multiple of the objective, e.g., $B_0 = |j(\boldsymbol{\alpha}_0)|I$. As Nocedal and Wright state [116, Ch. 6], an often effective heuristic is to scale the starting matrix *after* the first stepsize is accepted yet before the first BFGS update is performed. The first matrix iterate B_0 is set to

$$B_0 \coloneqq \frac{y^T y}{y^T t} I$$

with

$$t \coloneqq \boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k = s_k d_k, \qquad y^T \coloneqq j'(\boldsymbol{\alpha}_{k+1}) - j'(\boldsymbol{\alpha}_k),$$

Here, s_k is the stepsize in iteration k. Before stating the update rule for B, first recall the *Armijo-Wolfe conditions* for finding a stepsize. Let $\tau_1 \in (0,1)$ and $\tau_2 \in (\tau_1, 1)$, then find a stepsize s_k that fulfills

$$j(\boldsymbol{\alpha}_{k+1}) \leq j(\boldsymbol{\alpha}_k) + \tau_1 s_k j'(\boldsymbol{\alpha}_k) d_k \quad \text{(Armijo condition)}, \tag{4.3a}$$

$$j'(\boldsymbol{\alpha}_{k+1}) d_k \geq \tau_2 j'(\boldsymbol{\alpha}_k) d_k \qquad \text{(Wolfe condition)} \tag{4.3b}$$

for $\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + s_k d_k$. The first inequality is known as *Armijo rule* and ensures sufficient descent of the step. The second inequality is a slope condition, guaranteeing sufficient reduction of the slope. Convergence results for BFGS can already be proved for only the Armijo rule. But with the Wolfe stepsize – at least theoretically [166] – the updates of B_k are ensured to be positive definite. With the descent direction d_k from (4.1) and a stepsize s_k fulfilling (4.3) the matrix update is

$$B_{k+1} = B_k - \frac{(B_k t)(B_k t)^T}{t^T B_k t} + \frac{y y^T}{y^T t}. \tag{4.4}$$

In practice different approaches to B offer themselves up, since inverting B may be impractical. One different take is to work with the approximate $H = B^{-1}$ and have an

update rule for H_{k+1}. While Werner mentions it in his book [166, p. 201] that one loses control of whether the update matrix B_{k+1} is still "sufficiently positive definite", both he and Nocedal and Wright [116, p. 141] show, that if H_k is positive definite, then H_{k+1} is positive definite, too.

A different idea that is also memory efficient, computes, stores, and updates the Cholesky decomposition of B, namely the triangular matrix L.

The BFGS method will here be applied problem setting of implant position optimization. With the constraint of retaining the range of motion the spatial position of the head of the femur's implant is all but fixed, leaving three design variables $\boldsymbol{\alpha} = (\alpha_x, \alpha_y, \alpha_z)$ as was already argued in Sec. 2.1. This leads to a relatively small Hessian to be stored and inverted.

It is to mention, that the convergence of BFGS algorithm is largely proved for exact values of the objective function j and its derivative j'. As the computed values here come from discretized quantities, they are inherently inexact. For this setting there are yet few results available. Götschel [64, Ch. 5.3] proves that d_k is indeed a descent direction that fulfills the angle condition (4.2) in case of inexact derivative $\tilde{j}'_k$, if the symmetric positive definite matrix B_k has bounded condition number $\kappa(B_k)$ and the error in the derivative $e^k_{\tilde{j}'}$ is small enough

$$\|e^k_{\tilde{j}'}\|_2 \leq \frac{\varepsilon}{\kappa(B_k)^{1/2}} \|\tilde{j}'_k\|_2$$

with $\varepsilon < 1/2$. Götschel is even able to show that the BFGS update (4.4) still preserves symmetry and positive definiteness of B_k, even if quantities t and y are inexact. The subsequent convergence results are then given for convex problems.

Recent research by Xie, Byrd, and Nocedal [174] shows that with a slight modification of the BFGS algorithm, i.e., with a lengthening of the search direction if no stepsize s_k is found fulfilling the Armijo-Wolfe conditions (4.3), the optimization method still converges to a neighborhood of the solution at an R-linear rate. For the proofs the assumptions are required, that the function j is bounded from below with an M-Lipschitz continuous ($M > 0$) derivative, i.e.,

$$\left\|j'(\boldsymbol{\alpha}_1) - j'(\boldsymbol{\alpha}_2)\right\|_2 \leq M\|\boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_2\|_2, \text{ for all } \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in A.$$

Furthermore, the errors in function and gradient values need to be uniformly bounded, and lastly the function j needs to be m-strongly convex, with $0 < m \leq M$.

Though both summands in the target function Eq. (2.17) are strictly or m-strongly convex in l, the same needs not necessarily be true for $\boldsymbol{\alpha}$.

Thus it needs to be numerically evaluated whether the BFGS algorithm still converges for the setting of optimal positioning of the hip joint implant. Otherwise BFGS could be replaced with a simpler gradient descent algorithm. However, praxis has shown that BFGS is functional here.

Remark. *When discretizing a continuous problem (2.17) and (2.18), discretization errors are inevitable. When the discretization is changed – be it of the FE-grid $\mathcal{T}_h$, the sample-response pairs in Kriging introduced in Sec. 3.3, the drawn Monte Carlo samples introduced in Sec. 3.2 – the discretization error changes and with it the problem definition. These are so called* "differential crimes" *in the optimization, possibly leading to discontinuities in the evaluation of the (discrete) target function, thus prohibiting the monotone decrease of the objective function value.*

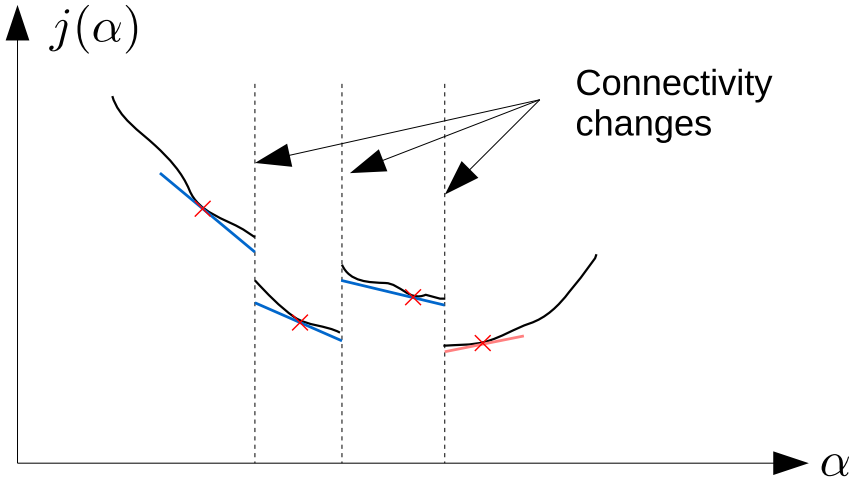

Fig. 8: *Discontinuous objective function due to topology changes.*

The "differential crimes" in Kriging and Monte Carlo can relatively easy be accounted for; the former just needs to reuse the previous samples. For the Monte Carlo integration, the loads l_i are only drawn randomly the first time MC is used and saved to be reused in every consecutive application of Monte Carlo as long as the discretization of the geometry Ω_h stays unaltered. Otherwise the Monte Carlo-samples will be reset.

Only a remeshing of the geometry will be inevitable when the position of the femur implant changes too much. The prevention of a differential crime is especially necessary for the step size estimation when evaluating the Armijo-Wolfe conditions (4.3). This will be accounted for by a special line search proposed in the next section.

4.2 Line search under grid topology changes

Position changes of the implant grid within the bone will, if the change is big enough, lead to a different meshing of bone and implant which ultimately leads to discontinuities in the objective function. See, e.g., Fig. 8 for a symbolic portrayal of such discontinuities. The alternatives are those of a strangely deformed grid or a grid with bad mesh properties which are both undesirable. A remeshing leads to a change in the discretization error and ultimately to a faulty optimization algorithm, if this is not accounted for. In order to prevent the commission of a differential crime, a fixed body transformation is computed and applied to the bone nodes. Also, a threshold value $\Delta\boldsymbol{\alpha}_{\max}$ is defined. Given a remeshing of the geometry took place at $\boldsymbol{\alpha}_*$, only $\Delta\boldsymbol{\alpha} \in B_{\Delta\boldsymbol{\alpha}_{\max}}(\boldsymbol{\alpha}_*)$ with $B_{\Delta\boldsymbol{\alpha}_{\max}}(\boldsymbol{\alpha}_*)$ defined as $B_{\Delta\boldsymbol{\alpha}_{\max}}(\boldsymbol{\alpha}_*) := \{\Delta\boldsymbol{\alpha} \in A : |\Delta\boldsymbol{\alpha} - \boldsymbol{\alpha}_*| < \Delta\boldsymbol{\alpha}_{\max}\}$ is feasible for an implant positioning without changing the mesh discretization. The set A represents all feasible angles.

Should the positioning of the implant deviate more than $\Delta\boldsymbol{\alpha}_{\max}$ from an initial position $\boldsymbol{\alpha}_*$, a remeshing is triggered by the algorithm thus controlling anticipated discontinuities in the objective. The Wolfe conditions Eq. (4.3), namely the first part – the Armijo condition – can only be tested in this small neighborhood $B_{\Delta\boldsymbol{\alpha}_{\max}}(\boldsymbol{\alpha}_*)$. That limits the stepsize to the boundary of the ball $B_{\Delta\boldsymbol{\alpha}_{\max}}$. Once a new iterate $\boldsymbol{\alpha}_{k+1}$ reaches the limit, a remeshing is initiated.

Another possibility is to test whether the descent direction is unchanged across the threshold, i.e., test if $\langle j'(\boldsymbol{\alpha}_{k+1})^T, d_k\rangle_2 < 0$ (cf. angle-condition (4.2)) holds true and sus-

pend the sufficient descent check for this duration which is the modus operandi here.

More on line search and Wolfe condition can be found in Yousefpour [178].

4.3 Derivative evaluation

In order to determine an optimized implant position, the target function (2.17) is minimized utilizing its derivative and employing it for the descent direction d_k as described in Eq. (4.1). For the evaluation of the objective function's derivative $j'(\boldsymbol{\alpha})$ start with the function itself

$$j(\boldsymbol{\alpha}) = \int_{l\in\mathcal{L}} \pi(l) \int_{x\in\Omega_{\text{bone}}(\boldsymbol{\alpha})} \frac{1}{2}\big\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\big\|_F^2 \,\mathrm{d}x + p\big(\Sigma_{\max}(l;\boldsymbol{\alpha})\big)\,\mathrm{d}l.$$

The control variable $\boldsymbol{\alpha} \in A$ is from a set of admissible angles A. The main difficulty lies with the control $\boldsymbol{\alpha}$ influencing the PDE domain $\Omega_{\text{bone}}(\boldsymbol{\alpha})$ and thus the Cauchy stress tensor $\Sigma(\boldsymbol{\alpha})$ that derives from the partial differential equation (2.18). This necessitates the use of the transformation rule as well as the implicit function theorem. Since this is the case, the function j will initially be stripped of the integral over the load domain $\mathcal{L}$ and of the penalty function p (2.14) whose derivative is straight forward.

Remark. *The easily computable derivative of the penalty p is indeed true for all appearing terms in Eq. (2.14) but for the maximum norm. As mention in Sec. 2.1.5, the maximum norm could be approximiated by the L^p-norm with sufficiently big exponent p. Of this norm, one can again compute a derivative with ease and use this as an approximation in the computation of $\frac{\partial}{\partial\boldsymbol{\alpha}}p$.*

Also an abbreviation $\Omega(\boldsymbol{\alpha})$ for $\Omega_{\text{bone}}(\boldsymbol{\alpha})$ will be used. The (reduced) target function now reads as

$$\tilde{j}(\boldsymbol{\alpha}) = \int_{\Omega(\boldsymbol{\alpha})} \frac{1}{2}\big\|\Sigma\big(u(\boldsymbol{\alpha})\big) - \bar{\Sigma}\big\|_F^2 \,\mathrm{d}x. \tag{4.5}$$

For the correct application of the transformation rule, the findings from Brandenburg et al. [19] are implemented. The domain transformation $\Omega(\boldsymbol{\alpha})$ or more precisely the positioning of the implant in the bone is realized through a transformation function

$$\tau(\boldsymbol{\alpha}) := \tau(x;\boldsymbol{\alpha}) = x + \boldsymbol{\alpha}\bar{u}(x) = x + \bar{u}^{\boldsymbol{\alpha}}(x) \tag{4.6}$$

which is defined as the deformation of an elastic body given the boundary displacements $\bar{u}$.

With $\boldsymbol{\alpha} = 0$, the reference configuration is given as $\Omega(0) =: \Omega$ and one computes $\Sigma := \Sigma(u(0))$. When $\boldsymbol{\alpha} \neq 0$ applies, the definition is adjusted to $\Omega(\boldsymbol{\alpha}) =: \Omega_{\boldsymbol{\alpha}}$, $\tau(\boldsymbol{\alpha}) =: \tau$, and $\Sigma(u(\boldsymbol{\alpha})) =: \Sigma(\boldsymbol{\alpha})$. For better readability and for better differentiation between the reference domain Ω and the deformed domain $\Omega_{\boldsymbol{\alpha}}$ introduce $\tau(x) =: \xi$ for $x \in \Omega$ and $\xi \in \Omega_{\boldsymbol{\alpha}}$.

As stated before, cf. Eq. (2.10), the function Σ is defined via u, i.e., with the Green-Lagrange stress tensor (2.9) $E(u)$, $E : \Omega \to \mathbb{S}^d$ ($\mathbb{S}$ being the space of symmetric tensors of second order in $\mathbb{R}^d$) and the isotropic St. Venant-Kirchhoff tensor $\mathcal{C}$

$$\Sigma(u) = \mathcal{C} : E(u).$$

Though the convention for stating the Green-Lagrange stress tensor is as in (2.9), for the computation of the derivative $\tilde{j}'(\boldsymbol{\alpha})$ of the function in Eq. (4.5) the mathematically correct notation $E(u) = \frac{1}{2}(u_x + u_x^T + u_x^T u_x)$ is needed and used.

Assumption 4.1. *Let the functions u and $\bar{u}$ fulfill the condition $u \in \left(H^1(\Omega_{\boldsymbol{\alpha}})\right)^d$ and $\bar{u} \in \left(C^1(\Omega)\right)^d$. Furthermore let $\tau \in \left(C^1(\Omega \times A)\right)^d$ hold.*

Remark. *Usually one can only assume a lower regularity $\bar{u} \in \left(H^1(\Omega)\right)^d$. The higher regularity assumption fails especially if one has a corner singularity. This case needs to be dealt with different means and requires theory found for example in Grisvard [74]. The geometry in the present studies does not contain a corner singularity and therefore needs not be considered here, such that a thorough in depth analysis would exceed the scope of this thesis.*

Higher regularity for τ needs to hold such that Schwarz's theorem can be applied.

Now the target function can be defined on the reference domain Ω as well as on the deformed domain $\Omega_{\boldsymbol{\alpha}}$

$$\tilde{j}(0) = \int_\Omega \frac{1}{2}\left\|\Sigma - \bar{\Sigma}\right\|_F^2 \mathrm{d}x, \qquad \tilde{j}(\boldsymbol{\alpha}) = \int_{\Omega_{\boldsymbol{\alpha}}} \frac{1}{2}\left\|\Sigma(\boldsymbol{\alpha}) - \bar{\Sigma}\right\|_F^2 \mathrm{d}\xi. \tag{4.7}$$

For the optimization one is interested in the gradient of (4.5) with respect to the control value $\boldsymbol{\alpha}$. The computed functions are defined on a transformed domain $\Omega_{\boldsymbol{\alpha}}$. So first one needs to transform back to the reference domain Ω applying the transformation rule. The result is stated in Lem. 4.1

Lemma 4.1. *Let Ass. 4.1 hold. Applying the transformation rule to $\tilde{j}(\boldsymbol{\alpha})$, the integral over $\Omega_{\boldsymbol{\alpha}}$ can be converted into an integral over Ω. With $\Sigma(\boldsymbol{\alpha})$ as given in Eq. (2.10) and $u = u(\boldsymbol{\alpha})$ the transformation result is*

$$\int_{\Omega_{\boldsymbol{\alpha}}} \frac{1}{2}\left\|\Sigma(\boldsymbol{\alpha}) - \bar{\Sigma}\right\|_F^2 \mathrm{d}\xi = \int_\Omega \frac{1}{2}\Bigg\| \underbrace{\frac{1}{2}\mathcal{C} : \left(u_x \tau_x^{-1} + \tau_x^{-T} u_x^T + \tau_x^{-T} u_x^T u_x \tau_x^{-1}\right)}_{=:\Sigma(\tau,\boldsymbol{\alpha})} - \bar{\Sigma}\Bigg\|_F^2 \det \tau_x \,\mathrm{d}x, \tag{4.8}$$

with $u_\xi(\tau(x)) = u_x(x)\tau_x(x)^{-1}$.

Proof. The proof will be given by equation transformation where in the last step the transformation rule for integrals is applied. Also the above stated identity $\tau(x) = \xi$ for $x \in \Omega$ and $\xi \in \Omega_{\boldsymbol{\alpha}}$ is used.

$$\begin{aligned}
&\int_{\Omega_{\boldsymbol{\alpha}}} \frac{1}{2}\left\|\Sigma(\boldsymbol{\alpha}) - \bar{\Sigma}\right\|_F^2 \mathrm{d}\xi = \\
&\qquad \int_{\Omega_{\boldsymbol{\alpha}}} \frac{1}{2}\left\| \frac{1}{2}\mathcal{C} : \left(u_\xi(\xi) + u_\xi(\xi)^T + u_\xi(\xi)^T u_\xi(\xi)\right) - \bar{\Sigma}\right\|_F^2 \mathrm{d}\xi \\
&= \int_{\Omega_{\boldsymbol{\alpha}}} \frac{1}{2}\left\| \frac{1}{2}\mathcal{C} : \left(u_\xi\big(\tau(x)\big) + u_\xi\big(\tau(x)\big)^T + u_\xi\big(\tau(x)\big)^T u_\xi\big(\tau(x)\big)\right) - \bar{\Sigma}\right\|_F^2 \mathrm{d}\xi \\
&= \int_\Omega \frac{1}{2}\left\| \frac{1}{2}\mathcal{C} : \Big(u_x(x)\tau_x(x)^{-1} + \tau_x(x)^{-T} u_x^T(x) \right. \\
&\qquad\qquad \left. + \tau_x(x)^{-T} u_x^T(x)\big(u_x(x)\tau_x(x)^{-1}\big)\Big) - \bar{\Sigma}\right\|_F^2 \det \tau_x(x)\,\mathrm{d}x.
\end{aligned}$$

□

With this, one can now approach the derivative $\tilde{j}'(\boldsymbol{\alpha})$. Eventually one is interested in $\tilde{j}'(\boldsymbol{\alpha})$ with $\boldsymbol{\alpha} \to 0$. For the reason why $\tilde{j}'(0)$ as opposed to $\tilde{j}'(\boldsymbol{\alpha})$ is of interest, it needs to be clarified, that the transformation $\tau(\boldsymbol{\alpha})$ from Eq. (4.6) is directly applied to the grid. While $\boldsymbol{\alpha}$ is the design variable in Eq. (2.17), here in the computation of the derivative $\tilde{j}'$ the angle is fixed, say $\boldsymbol{\alpha}_0$, and one is interested in the function's derivative for the current implant position. In order to derive and use the correct quantities, an arbitrary $\boldsymbol{\alpha}$ is assumed to accurately apply the transformation rule changing from $\Omega_{\boldsymbol{\alpha}} \to \Omega$, but then the transition $\boldsymbol{\alpha} \to 0$ has to be performed to compute the derivative of $\tilde{j}$ in the current position $\boldsymbol{\alpha}_0$. Hence, $\tilde{j}(0)$ here refers to the objective value on the current grid configuration.

The derivation of the derivative will be given in two steps. For $\tilde{j}'(0)$ one needs $u_{\boldsymbol{\alpha}}(0)$. The computation of which will be derived the first step. In the second step the derivative of $\tilde{j}$ with respect to $\boldsymbol{\alpha}$ is computed.

Its computation of $u_{\boldsymbol{\alpha}}(0)$ will be stated in the following Lemma:

Lemma 4.2. *Let Ass. 4.1 hold. Define the functional*

$$W(v, \boldsymbol{\alpha}) \coloneqq \int_\Omega w\big(v_x(x)\tau_x^{-1}(x), x\big) \det\big(\tau_x(x)\big) \mathrm{d}x.$$

Then $u_{\boldsymbol{\alpha}}(0)$ solves the linear system of equations

$$W_{v,v}\big(u(0), 0\big) u_{\boldsymbol{\alpha}}(0) + W_{v,\boldsymbol{\alpha}}\big(u(0), 0\big) = 0. \tag{4.9}$$

Proof. Starting in the transformed domain $\Omega_{\boldsymbol{\alpha}}$, the displacement $\tilde{u}$ due to boundary forces is computed solving a variational problem. Specifically, $\tilde{u}$ solves

$$\tilde{u} = \operatorname{argmin}_{\tilde{v} \in H^1(\Omega_{\boldsymbol{\alpha}})} \int_{\Omega_{\boldsymbol{\alpha}}} w\big(\tilde{v}_\xi(\xi), \tau^{-1}(\xi)\big) \mathrm{d}\xi =: \tilde{W}(\tilde{v}).$$

Transferring this to Ω using Lem. 4.1 yields

$$\begin{aligned} u(\boldsymbol{\alpha}) &= \operatorname{argmin}_{v \in H^1(\Omega)} W(v, \boldsymbol{\alpha}), \\ W(v, \boldsymbol{\alpha}) &= \int_\Omega w\big(v_x(x)\tau_x^{-1}(x), x\big) \det\big(\tau_x(x)\big) \mathrm{d}x, \end{aligned}$$

with $W(v, \boldsymbol{\alpha})$ as defined in the Lemma. This means, that $u(\boldsymbol{\alpha})$ solves $W_v\big(u(\boldsymbol{\alpha}), \boldsymbol{\alpha}\big) = 0$. From this one can infer that $u_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$ solves

$$\frac{\partial}{\partial \boldsymbol{\alpha}} W_v\big(u(\boldsymbol{\alpha}), \boldsymbol{\alpha}\big) = W_{v,v}\big(u(\boldsymbol{\alpha}), \boldsymbol{\alpha}\big) u_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}) + W_{v,\boldsymbol{\alpha}}\big(u(\boldsymbol{\alpha}), \boldsymbol{\alpha}\big) = 0.$$

For the second derivative of W with respect to v one derives

$$W_{v,v}\big(u(\boldsymbol{\alpha}), \boldsymbol{\alpha}\big)[\phi, \psi] = \int_\Omega w_{v_x, v_x}\big(u_x(\boldsymbol{\alpha})\tau_x^{-1}, \cdot\big)\big[\phi_x\tau_x^{-1}, \psi_x\tau_x^{-1}\big] \det \tau_x \mathrm{d}x,$$

while the derivative $W_{v,\boldsymbol{\alpha}}$ computes to

$$\begin{aligned} W_{v,\boldsymbol{\alpha}}\big(u(\boldsymbol{\alpha}), \boldsymbol{\alpha}\big)\phi = \int_\Omega \Big(& w_{v_x, v_x}\big(u_x(\boldsymbol{\alpha})\tau_x^{-1}, \cdot\big)\big[\phi_x\tau_x^{-1}, u_x(\boldsymbol{\alpha}) \tfrac{\partial}{\partial \boldsymbol{\alpha}}(\tau_x^{-1})\big] \\ & + w_{v_x}\big(u_x(\boldsymbol{\alpha})\tau_x^{-1}, \cdot\big)\big[\phi_x \tfrac{\partial}{\partial \boldsymbol{\alpha}}(\tau_x^{-1})\big] \Big) \det \tau_x + w_{v_x}\big(u_x(\boldsymbol{\alpha})\tau_x^{-1}, \cdot\big)\big[\phi_x\tau_x^{-1}\big] \operatorname{tr}\big(\tau_x^{-1}\bar{u}_x\big) \det \tau_x \, \mathrm{d}x. \end{aligned}$$

The equality $\frac{\partial}{\partial \boldsymbol{\alpha}} \tau_x^{-1} = -\tau_x \bar{u}_x \tau_x^{-1}$ holds. Using the assumptions for τ, the order of differentiation $(\tau_{\boldsymbol{\alpha}})_x = (\tau_x)_{\boldsymbol{\alpha}}$ is exchangeable using Schwarz's theorem.

Now one derives $u_{\boldsymbol{\alpha}}(0)$ from the above equations by setting $\boldsymbol{\alpha} = 0$ and $\tau = \mathrm{id}$. This yields

$$W_{v,v}\big(u(0),0\big)u_{\boldsymbol{\alpha}}(0) + W_{v,\boldsymbol{\alpha}}\big(u(0),0\big) = 0$$

with

$$W_{v,v}\big(u(0),0\big)[\phi,\psi] = \int_\Omega w_{v_x,v_x}\big(u_x(0),\cdot\big)\big[\phi_x,\psi_x\big]\,\mathrm{d}x,$$

and

$$\begin{aligned} W_{v,\boldsymbol{\alpha}}\big(u(0),0\big)\phi = \int_\Omega -\Big(& w_{v_x,v_x}\big(u_x(0),\cdot\big)\big[\phi_x, u_x(0)\bar{u}_x\big] \\ &+ w_{v_x}\big(u_x(0),\cdot\big)\big[\phi_x\bar{u}_x\big]\Big) + w_{v_x}\big(u_x(0),\cdot\big)[\phi_x]\mathrm{div}(\bar{u})\,\mathrm{d}x. \end{aligned}$$

This completes the proof. □

Lemma and proof have left out the contribution of the boundary parts. Its derivation works in similar and easier way. The resulting addition is the following:

To $W_{v,v}\big(u(0)),0\big)$ one has to add

$$W_{v,v}\big(u(0),0\big)_\Gamma[\phi,\psi] = \int_{\Gamma_D} \gamma[\phi,\psi]\,\mathrm{d}s,$$

where $\gamma \in \mathbb{R}$ is the penalty parameter with which the Dirichlet boundary is enforced. To $W_{v,\boldsymbol{\alpha}}\big(u(0),0\big)$ one needs to add

$$W_{v,\boldsymbol{\alpha}}\big(u(0),0\big)_\Gamma\phi = \int_{\Gamma_D} \gamma u(0)\phi\big(t^T\bar{u}_x t\big)\,\mathrm{d}s - \int_{\Gamma_N} g_N\phi\big(t^T\bar{u}_x t\big)\,\mathrm{d}s,$$

with t being the tangent to the boundary. The missing contact boundary contribution on the other hand is more intricate and will be included using finite differences as opposed to the inclusion by analytical means. Its inclusion into the derivative $\tilde{j}'$ will be explained in the Sec. 5.

With this at hand, one can now turn to the actual derivative $\tilde{j}'(0)$.

Lemma 4.3. *Let Ass. 4.1 hold and let $u_{\boldsymbol{\alpha}}$ as in Lem. 4.2 be given. Then the gradient $\tilde{j}'(0)$ of the target function (4.5) is given by*

$$\tilde{j}'(0) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\alpha}}\frac{1}{2}\int_{\Omega_{\boldsymbol{\alpha}}} \big\|\Sigma - \bar{\Sigma}\big\|_F^2\,\mathrm{d}\xi = \int_\Omega \frac{1}{2}\big\|\Sigma - \bar{\Sigma}\big\|_F^2\mathrm{div}(\bar{u}) + \big\langle \Sigma - \bar{\Sigma}, \Sigma_{\boldsymbol{\alpha}}(\mathrm{id},0)\big\rangle_F\,\mathrm{d}x, \quad (4.10)$$

with

$$\begin{aligned} \Sigma_{\boldsymbol{\alpha}}(\mathrm{id},0) = \frac{1}{2}\mathcal{C} : \Big[& u_{x,\boldsymbol{\alpha}} - u_x\bar{u}_x + \big(u_{x,\boldsymbol{\alpha}} - u_x\bar{u}_x\big)^T \\ &+ u_{x,\boldsymbol{\alpha}}^T u_x - \bar{u}_x^T u_x^T u_x + \big(u_{x,\boldsymbol{\alpha}}^T u_x - \bar{u}_x^T u_x^T u_x\big)^T\Big]. \end{aligned} \quad (4.11)$$

Proof. In computing the derivative of $\tilde{j}$ one uses $\frac{\partial}{\partial \boldsymbol{\alpha}}\tau = \bar{u}$ with τ as given in Eq. (4.6). Note as before, that the order of differentiation $(\tau_{\boldsymbol{\alpha}})_x = (\tau_x)_{\boldsymbol{\alpha}}$ is exchangeable using Schwarz's theorem. This gives

$$\begin{aligned}\tilde{j}'(\boldsymbol{\alpha}) =& \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\alpha}} \frac{1}{2} \int_\Omega \left\|\Sigma(\tau, \boldsymbol{\alpha}) - \bar{\Sigma}\right\|_F^2 \det \tau_x \,\mathrm{d}x \\ =& \int_\Omega \frac{1}{2} \left\|\Sigma(\tau, \boldsymbol{\alpha}) - \bar{\Sigma}\right\|_F^2 \mathrm{tr}\left(\tau_x^{-1} \bar{u}_x\right) \det \tau_x + \left\langle \Sigma(\tau, \boldsymbol{\alpha}) - \bar{\Sigma}, \Sigma_{\boldsymbol{\alpha}}(\tau, \boldsymbol{\alpha})\right\rangle_F \det \tau_x \,\mathrm{d}x.\end{aligned} \tag{4.12}$$

The term $\Sigma_{\boldsymbol{\alpha}}(\tau, \boldsymbol{\alpha})$ equals to

$$\begin{aligned}\Sigma_{\boldsymbol{\alpha}}(\tau, \boldsymbol{\alpha}) = \frac{1}{2}\mathcal{C} : \Big(& (u_x)_{\boldsymbol{\alpha}} \tau_x^{-1} - u_x \tau_x \bar{u}_x \tau_x^{-1} + \left((u_x)_{\boldsymbol{\alpha}} \tau_x^{-1} - u_x \tau_x \bar{u}_x \tau_x^{-1}\right)^T \\ & + \tau_x^{-T} (u_x^T)_{\boldsymbol{\alpha}} u_x \tau_x^{-1} - \tau_x^{-T} \bar{u}_x^T \tau_x^T u_x^T u_x \tau_x^{-1} \\ & + \left(\tau_x^{-T} (u_x^T)_{\boldsymbol{\alpha}} u_x \tau_x^{-1}\right)^T - \tau_x^{-T} \bar{u}_x^T \tau_x^T u_x^T u_x \tau_x^{-1} \Big).\end{aligned} \tag{4.13}$$

The equality $\frac{\partial}{\partial \boldsymbol{\alpha}} \tau_x^{-1} = -\tau_x \bar{u}_x \tau_x^{-1}$ holds.

When computing the derivative $\tilde{j}'$ for $\boldsymbol{\alpha} = 0$, the transformation τ yields the identity. This simplifies the equations (4.12) and (4.13) admitting $\tilde{j}'(0)$ and $\Sigma_{\boldsymbol{\alpha}}(\mathrm{id}, 0)$ as stated in the Lemma. □

Now returning to the full target function (2.17) including the integral over the load domain $\mathcal{L}$ and the penalty function p, one gets for j'

$$j'(0) = \int_{l \in \mathcal{L}} \pi(l) \int_{\Omega_{\mathrm{bone}}(\boldsymbol{\alpha})} \frac{1}{2} \left\|\Sigma - \bar{\Sigma}\right\|_F^2 \mathrm{div}(\bar{u}) + \left\langle \Sigma - \bar{\Sigma}, \Sigma_{\boldsymbol{\alpha}}(\mathrm{id}, 0)\right\rangle_F \mathrm{d}x + \frac{\partial}{\partial \boldsymbol{\alpha}} p\big(\Sigma_{\max}(l; 0)\big) \,\mathrm{d}l. \tag{4.14}$$

4.4 Optimization algorithm

Bringing all of the above together, the optimization algorithm is given as a whole. The algorithm needs stopping criteria. With $\varepsilon_j > 0$ and $\varepsilon_{\text{deriv}} > 0$ one controls, if objective value or the value of the derivative is sufficiently small. With $\varepsilon_{\Delta j} > 0$ and $\varepsilon_{\Delta \boldsymbol{\alpha}} > 0$ one examines, whether the change of the objective value or the change in the control variable is small enough. Define $\delta \in (0,1)$ for stepsize reduction. For a given initial implant position $\boldsymbol{\alpha}_0$, initial symmetric, positive definite matrix B_0, $\tau_1 \in (0,1)$, and $\tau_2 \in (\tau_1, 1)$ the optimization is as in Alg. 3.

Algorithm 3 Optimize Implant Position

procedure OPTIMIZEPOSITION($\boldsymbol{\alpha}_0$,B_0,$\varepsilon_{\text{DERIV}}$,$\varepsilon_j$, $\varepsilon_{\Delta j}$, $\varepsilon_{\Delta \boldsymbol{\alpha}}$, δ, τ_1, τ_2)
  Set $k := 0$, $l := 0$
  Compute $j(\boldsymbol{\alpha}_k)$
  if $j(\boldsymbol{\alpha}_k) < \varepsilon_j$ **then return** $\boldsymbol{\alpha}_k$
  Compute $j'(\boldsymbol{\alpha}_k)$
  if $\|j'(\boldsymbol{\alpha}_k)\|_2 < \varepsilon_{\text{deriv}}$ **then return** $\boldsymbol{\alpha}_k$
  $d_k = -B_k^{-1} j'(\boldsymbol{\alpha}_k)^T$
  $s_k = \delta^l$
  $\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + s_k d_k$
  Compute $j(\boldsymbol{\alpha}_{k+1})$ and $j'(\boldsymbol{\alpha}_{k+1})$
  if $j(\boldsymbol{\alpha}_k) < \varepsilon_j$ or $\|j'(\boldsymbol{\alpha}_k)\|_2 < \varepsilon_{\text{deriv}}$ **then return** $\boldsymbol{\alpha}_{k+1}$
  Test $j'(\boldsymbol{\alpha}_{k+1}) d_k \geq \tau_2 j'(\boldsymbol{\alpha}_k) d_k$ and
    $j(\boldsymbol{\alpha}_{k+1}) \leq j(\boldsymbol{\alpha}_k) + \tau_1 s_k j'(\boldsymbol{\alpha}_k) d_k$
  if grid connectivity has changed (due to remeshing) **then**
    replace Armijo rule by: $\langle -j'(\boldsymbol{\alpha}_k), j'(\boldsymbol{\alpha}_{k+1}) \rangle_2 < 0$
  if test results all *true* **then**
    if $\|j(\boldsymbol{\alpha}_{k+1}) - j(\boldsymbol{\alpha}_k)\|_2 < \varepsilon_{\Delta j}$ **then return** $\boldsymbol{\alpha}_{k+1}$
    if $\|\boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k\|_2 < \varepsilon_{\Delta \boldsymbol{\alpha}}$ **then return** $\boldsymbol{\alpha}_{k+1}$
    Set $t := \boldsymbol{\alpha}_{k+1} - \boldsymbol{\alpha}_k$ and $y^T := j'(\boldsymbol{\alpha}_{k+1}) - j'(\boldsymbol{\alpha}_k)$
    Update $B_{k+1} := B_k - \frac{(B_k t)(B_k t)^T}{t^T B_k t} + \frac{y y^T}{y^T t}$
    $k = k + 1$, $l = 0$, **goto** 3
  else
    $l = l + 1$, **goto** 8

5 Numerical Evaluation

Before the numerical performance of the interplay of algorithms 1, 2, and 3 is presented, first some more information to the setup will be given. After that the relative computing times of solving the PDE (2.18) for the initial Kriging samples in Tab. 4, setting up Kriging interpolation, and Monte Carlo integration will be compared in Sec. 5.2. Finally, in Sections 5.3–5.5 the Kriging method, Monte Carlo method and the optimization are evaluated.

5.1 The setup

The start of the numerical section is marked by clarifying the origin of some data and algorithms in the background.

The anonymized *bone geometry data*, the femur and pelvis grid, were provided by the work group "Computational Diagnosis and Therapy Planning" from Zachow at the Zuse Institute Berlin. More of their work on computational geometry and virtual anatomy can be found in Zachow et al. [179].

The testing of algorithms started for computing time reasons with a *2D setup* of a hip joint with and without an implant. See Fig. 9 and Fig. 10 for the respective (coarse) FE-grids. The *(re-)meshing* of the grids is performed by *Triangle*[‖] from Shewchuk [142]. With the help of this software, a conforming Delaunay mesh generation with no angles smaller than 20 degrees is realized.

In the 2D setup the *number of control variables* $\boldsymbol{\alpha}$ is reduced from three entries to a scalar α. For this design variable the admissible space of positions A is defined as follows: From a centered setup of the femur implant inside the bone an angle interval was estimated in such a way, that the implant stem is allowed to touch the cortical bone and must not go further. The resulting interval is $A \coloneqq [-3.5°, 1.5°]$, cf. Fig. 11.

The defined *boundaries* for the bone and the implant geometry are illustrated in Fig. 12. The contact boundary is depicted in red and the Neumann boundary in green. All non-colored parts belong to the Dirichlet boundary. While on the pelvis the Dirichlet boundary is strongly enforced, thus rendering the pelvis a fixed obstacle, it is only weakly enforced on the femur.

The setup and implementation of a roughly *realistic bone-implant* environment, material parameters for (healthy) bone, a common implant alloy in Tab. 1 in Sec. 18, and yield stress as given below have been compiled from [18, 76, 112, 120, 124]. More on hip joint implants, different fixation methods, long, middle, and short stem implants, a variety of alloys, their mechanical properties, and different casting techniques can be found in [18, 47, 117].

Getting an average estimate of the *frequency of daily motions* $w(m)$ as introduced in Eq. (2.21), the report [110] was utilized. See Fig. 13 for a visualization of the complete load probability map $\pi(l)$ on the discretized load domain $F_x \times F_z$ (the discretization will be specified in the next paragraph). For an improved visualization, the logarithm of the load probability $\log\big(\pi(l)\big)$ is depicted, where values of $\pi(l) = 0$ were changed to 10^{-18} before applying the logarithm. In the bottom image the red crosses mark the initial load samples from Tab. 4.

The *solution of the PDE* (2.18) was implemented in *Kaskade7*[**] applying a SQP

[‖] https://www.cs.cmu.edu/~quake/triangle.html

[**] https://www.zib.de/projects/kaskade7-finite-element-toolbox

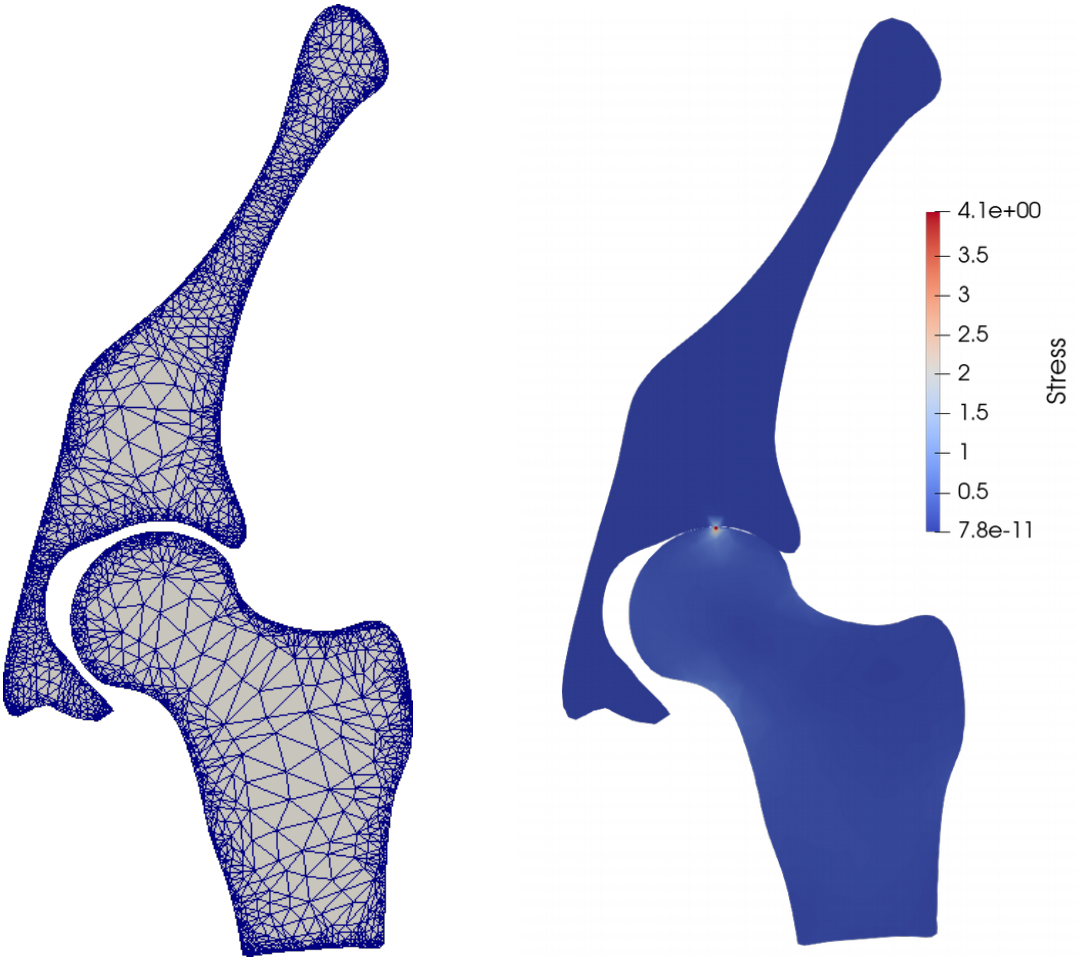

Fig. 9: *Bone grid (left,* 20,040 *DOF) without implant for the computation of the stress distribution* $\bar{\Sigma}$ *(right).*

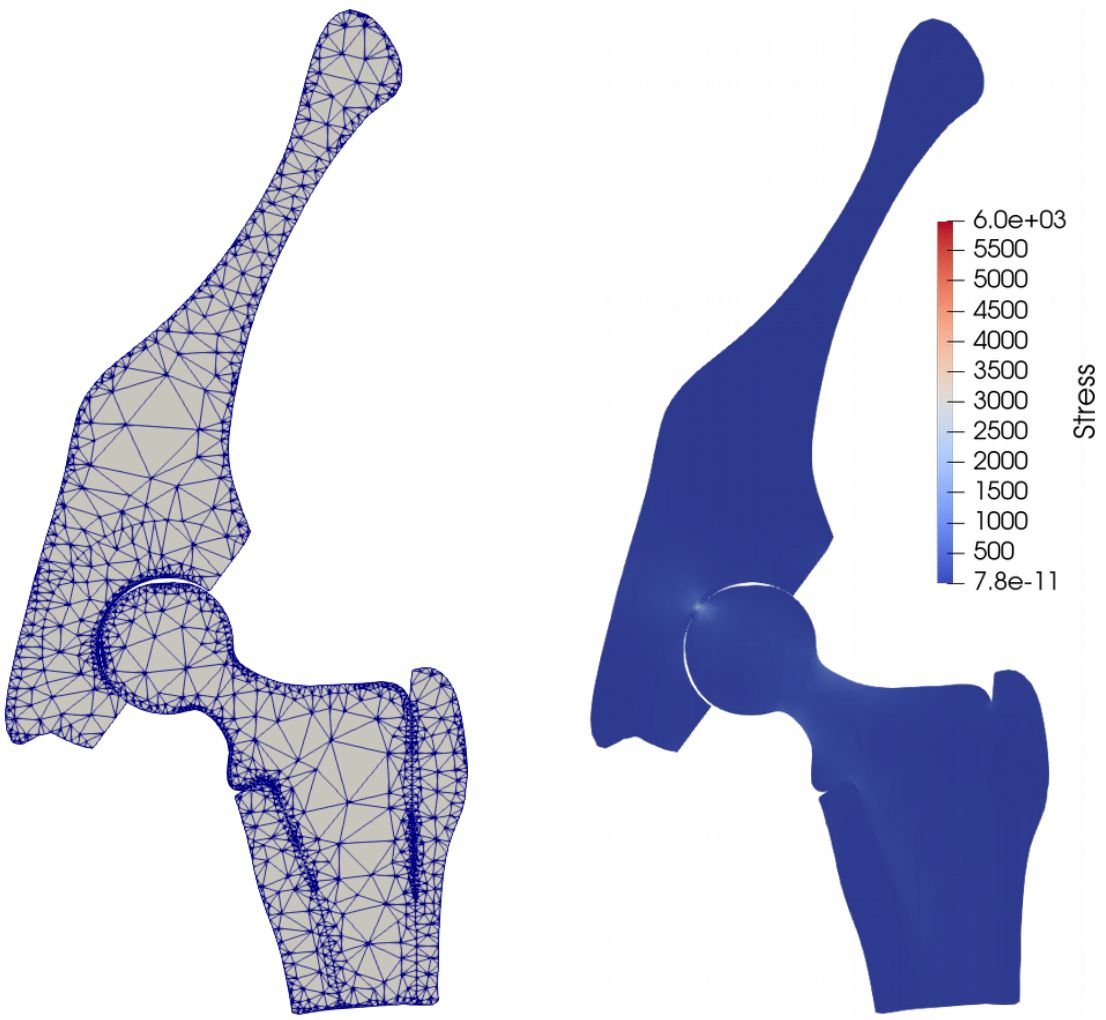

Fig. 10: *Bone-implant grid (left,* 2197 *DOF) for the computation of the stress distribution* Σ *(right).*

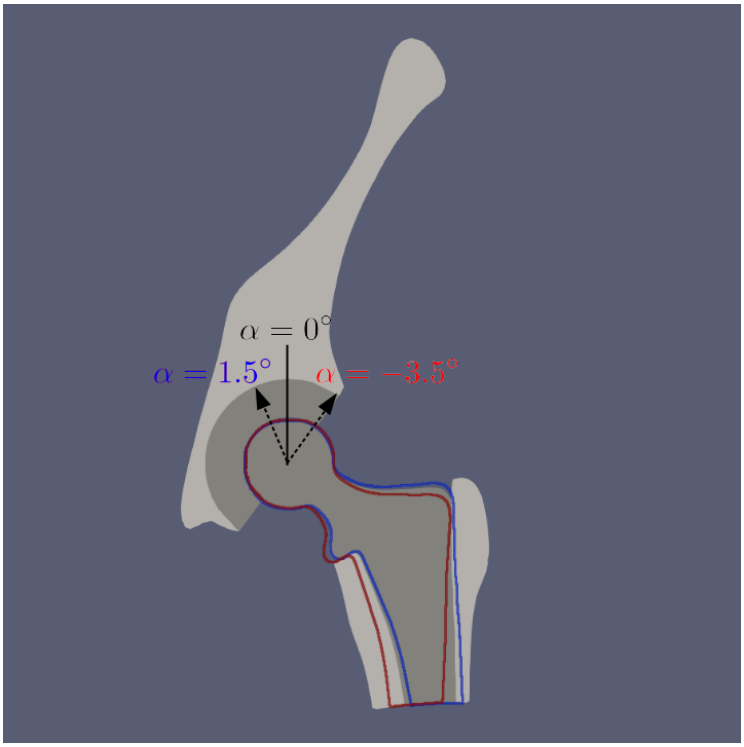

Fig. 11: *The 2D hip joint with a centered femur implant position* $\alpha = 0°$ *and the possible changes within* $[-3.5°, 1.5°]$.

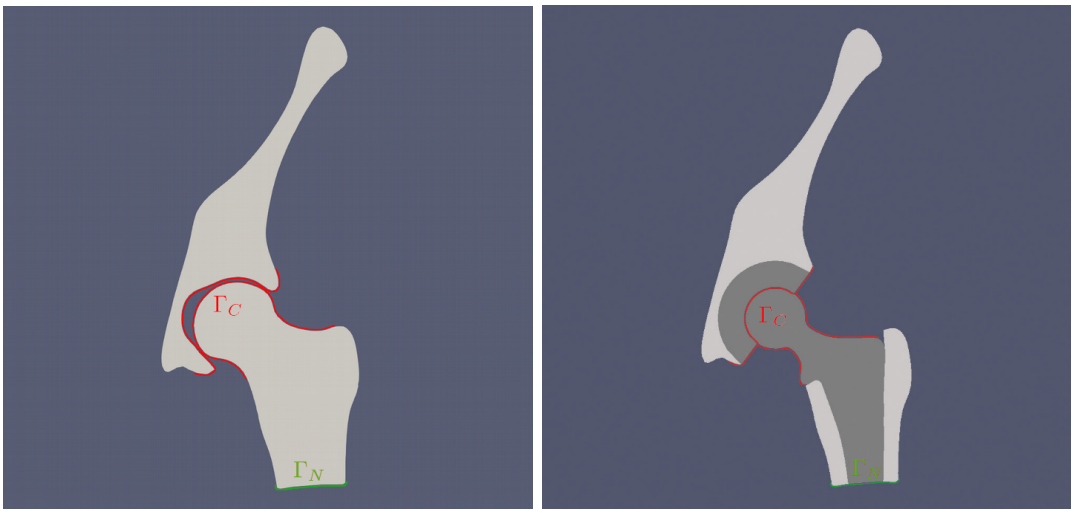

Fig. 12: *The defined boundaries are the contact boundary (red), the Neumann boundary (green at the bottom), and everything else is the Dirichlet boundary.*

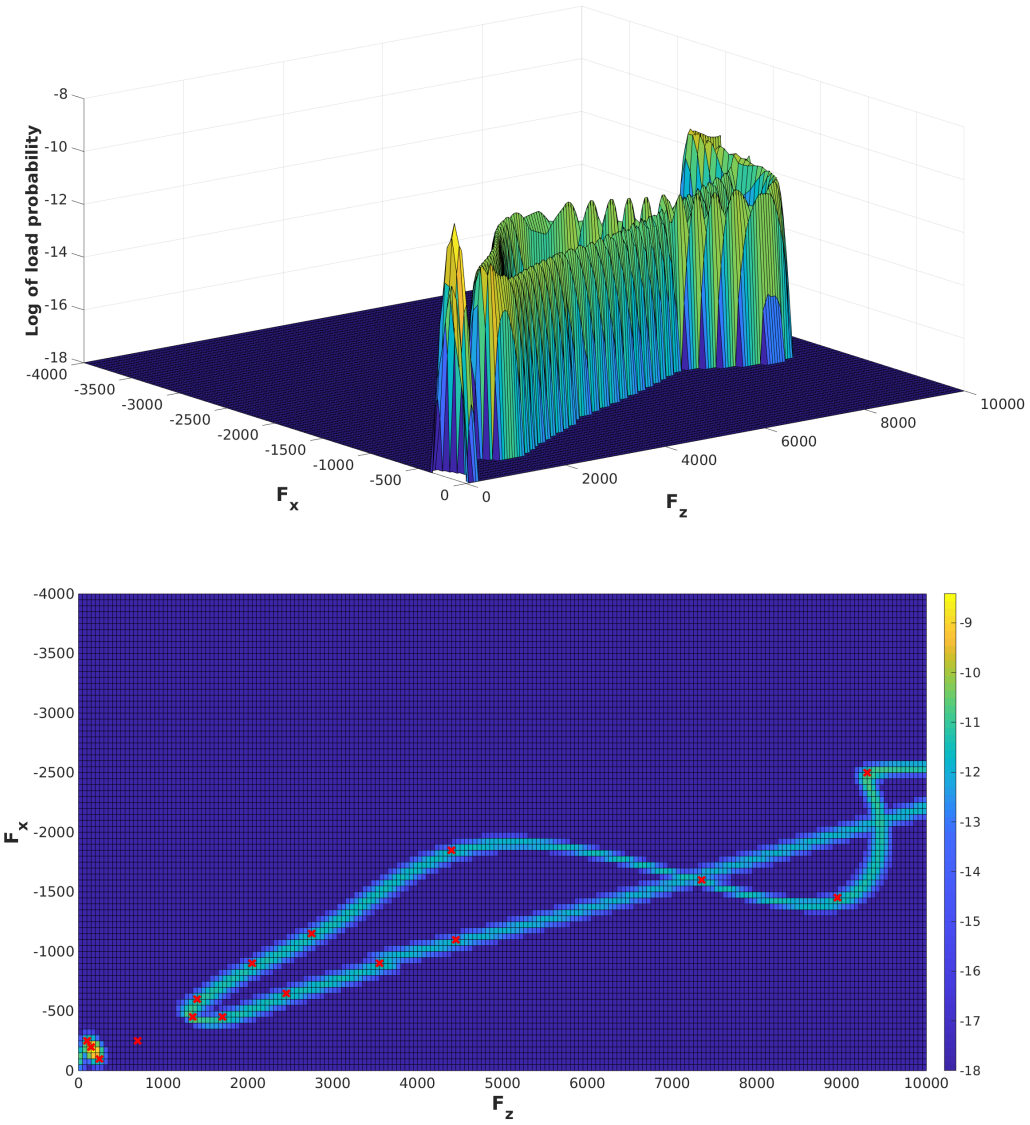

Fig. 13: *The logarithm of the load probability map* $log\,(\pi(l))$ *computed on* $\mathcal{L}_{50}$ *using information from report [110] for an estimate on* $w(m)$ *with marked initial load samples (red crosses) as in Tab. 4; if* $\pi(l) = 0$ *then this value was changed to* 10^{-18} *before applying the logarithm.*

method as explained in Sec. 3.1. Kaskade7 is a finite element toolbox for the solution of stationary or transient systems of partial differential equations [66, 68] that also allows for the inclusion of contact constraints.

Stress responses on the load domain $\mathcal{L}$. As stated in Sec. 1, the *load domain* $\mathcal{L}$ is set up using information from the database OrthoLoad by Bergmann and Duda, cf. [13, 14, 39]. The standard loads on the hip joint are mostly taken from daily activities, e.g., walking, sitting down, or going up the stairs, and measurements of up to ten different people per activity were recorded. The activities were provided for both, the average load case (standardized to a 75 kg person) and the high load case (standardized to a 100 kg person). This defines a domain of daily loads (recall Eq. (1.1))

$$\begin{aligned}\mathcal{L} &= F_x \times F_y \times F_z \times M_x \times M_y \times M_z \\ &= [-4000, 0] \times [-200, 3000] \times [0, 10{,}000] \times [-4, 5] \times [-2, 3] \times [-2, 2]\end{aligned}$$

(in N for forces and Nm for moments) with the coordinate system as given in Fig. 3. In 2D the load domain reduces to

$$\mathcal{L}_{2\mathrm{D}} \coloneqq F_x \times F_z \times M_y = [-4000, 0] \times [0, 10{,}000] \times [-2, 3]$$

which will in the following be referred to as $\mathcal{L}$.

The load domain $\mathcal{L}$ was discretized with a mesh width of 50 resulting in 81×201 grid nodes (and one fixed moment $m_y \in M_y$), on which the stress difference, maximum stresses, penalty function values, and the load probability will be evaluated. Computations on this grid will be indicated as computations on $\mathcal{L}_{50}$. The implant will be fixed in position $0°$.

In Fig. 14, one can see the result of the integrated stresses for the remaining bone, with $\Omega_{\mathrm{bone}}(0°)$ as defined in Eq. (2.1), in the bone-implant setup computed on $\mathcal{L}_{50}$ while Fig. 15 depicts the integrated stress of the healthy bone setup but only incorporating those bone parts that remain after implantation. In Fig. 16 on the other hand, one sees the result of the integrated stress differences $\int_{\Omega_{\mathrm{bone}}(0°)} \frac{1}{2}\|\Sigma - \bar{\Sigma}\|_F^2 \,\mathrm{d}x$ again on $\mathcal{L}_{50}$. The result of maximum stresses $\Sigma_{\max}$ on the discretized grid $\mathcal{L}_{50}$ can be found in Fig. 17.

Unfortunately, one sees highly noisy data in Fig. 16 which prohibits the use of Kriging interpolation as argued in Sec. 3.3. The noisy data seems to originate from subtle bugs in the implementations that could not be fixed in due time. Recall from Sec. 3.3.4 the two functions $z_{\Delta\Sigma}(l;\alpha)$ and $z_{\max}(l;\alpha)$ to be interpolated. In light of the noisy data, they will be substituted from Sec. 5.3 on. In the succeeding Sec. 5.2, where the computing times of the subalgorithms will be evaluated, we will solve the PDE (2.18) and will set up Kriging functions using arbitrary kernel and drift models which will thereafter be implemented in the Monte Carlo integration. Only afterwards, starting from Sec. 5.3, one is interested in the correct interpolation, interpolation error, and decrease thereof when adaptively adding samples, as opposed to abstract computing times. The target functions $z_{\Delta\Sigma}$ and $z_{\max}$ will then be substituted by

$$\begin{aligned}\tilde{z}_{\Delta\Sigma}(l;\alpha) &\coloneqq (-f_x + f_z)\big(\sin(-f_x) + \sin(f_z) + 2\big)\big[(\alpha+2)(\alpha-1) + 2.25\big], \\ \tilde{z}_{\max}(l;\alpha) &\coloneqq 350 f_z\big[(\alpha+1.5)(\alpha+1.2) + 0.15^2\big].\end{aligned} \tag{5.1}$$

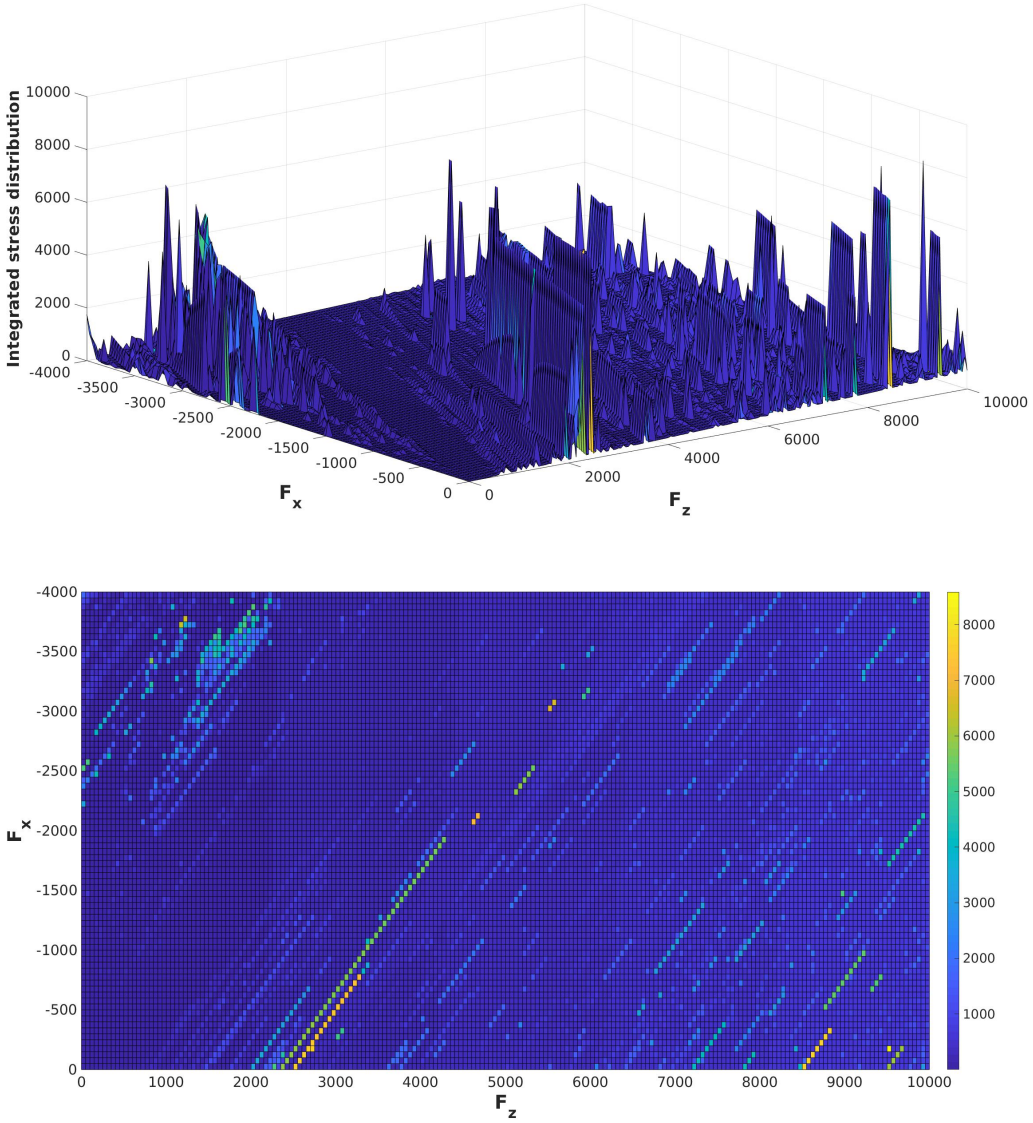

Fig. 14: *The integrated stress response* $\int_{\Omega_{bone}(0^\circ)} \Sigma(l)\,\mathrm{d}x$ *in the remaining bone of the bone-implant setup computed on* $\mathcal{L}_{50}$.

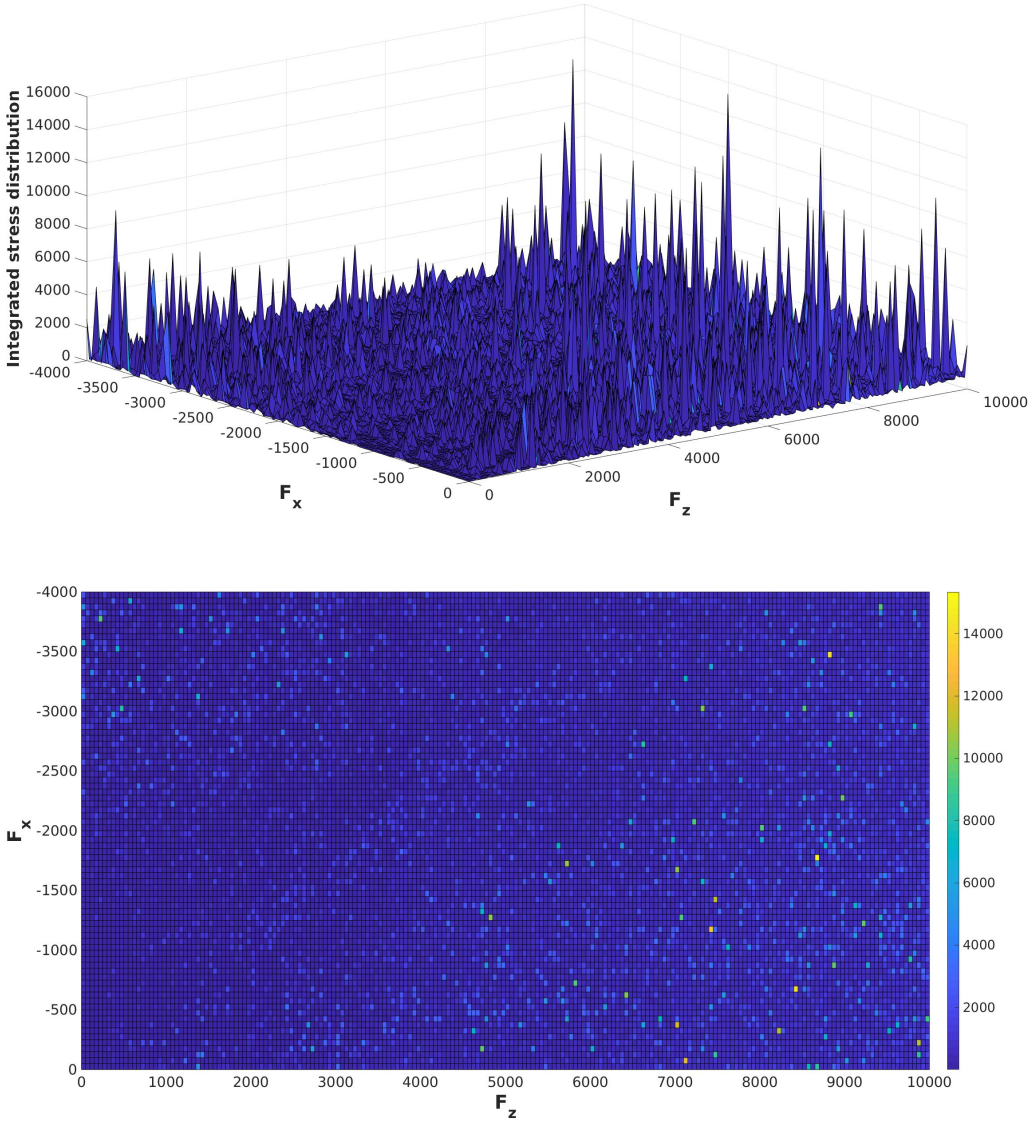

Fig. 15: *The integrated stress response* $\int_{\Omega_{bone}(0^\circ)} \bar{\Sigma}(l)\,\mathrm{d}x$ *in the healthy bone in what is to be the remaining bone of the bone-implant setup computed on* $\mathcal{L}_{50}$.

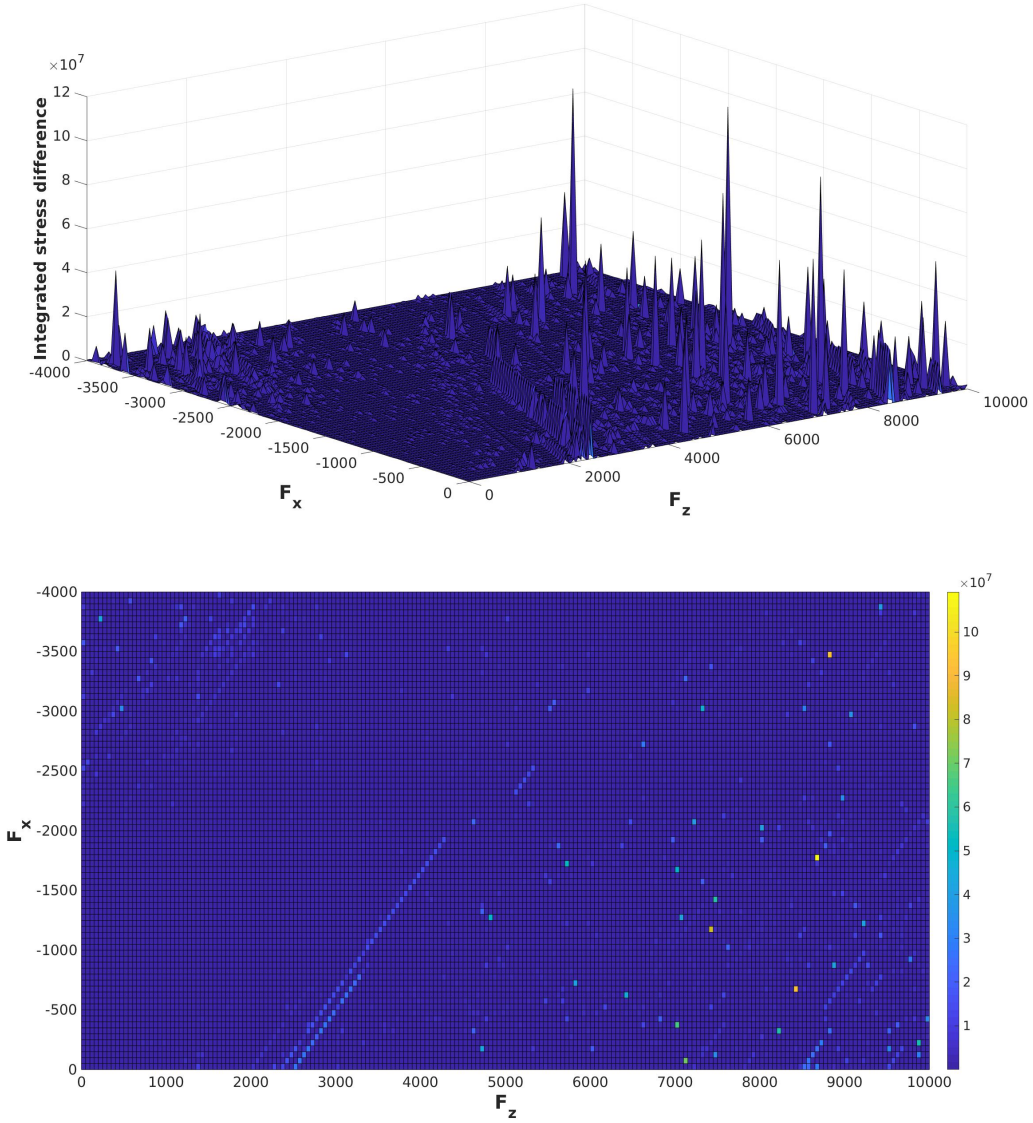

Fig. 16: *The integrated stress difference* $\int_{\Omega_{bone}(0^\circ)} \frac{1}{2}\|\Sigma(l) - \bar{\Sigma}(l)\|_F^2 \,\mathrm{d}x$ *in the remaining bone of the bone-implant setup computed on* $\mathcal{L}_{50}$.

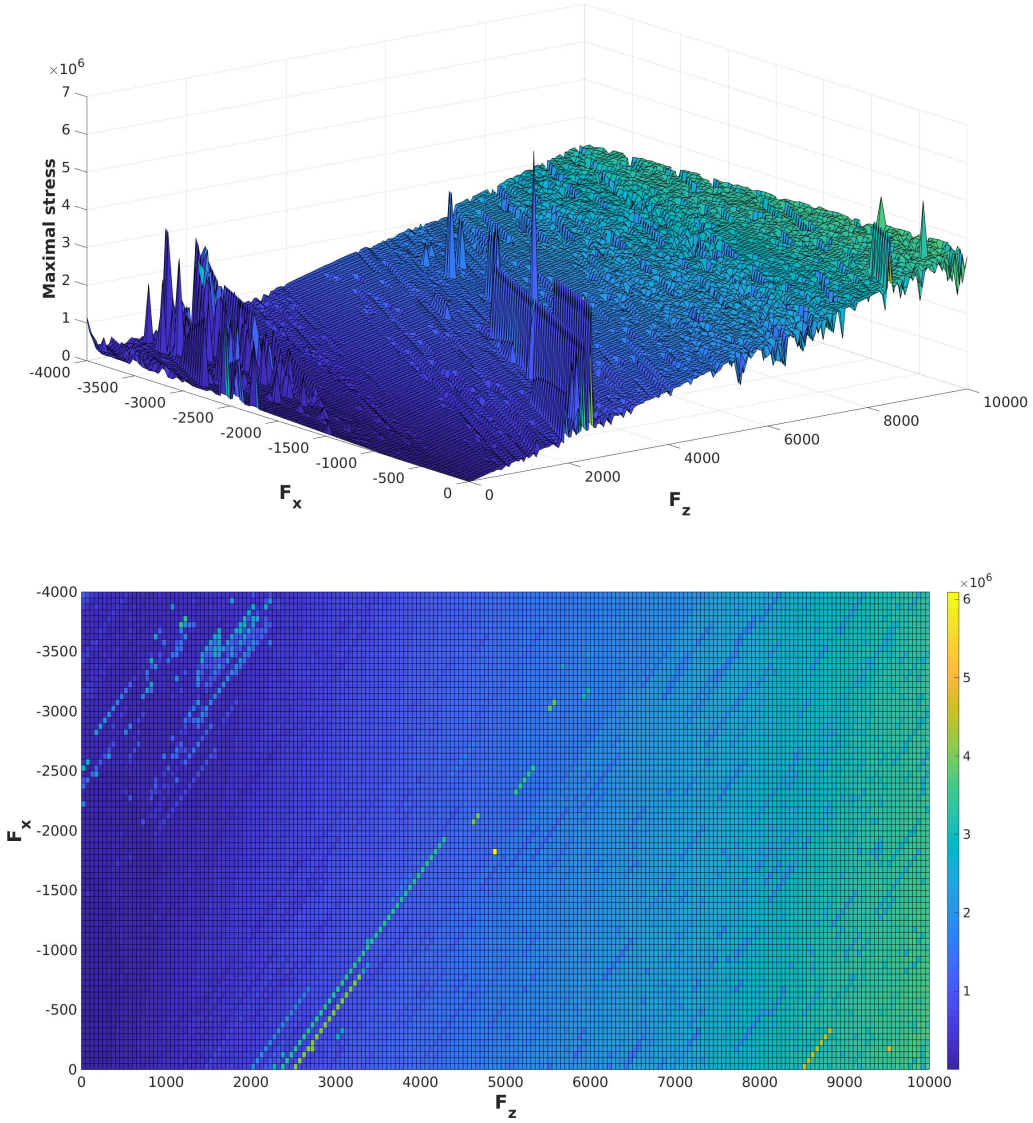

Fig. 17: *The maximum stress values Σ_{max} in the remaining bone of the bone-implant setup computed on $\mathcal{L}_{50}$.*

Fixing the parameters of the penalty function. The *yield stress* for cortical bone is approximately 140 MPa in uniaxial, i.e., in F_z direction and 55 MPa in transverse to uniaxial direction (on compression) [18, 56, 120, 131, 155, 170] which roughly equates to total loads of $f_z = 280$ kN and $f_x = 110$ kN, respectively, above which the stability of bone can not be guaranteed any more.

In a realistic model, one might set Σ_{thres} from Sec. 2.1.5 to accommodate these yield stress values. Combining the surgeon's experience with an analysis of the bone structure, which may be weakened due to osteoporosis, the yield stress may even be lower than the average value in a healthy bone. Here, the set up model is a demonstrative one, to prove the function of the proposed optimization with its subalgorithms. For this purpose, the quantity Σ_{thres} will be set to a value relevant for the present evaluation and the parameters in the penalty function will be adjusted accordingly. This will be done utilizing the peak load during stumbling – a bad case scenario included in the OrthoLoad database – and an average high load sample. Also, the interest lies in the highest stress in the remaining bone $\Omega_{\text{bone}}(\alpha)$ as opposed to the overall highest occurring stress. After the surgery, the highest appearing stress would be in the implant, since the implant parts are in contact with each other and in the area of contact, the stress value is maximized (cf. Fig. 20). But as the implant is much stiffer than the (possibly diseased) bone, penalizing high stresses in the remaining bone is of greater interest.

Defining the peak load from the recorded stumbling

$$l_{\text{peak}} \coloneqq (f_x, f_z) = (2462.7, -10{,}608.5) \tag{5.2}$$

(no moments were recorded) and an average high load $\tilde{l} \coloneqq (1010, -3451)$, the computed maximum stress in the remaining bone $\Omega_{\text{bone}}(\alpha)$ of the bone-implant setup for the peak load (5.2) resulted in 3.9 MPa and 1.3 MPa for an average high load case $\tilde{l}$.

Thus, the parameter Σ_{thres} is set to 2.9 MPa with a bias towards the non-harmful loads and a sharp increase of the penalty function thereafter. With this value, the algorithm penalizes loads resulting in high stress responses in the bone well before the malign case of peak stumbling stress is reached. Setting the parameters $b = 10^{-15}$ and $c = 10^{-4}$ finalizes the penalty function (2.14).

The logarithm of penalty function values, i.e., $\log\Big(p\big(\Sigma_{\max}(l; 0°)\big)\Big)$, over $\mathcal{L}$ with femur implant in position $\alpha = 0°$ are depicted in Fig. 18. Values of 0 were changed to 1 before applying the logarithm.

Now keeping a fixed load $l_0 \coloneqq (-200, 9000)$, the logarithm of the penalty function for different implant positions $\boldsymbol{\alpha}$ is shown in Fig. 19, again setting 0 values to 1 before applying the logarithm. For this, the interval of admissible implant position angles A was discretized by $\Delta\alpha = (10^{-4})°$. When the change in position of the implant adds up 0.1° the bone-implant mesh was reset, leading to a different FE-discretization and subsequently to a discontinuity in the objective. Below the value 0.1°, fixed body transformation was applied. Thus, discontinuities are expected to occur every 0.1°. The depicted discontinuities within the 0.1° intervals may originate from a coarse FE-grid and a fluctuation between elements in which the maximum stress $\Sigma_{\max}$ is attained.

The Figures 18 and 19 depict the logarithm of the penalty function because otherwise the figures would only show the single highest peak of the graph or surface, respectively. With the approach chosen here, one achieves a better understanding of the penalty function.

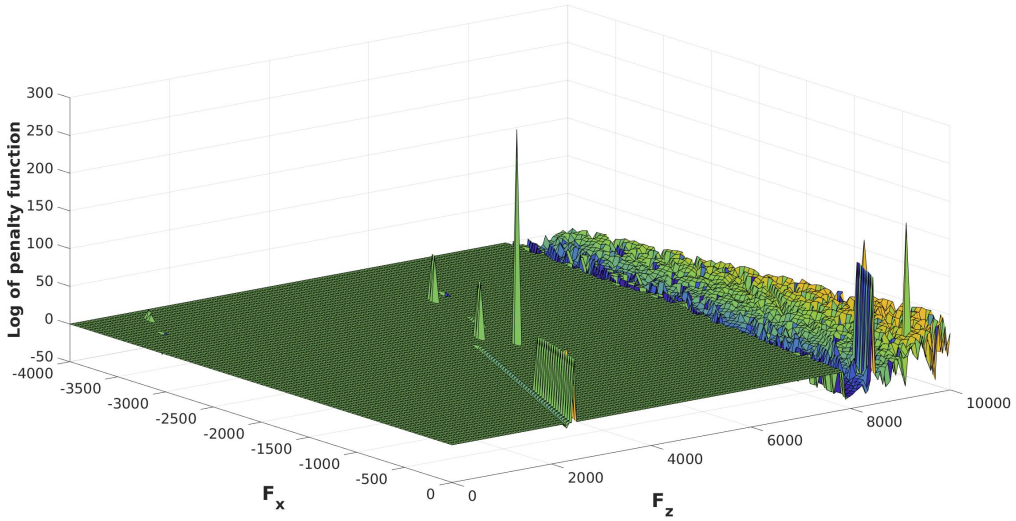

Fig. 18: *Graph of the logarithm of the penalty function* (2.14) $\log\Big(p\big(\Sigma_{\max}(l;0^\circ)\big)\Big)$ *computed on* $\mathcal{L}_{50}$.

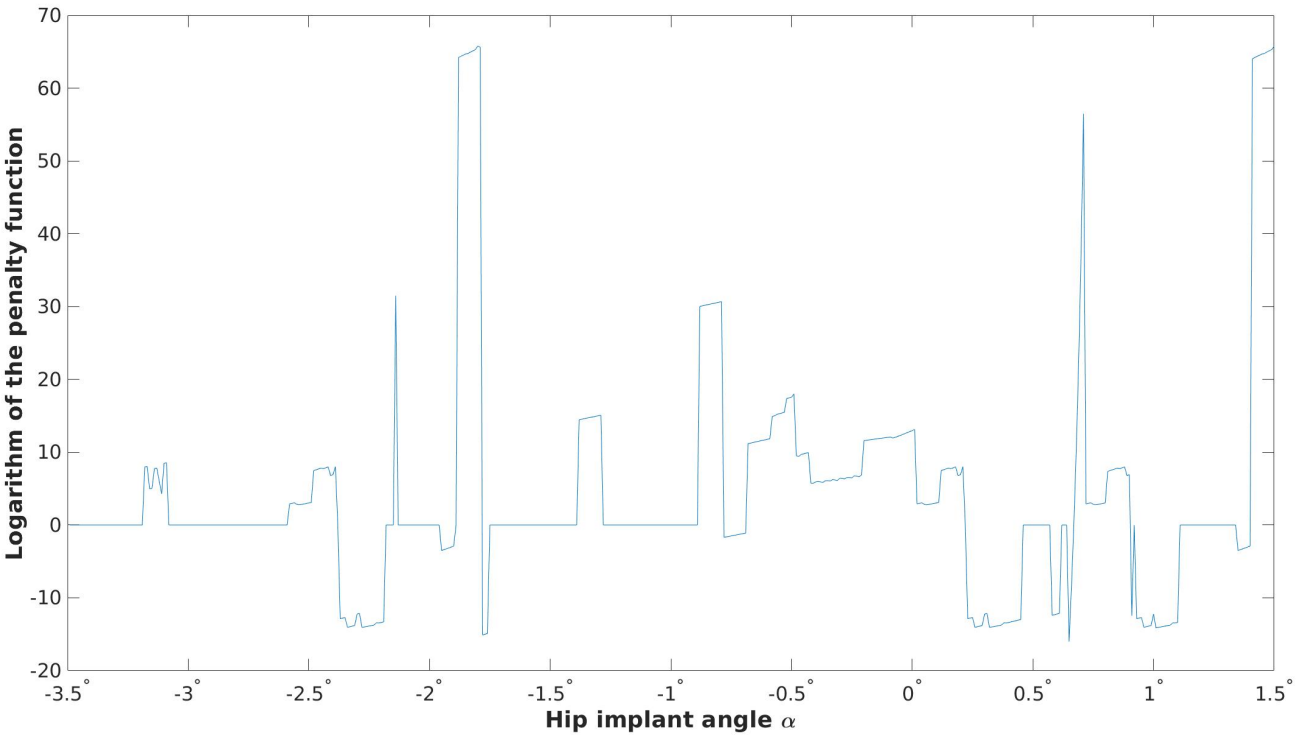

Fig. 19: *Graph of the logarithm of the penalty function* (2.14) $\log\Big(p\big(\Sigma_{\max}(l_0;\alpha)\big)\Big)$ *with fixed load* $l_0 = (-200, 9000)$ *for different implant positions with a discretization of* $\Delta\alpha = \big(10^{-4}\big)^\circ$ *of the admissible angle interval* A*; the apparent discontinuities are elaborated in the text.*

Visualization of *stress shielding*. The result of the stress distribution in the healthy bone for the case of l_{peak} from Eq. (5.2) can be seen in Fig. 20 on the left. The same load applied to the bone-implant geometry with the femur implant in position $\alpha = 0°$ computes to the result in Fig. 20. In Fig. 21, the difference in stress distribution with respect to the remaining bone after hip joint replacement is depicted. The minimization of this stress difference, i.e., the mean across all loads $l \in \mathcal{L}$, with the added value of the penalty function (2.14) is the aim of optimization.

Objective function. Without optimization, the interval of admissible implant position angles A was discretized by $\Delta\alpha = \left(10^{-4}\right)°$. When the change in position of the implant adds up to 0.1° the bone-implant mesh was reset, leading to a different mesh discretization and subsequently to a discontinuity in the objective. Below the value 0.1°, rigid body transformation was applied. In these sections, the objective function is continuous and allows for optimization. Using the exemplary load $l = (-200, 250)$ the resulting graph of the stress difference integral $\int_{\Omega_{\text{bone}}(\alpha)} \frac{1}{2}\|\Sigma(l) - \bar{\Sigma}(l)\|_F^2 \, \mathrm{d}x$ can be seen in Fig. 22, the resulting penalty function values in Fig. 19.

The logarithm $\log\Big(\int_{x\in\Omega_{\text{bone}}(0°)} \frac{1}{2}\big\|\Sigma(l;0°) - \bar{\Sigma}(l)\big\|_F^2 \, \mathrm{d}x + p\big(\Sigma_{\max}(l;0°)\big)\Big)$ of both the stress differences depicted in Fig. 16 and the penalty function shown in Fig. 18 computed on $\mathcal{L}_{50}$ is given in Fig. 23. The inclusion of the penalty function necessitates taking the logarithm of the sum of stress difference integral and the penalty function. Note that this figure still leaves out the probability function $\pi(l)$. Finally, Fig. 24 shows the response surface of the logarithm of the target function at $\alpha = 0°$

$$\log\Big(\pi(l)\big[\int_{x\in\Omega_{\text{bone}}(0°)} \frac{1}{2}\big\|\Sigma(l;0°) - \bar{\Sigma}(l)\big\|_F^2 \, \mathrm{d}x + p\big(\Sigma_{\max}(l;0°)\big)\big]\Big),$$

computed on $\mathcal{L}_{50}$. If $\pi(l)\big[\int_{x\in\Omega_{\text{bone}}(0°)} \frac{1}{2}\big\|\Sigma(l;0°) - \bar{\Sigma}(l)\big\|_F^2 \, \mathrm{d}x + p\big(\Sigma_{\max}(l;0°)\big)\big] = 0$ for one of the depicted $l \in \mathcal{L}_{50}$ then this value was changed to 10^{-18} before applying the logarithm. Leaving out the logarithm that was taken here for demonstrative purpose, and integrating over the load domain $\mathcal{L}$ gave the objective function value $j(0°)$, Eq. (2.17).
The shown surfaces in Fig. 16 and Fig. 18 are the response surfaces to be interpolated with Kriging.

Set up Kriging and Monte Carlo. The initial sample loads for Kriging were chosen considering the load probability map $\pi(l)$, cf. Fig. 13. The initial samples l_i are found in Tab. 4. For the moment space M_y a fixed moment m was chosen. For the Monte Carlo integration N_{MC} was set to 5000.

5.2 Computing times of the separate subalgorithms

Two FE-grids of the bone (Fig. 9) and the bone-implant setup (Fig. 10) with different discretizations will be used to show the impact on computation time of the separate subalgorithms. The first is the non-refined, coarse grid ($N = 0$), the second is the *fine grid* ($N = 1$) which is once uniformly refined.

As mentioned at the end of Sec. 1.3, the FE-grid of the pure bone geometry is finer discretized than the bone-implant geometry because the PDE needs to only be solved once per applied load for this geometry. On the other hand, the contact boundary of

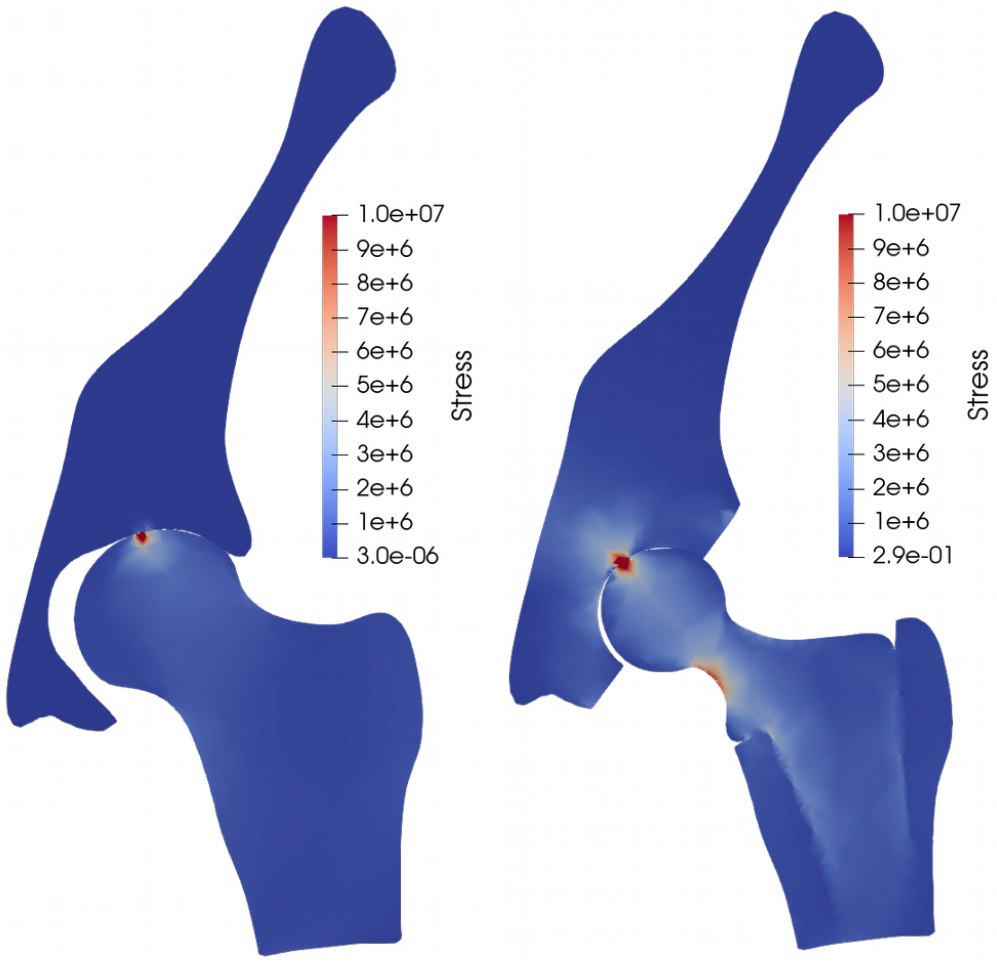

Fig. 20: *Resulting stress distribution in the bone (left) and the bone-implant geometry (right) with the femur implant in position $\alpha = 0°$ when the measured peak load l_{peak} (cf. Eq. (5.2)) from stumbling is applied.*

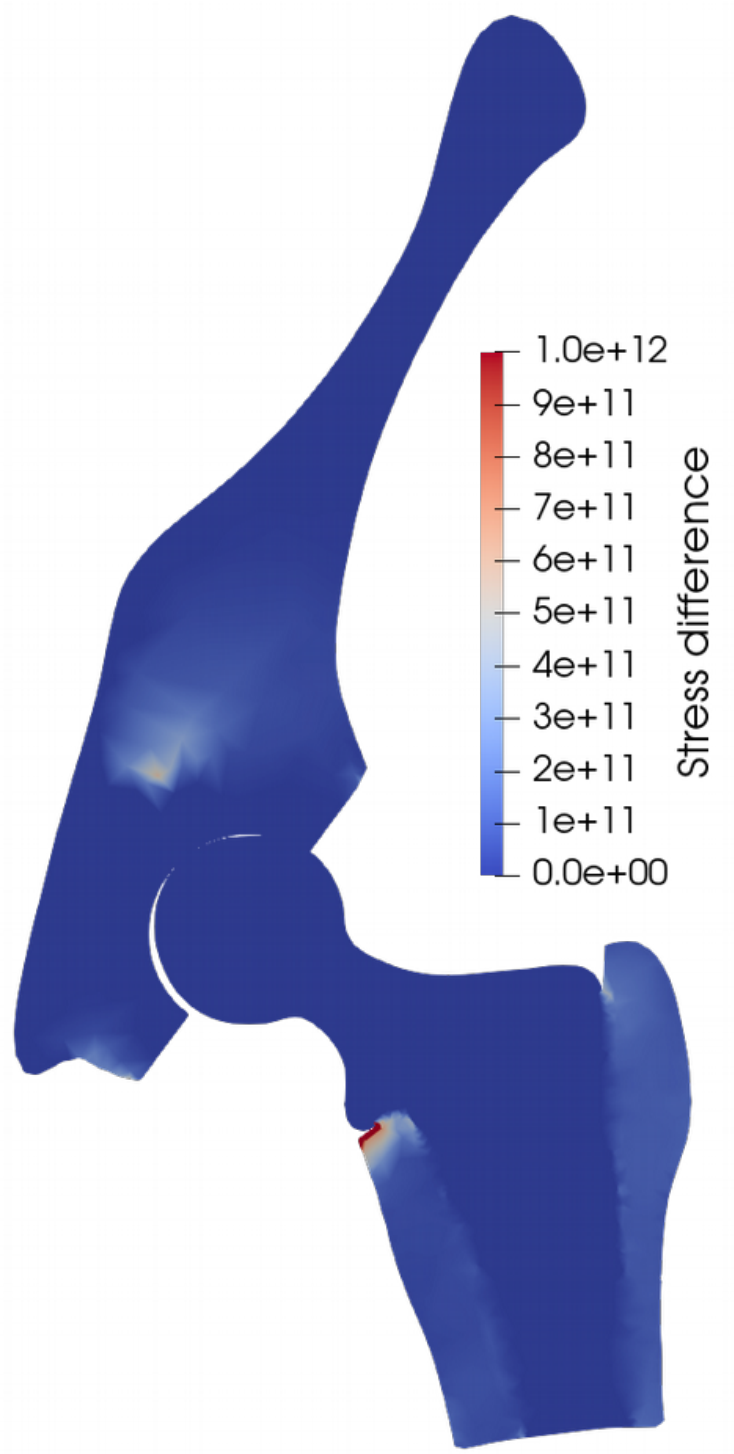

Fig. 21: *Resulting difference in stress distribution* $\frac{1}{2}\left\|\Sigma(l_{peak}) - \bar{\Sigma}(l_{peak})\right\|_F^2$ *in both the pelvis and the femur geometry when the measured peak load* l_{peak} *(Eq. (5.2)) from stumbling is applied.*

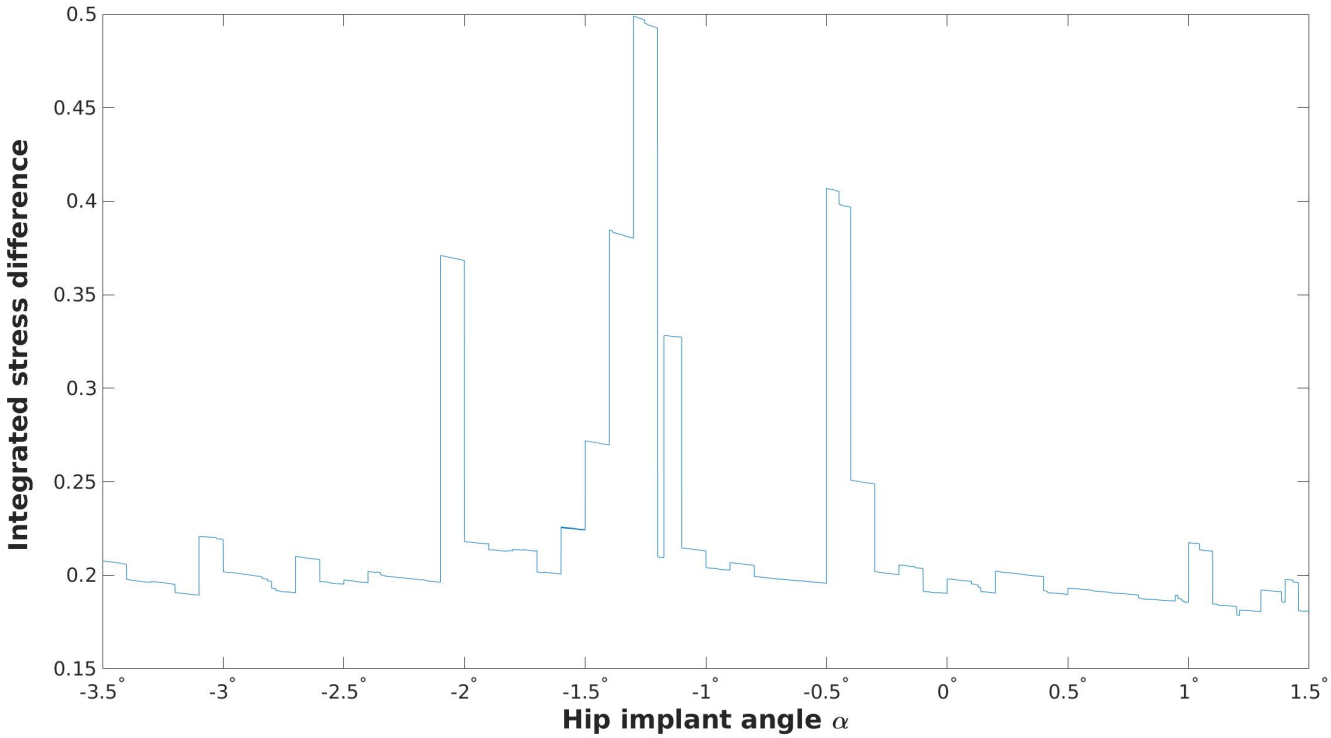

Fig. 22: *Graph of the stress difference integral* $\int_{\Omega_{bone}(\alpha)} \frac{1}{2}\|\Sigma(l) - \bar{\Sigma}(l)\|_F^2 \, \mathrm{d}x$ *exemplary for load* $l = (-200, 250)$ *for different implant positions with a discretization of* $\Delta\alpha = \left(10^{-4}\right)^\circ$ *of the admissible angle interval* A.

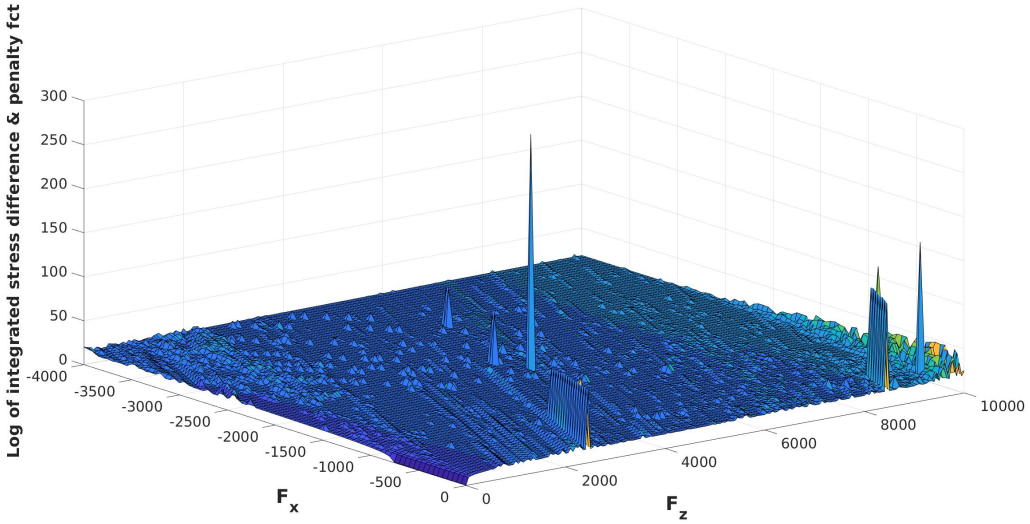

Fig. 23: *The surface of* $\log\Big(\int_{x\in\Omega_{bone}(0^\circ)} \frac{1}{2}\big\|\Sigma(l;\boldsymbol{\alpha}) - \bar{\Sigma}(l)\big\|_F^2 \, \mathrm{d}x + p\big(\Sigma_{\max}(l;0^\circ)\big)\Big)$ *(without probabilities* $\pi(l)$*) computed on* $\mathcal{L}_{50}$.

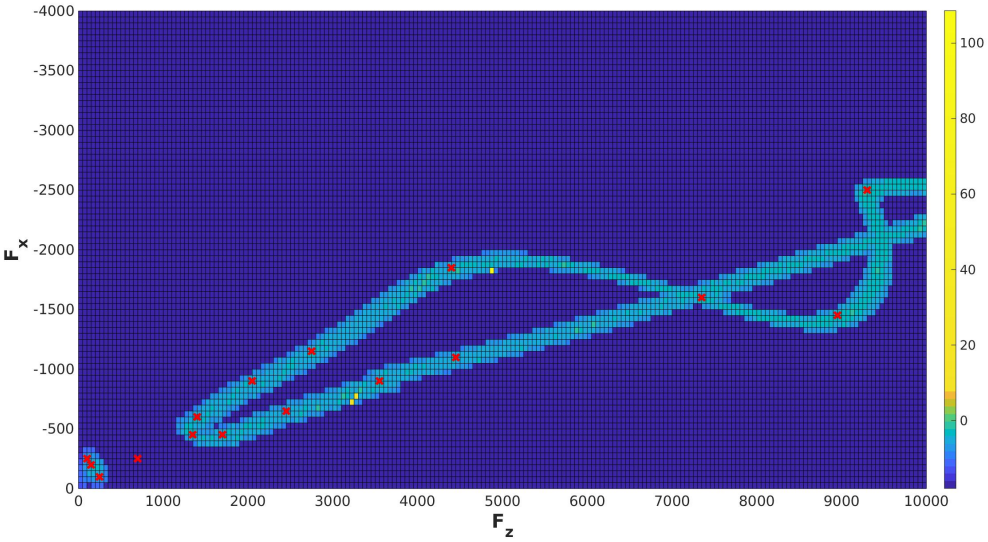

Fig. 24: *The surface of* $\log\Big(\pi(l)\big[\int_{x\in\Omega_{bone}(0^\circ)}\frac{1}{2}\big\|\Sigma(l;0^\circ)-\bar{\Sigma}(l)\big\|_F^2\,\mathrm{d}x+p\big(\Sigma_{\max}(l;0^\circ)\big)\big]\Big)$ *(*with *probabilities* $\pi(l)$*) computed on* $\mathcal{L}_{50}$ *with marked initial samples load samples (red crosses) as in Tab. 4; if* $\pi(l)\big[\int_{x\in\Omega_{bone}(0^\circ)}\frac{1}{2}\big\|\Sigma(l;0^\circ)-\bar{\Sigma}(l)\big\|_F^2\,\mathrm{d}x+p\big(\Sigma_{\max}(l;0^\circ)\big)\big]=0$ *then this value was changed to* 10^{-18} *before applying the logarithm.*

Table 4: *Initial Kriging samples chosen in regions where the load probability* $\pi(l)$ *is maximized and two in the subdomain of* $\mathcal{L}$ *where the penalty function becomes relevant (cf. Fig. 13 and 24)*

$l_i=(f_x,f_z)_i=$	$(-100,250)$,	$(-650,2450)$,
	$(-200,150)$,	$(-900,2050)$,
	$(-250,100)$,	$(-900,3550)$,
	$(-250,700)$,	$(-1100,4450)$,
	$(-450,1350)$,	$(-1150,2750)$,
	$(-450,1700)$,	$(-1600,7350)$,
	$(-600,1400)$,	$(-1850,4400)$,
	$(-1450,8950)$,	$(-2500,9300)$.

the bone-implant setup is finer discretized than the interior of the domain and the non-contact boundary. Hence, in a coarsely discretized setup excluding the implant with 11,832 DOF in the pelvis and 8208 DOF in the femur, see Fig. 9, the PDE solves take up 28% of the overall computation time. In the coarse discretization of the geometry that includes the implants (Fig. 10), the pelvis accounts for 915 DOF and the femur for 1282 DOF. In this setup, the PDE solves take up 11% of the overall computation time. Together they account for 39% of total computation time, while the estimation of Kriging parameters and necessary matrices, that can be precomputed, takes up $8 \cdot 10^{-5}$%, and Monte Carlo integration accounts for 61% of the almost 12 minutes total computing time. The relatively high share of Monte Carlo in the total computation time is mostly due to the coarse discretization of the FE-grid $\mathcal{T}_h$, which makes one PDE solve in this test setup much cheaper than in a realistic 3D and quantitatively reliable discretization. Also, the PDE is only solved for a low number of 16 initial loads and adaptivity is at this point not yet included.

With a once uniformly refined FE-grid (32,832 DOF in the pelvis, 47,328 DOF in the femur, 3123 DOF and 4680 DOF including the implant, resp.) one gets 48.5% for the response computation in the pure bone and 24.7% for the response in the bone-implant setup. Together they make up 73.3%, as against $6 \cdot 10^{-4}$% for Kriging, and 26.8% for Monte Carlo (a little over 26 minutes computing time).

With unchanged algorithms for the setup of Kriging and Monte Carlo integration, there is little variation in the computing times between the aforementioned two cases. They are not dependent on the number of degrees of freedom, as opposed the solution of the PDE. This is why with increasing precision of the PDE solution, one solve becomes quickly very expensive in contrast to the setup of the Kriging interpolation scheme and the Monte Carlo integration.

This proves the concept of wanting to reduce the PDE solves and rely on interpolation and comparatively quick integration schemes – here Kriging and Monte Carlo – to compute the integral value of the target function (2.17).

5.3 Evaluation of Kriging

As stated before, Kriging will be applied to the substitute functions

$$\begin{aligned}\tilde{z}_{\Delta\Sigma}(l;\alpha) &:= (-f_x + f_z)\big(\sin(-f_x) + \sin(f_z) + 2\big)\big[(\alpha+2)(\alpha-1) + 2.25\big],\\ \tilde{z}_{\max}(l;\alpha) &:= 350 f_z\big[(\alpha+1.5)(\alpha+1.2) + 0.15^2\big],\end{aligned}$$

previously given in Eq. (5.1) for function values corresponding with the initial samples l_i from Tab. 4. A depiction of the substitute response surfaces computed on $\mathcal{L}_{50}$ can be seen in Fig. 25. The resulting logarithm of the target response surface with $\alpha = 0°$

$$\log\Big(\pi(l)\big[\tilde{z}_{\Delta\Sigma}(l;0°) + p\big(\tilde{z}_{\max}(l;0°)\big)\big]\Big),$$

is shown in Fig. 26, where for loads $l \in \mathcal{L}_{50}$ with

$$\pi(l)\big[\tilde{z}_{\Delta\Sigma}(l;0°) + p\big(\tilde{z}_{\max}(l;0°)\big)\big] = 0,$$

the value 0 was exchanged with 10^{-18} before applying the logarithm. As it is yet unclear, which kernel model in combination with which drift model best reproduces either of the two response surfaces of the stress differences $\tilde{z}_{\Delta\Sigma}$ and of the maximum stresses $\tilde{z}_{\max}$ different measures are presented and partially evaluated to identify the best fitting model combination.

Remark. *The performance of the Matérn kernel (cf. Tab. 2) will not be investigated. This is due to the complexity of the analytical treatment of the function and its derivatives as well as the fact, that this model does not behave quantitatively different than the other kernels.*

A priori variogram model estimation. Two model-fit measures will be presented here, that one may use to determine, which variogram-drift model combination best reproduce the sought after response surface.

Since the model parameters of the variogram are computed by solving a restricted maximum likelihood (REML) problem, cf. Eq. (3.13), the first way of measuring the model performance is to evaluate the final value of the (REML)-function.

A second way of estimating the model fit is to compute the sum of distances between the responses $z(l_i)$ from the sample-response pairs $\big(l_i, z(l_i)\big)$ and the function values of the interpolation function $z^*(l_i)$ in the Euclidean norm

$$\mathcal{E}_{\mathrm{I}} \coloneqq \left(\sum_i \big(z(l_i) - z^*(l_i) \big)^2 \right)^{1/2}. \tag{5.3}$$

This measure will be called *interpolation error* $\mathcal{E}_{\mathrm{I}}$. Since Kriging is an exact interpolator, one must not insert l_i into $\mathcal{E}_{\mathrm{I}}$, when the Kriging model z^* is also trained with $\big(l_i, z(l_i)\big)$. Usually, the quantity $\mathcal{E}_{\mathrm{I}}$ is evaluated utilizing cross-validation, as previously mentioned in Sec. 3.4 as a possibility to compute $\mathcal{E}_{\mathrm{K}}$. In the present case with surrogate response functions $\tilde{z}_{\Delta\Sigma}$ and $\tilde{z}_{\max}$, the evaluation of $z(l_i)$ is cheap. Therefore, here, too, one can employ Monte Carlo sampling to draw $N_{\mathrm{K},\mathcal{E}} \in \mathbb{N}$ samples using the probability density $\pi(l)$ from Eq. (2.21) as one does for the computation of $\mathcal{E}_{\mathrm{K}}$.

Further investigated ideas to estimate the performance of the separate Kriging models were to compute the L^2-norm of the misfit between precomputed values of z on $\mathcal{L}_{50}$ and interpolated values from z^*. This idea suffered from two drawbacks: on the one hand, in a realistic scenario, the computed values of z may not exist, may not be computable at all or be too expensive to compute. On the other hand, with only few initial samples l_i (16 here) the training data for interpolation is scarce, making the error big and almost equal for all model combinations thus rendering this error measure futile.

Another idea was to compute the sum of model variance values from Eq. (3.28) for each sample l_i which the model is trained with using cross-validation, or for each of the sampled l_i using importance sampling. This is similar to the first step in computing the Kriging error estimate $\mathcal{E}_{\mathrm{K}}$ in Eq. (3.34) used in the adaptive Kriging scheme. Here, the information gain compared to the interpolation error $\mathcal{E}_{\mathrm{I}}$ was negligible and thus discarded.

Bounds on kernel parameters $\boldsymbol{\theta}$. The kernel parameters a and b are determined by computing the minimum of $rL(\boldsymbol{\theta})$, Eq. (3.12). When evaluating rL the samples are always the same l_i as in Tab. 4. For the responses one needs to differentiate between $\tilde{z}_{\Delta\Sigma}$ and $\tilde{z}_{\Sigma_{\max}}$. Also, one needs to consider the different variogram kernels (bounded linear, exponential, Gauss – see Tab. 2) and drift models (constant, linear, and quadratic). Further, the parameters have a physical meaning that necessitate an imposition of parameter bounds.

For the sill parameter b one anticipates values that are of similar magnitude as the average value of absolute sample responses. This makes for a good starting value of the sill b. By definition, b needs to be greater than 0. When b comes close to 0, the $\tilde{\Gamma}$ matrix

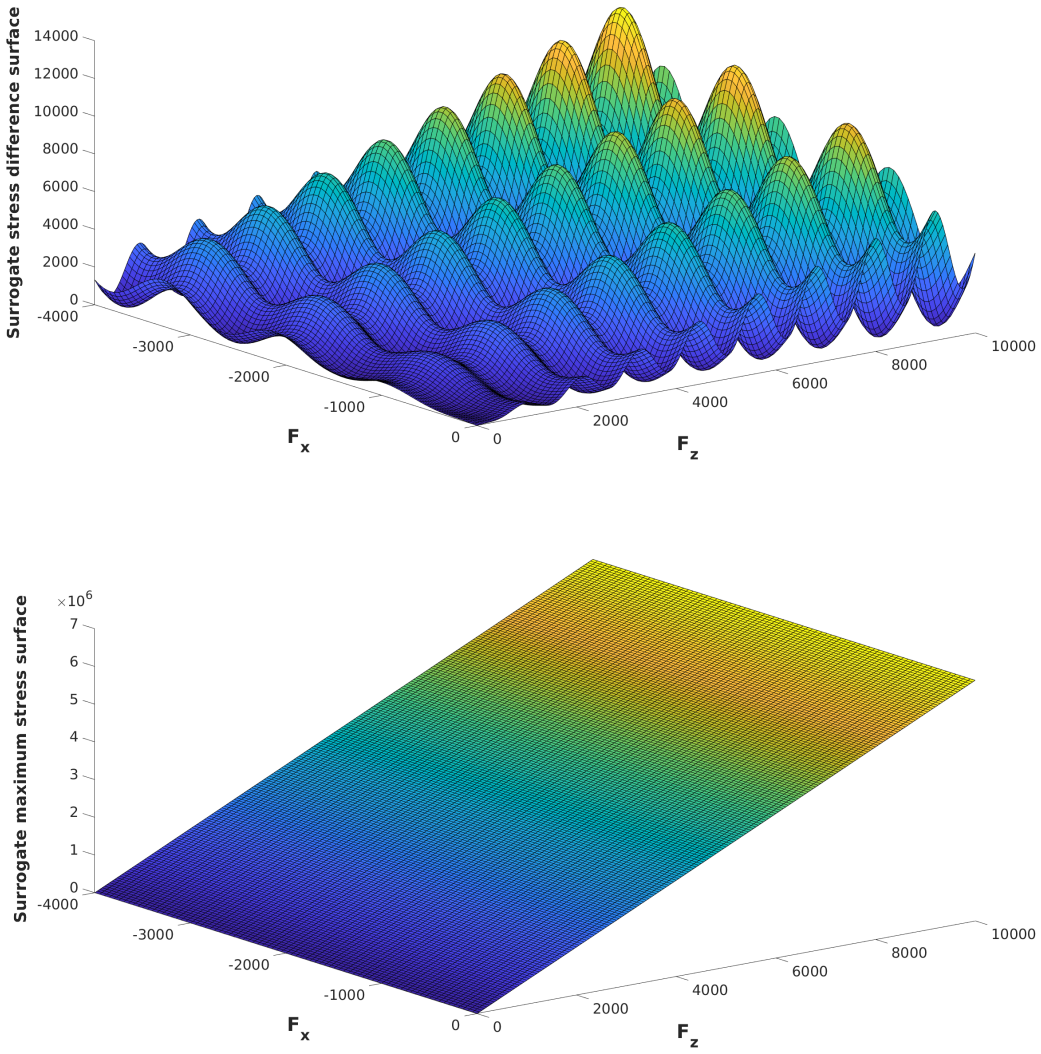

Fig. 25: *Substitute response surfaces of $\tilde{z}_{\Delta\Sigma}$ (top) and $\tilde{z}_{max}$ (bottom) as given in Eq. (5.1) computed on $\mathcal{L}_{50}$.*

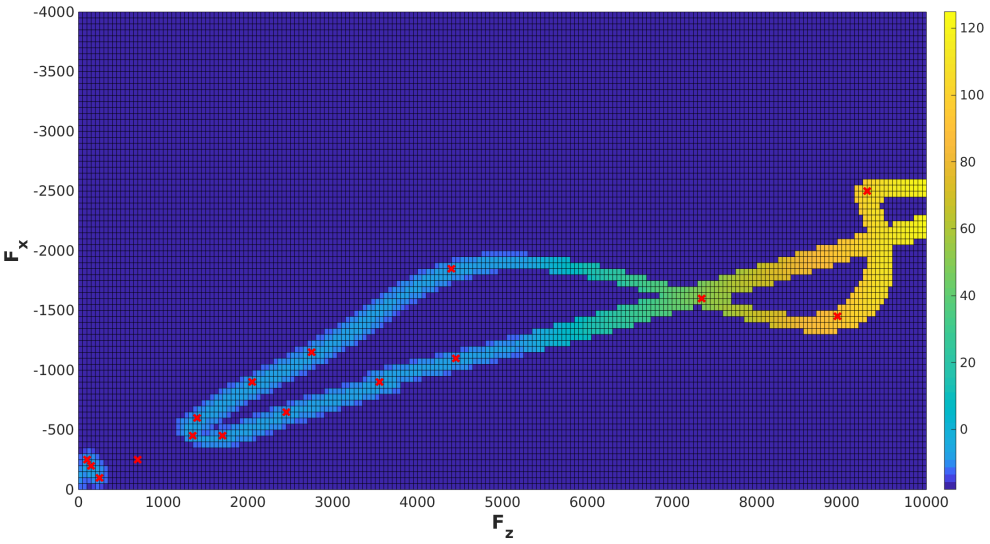

Fig. 26: *The logarithm of the target function* $j(0°)$ *(Eq. (2.15))* $\log\Big(\pi(l)\big[\tilde{z}_{\Delta\Sigma}(l;0°) + p\big(\tilde{z}_{\max}(l;0°)\big)\big]\Big)$ *computed on* $\mathcal{L}_{50}$ *with surrogate functions* $\tilde{z}_{\Delta\Sigma}$ *and* $\tilde{z}_{\max}$*; the red crosses indicate the initial load sample positions* l_i *as in Tab. 4; if* $\pi(l)\big[\tilde{z}_{\Delta\Sigma}(l;0°)+p\big(\tilde{z}_{\max}(l;0°)\big)\big] = 0$ *then this value was changed to* 10^{-18} *before applying the logarithm.*

becomes singular and the interpolation model breaks down. Therefore, the sill value b will be bounded from below by 10 to make the optimization problem well-posed for all instances. This corresponds with the behavior of rL for $b \to 0$ where $|rL| \to \infty$, cf. Fig. 27 and Fig. 28. For the chosen model combinations below, this safety measure was not used by the algorithm when determining parameter b.

Shifting the focus to the range parameter a, recall that the range a determines up to what distance between samples there exists a (meaningful) co-/variance (cf. end of Sec. 3.3.2). If this parameter got too low, each sample-response pair became isolated meaning there is no correlation at all between neighboring samples. Numerics showed, that the bound of $a > 0$ suffices.

Similarly, with a domain as large as $\mathcal{L}$, it is implausible for samples across the domain to have a co-/variance. Thus the parameter a is bounded from above to half the diameter of the domain $\mathcal{L}$, as is suggested in Wackernagel [157, Ch. 4]. With the present definition of $\mathcal{L}$, this gives an upper bound of 5500 on the range parameter a. Note, that only for the bounded linear variogram the range parameter translates directly to the effective range. The *effective range* [161, Ch. 5.2] or *practical range* [157, Ch. 6] describes the distance h at which the chosen variogram model reaches 95% of the sill value b – or similarly the covariance has decreased by 95% from the sill value b. For the exponential variogram the effective range is $h = 3a$, for the Gaussian model, it is $h = \sqrt{3}a$. This affects the upper bound on a to be approximately 1850 for the exponential variogram and approximately 3200 for the Gaussian variogram. These comments on a also correspond with the examination of rL for $\tilde{z}_{\Sigma_{\max}}$ with the Gauss kernel and the linear or quadratic drift models. The value of rL falls below 0 for values of a above the effective range, see bottom of Fig. 28.

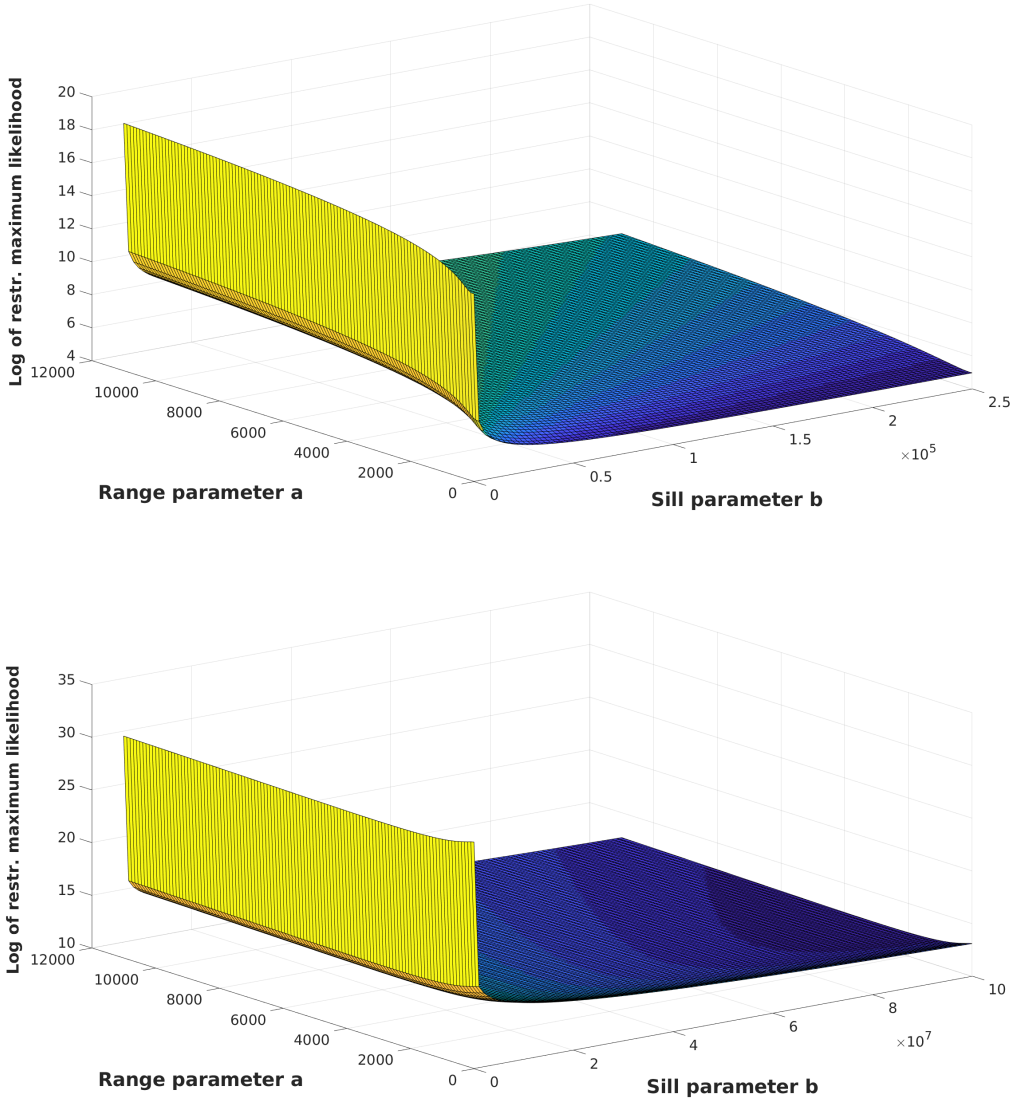

Fig. 27: *Logarithm of restricted maximum likelihood functions* $\log\big(rL(\boldsymbol{\theta})\big)$*, Eq.* (3.12)*, for samples* l_i *from Tab.* 4*, with responses from* $\tilde{z}_{\Delta\Sigma}$ *(top) using an exponential kernel and a quadratic drift, and with responses from* $\tilde{z}_{\Sigma_{\max}}$ *(bottom) using an exponential kernel and a linear drift.*

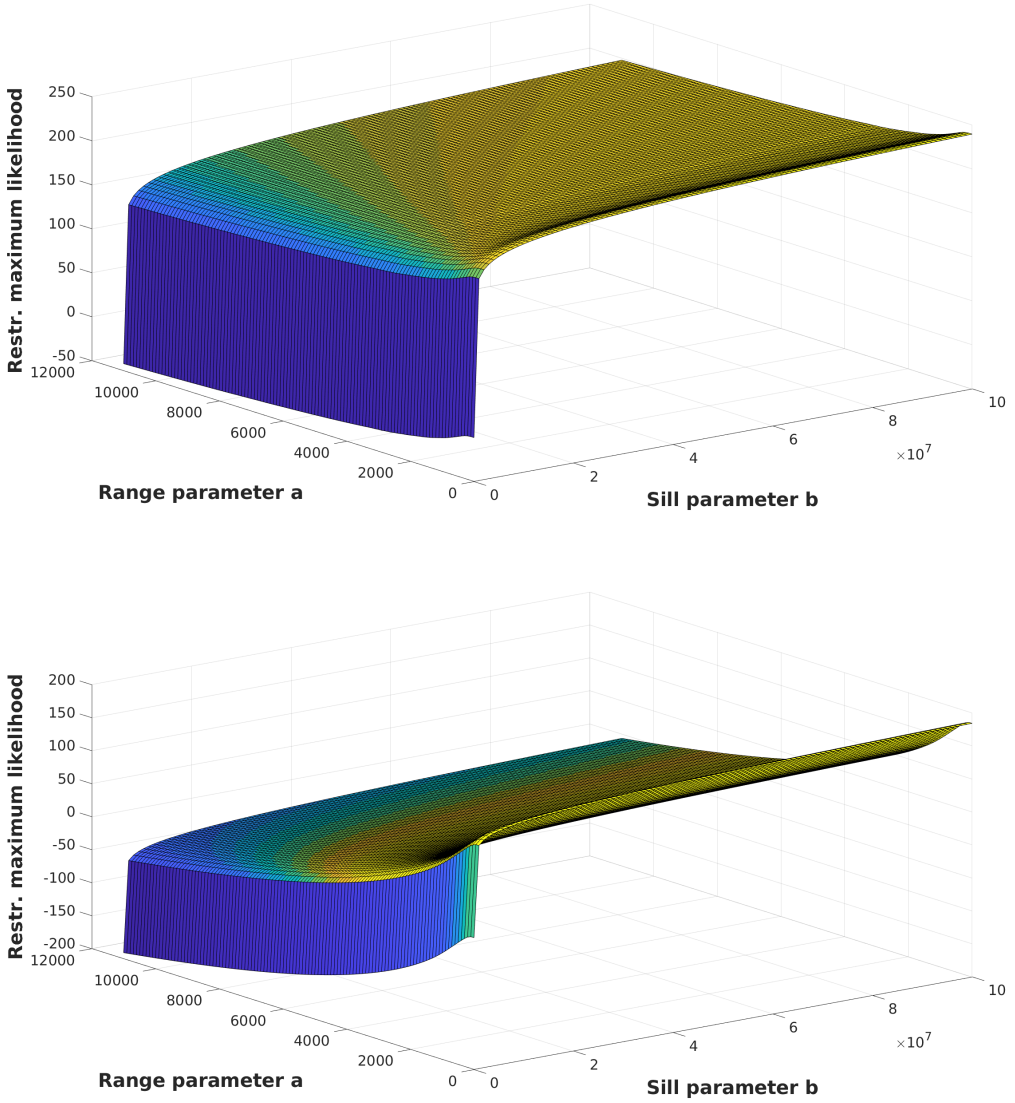

Fig. 28: *Restricted maximum likelihood functions $rL(\boldsymbol{\theta})$, Eq. (3.12), for samples l_i from Tab. 4, with responses from $\tilde{z}_{\Sigma_{\max}}$ using a bounded linear kernel and a linear drift (top), and using a Gaussian kernel and a quadratic drift (bottom).*

Table 5: *Restricted maximum likelihood (REML) values of Eq. (3.12) after the parameters a and b are evaluated for the different drift (constant, linear, quadratic) and variogram models (bounded linear, exponential, Gauss – see Tab. 2) for 16 stress difference response values* $\int_{x\in\Omega_{bone}(0°)} \frac{1}{2}\left\|\Sigma(l_i;0°)-\bar{\Sigma}(l_i)\right\|_F^2 \mathrm{d}x$ *(cf. Eq. (3.8)) belonging to the initial load samples given in Tab. 4; the bold entry marks the one with the lowest value.*

(REML)	Constant	Linear	Quadratic
Bounded linear model	230.8	193.6	$1.697 \cdot 10^4$
Exponential model	232.7	199.2	**151.3**
Gauss model	$4.669 \cdot 10^4$	$4.569 \cdot 10^4$	$3.604 \cdot 10^4$

Table 6: *Restricted maximum likelihood (REML) values of Eq. (3.12) after the parameters a and b are evaluated for the different drift (constant, linear, quadratic) and variogram models (bounded linear, exponential, Gauss – see Tab. 2) for 16 stress difference response values* $\Sigma_{\max}(l_i;0°)$ *(cf. Eq. (3.8)) belonging to the initial load samples given in Tab. 4; the bold entry is the one with the lowest value.*

(REML)	Constant	Linear	Quadratic
Bounded linear model	$4.901 \cdot 10^9$	183.6	90.64
Exponential model	$2.776 \cdot 10^9$	194.0	147.2
Gauss model	1489	126.4	**86.82**

Evaluation of the kernel-drift model performance. The starting parameters were $\boldsymbol{\theta}_0 = \left(500, 10^7\right)$ for $z^*_{\Delta\Sigma}$ and $\boldsymbol{\theta}_0 = \left(200, 10^7\right)$ for $z^*_{\Sigma_{\max}}$. For some model combinations the initial parameters needed to be altered, as otherwise the appearing matrix inversions could not be performed due to matrices being singular – already indicating that these models are not well fit for the interpolation at hand. With the responses from $z^*_{\Delta\Sigma}$ using the bounded linear model and quadratic drift, b_0 was set to 10^4. The same model combination applied to the responses from $z^*_{\Sigma_{\max}}$ started with $b_0 = 10^5$. When interpolating $z^*_{\Sigma_{\max}}(l_i, 0°)$ with a constant drift model, $\boldsymbol{\theta}_0$ was set to $\left(500, 10^4\right)$. For the number of samples $N_{\mathrm{K},\mathcal{E}}$ included in the importance sampling used to evaluate $\mathcal{E}_{\mathrm{I}}$, set $N_{\mathrm{K},\mathcal{E}} = 1000$.

For the case of stress differences $\Delta\Sigma$, the values for (REML) are found in Tab. 5. These values indicate, that a quadratic drift is best suited – except for the bounded linear kernel joined with quadratic drift –, the constant mean constitutes the worst fit, and the Gauss model is an unsuitable kernel for the given interpolation problem.

For the case of maximum stresses $\Sigma_{\max}$, the (REML) estimates are in Tab. 6. The values for the constant mean approach indicate, that this drift model is utterly wrong. This makes sense, looking at the bottom of Fig. 25 where one would assume a linear drift. But again the (REML) values for the quadratic drift model are for all kernels better than in the case of linear drift. This does not seem conclusive and the focus turns to the interpolation error $\mathcal{E}_{\mathrm{I}}$.

The results of $\mathcal{E}_{\mathrm{I}}$ for $\Delta\Sigma$ are found in Tab. 7. The exponential kernel shows the lowest error for all drift models. As for this kernel the lowest error value lies in the combination with the constant mean, this conjunction will be chosen. The resulting parameters for

Table 7: *Interpolation error $\mathcal{E}_I$ for the responses of the stress difference values $\int_{x\in\Omega_{bone}(0°)} \frac{1}{2}\left\|\Sigma(l;0°) - \bar{\Sigma}(l)\right\|_F^2 \mathrm{d}x$ and the interpolation values at the same sample locations computed for the different drift (constant, linear, quadratic) and variogram models (bounded linear, exponential, Gauss – see Tab. 2); the bold entry is the one with the lowest value.*

$\mathcal{E}_\mathrm{I}$	Constant	Linear	Quadratic
Bounded linear model	$1.70 \cdot 10^4$	$1.79 \cdot 10^4$	$1.77 \cdot 10^4$
Exponential model	$\mathbf{1.69 \cdot 10^4}$	$1.75 \cdot 10^4$	$1.71 \cdot 10^4$
Gauss model	$2.06 \cdot 10^4$	$2.03 \cdot 10^4$	$2.03 \cdot 10^4$

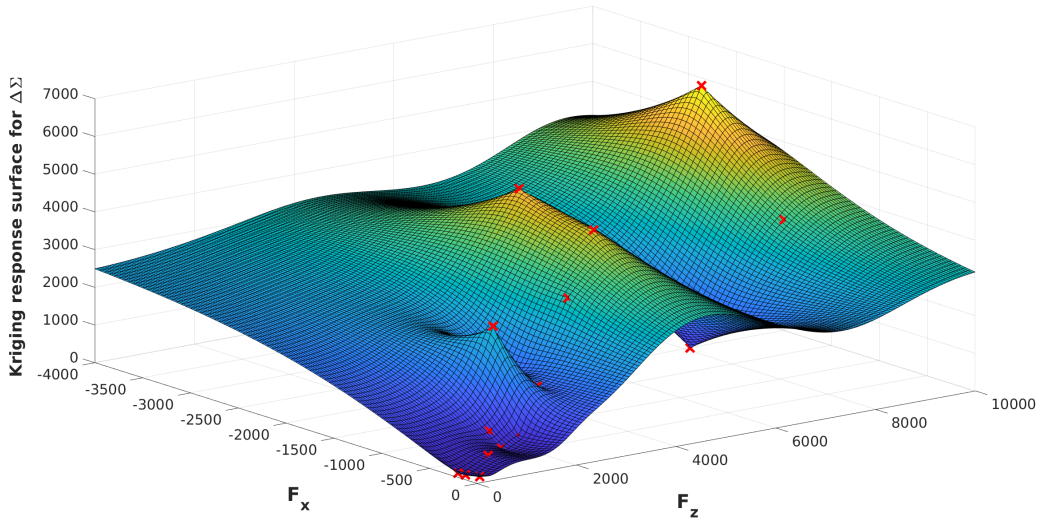

Fig. 29: *Interpolation result $z^*_{\Delta\Sigma}$ of substitute response surfaces of $\tilde{z}_{\Delta\Sigma}$ as given in Eq. (5.1) computed on $\mathcal{L}_{50}$; the markers show the initial sample locations l_i as in Tab. 4 with their respective responses $\tilde{z}_{\Delta\Sigma}(l_i)$.*

this choice of model are $(a, b) = \left(1803, 10^7\right)$. The function $z^*_{\Delta\Sigma}(l)$ with $l \in \mathcal{L}_{50}$ is shown in Fig. 29 with the depicted red crosses being the initial sample-response pairs.

Evaluating $\mathcal{E}_\mathrm{I}$ between the responses $\tilde{z}_{\max}(l_i)$ and interpolation $z^*_{\max}(l_j)$, the results are found in Tab. 8. The error values for the constant drift model show, that a constant mean ansatz is unsuitable. Thus, the constant drift is discarded here. Again, the exponential kernel shows the best performance among all three kernel models. The lowest error is attained when combining the exponential kernel with linear drift. The resulting parameters for this conjoined model are $(a, b) = \left(1798, 9.9996 \cdot 10^6\right)$. The function $z^*_{\max}(l)$ with $l \in \mathcal{L}_{50}$ is shown in Fig. 30 with the depicted red crosses being the initial sample-response pairs.

Closing, one is to conclude that the examination of the (REML) values for kernel-drift estimation is inconclusive. The interpolation error $\mathcal{E}_\mathrm{I}$ on the other hand is a constructive measure. With the model combinations chosen one can now turn to the evaluation of adaptive Kriging.

Table 8: *Interpolation error $\mathcal{E}_I$ for the responses the maximum stress values $\Sigma_{\max}(l; 0°)$ and the interpolation values at the same sample locations computed for the different drift (constant, linear, quadratic) and variogram models (bounded linear, exponential, Gauss – see Tab. 2); the bold entry is the one with the lowest value.*

$\mathcal{E}_\mathrm{I}$	Constant	Linear	Quadratic
Bounded linear model	$3.51 \cdot 10^{7}$	$5.54 \cdot 10^{-8}$	$1.13 \cdot 10^{-7}$
Exponential model	$1.68 \cdot 10^{7}$	$\mathbf{3.19 \cdot 10^{-8}}$	$4.98 \cdot 10^{-8}$
Gauss model	$2.30 \cdot 10^{7}$	$1.46 \cdot 10^{-3}$	$1.17 \cdot 10^{-3}$

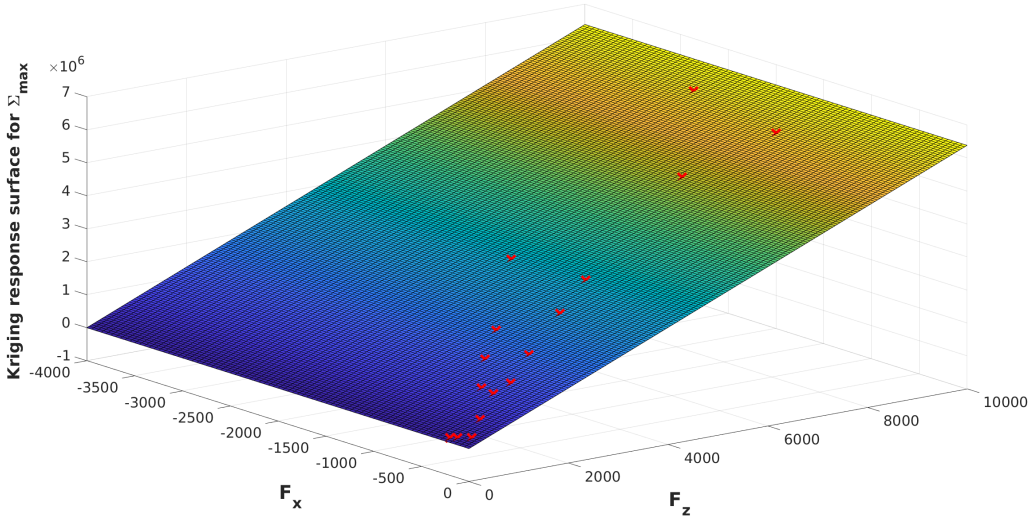

Fig. 30: *Interpolation result z^*_{max} of substitute response surfaces of $\tilde{z}_{max}$ as given in Eq. (5.1) computed on $\mathcal{L}_{50}$; the markers show the initial sample locations l_i as in Tab. 4 with their respective responses $\tilde{z}_{\max}(l_i)$.*

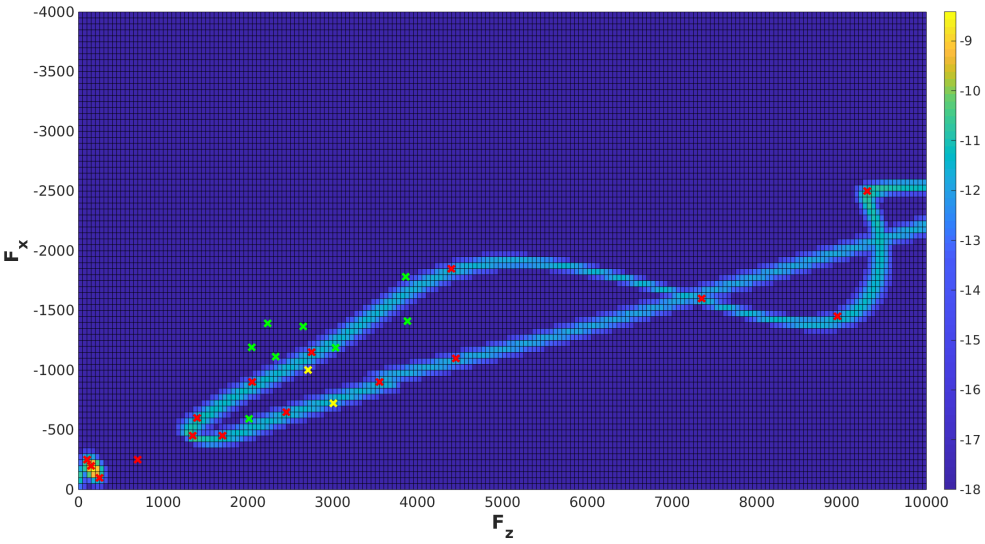

Fig. 31: *Ten adaptively added samples shown on the surface of* $\log\big(\pi(l)\big)$ *computed on* $\mathcal{L}_{50}$ *with the initial samples from Tab. 4 (red markers); the yellow markers indicate those added samples when local exploitation was favored, the green markers where added with global exploration being predominant; if* $\pi(l) = 0$ *then this value was changed to* 10^{-18} *before applying the logarithm.*

5.3.1 Adaptivity in Kriging

Now, the performance of adaptivity in Kriging as described in Sec. 3.4 will be evaluated. In the optimization, the adaptive scheme will be triggered, when the Kriging error $\mathcal{E}_{\mathrm{K}}$ (3.34) is above the minimum required precision $\tau_h/3$, see Eq. (3.35). When that is the case, Eq. (3.31) will be evaluated to find the next added load. Since in this section for the precision criterion (3.35) there is no τ_h given, the number of adaptively added samples will here be set to ten samples without a precision check. The ten samples are consecutively added and afterwards it will be scrutinized whether global exploration or local exploitation is favored and how the respective interpolation errors $\mathcal{E}_{\mathrm{I}}$ and the Kriging error $\mathcal{E}_{\mathrm{K}}$ change. The number $N_{\mathrm{K},\mathcal{E}}$ was set to 1000 in the previous section.

The progression of the β value (cf. Eq. (3.30)) is for two of the ten added samples greater than or equal to 0.5 (yellow markers in Fig. 31), favoring local exploitation, and for the other samples $\beta < 0.5$ (green markers in Fig. 31), thus granting global exploration the bigger influence. Therefore, global exploration and local exploitation are both impacting the determination of new samples with a bias towards global exploration.

For the ten added samples, depicted in Fig. 31 by the yellow and green markers, one finds that the interpolation error $\mathcal{E}_{\mathrm{I}}$ actually increases for both function interpolations $z^*_{\Delta\Sigma}$ and $z^*_{\max}$. When interpolating $\Delta\Sigma$, the error increases from $1.691 \cdot 10^4$ to $1.693 \cdot 10^4$, and for $\Sigma_{\max}$ the error goes up from $3.189 \cdot 10^{-8}$ to $1.760 \cdot 10^{-7}$. While this is not necessarily wanted, it is not harmful either, since the interpolation error is only evaluated and needed in the determination of suitable kernel and drift models.

The initially measured Kriging error $\mathcal{E}_{\mathrm{K}}$ of 1.824 decreased to 1.767 after ten added samples. This is the important measure that needs to decrease and indeed it does.

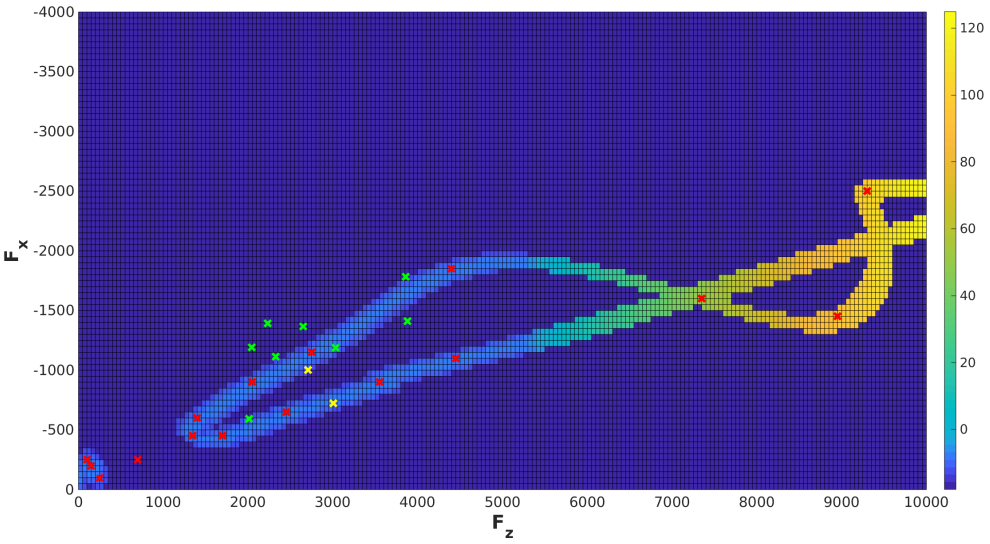

Fig. 32: *The surface of* $\log\Big(\pi(l)\big[z^*_{\Delta\Sigma}(l;0°) + p\big(z^*_{\max}(l;0°)\big)\big]\Big)$ *computed on* $\mathcal{L}_{50}$ *using Kriging interpolation for* $\Delta\Sigma$ *and* $\Sigma_{\max}$ *set up with initial* and *adaptively added samples; the yellow markers indicate those added samples when local exploitation was favored, the green markers were added with global exploration being predominant; if* $\pi(l)\big[z^*_{\Delta\Sigma}(l;0°) + p\big(z^*_{\max}(l;0°)\big)\big] = 0$ *then this value was changed to* 10^{-18} *before applying the logarithm.*

In Fig. 32 one also sees the surface of $\log\Big(\pi(l)\big[z^*_{\Delta\Sigma}(l;0°) + p\big(z^*_{\max}(l;0°)\big)\big]\Big)$ with interpolating functions $z^*_{\Delta\Sigma}$ and $z^*_{\max}$ now trained with 26 sample-response pairs. Again, when $\pi(l)\big[z^*_{\Delta\Sigma}(l;0°) + p\big(z^*_{\max}(l;0°)\big)\big] = 0$, this value was changed to 10^{-18} before applying the logarithm.

5.4 Evaluation of Monte Carlo integration

The setting of Monte Carlo integration – the randomized drawing of loads – is prone to committing a *differential crime*. How to compare two consecutive estimated objective values, when the MC samples are different? For the presently examined algorithm, one can set a seed for the random generator. All drawn loads are pseudo-random, i.e., given a seed all loads follow a deterministic sequence as is the case for all programmed random number generators. This deficiency of a "random" number generator will be taken advantage of here, allowing the prevention of the inherent differential crime in the Monte Carlo integration.

For the evaluation of Monte Carlo integration, $z^*_{\Delta\Sigma}$ and $z^*_{\max}$ where set up with responses from the surrogate functions $\tilde{z}_{\Delta\Sigma}$ and $\tilde{z}_{\max}$ from Eq. (5.1) with $\alpha = 0°$. In the analysis, the progression of the objective value $j_{N_{\mathrm{MC}}}$ and of the Monte Carlo error $\mathcal{E}_{\mathrm{MC}}$ are scrutinized for both the simple interpolation model set up with 16 data points for training and the adaptively enriched model with 26 load samples.

When using the Kriging model that is trained with 16 initial samples after 5000 MC draws the computed objective value is $j_{N_{\mathrm{MC}}} = 547.2$ after 5000 draws and $j_{N_{\mathrm{MC}}} = 545.7$ after 10,000 MC draws. The Monte Carlo error $\mathcal{E}_{\mathrm{MC}}$ is 1.39% when using 5000 draws. When doubling the MC draws, the error decreases to 0.989% by a factor of almost $\sqrt{2}$.

With an enriched the Kriging model, again starting with 5000 draws for Monte Carlo, the initial relative root mean squared error is 1.32% of the objective value $j_{N_{\mathrm{MC}}} = 509.6$. As previously with 10,000 draws, the resulting values are $\mathcal{E}_{\mathrm{MC}} = 0.937\%$ and $j_{N_{\mathrm{MC}}} = 509.1$. Therefore, the MC error decreases with a factor of approximately $\sqrt{2}$ and the asymptotic convergence is attained.

This analysis shows, that the factor of error reduction attains the asymptotic rate of convergence for $N_{\mathrm{MC}} = 5000$. In this case, with a given target accuracy of $\frac{\tau_h}{3}$ and a computed Monte Carlo error of $\mathcal{E}_{\mathrm{MC}}$ that is above this target accuracy, one can compute the ratio between current and target error $\mathcal{E}_{\mathrm{MC}}/(\tau_h/3)$. By setting the total MC draws to the square of this ratio times the initial MC draws

$$N_{\mathrm{MC}}^{\mathrm{total}} := \left(9\mathcal{E}_{\mathrm{MC}}^2/\tau_h^2\right) \cdot N_{\mathrm{MC}},$$

one achieves the wanted integration accuracy.

Should the number of total MC draws be too big, e.g., above one million draws, it is advisable to mistrust the long-time prognosis. In doing so, the progression of the error reduction is to be monitored and the error will be computed after a predefined multiple of draws, e.g., after every time 10^6 MC draws are added to the MC estimate. Should the error progression deviate from the anticipated path, the number $N_{\mathrm{MC}}^{\mathrm{total}}$ is adjusted accordingly.

5.5 Evaluation of the optimization

For the evaluation of the optimization, the initial implant position was first set to an angle of $\alpha_0 = 0°$, then $\alpha_0 = -1.5°$ and $\alpha_0 = 1°$ were analyzed as starting positions.

In 2D the control variable α is a scalar, therefore the storage and inversion of B_k, cf. Eq. (4.1), poses no problem. The constants τ_1 and τ_2 are set to $\tau_1 = 10^{-4}$ and $\tau_2 = 0.9$, see, e.g., [116].

For the different stopping criteria, set $\varepsilon_j = 10^{-3}$ for the objective value, $\varepsilon_{\mathrm{deriv}} = 10^{-3}$ for the derivative, $\varepsilon_{\Delta j} = 10^{-3}$ for the change in the objective function, and $\varepsilon_{\Delta\alpha} = 10^{-4}$ if the change in angle α and thus its precision becomes small enough.

Since the response functions $z_{\Delta\Sigma}$ and $z_{\Sigma_{\max}}$ computed from the PDE (2.18) proved faulty, they were exchanged with $\tilde{z}_{\Delta\Sigma}$ and $\tilde{z}_{\Sigma_{\max}}$, see Eq. (5.1), of which the derivatives with respect to α as well as $\frac{\delta}{\delta\alpha}j$ are easily computed:

$$\begin{aligned}\frac{\delta}{\delta\alpha}j(l;\alpha) = (-f_x + f_z)\big(\sin(-f_x) + \sin(f_z) + 2\big)(2\alpha + 1) \\ + bc \cdot \exp\Big(c\big(\tilde{z}_{\max}(l;\alpha) - \Sigma_{\mathrm{thres}}\big)\Big) \cdot 350 f_z(2\alpha + 2.7).\end{aligned} \tag{5.4}$$

As there is at this point no means implemented to determine the influence of the grid discretization on the accuracy of the PDE solution, the Kriging interpolation, and the Monte Carlo integration, a tolerance of $\frac{\tau_h}{3} := 8$ is defined. With this tolerance the Monte Carlo error always proved good enough and the adaptive Kriging scheme is triggered in all three investigated initial implant scenarios. For reasons elaborated at the end of this section, if $\mathcal{E}_{\mathrm{K}} > 20$, the adaptively added samples were capped at ten. When $\mathcal{E}_{\mathrm{K}} \leq 20$, the maximum of added Kriging samples was capped at 300. For demonstration purposes on the development of the Kriging error $\mathcal{E}_{\mathrm{K}}$ (3.34) this will suffice.

Initial position $\alpha_0 = 0°$. With $\alpha_0 = 0°$ as the starting position, the Alg. 3 converged in three iterations to $\alpha = -0.5000°$. The objective value j (2.17) approximated with j_{MC} (3.5) decreased from 547 to $1.96 \cdot 10^{-6}$ and the derivative (5.4), also computed with MC, decreased from $5.48 \cdot 10^3$ to 0.146. As the objective value was smaller than $\varepsilon_j = 10^{-3}$, the optimization was terminated.

In the first iteration, the initial Kriging error $\mathcal{E}_{\mathrm{K}}$ (3.34) is 1.82. Therefore, no adaptivity was needed. In the second iteration, the Kriging error started with a value of 15.4. After 300 samples were added, this error decreased to 10.7. Here, 168 samples were added with $\beta > 0.5$, i.e., with the local exploitation strategy being predominant. In the third and last iteration, the Kriging error was 21.92 in the beginning and reduced to 21.90 after ten samples were added. Of those samples, seven were added with $\beta > 0.5$. As the angle difference between α_2 and α_1 was below 0.1°, the 500 added samples from the second iteration were kept and the total Kriging samples was 526 after adaptive refinement in the third iteration.

The Monte Carlo error $\mathcal{E}_{\mathrm{MC}}$ (3.22) was 1.39% in the first iteration, therefore no adaptive increase of the drawn Monte Carlo samples was needed. With a descent direction of $-10.06°$ and because $-10.06°$ is outside the admissible angle range, the step length computation advanced directly to $-2.502°$. Afterwards the angles $-1.251°$ and $-0.6254°$ ensued to which the final angle of this iteration, $\alpha_1 = -0.3127°$, was computed. For this angle, the Armijo-Wolfe conditions (4.3) were fulfilled.

In the second and third iteration, the MC error was 1.09%, respectively. Therefore, the adaptive increase of the drawn Monte Carlo samples was unnecessary again. The descent direction in the second iteration of optimization was $-6.813° \cdot 10^{-2}$. Here, the unit step length sufficed, α_2 was set to $-0.3808°$, and the optimization went to the next loop.

The descent direction in the third iteration of optimization was $-0.1191°$. With this, again the unit step length was sufficient and the final angle $\alpha_3 = -0.5000°$ was estimated.

Initial position $\alpha_0 = -1.5°$. With an initial angle of $\alpha_0 = -1.5°$, the algorithm converged the control variable α in three iterations to $-0.5000°$. The objective value decreased from $5.48 \cdot 10^3$ to $3.82 \cdot 10^{-6}$ and the derivative from $-5.43 \cdot 10^3$ to -0.176. The algorithm stopped, since the objective value was below the tolerance $\varepsilon_j = 10^{-3}$.

The Kriging error $\mathcal{E}_{\mathrm{K}}$ was 0.519 in the first iteration. Therefore, no samples were added to the Kriging model. In the second iteration, the initial Kriging error was 6398 and reduced to 5495 after ten added samples. Of those ten samples, eight were added with $\beta > 0.5$. For the third iteration, the error of Kriging interpolation started at $1.01 \cdot 10^5$ and reduced to $9.28 \cdot 10^4$ with ten samples added, of which eight samples were chosen with the local exploitation strategy.

In the first iteration of the optimization, the descent direction was 0.9908°, such that the unit step length sufficed to reach the next iterated angle of $\alpha_1 = -0.5092°$. The Monte Carlo error $\mathcal{E}_{\mathrm{MC}}$ was 0.53% at $\alpha_0 = -1.5°$.

In the second iteration, the descent direction computed to $6.950° \cdot 10^{-3}$. Here, as well as in the third iteration with descent direction $2.201° \cdot 10^{-3}$, the unit step length was sufficient, estimating $\alpha_2 = -0.5022°$ and $\alpha_3 = -0.5000°$. The Monte Carlo error $\mathcal{E}_{\mathrm{MC}}$ was 1.31% in the second iteration and 1.32% in the third iteration. The optimization finished, since the objective value became small enough.

Initial position $\alpha_0 = 1°$. With an initial angle of $\alpha_0 = 1°$ the algorithm converged to $\alpha = -0.4999°$ in four iterations with a small change to the algorithm elaborated below. The initial objective value of $4.78 \cdot 10^{189}$ was reduced to $3.29 \cdot 10^{-5}$ and the derivative decreased from $3.13 \cdot 10^{192}$ to 0.515. Eventually, the optimization stopped due to the objective value being below the target precision $\varepsilon_j = 10^{-3}$.

Here, with the choice of the penalty function (2.14), its parameters $b = 10^{-15}$, $c = 10^{-4}$, and $\Sigma_{\text{thres}} = 2.9 \cdot 10^6$ and the surrogate function $\tilde{z}_{\max}$ in Eq. (5.1), the starting value of the objective is in $\mathcal{O}\left(10^{189}\right)$ and the derivative value in $\mathcal{O}\left(10^{192}\right)$ which are both extremely high. The reason behind magnitude of these values as well as the incomputability of the MC error will be discussed at the end of this section.

Since through the updated approximate Hessian matrix B_{k+1}, see Eq. (4.4), the last derivative value $j'(\alpha_k)$ influences the new direction d_{k+1}. Without any changes to the algorithm, the descent direction d_{k+1} in the second iteration had a magnitude of $\mathcal{O}\left(10^{-190}\right)$ which is far below the tolerance of $\varepsilon_{\Delta\alpha} = 10^{-4}$. Therefore, the algorithm would stop in a suboptimal implant position. To prevent this situation, another check was added before line 20 of Alg. 3. If $j'(\alpha_k) > 10^{10}$ then $B_{k+1} = |j(\alpha_{k+1})|I$, similar to the definition of B_0. With this adaptation to the algorithm, the optimization was able to perform three iterations rather than only one.

In the first iteration from α_0 to α_1, the Kriging error was $3.81 \cdot 10^{-187}$. In the second iteration, the initial Kriging error of 11.17 decreased to 7.867 after adding 47 samples. Of those 47 samples, 29 were added with the global exploration strategy being predominant. In the third iteration that determined α_3, the initial Kriging error was 88.62 which was reduced to 76.11 after ten adaptively added samples. Of those, eight had a value $\beta > 0.5$. In the final iteration α_4 was determined and the Kriging error was 2962 in the beginning. As $|\alpha_4 - \alpha_3| < 0.1°$ held, in a real case scenario the implant would be replaced in the bone using rigid body transformation and the model refinements from the previous iteration are kept. Thus, this iteration started with 26 Kriging samples and ten more were added. The final Kriging error was $\mathcal{E}_{\text{K}} = 2726$ with 36 training data. Again, eight were added with $\beta > 0.5$.

For the initial implant position of $\alpha_0 = 1°$, the MC error was incomputable due to infinitely high penalty values. With a descent direction of $-655.2°$, the step length was estimated in two iterations. The determined angles were $\alpha = -1.559°$ and then $\alpha_1 = -0.2796°$. The determination of the step length ended here and the optimization went into the second loop. In the second iteration of the optimization, where the MC error was 1.12%, the computed descent direction was $-9.106°$. The step length was determined after five iterations and the final angle in this iteration is $\alpha_2 = -0.4219°$. It is to note, that objective value for the first admissible angle $-2.556°$ in the step length estimation was in $\mathcal{O}(10^4)$, which indicates a steep rise of the penalty function taking effect here. In the third iteration of the optimization, the descent direction was $-6.503° \cdot 10^{-2}$. Thus, with a step length of 1 the second to last angle of $\alpha_3 = -0.4869°$ was reached. The MC error accounted for 1.31%. In the final iteration, the descent direction was $-1.297° \cdot 10^{-2}$ and the MC error was 1.32%. Again, the unit step length was accepted. Since the objective value was below the tolerance of 10^{-3}, the optimization stopped here.

Conclusion of the optimization. It is apparent, that the implant is for $\alpha \geq 0.5°$, but also for $\alpha < -3°$, in an inadmissible position concerning the maximum stresses, which is therefore punished by the penalty function p, Eq. (2.14). With objective and derivative values as high as $\mathcal{O}\left(10^{86}\right)$ or higher for $\alpha > 0.5°$, one realizes, that the penalization of high local stresses in these cases needs to be enforced with an inequality constraint rather than the penalty function. The use of a hybrid from of the penalty function and an inequality constraint should be considered as well. That is, use the penalty function for $p'(\alpha) < 10^{10}$ and switch to using an inequality constraint otherwise.

For the given surrogate functions $\tilde{z}_{\Delta\Sigma}$ and $\tilde{z}_{\max}$ and the prescribed interpolation and integration algorithms used in Alg. 3, the optimization algorithm proved to converge to the minimum of $j(\alpha)$ within a few steps. With the penalty function p and surrogate function $\tilde{z}_{\max}$ as defined, issues in computing MC error and concerning termination of the optimization in suboptimal implant positions were encountered. This could be circumvented here with a slight adjustment of the algorithm, as described in the previous paragraph. This allowed the algorithm to deal with the very high values produced by the penalty function for unfavorable implant positions α. However, this would not work, if the objective and derivative values were altogether incomputable.

As was seen in the second iteration of optimization for $\alpha_0 = 1°$, the number of adaptively added samples to the Kriging model was moderate, while in the second iteration of optimization for $\alpha_0 = 0°$, the number of needed samples became high. For Kriging errors above 20, the necessary samples to achieve a target precision below 8 would have been extremely high. For this reason the number of added samples was capped at ten in these instances. On the other hand, the numerical analysis showed that the optimization was not inhibited due to this limit. Therefore one can conclude, also considering the fact that the Monte Carlo error was in $\mathcal{O}\left(10^{-2}\right)$ or of even lower order, that the error equilibration introduced at the end of Sec. 3.4 needs refinement.

Since the precision of Monte Carlo integration is more cheaply improved than that of the Kriging interpolation, and with the MC error being low to start with, the tolerance applied here can be reduced. Contrary, what with the optimization converging even if the Kriging interpolation has not reached the target accuracy, one can handle the tolerance here more freely. The biggest improvement of the Kriging error was investigated for the first 10-40 adaptively added samples. The gain in precision can theoretically be quantized with a gain-effort model which then allows an estimation of optimal distribution of the tolerance τ_h to the three processes of PDE-solution with $\mathcal{E}_{\mathrm{PDE}}$, the Kriging interpolation with $\mathcal{E}_{\mathrm{K}}$, and the Monte Carlo integration with $\mathcal{E}_{\mathrm{MC}}$. Approaches to such models are covered in Deuflhard and Weiser [42].

Incidentally, if one set $\frac{\tau_h}{3} = 1.73$ for $\alpha_0 = 0°$, 25 Kriging samples are added in the first iteration. With this, the descent direction computes to $-4°$ and with three step length adjustments the final angle of $-0.4998°$ is reached in one optimization step. This underlines the importance of an adequate tolerance estimation. A higher precision in the early stages may save considerable computation time.

The successful operation of both, the computed derivative in Eq. (4.14) and the special line search devised in Sec. 4.2, could not yet be analyzed, due to the technical difficulties explained in this section. Besides that, the above numerical experiments showed the successful operation of Alg. 3 and the algorithms used within.

6 Summary

In this thesis, adaptive algorithms in optimization under PDE constraints have been investigated. In its application, the aim of optimization is to increase the longevity of implants, namely the hip joint implant, and in doing so to minimize stress shielding and simultaneously minimize the influence of locally high stresses, that, above a threshold value, are malign to the bone structure. Under the constraint of the equilibrium of forces, describing an elastodynamic setup, coupled with a contact inequality condition, a computationally expensive problem formulation is given.

The first step to make the solution of the given problem possible and efficient was to change over to the spatial equilibrium equation, thus rendering an elastostatic setup. Subsequently the intrinsically dynamic motions – trajectories in the load domain – were converted to the static setup. Thus, the trajectories are marginalized to the load domain and characterized with probability distributions. Therefore the solving of the PDE constraint, the contact problem, is simplified.

Yet in the whole optimization process, the solving of the PDE, the spatial equilibrium equation (2.18a)–(2.18c) together with the contact condition (2.18d) has the most expensive contribution still and hence needed further reduction. This was achieved by application of Kriging interpolation to the load responses of the integrated distribution of stress difference and the maximum stresses. The interpolation of the two response surfaces only needs comparatively few PDE solves to set up the models. Moreover, the Kriging models can be adaptively extended by sequentially adding sample-response pairs. For this the Kriging inherent variance is used to estimate ideal new sample locations with maximum variance values. In doing so, the overall interpolation variance and therefore the interpolation error is reduced.

For the integration of the integrated stress differences and penalty values on the relative high dimensional load domain Monte Carlo integration was implemented, averting the curse of dimension. Here, the motion's probability distribution combined with patient specific data of motion frequencies is taken advantage of, making obsolete the use of the otherwise necessary importance sampling.

Throughout the optimization, the FE-discretization error and the subsequently attached errors entering the solution process via PDE discretization and approximative solving of the PDE, Kriging interpolation and Monte Carlo integration need to decrease. While the FE-discretization error and the solution of the elastostatic contact problem were assumed precise enough, numerics showed, that the interpolation and integration errors can be controlled by adaptive refinement of the respective methods. For this purpose comparable error quantities for the particular algorithms were introduced and effectively put to use.

For the implant position's optimization, the derivative of the objective function was derived using the implicit function theorem. As the FE-discretization changes with implant position modifications big enough, a special line search had to be used to deal with the discontinuities in the objective function.

The interplay and performance of the subalgorithms was demonstrated numerically on a reduced 2D setup of a hip joint with and without the implant. Consequently the load domain and the control variable were also limited to the 2D case.

Outlook. The closure of the thesis is dedicated to the exploration of model refinements and extensions.

First of all, the subtle bugs in the implementation mentioned in Sec. 5.1 need to be eliminated, such that the optimization with its subalgorithms, but also the analytic derivative of the objective function with respect to the control variable $\boldsymbol{\alpha}$, Eq. (4.14), and the special line search introduced in Sec. 4.2 can be tested.

Then, the truncated errors of FE-discretization and the PDE solution need to be included and their influence on the subsequent errors investigated properly.

Further, the optimization needs to be extended to a 3D setting. Here, the memory demand and computation effort increase. This demand can be countered by a reduced model approach, for example, by either employing a reduced model for 3D deformable objects obeying the St. Venant-Kirchhoff material law as investigated in [8], or by using a POD model reduction ansatz as, e.g., researched by Janon et al. [84]. They present a POD approach for a Monte Carlo estimator, applying a reduced-basis method and deriving a goal-oriented error estimator which could also be applied in adaptivity. On the other hand, the application of sparse grids combined with Monte Carlo as in Bungartz et al. [24] can be further investigated. Sparse grids also offer for adaptivity.

The remeshing on hip implant repositioning may be refined using results from Zilske et al. [181].

The line search approach in optimization could be exchanged for trust-region to see if there are better runtime- and convergence results.

Moreover, multilevel optimization could be added in either the solution of the PDE or the optimization itself. This can be in the form of either a preconditioner, which is favorable in both contexts to either quicker estimate a solution to the contact problem or to find a good starting position of the implant, thus having superlinear convergence of the BFGS algorithm once it is applied; or multilevel optimization could replace the semismooth Newton method in the solving of the contact problem, see, e.g., [69, 91, 92, 169], or it could replace the BFGS algorithm for optimization itself.

Once turning to real-life application of the presented optimization, the modeling of the involved joints and implants needs to become more realistic. Modeling the bone and implant material in more detail, i.e., incorporating anisotropies as compact and cancellous bone, one may incorporate multi-material mesh generation as it was investigated in [86].

The inclusion of patient specific bone girds through statistical shape meshes [97, 139] should also be considered, thus setting up a wholly pipeline for use in orthopedic surgery.

Bibliography

[1] P. Abrahamsen. *A Review of Gaussian Random Fields and Correlation Functions*. Norsk Regnesentral/Norwegian Computing Center, 1997.

[2] W. Alt. *Nichtlineare Optimierung: Eine Einführung in Theorie, Verfahren und Anwendungen*. vieweg studium; Aufbaukurs Mathematik. Vieweg+Teubner Verlag, 2013.

[3] H. Altenbach. *Kontinuumsmechanik*. Springer Vieweg, 2nd edition, 2012.

[4] A. E. Anderson, B. J. Ellis, S. A. Maas, C. L. Peters, and J. A. Weiss. Validation of finite element predictions of cartilage contact pressure in the human hip joint. *J Biomech Eng*, 130(5), 2008.

[5] L.-E. Andersson. A quasistatic frictional problem with normal compliance. *Nonlinear Analysis: Theory, Methods & Applications*, 16(4):347–369, 1991.

[6] P. F. Antonietti, C. Marcati, I. Mazzieri, and A. Quarteroni. High order discontinuous Galerkin methods on simplicial elements for the elastodynamics equation. *Numerical Algorithms*, 71(1):181–206, 2016.

[7] O. M. Badr, F. Barlat, B. Rolfe, M.-G. Lee, P. Hodgson, and M. Weiss. Constitutive modelling of high strength titanium alloy Ti-6Al-4V for sheet forming applications at room temperature. *International Journal of Solids and Structures*, 80:334–3470, 2016.

[8] J. Barbič and D. James. Real-time subspace integration for St. Venant-Kirchhoff deformable models. *ACM Trans. Graph.*, 24:982–990, 2005.

[9] J. Barrett and C. Elliott. Finite Element Approximation of the Dirichlet Problem Using the Boundary Penalty Method. *Numerische Mathematik*, 1986.

[10] R. Becker, H. Kapp, and R. Rannacher. Adaptive finite element methods for optimal control of partial differential equations: Basic concept. *SIAM J. Control and Optimization*, 39:113–132, 2000.

[11] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica*, 10:1–102, 2001.

[12] M. Benzi, G. Golub, and J. Liesen. Numerical solution of saddle point problems. *Acta Numerica*, 14:1–137, 2005.

[13] G. Bergmann, A. Bender, J. Dymke, G. Duda, and P. Damm. Standardized loads acting in hip implants. *PLoS One*, 11(5), 2016.

[14] G. Bergmann, F. Graichen, A. Rohlmann, A. Bender, B. Heinlein, G. Duda, M. O. Heller, and M. M. Morlock. Realistic loads for testing hip implants. *Biomed Mater Eng.*, 20, 2010.

[15] H. Bobzin, A. Neumann, and R. Tremmel. Qualitätsreport 2015. Technical report, Institut für Qualitätssicherung und Transparenz im Gesundheitsweisen (i.A. Gemeinsamer Bundesausschuss), 2016.

[16] P. Boieri, F. Gastaldi, and D. Kinderlehrer. Existence, uniqueness, and regularity results for the two-body contact problem. *Applied Mathematics and Optimization*, 15(1):251–277, Jan 1987.

[17] J. Bonet and R. D. Wood. *Nonlinear Continuum Mechanics for Finite Element Analysis*. Cambridge University Press, 2008.

[18] M. Bram. *Pulvermetallurgische Herstellung von porösem Titan und von NiTi-Legierungen für biomedizinische Anwendungen*. Priv. doz., Ruhr Univ. Bochum, 2012. Ruhr Univ. Bochum, Habil., 2013.

[19] C. Brandenburg, F. Lindemann, M. Ulbrich, and S. Ulbrich. *A Continuous Adjoint Approach to Shape Optimization for Navier Stokes Flow*, volume 158, pages 35–56. Birkhäuser Verlag Basel, 2009.

[20] B. Braun. Knie- und Hüft-(Total-) Endoprothesen 2008 bis 2012. *hkk Gesundheitsreport*, 2013.

[21] S. C. Brenner and L. R. Scott. *The Mathematical Theory of Finite Element Methods*, volume 15 of *Texts in Applied Mathematics*. Springer, 2008.

[22] J. Brousek, P. Frankova, M. Hanuš, H. Kopincová, R. Kuzel, R. Tezaur, P. Vanek, and Z. Vastl. An overview of multilevel methods with aggressive coarsening and massive polynomial smoothing. *Electronic transactions on numerical analysis ETNA*, 44:401–442, 2015.

[23] H.-J. Bungartz and S. Dirnstorfer. Multivariate quadrature on adaptive sparse grids. *Computing*, 71(1), 2003.

[24] H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.

[25] A. Capatina. *Variational Inequalities and Frictional Contact Problems*, volume 31. Springer, Cham, 2014.

[26] A. Capatina. A quasistatic frictional contact problem with normal compliance and unilateral constraint. *Mathematics and Mechanics of Solids*, 22(11):2135–2155, 2017.

[27] C. Carstensen and J. Hu. An optimal adaptive finite element method for an obstacle problem. *Computational Methods in Applied Mathematics*, 15(3), 2015.

[28] C. Charbonnier, L. Assassi, P. Volino, and N. Magnenat-Thalmann. Motion study of the hip joint in extreme postures. *The Visual Computer*, 25(9):873–882, 2009.

[29] C. Charbonnier, S. Chagué, M. Ponzoni, M. Bernardoni, P. Hoffmeyer, and P. Christofilopoulos. Sexual activity after total hip arthroplasty: A motion capture study. *The Journal of Arthroplasty*, 29(3):640–647, 2014.

[30] C. Charbonnier, S. Chagué, J. Schmid, F. C. Kolo, M. Bernardoni, and P. Christifilopoulos. Analysis of hip range of motion in everyday life: a pilot study. *Hip Int*, 25(1):82–90, 2015.

[31] S. Chaturantabut and D. Sorensen. Application of POD and DEIM on dimension reduction of non-linear miscible viscous fingering in porous media. *Mathematical and Computer Modelling of Dynamical Systems*, 17:337–353, 2011.

[32] A. Chkifa, A. Cohen, and C. Schwab. High-dimensional adaptive sparse polynomial interpolation and applications to parametric PDEs. *Foundations of Computational Mathematics*, 14(4):601–633, 2014.

[33] P. G. Ciarlet. *Finite Element Method for Elliptic Problems*. Elsevier Science, 1978.

[34] P. G. Ciarlet. *Mathematical Elasticity, Volume I: Three-Dimensional Elasticity*. Studies in Mathematics and its Applications, 1988.

[35] M. Cilla, E. Borgiani, J. Martínez, G. N. Duda, and S. Checa. Machine learning techniques for the optimization of joint replacements: Application to a short-stem hip implant. *PLoS One*, 12(9), 2017.

[36] K. Colic, A. Sedmak, A. Grbovic, U. Tatic, S. Sedmak, and B. Djordjevic. Finite element modeling of hip implant static loading. *Procedia Engineering*, 149:257–262, 2016.

[37] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust Region Methods*. Society for Industrial and Applied Mathematics, 2000.

[38] G. E. Cook, S. Samiezadeh, Z. Marison, M. S. Aziz, H. Bougherara, R. Zdero, and E. H. Schemitsch. Biomechanical optimization of the angle and position for surgical implantation of the depuy silent hip implant. *CMBES Proceedings*, 37, 2014.

[39] P. Damm, S. Reitmaier, S. Hahn, V. Waldheim, A. Firouzabadi, and H. Schmidt. In vivo hip and lumbar spine implant loads during activities in forward bent postures. *Journal of Biomechanics*, 102, 2020. 3rd International Workshop on Spine Loading and Deformation.

[40] P. Deuflhard and A. Hohmann. *Numerical Analysis in Modern Scientific Computing*. De Gruyter, 4th edition, 2008. German title: Numerische Mathematik 1: Eine algorithmisch orientierte Einführung.

[41] P. Deuflhard, R. Krause, and S. Ertel. A contact-stabilized Newmark method for dynamical contact problems. *International Journal for Numerical Methods in Engineering*, 73:1274–1290, 2008.

[42] P. Deuflhard and M. Weiser. *Adaptive Numerical Solution of PDEs*. De Gruyter, 2011. German title: Numerische Mathematik 3: Adaptive Lösung partieller Differentialgleichungen.

[43] C. Dick. *Computational Steering for Implant Planning in Orthopedics*. PhD thesis, Technische Universität München, 2012.

[44] C. R. Dietrich and M. R. Osborne. Estimation of covariance parameters in kriging via restricted maximum likelihood. *Mathematical Geology*, 23:119–135, 1991.

[45] T. Driscoll and A. Heryudono. Adaptive residual subsampling methods for radial basis function interpolation and collocation problems. *Computers & Mathematics with Applications*, 53:927–939, 2007.

[46] C. Eck. *Existenz und Regularität der Lösungen für Kontaktprobleme mit Reibung*. PhD thesis, Universität Stuttgart, 1996.

[47] H. Effenberger. Hüftendoprothetik: Konstruktion, Klassifikation, Implantate, Ergebnisse. https://docplayer.org/17409608-Hueftendoprothetik-konstruktion-klassifikation-implantate-ergebnisse-h-effenberger.html (last checked: 2020-11-09), 2007.

[48] I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. Classics in Applied Mathematics. Society for Industrial and Applied Mathematics, 1999.

[49] V. Elvira, L. Martino, and C. P. Robert. Rethinking the effective sample size. arXiv:1809.04129, 2018.

[50] J. B. Erway, P. E. Gill, and J. D. Griffin. Iterative methods for finding a trust-region step. *SIAM Journal on Optimization*, 20(2):1110–1131, 2009.

[51] P. R. Fernandes, H. Rodrigues, and C. R. Jacobs. A model of bone adaptation using a global optimisation criterion based on the trajectorial theory of Wolff. *Computer Methods in Biomechanics and Biomedical Engineering*, 2(2):125–138, 1999.

[52] E. Fonseca, M. Lima, and L. Barreira. Human femur assessment using isotropic and orthotropic materials dependent of bone density. In *Proceedings of IRF 2009 - 3rd International Conference on Integrity, Reliability and Failure*, pages 20–24, 2009.

[53] J. Forsberg and L. Nilsson. On polynomial response surfaces and Kriging for use in structural optimization of crashworthiness. *Structural and Multidisciplinary Optimization*, 29(3):232–243, 2005.

[54] S. Foss, D. Korshunov, and S. Zachary. *An introduction to heavy-tailed and subexponential distributions*. Springer, New York, NY, 2013.

[55] M. Fraldi, L. Esposito, G. Perrella, A. Cutolo, and S. C. Cowin. Topological optimization in hip prosthesis design. *Biomechanics and Modeling in Mechanobiology*, 9(4):389–402, 2010.

[56] P. Fratzl. *Collagen: Structure and Mechanics, an Introduction*. Springer, 2008.

[57] F. Galloway, M. Kahnt, H. Ramm, P. Worsley, S. Zachow, P. Nair, and M. Taylor. A large scale finite element study of a cementless osseointegrated tibial tray. *Journal of Biomechanics*, 46(11):1900–1906, 2013.

[58] D. Y. Gao and E. Hajilarov. *Analytic Solutions to 3-D Finite Deformation Problems Governed by St Venant–Kirchhoff Material*, pages 69–88. Springer International Publishing, Cham, 2017. Ed. David Yang Gao and Vittorio Latorre and Ning Ruan.

[59] T. Gerstner and M. Griebel. Sparse grids. In R. Cont, editor, *Encyclopedia of Quantitative Finance*. John Wiley and Sons, 2010.

[60] F. Ghavamian, P. Tiso, and A. Simone. POD-DEIM model order reduction for strain softening viscoplasticity. *Computer Methods in Applied Mechanics and Engineering*, 317:458–479, 2016.

[61] D. Ginsbourger, D. Dupuy, A. Badea, L. Carraro, and O. Roustant. A note on the choice and the estimation of Kriging models for the analysis of computer experiments. *Applied Stochastic Models in Business and Industry*, 25, 2009.

[62] A. A. Giunta and L. T. Watson. A comparison of approximation modeling techniques: Polynomial versus interpolating models. Technical report, American Institute of Aeronautics and Astronautics, 1998.

[63] G. H. Golub and C. F. van Loan. *Matrix Computations*. JHU Press, 4th edition, 2013.

[64] S. Götschel. *Adaptive Lossy Trajectory Compression for Optimal Control of Parabolic PDEs*. PhD thesis, Freie Universität Berlin, 2015.

[65] S. Götschel, N. Chamakuri, K. Kunisch, and M. Weiser. Lossy compression in optimal control of cardiac defibrillation. *J. Sci. Comput.*, 60(1):35–59, 2014.

[66] S. Götschel, A. Schiela, and M. Weiser. Kaskade 7 – a flexible finite element toolbox. *Computers and Mathematics with Applications*, 2020.

[67] S. Götschel and M. Weiser. Lossy compression for PDE-constrained optimization: Adaptive error control. *Comput. Optim. Appl.*, 62(1):131–155, 2015.

[68] S. Götschel, M. Weiser, and A. Schiela. Solving optimal control problems with the Kaskade 7 finite element toolbox. In A. Dedner, B. Flemisch, and R. Klöfkorn, editors, *Advances in DUNE*, pages 101–112. Springer, 2012.

[69] C. Gräser and R. Kornhuber. Multigrid methods for obstacle problems. *International Journal of Computer Mathematics*, 27:1–44, 2009.

[70] C. Gräser, U. Sack, and O. Sander. Truncated nonsmooth Newton multigrid methods for convex minimization problems. In M. Bercovier, M. J. Gander, R. Kornhuber, and O. Widlund, editors, *Domain Decomposition Methods in Science and Engineering XVIII*, pages 129–136, Berlin, Heidelberg, 2009. Springer.

[71] S. Gratton, A. Sartenaer, and P. Toint. Recursive trust-region methods for multiscale nonlinear optimization. *SIAM Journal on Optimization*, 19:414–444, 2008.

[72] A. Griewank and A. Walther. Treeverse: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math. Software*, 26:200–0, 1997.

[73] A. Grimberg, V. Jansson, O. Melsheimer, and A. Streinbrück. Jahresbericht 2019 – Mit Sicherheit mehr Qualität. Technical report, EPRD Deutsche Endoprothesenregister gGmbH, 2019.

[74] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*. Society for Industrial and Applied Mathematics, 2011.

[75] H. Y. He and A. B. Owen. Optimal mixture weights in multiple importance sampling. arXiv:1411.3954, 2014.

[76] H. Heger. *Modellierung der viskoelastischen Eigenschaften des menschlichen Hüftgelenkes*. PhD thesis, Friedrich-Schiller-Universität Jena, 2007.

[77] S. Herrmann, D. Kluess, M. Kaehler, R. Grawe, R. Rachholz, R. Souffrant, J. Zierath, R. Bader, and C. Woernle. A novel approach for dynamic testing of total hip dislocation under physiological conditions. *PloS One*, 10, 12 2015.

[78] M. Hintermüller and M. Ulbrich. A mesh-independence result for semismooth Newton methods. *Mathematical Programming*, 101(1):151–184, 2004.

[79] M. Hinze and S. Volkwein. Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: Error estimates and suboptimal control. In P. Benner, D. C. Sorensen, and V. Mehrmann, editors, *Dimension Reduction of Large-Scale Systems*, pages 261–306, Berlin, Heidelberg, 2005. Springer.

[80] C. P. Ho, M. Kocvara, and P. Parpas. Newton-type multilevel optimization method, 2019.

[81] C. P. Ho and P. Parpas. Multilevel optimization methods: Convergence and problem structure, 2016.

[82] K. Ito and K. Kunisch. The augmented Lagrangian method for equality and inequality constraints in Hilbert spaces. *Math. Program.*, 46(3):341–360, 1990.

[83] E. Iuliano. Efficient design optimization assisted by sequential surrogate models. *International Journal of Aerospace Engineering*, 2019:34 pages, 2019.

[84] A. Janon, M. Nodet, and C. Prieur. Goal-oriented error estimation for the reduced basis method, with application to sensitivity analysis. *Journal of Scientific Computing*, 68, 2013.

[85] Y. Kagiyama, I. Otomaru, F. Yokota, T. Okada, M. Nakamoto, M. Takao, N. Sugano, Y. Tada, N. Tomiyama, and Y. Sato. Optimization of implant compatibility and hip joint function based on statistical atlas for automated 3D surgical planning of total hip arthroplasty. In *9th World Congress on Structural and Multidisciplinary Optimization*, 2011.

[86] M. Kahnt, H. Ramm, H. Lamecker, and S. Zachow. Feature-preserving, multimaterial mesh generation using hierarchical oracles. In J. A. Levine, R. R. Paulsen, and Y. Zhang, editors, *Proc. MICCAI Workshop on Mesh Processing in Medical Image Analysis (MeshMed)*, volume 7599, pages 101–111, 2012.

[87] N. Kikuchi and J. T. Oden. *Contact Problem in Elasticity*. SIAM, 1988.

[88] I. M. Kiš. Comparison of Ordinary and Universal Kriging interpolation techniques on a depth variable (a case of linear spatial trend), case study of the Šandrovac Field. *The Mining-Geology-Petroleum Engineering Bulletin*, 31(2):41–58, 2016.

[89] C. Klapproth. *Adaptive numerical integration for dynamical contact problems*. PhD thesis, Freie Universität Berlin, 2012.

[90] R. Kornhuber and R. Krause. Adaptive multigrid methods for Signorini's problem in linear elasticity. *Computing and Visualization in Science*, 4:9–20, 11 2001.

[91] R. Kornhuber, R. Krause, O. Sander, P. Deuflhard, and S. Ertel. A monotone multigrid solver for two body contact problems in biomechanics. *Computing and Visualization in Science*, 11:3–15, 12 2007.

[92] R. Kornhuber, C. Schwab, and M.-W. Wolf. Multilevel Monte Carlo finite element methods for stochastic elliptic variational inequalities. *SIAM Journal on Numerical Analysis*, 52:1243–1268, 2014.

[93] D. P. Kouri, M. Heinkenschloss, D. Ridzal, and B. G. van Bloemen Waanders. A trust-region algorithm with adaptive stochastic collocation for PDE optimization under uncertainty. *SIAM Journal on Scientific Computing*, 35(4):A1847–A1879, 2013.

[94] D. P. Kouri and D. Ridzal. *Inexact Trust-Region Methods for PDE-Constrained Optimization*, pages 83–121. Springer New York, 2018.

[95] D. G. Krige. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *J. of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, 1951.

[96] H. Lamecker. *Variational and statistical shape modeling for 3D geometry reconstruction.* PhD thesis, Freie Universität Berlin, 2008.

[97] H. Lamecker, T. Mansi, J. Relan, F. Billet, M. Sermesant, N. Ayache, and H. Delingette. Adaptive tetrahedral meshing for personalized cardiac simulations. In *Proc. MICCAI Workshop on Cardiovascular Interventional Imaging and Biophysical Modelling (CI2BM)*, pages 149–158, 2009.

[98] T. A. Laursen. *Computational Contact and Impact Mechanics.* Springer, 2003.

[99] R. Lewis and S. Nash. Using inexact gradients in a multilevel optimization algorithm. *Computational Optimization and Applications*, 56, 2013.

[100] H. Liu, J. Cai, and Y.-S. Ong. An adaptive approach for Kriging metamodeling by maximizing expected prediction error. *Comp. and Chem. Engin.*, 106:171–182, 2017.

[101] L. Lubkoll. Optimal control in implant shape design. Master's thesis, ZIB, TU Berlin, 2011.

[102] L. Lubkoll. *An Optimal Control Approach to Implant Shape Design: Modeling, Analysis and Numerics.* PhD thesis, Universität Bayreuth, 2015.

[103] L. Lubkoll, A. Schiela, and M. Weiser. An optimal control problem in polyconvex hyperelasticity. *SIAM J. Control. Optim.*, 52:1403–1422, 2014.

[104] L. Lubkoll, A. Schiela, and M. Weiser. An affine covariant composite step method for optimization with PDEs as equality constraints. *Optimization Methods and Software*, 32:1132–1161, 2017.

[105] Y.-T. Luo and Z. Chen. Elasticity and Anisotropy of Titanium Oxide Ti_xO_y:. *MATEC Web of Conferences*, 67:06014, 2016.

[106] B. P. Marchant and R. M. Lark. Robust estimation of the variogram by residual maximum likelihood. *Geoderma*, 140:62–72, 2007.

[107] G. Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, 1963.

[108] G. Matheron. *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature.* Masson, 1965.

[109] A. Meister. *Numerik linearer Gleichungssysteme.* Vieweg, 3rd edition, 2008.

[110] D. Meusch. Beweg Dich, Deutschland! – TK-Studie zum Bewegungsverhalten der Menschen in Deutschland. Technical report, Techniker Krankenkasse, 2013.

[111] S. Mishra and C. Schwab. Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. *Math. Comput.*, 81:1979–2018, 2012.

[112] K. Nagarchi and S. H. Saheb. Osteometric study of human femur. *International Journal of Research in Medical Sciences*, 2:104–107, 2014.

[113] S. G. Nash. A multigrid approach to discretized optimization problems. *Optimization Methods & Software*, 14:99–116, 2000.

[114] S. G. Nash. Properties of a class of multilevel optimization algorithms for equality-constrained problems. *Optimization Methods and Software*, 29(1):137–159, 2014.

[115] M. E. Nixon, C. Oana, and R. A. Lebensohn. Anisotropic response of high-purity α-titanium: Experimental characterization and constitutive modeling. *International Journal of Plasticity*, 26(4):516–532, 2010.

[116] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2006.

[117] C. Oldani and A. Dominguez. *Titanium as a Biomaterial for Implants*, chapter 9. IntechOpen, 2012.

[118] A. B. Owen. Monte Carlo theory, methods and examples. https://statweb.stanford.edu/ owen/mc/ (last checked: 2020-11-09), 2013.

[119] A. B. Owen and Y. Zhou. Safe and effective importance sampling. *Journal of the American Statistical Association*, 95(449):135–143, 2000.

[120] S. Pal. *Design of Artificial Human Joints & Organs*, pages 75–100. Springer, Boston, MA, 2014.

[121] E. Pardo-Igúzquiza and P. Dowd. The second-order stationary universal Kriging model revisited. *Mathematical Geology*, 30:347–378, 1998.

[122] K. B. Petersen and M. S. Pedersen. The matrix cookbook, 2012. https://www.math.uwaterloo.ca/ hwolkowi/matrixcookbook.pdf (last checked: 2020-11-09).

[123] N. Pop, L. Vladareanu, and P. Pop. Finite element analysis of quasistatic frictional contact problems with an incremental-iterative algorithm. In K. Gao, G. Kouzaev, and L. Vladareanu, editors, *Advanced Applications of Electrical Engineering*, pages 173–178, 2009. Proceedings of the 8th International Conference on Applications of Electrical Engineering.

[124] A. Porter. Analyses of measurements taken from adult femurs of a british population. *International Journal of Osteoarcheology*, 5:305–323, 1995.

[125] R. Putz and R. Pabst. *Sobotta, Anatomie des Menschen – Der komplette Atlas in einem Band*. Elsevier – Urban & Fischer, 22nd edition, 2007.

[126] C. Quental, P. R. Fernandes, J. Monteiro, and J. Folgado. Bone remodelling of the scapula after a total shoulder arthroplasty. *Biomechanics and Modeling in Mechanobiology*, 13(4):827–838, 2014.

[127] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, 2006.

[128] R. Rocca and M. Cocou. Numerical analysis of quasi-static unilateral contact problems with local friction. *Siam Journal on Numerical Analysis*, 39, 2001.

[129] R. T. Rockafellar. Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming. *SIAM Journal on Control*, 12(2):268–285, 1974.

[130] A. Rösch and D. Wachsmuth. Semi-smooth Newton method for an optimal control problem with control and mixed control-state constraints. *Optimization Methods and Software*, 26(2):169–186, 2011.

[131] J. Rösler, H. Harders, and M. Bäker. *Mechanisches Verhalten der Werkstoffe*. Vieweg + Teubner, 2016.

[132] G. Sagnol, H.-C. Hege, and M. Weiser. Using sparse kernels to design computer experiments with tunable precision. In *22nd Intern. Conf. on Computational Statistics - COMPSTAT 2016, Oviedo, Spain, 23-26 August 2016, Proceedings ISBN 978-90-73592-36-0*, pages 397–408, 2016.

[133] S. Sakata and F. Ashida. Integral estimation with the ordinary Kriging method using the Gaussian semivariogram function. *International Journal for Numerical Methods in Biomedical Engineering*, 27:1235–1251, 2011.

[134] O. Sander. *Multidimensional Coupling in a Human Knee Model*. PhD thesis, Freie Universiät Berlin, 2008.

[135] O. Sander, C. Klapproth, J. Youett, R. Kornhuber, and P. Deuflhard. Towards an efficient numerical solution of complex 3D knee joint motion. *Comput. Vis. Sci.*, 16(3):119–138, 2013.

[136] E. Saputra, I. B. Anwar, J. Jamari, and E. van der Heide. Finite element analysis of artificial hip joint movement during human activities. *Procedia Engineering*, 68:102–108, 2014.

[137] A. Schiela and A. Günther. An interior point algorithm with inexact step computation in function space for state constrained optimal control. *Numerische Mathematik*, 119:373–407, 2011.

[138] H. R. Schwarz and N. Köckler. *Numerische Mathematik*. Vieweg+Teubner, 7th edition, 2009.

[139] H. Seim, D. Kainmueller, M. Heller, H. Lamecker, S. Zachow, and H.-C. Hege. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. In *Eurographics Workshop on Visual Computing for Biomedicine (VCBM)*, pages 93–100, 2008.

[140] A. Serna and C. Bucher. Advanced surrogate models for multidisciplinary design optimization. In *6th Optimization and Stochastic Days*, 2009.

[141] A. Shapiro and T. Homem-de-Mello. On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs. *Siam Journal on Optimization - SIAMJO*, 11:70–86, 2000.

[142] J. R. Shewchuk. Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In M. C. Lin and D. Manocha, editors, *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science*, pages 203–222. Springer-Verlag, 1996. From the First ACM Workshop on Applied Computational Geometry.

[143] M. Shillor. Quasistatic problems in contact mechanics. *International Journal for Applied Mathematics and Computer Science*, 11(1):189–204, 2001.

[144] K. Shimoyama, S. Kawai, and J. Alonso. Dynamic adaptive sampling based on Kriging surrogate models for efficient uncertainty quantification. In *Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, 2013.

[145] M. Sofonea and A. Matei. *Mathematical Models in Contact Mechanics*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2012.

[146] G. E. Stavroulakis, editor. *Recent Advances in Contact Mechanics*. Springer, 2013. Papers Collected at the 5th Contact Mechanics International Symposium (CMIS2009), April 28-30, 2009, Chania, Greece.

[147] O. Steinbach, B. Wohlmuth, and L. Wunderlich. Trace and flux a priori error estimates in finite element approximations of Signorni-type (sic) problems. *IMA Journal of Numerical Analysis*, 36(3), 2014.

[148] A. Touzaline. A quasistatic frictional contact problem with normal compliance and finite penetration for elastic materials. *Glasnik Matematicki*, 45:109–124, 2010.

[149] G. L. Trigg. *Mathematical Tool for Physicists*. Wiley-VCH, 2005.

[150] F. Tröltzsch. *Optimal Control of Partial Differential Equations: Theory, Methods, and Applications*. American Mathematical Society, 2010.

[151] G. A. Turley, S. M. Ahmed, M. A. Williams, and D. R. Griffin. Establishing a range of motion boundary for total hip arthroplasty. *Proceedings of the Institution of Mechanical Engineers. Part H, Journal of engineering in medicine*, 225(8):769–782, 2011.

[152] G. A. Turley, M. A. Williams, R. M. Wellings, and D. R. Griffin. Evaluation of range of motion restriction within the hip joint. *Medical & Biological Engineering & Computing*, 51(4):467–477, 2013.

[153] M. Ulbrich. Semismooth newton methods for operator equations in function spaces. *SIAM Journal on Optimization*, 13(3):805–841, 2002.

[154] S. Ulbrich and J. C. Ziems. Adaptive multilevel trust-region methods for time-dependent PDE-constrained optimization. *Portugaliae Mathematica*, 74:37–67, 2017.

[155] J. W. Vahey, L. J. L., and R. Vanderby Jr. Elastic moduli, yield stress, and ultimate stress of cancellous bone in the canine proximal femur. *Journal of Biomechanics*, 20(1):29–33, 1987.

[156] S. Volkwein. Model reduction using proper orthogonal decomposition. In *CMCS – Modelling and Scientific Computing MATHICSE*, 2011.

[157] H. Wackernagel. *Multivariate Geostatistics*. Springer-Verlag, 1995.

[158] W. Wang, R. Tuo, and C. F. J. Wu. On prediction properties of Kriging: Uniform error bounds and robustness. *Journal of the American Statistical Association*, 115(530):920–930, 2020.

[159] X. Wang, S. Xu, S. Zhou, W. Xu, M. Leary, P. Choong, M. Qian, M. Brandt, and Y. M. Xie. Topological design and additive manufacturing of porous metals for bone scaffolds and orthopaedic implants: A review. *Biomaterials*, 83(11):127–141, 2016.

[160] Y. Wang, S. Arabnejad, M. Tanzer, and D. Pasini. Hip implant design with three-dimensional porous architecture of optimized graded density. *J. Mech. Des.*, 140(11), 2018.

[161] R. Webster and M. A. Oliver. *Geostatistics for Environmental Scientists*. John Wiley & Sons, Ltd., 2nd edition, 2007.

[162] M. Weiser. On goal-oriented adaptivity for elliptic optimal control problems. *Opt. Meth. Softw.*, 28(13):969–992, 2013.

[163] M. Weiser, P. Deuflhard, and B. Erdmann. Affine conjugate adaptive Newton methods for nonlinear elastomechanics. *Optimization Methods and Software*, 22(3):413–431, 2007.

[164] M. Weiser and S. Götschel. State trajectory compression for optimal control with parabolic PDEs. *SIAM J. Sci. Comput.*, 34(1):A161–A184, 2012.

[165] A. Wengler, U. Nimptsch, and T. Mansky. Hip and knee replacement in Germany and the USA – analysis of individual inpatient data from German and US hospitals for the years 2005 to 2011. *Dtsch Arztebl Int*, 111:407–16, 2014.

[166] J. Werner. *Numerische Mathematik 2: Eigenwertaufgaben, lineare Optimierungsaufgaben, unrestringierte Optimierungsaufgaben*. Vieweg+Teubner Verlag, 1992.

[167] D. C. Wirtz, N. Schiffers, T. Pandorf, K. Radermacher, D. Weichert, and R. Forst. Critical evaluation of known bone material properties to realize anisotropic FE-simulation of the proximal femur. *J Biomech*, 33:1325–30, 2000.

[168] B. Wohlmuth. Variationally consistent discretization schemes and numerical algorithms for contact problems. *Acta Numerica*, 20:569–734, 2011.

[169] B. Wohlmuth and R. Krause. Monotone multigrid methods on nonmatching grids for nonlinear multibody contact problems. *Siam Journal on Scientific Computing*, 25(1):324–347, 2003.

[170] U. Wolfram and J. Schwiedrzik. Post-yield and failure properties of cortical bone. *BoneKEy Reports*, 5:829, 2016.

[171] P. Wriggers. *Nonlinear Finite Element Methods*, volume 1. Springer, 2008.

[172] P. Wriggers. Discretization methods for solids undergoing finite deformations. In J. Schröder and P. Wriggers, editors, *Advanced Finite Element Technologies*, pages 17–51. Springer, Cham, 2016.

[173] Z. Wu and R. Schaback. Local error estimates for radial basis function interpolation of scattered data. *IMA J. Numer. Anal*, 13:13–27, 1992.

[174] Y. Xie, R. H. Byrd, and J. Nocedal. Analysis of the BFGS method with errors. *SIAM Journal on Optimization*, 30(1):182–209, 2020.

[175] S. J. Yakowitz and F. Szidarovszky. A comparison of Kriging with nonparametric regression methods. *Journal of Multivariate Analysis*, 16(1):21–53, 1985.

[176] X. Yang, D. A. Barajas-Solano, G. Tartakovsky, and A. M. Tartakovsky. Physics-informed CoKriging: A Gaussian-process-regression-based multifidelity method for data-model convergence. *J. Comput. Physics*, 395:410–431, 2018.

[177] J. Youett, O. Sander, and R. Kornhuber. A globally convergent filter-trust-region method for large deformation contact problems. *SIAM Journal on Scientific Computing*, 41, 2019.

[178] R. Yousefpour. Combination of steepest descent and BFGS methods for nonconvex nonsmooth optimization. *Numerical Algorithms*, 72:57–90, 2016.

[179] S. Zachow, M. Zilske, and H.-C. Hege. 3D reconstruction of individual anatomy from medical image data: Segmentation and geometry processing. *Proceedings of the CADFEM Users Meeting*, 2007.

[180] J. C. Ziems and S. Ulbrich. Adaptive multilevel inexact SQP methods for PDE-constrained optimization. *SIAM J. on Optimization*, 21(1):1–40, 2011.

[181] M. Zilske, H. Lamecker, and S. Zachow. Adaptive remeshing of non-manifold surfaces. In *Eurographics 2008 Annex to the Conf. Proc.*, pages 207–211, 2008.

Acknowledgments

First of all, I would very much like to thank Dr. Martin Weiser, head of the Computational Medicine group and the Modeling and Simulation of Complex Processes department at the Zuse Institute Berlin (ZIB), for his continuous support throughout the thesis development and his redoubled efforts towards the end. His scientific insight into various areas, his always open door and ears, and his unshaken believe in the completion of this work has helped immensely, while his good-humored, well-distributed comments have taken off the edge of some dull stages.

Second, I would like to extend my greatest thanks to former colleague and, I daresay, friend Sebastian Götschel, for his scientific, mental, emotional, albeit not physical support. Not only did he sit down with me to look over the program to pin down the implementation errors that escaped my attention – partly successful, partly not –, but his persisting questioning forced me to wholly grasp the diverse mathematical concepts in this thesis in order to then be able to formally recount and explain. That being said, if he hasn't understood Kriging and the concept of a variogram yet, I fear I'm at the end of my wits to do so. In some downhearted moments he was there to uplift and to push me on with his patronage. Finally, I am still astonished about his repeated and concise revision of my thesis, worming into everything and demanding further precision and clarification that greatly aided in finalizing this work.

Furthermore, I want to thank both my referees Prof. Dr. Ralf Kornhuber, Freie Universität Berlin, and Prof. Dr. Anton Schiela, Universität Bayreuth, for their readiness to scrutinize my thesis.

Towards current and former members of the work group and further colleagues from the Zuse Institute I would like to extend my great thanks for various fruitful, joyous, and in-depth discussions about what matters in life and what is secondary, next to – of course – mathematical topics.

Further I would like to mention my friends and mates from my the fencing club, the Fecht-Club Grunewald Berlin. Physical exertion, mental maneuvering, and personal fulfillment through voluntary work in a team were a good complement to the theoretically coined desk work; and the countless occasions for an after-sports cold beverage were not unwelcome and seldom passed up.

Last but not least, I would like to extend my gratitude to my friends, such as Torsten Siebert, Dag Freyer, Sebastian Funk, and Sophie Lipschütz, the last of whom I wish the best of luck with her PhD thesis, my family – Silvia, Martin, Patrick, Karro, and Karin Moldenhauer – and my partner Garik Schall for their never-ending support, caring, encouragement, and patience.

Zusammenfassung

Durch steigendes Durchschnittsalter einer Bevölkerung werden Prothesen z.B. in der Hüfte oder am Knie immer notwendiger; und durch Abnutzungen, welche schmerzhaften Abrieb und dadurch Entzündungen, aber auch Lockerung verursachen, werden Revisionsoperationen immer häufiger. Da sich Implantate aber auch lockern, weil sich der umgebende Knochen durch die geänderte Druckverteilung deformiert, müssen Implantate so eingepasst werden, dass sie in einer optimalen Position die Druckänderung, das sogenannte *stress shielding*, minimieren. Gleichzeitig sollen aber auch lokal hohe Maximaldrücke vermieden werden, da sie durch das Übersteigen eines schadhaften Schwellwerts dazu führen können, dass der gegebenenfalls poröse Knochen bricht.

Hierzu wird ein elastomechanisches Kontaktproblem unter Berücksichtigung der täglichen bzw. aller möglichen Bewegungen, welche Trajektorien in einem relativ hochdimensionalen Lastraum darstellen, gelöst. Der Green-St. Venant Verzerrungstensor muss zwingend inklusive der geometrischen Nichtlinearität modelliert werden, da die beteiligten Körper in den Gelenken rotieren können. Um die zu erwartenden hohen Rechenanforderungen zu reduzieren und das Lösen des Optimalsteuerungsproblems überhaupt möglich zu machen, werden effiziente Methoden für Unterprobleme untersucht und angewandt.

Dafür soll zunächst die Anzahl der Lösungen der partiellen Differentialgleichung (PDG) verringert werden, da bei fein diskretisiertem Finite Elemente-Gitter die Lösung der PDG den teuersten Anteil ausmacht. Hierfür wird Kriging als Interpolationsmethode vorgestellt und implementiert. Kriging erlaubt das sequentielle Einfügen von Testlasten und den daraus resultierenden Drucklast-Antworten. Außerdem kann die modell-inhärente Varianz ausgenutzt werden, um zum einen die Interpolationsgenauigkeit zu prüfen und zum anderen an Stellen, wo das Interpolationsmodell die größte Varianz aufweist, dem Modell adaptiv hinzuzufügende Lasten zu bestimmen.

Weiterhin wird genutzt, dass die Bewegungen auf den Lastraum marginalisiert und durch eine Wahrscheinlichkeitsverteilung charakterisiert werden können. Dazu werden gemessene Lastdaten der öffentlich zugänglichen Datenbank OrthoLoad mit einer patientenspezifischen Gewichtung der Bewegungen kombiniert, sodass einerseits die PDG quasistatisch gerechnet werden kann und andererseits die Wahrscheinlichkeit in der Berechnung des Integrals über den Lastraum mittels Monte Carlo Integration zur Anwendung kommt.

Abhängig von der FE-Diskretisierung ergibt sich eine Fehlerordnung, anhand derer die Genauigkeit der PDG-Lösung, der Kriging Interpolation und der Monte Carlo Integration justiert und bei mangelhafter Genauigkeit adaptiv verfeinert werden können.

Zuletzt wird für das Optimierungsverfahren die nötige Ableitung des Zielfunktionals mithilfe des impliziten Funktionentheorems hergeleitet. Außerdem muss für die Schrittweitensteuerung eine spezielle Liniensuche eingeführt werden, sodass bei sich änderndem FE-Diskretisierungsfehler durch Verschiebung der Implantatsposition, die auftretenden Unstetigkeiten im Zielfunktionswert die Optimierung nicht zwecklos machen.

Das funktionierende Zusammenwirken der Algorithmen wird numerisch ausgewertet und nachgewiesen.