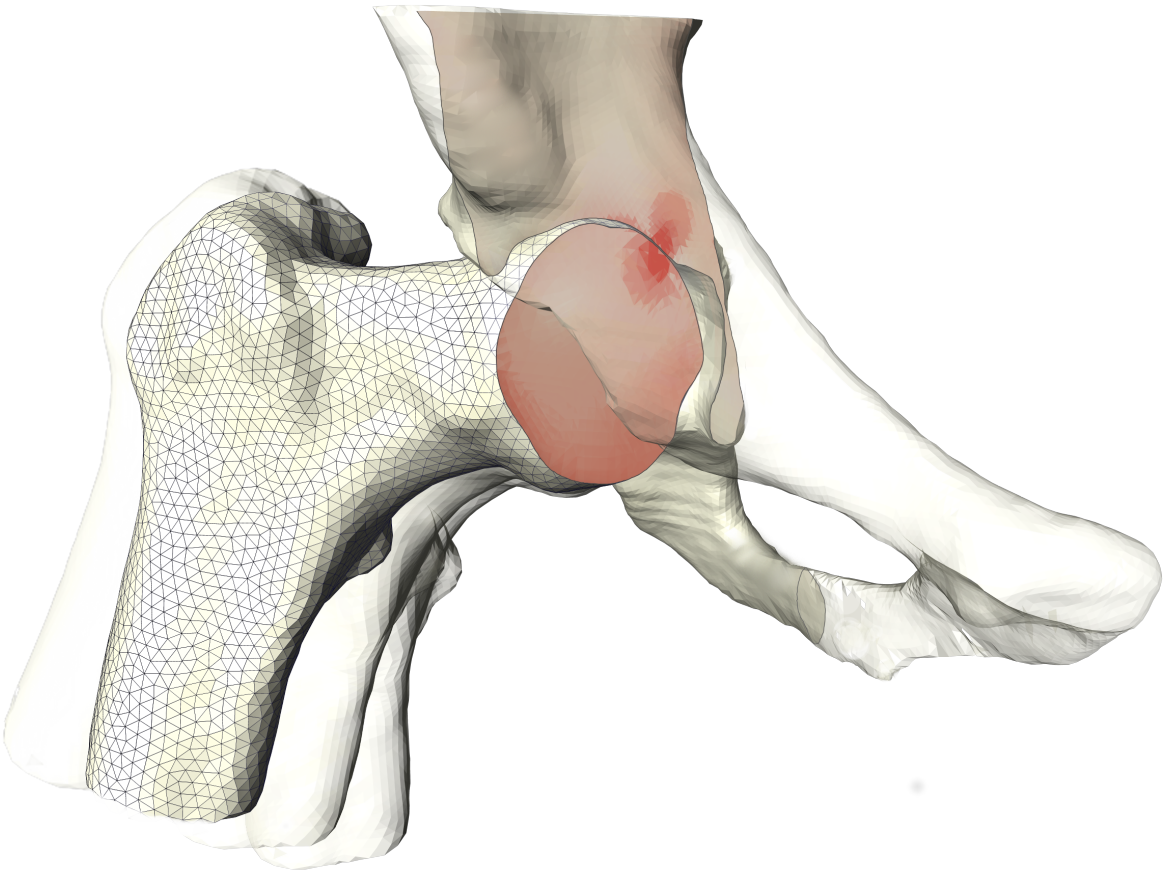
Dynamic large deformation contact problems and applications in virtual medicine

Inauguraldissertation
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
am Fachbereich Mathematik und Informatik
der Freien Universität Berlin



vorgelegt von
Jonathan William Youett

Berlin, 16. Dezember 2015

Betreuer:	**Prof. Dr. Ralf Kornhuber**	(Freie Universität Berlin)
Erstgutachter:	**Prof. Dr. Ralf Kornhuber**	(Freie Universität Berlin)
Zweitgutachter:	**Prof. Dr. Rolf Krause**	(USI Lugano)
Tag der Disputation:	18. Mai 2016	

This work was supported by the MATHEON funded through the Einstein Centre and the Berlin mathematical school.

Contents

Introduction

In 2006 over 160.000 hip replacement surgeries were counted in Germany, vastly increasing regarding the ageing society. Of these operations almost 20.000 were secondary interventions, requisite due to wear, dislocation or fixation failure of the implant [8]. Although not entirely, the success of a hip replacement and a pain-free long-time behaviour of the joint depends heavily on the experience of the surgeon and a good pre-operational planning. This planning is usually done based on X-ray data only. The anatomical centre of rotation and implant size and position are settled by using simple templates [76]. Recent approaches to digitalise this templating exist [98] but they do not close the clear gap between the purely geometric construction in 2D and the three dimensional dynamic reality. First software tools based on CT or MRI data were developed recently. These allow to perform a virtual range of motion analysis based on the assumption that the hip joint is a perfect ball and socket joint and hence the centre of rotation stays fixed [24]. These approaches have the disadvantage that no proper stability analysis of the joint centre under loads can be performed and thus detection of possible dislocation cannot be done.
With the work in this thesis we want to contribute to the improvement of present surgery planning by introducing a patient-specific finite element hip joint model for virtual testing. Based on the statistical shape model [85] the model provides the possibility of an automatic construction from CT data. In this thesis we focus on using this model for a range of motion and impingement (contact) analysis. The most important part of such a virtual testing framework is the modelling of contact and its fast and robust numerical solution, which is the main focus of the first part of this thesis.

Although dynamic large deformation contact problems arise in many industrial applications like auto mobile engineering or biomechanics, only very few methods exist for their solution, all having their advantages and disadvantages. Most of these methods are based on the continuous contact model by Laursen and Simo [90]. In this approach the contact traction forces are explicitly included in the weak formulation and treated as additional unknowns. In this thesis we will follow a different path that first appeared in [70]. Using non-smooth calculus [99] we will properly derive a differential inclusion depending only on the deformation. The idea is to incorporate the non-penetration constraints directly into Hamilton's principle from which the weak equations of motion are derived. This approach can also be used for general constrained Hamiltonian systems as it does not depend on the explicit form of the constraints. Hence also more general contact descriptions, e.g. only demanding the intersection of two bodies to be empty, can be treated formally. Analysing

this abstract description in the framework of subdifferential calculus might even lead to new existence results, which have been derived recently for a similar formulation of Signorini-type problems by Schuricht [122].

The discretisation of large deformation contact problems is complicated by the non-linearity, geometric and material, of the equations of motion and the non-penetration constraint.
In the framework of linearised contact, i.e. displacements are assumed to be small, many successful time discretisation schemes based on the Newmark family were developed [34, 78]. Unfortunately, for the large deformation model one can observe that, even in the absence of contact, most of these classical integrators loose their unconditional stability [125]. The reason for this is that spectral stability is only a necessary condition for stability in non-linear structural mechanics. Belytschko and Schoeberle showed that in non-linear mechanics the conservation or decay of the total energy is a sufficient criterion for the stability of implicit schemes [18]. In combining this conclusion with classical time integrators, Simo and Tarnow introduced the *Energy–Momentum* method [125], which is the prevalent basis of most time integrators for large deformation contact problems. A more detailed overview of present time discretisations can be found in Section 2.1. The idea of this method is to algorithmically enforce conservation of energy by adjusting the implicit midpoint rule suitably in each time step. One drawback of this approach is that the conservation is achieved at the cost of loosing the symmetry of the underlying tangent stiffness operator [84]. Now our motivation for investigating contact problems has its origin in virtual medicine where short computing times and hence a fast numerical solution is of importance. On the other hand the evolutions of interest, e.g. the range of motion of the hip joint, need to be resolved in very high detail to predict the resulting stresses accurately. With this mind, our first experimental choice for the discretisation of the non-standard differential inclusion, is the implicit midpoint rule. As clinical application in general require a high resolution of the trajectories we neglect the restricted stability. In the presence of contact numerical experiments showed that the midpoint rule is generating energy. Motivated by the predictor-corrector method in [70], we propose to stabilise the midpoint rule by modifying the generalised gradient term, which corresponds to the implicit contact forces. This results in a dissipative-like behaviour. In Section 4.4.3 we investigate the scheme numerically and suggest further stabilising steps to improve the method further.

A second difficulty of contact problems is the discretisation of the non-penetration constraint which involves identifying the discrete contact boundaries with each other, see Section 2.2 for a more thorough overview. A widely used approach is the *Node-to-Segment* method [7, 65]. Therein, each node of one surface is identified with a segment on the opposing surface and non-penetration is enforced for each of these nodes. Recently the use of mortar methods became more and more popular [79, 93]. The idea is to enforce the contact constraints in a variationally consistent manner which improves the stability of the underlying methods significantly [93]. Krause and Wohlmuth proposed to use dual basis functions for the mortar discretisation of linearised contact problems, which

enabled them to decouple the non-penetration constraints and to apply fast monotone multigrid solvers for the solution of the algebraic problems [79]. In this thesis we extend this approach to large deformation contact problems. First, we note that when dealing with large deformations, the mortar discretisation requires to integrate the constraints over the *deformed* configuration to account for the changing contact boundaries [65]. This prevents a simple decoupling as in the linearised case [79].
As mentioned before, methods that are based on the Energy–Momentum scheme do not inherit the symmetry of the continuous tangent stiffness operator. The resulting algebraic non-linear systems are usually solved using an active set strategy [68, 62, 66] or penalty methods [93], see also Chapter 4. In our approach we exploit that the resulting spatial problems of the stabilised midpoint rule can be reformulated as constrained minimisation problems. In this thesis we develop a solver based on sequential quadratic programming (SQP), i.e. the successive local approximation of the non-linear functional and constraints by quadratic models. As hyperelastic materials are in general non-convex, we apply a trust–region globalisation strategy to cope with the possible unboundedness of the local problems. In the case of non-linear constraints further adjustments of the trust–region method have to be made, as the intermediate iterates are in general not feasible w.r.t the non-linear constraints. A novel approach that is avoiding any penalty parameters is the *filter method* [45]. The idea is to accept a trial point if it either has a lower energy or a better feasibility w.r.t. the non-linear constraints than previous iterates. Under suitable conditions this method is shown to converge globally [46]. When applying this algorithm to large deformation contact problems, the linearised constraints are almost of the form as in the small strain framework, which could be decoupled into bound constraints [79]. Therefore, we extend the standard filter–trust–region method to allow inexact constraint Jacobians, namely the portion which allows to employ the same decoupling strategy. The mechanism of this inexact filter method works as follows: Solve inexact local problems until an approximate convergence criterion is fulfilled. Then, if the exact convergence criterion is not fulfilled, switch to solving exact local problems.
The inexact local problems are quadratic, non-convex and subject to simple bound constraints. To this end, we extend the monotone multigrid method [79, 54, 55] and prove global convergence for this class using the framework of upper hemi-continuity, see Section 3.1. A short overview of solvers for non-linear obstacle problems can be found in Section 4.4.1. In Section 3.4 we also prove that the inexact filter method converges globally when the approximation error, that we make by using inexact Jacobians, is controlled. Numerical results showed that already for small problems the computation time could be reduced up to 30% when the inexact filter method is used, see Chapter 4.

The second part of this thesis is concerned about the description of a heterogeneous hip joint model and a computational framework for a range of motion and femoroacetabular impingement analysis. The mathematical model is a direct extension of the knee joint model, introduced in [117, 120], to large deformation contact. The model comprises the pelvis and femur bones and the three major ligaments. Further, cartilage can be incorporated but is not considered here due to the usual lack of MRI data in practical applications. The bones are assumed to be three dimensional hyperelastic materials and

large deformation contact is enforced between the femoral head and the acetabulum. As proposed by Sander [117], we use a one-dimensional Cosserat rod model for the ligaments. The main assumption in this model is, that the cross-sections of the rod do not deform and hence can thus be considered to be rigid objects. The configuration space of a Cosserat rod is a manifold, the special Euclidean group, and standard finite elements cannot be applied any more. One way to handle this difficulty is to exchange the linear interpolation by interpolation along geodesics, leading to *geodesic finite elements*, introduced by Sander in [117]. The range of motion of the hip joint can only be estimated correctly if additionally non-penetration between the bones and ligaments is enforced. We neglect the traction forces that the ligaments exert on the bones and consider Signorini-type problems for the ligaments. In this thesis we introduce a novel contact model for Cosserat rods and propose a first nodal discretisation. For the numerical solution of the corresponding algebraic problem we combine the Riemannian trust–region method [116] with a filter mechanism. The coupling of ligaments and bones lead to transmission conditions on the coupling interface, derived by Sander in [117]. The resulting heterogeneous problems is solved using a Dirichlet–Neumann algorithm [115]. This has the advantage that the heterogeneous system decouples and successively large deformation contact problems for the bones and Signorini-type problems for the ligaments are solved using the methods described in this thesis.

The thesis is organised as follows. Chapter 1 covers a detailed introduction to large deformation contact between hyperelastic continua and a derivation of the weak equations of motion in the classical way. We then introduce a non-smooth Hamilton principle and construct a non-standard differential inclusion. We close the chapter by showing that classical solutions fulfil the differential inclusion. Chapter 2 is concerned about the temporal and spatial discretisation of the inclusion. We apply the implicit midpoint rule and describe the modification leading to an improved energy behaviour. Then we shortly depict the spatial discretisation using dual mortar elements and the derivation of the algebraic problems. In Chapter 3 we first introduce an extension of the monotone multigrid method to non-convex problems subject to box-constraints. We prove global convergence of the scheme by reformulating it as successive directional minimisation algorithm. Then we describe the filter trust–region method, incorporating inexact constraint Jacobians. Expanding the original proof for the exact filter method, we show that global convergence still holds if the approximation error is controlled. In chapter Chapter 4 we then apply the filter method to large deformation contact problems and shortly describe how the monotone multigrid is adjusted to handle the linearised non-penetration constraints. We close the chapter by showing some numerical results that confirm the theoretical convergence proofs. The proposed modified midpoint scheme is investigated numerically and further extension to improve the method are suggested. Finally in the last chapter 5 the heterogeneous hip joint model is described and first results of an advanced range of motion analysis are shown. The appendix of this thesis comprises a short introduction to concepts from non-linear optimisation that are used within this thesis and the detailed steps of the global convergence proof of the filter–trust–region method.

Acknowledgement During the last years I received a lot of support by a vast amount of people that helped me finishing this thesis and make me the person that I am now. Among these I especially want to thank Prof. Dr. Ralf Kornhuber for his constant encouragement, guidance and valuable advise, for convincing me as student not to go to economy and instead giving me the opportunity to work in such a cheerful environment of brilliant people. Further, my sincere gratitude goes to Prof. Dr. Oliver Sander for his continuous open ear and interest in my work and for getting my research started with many good ideas. As everybody who gets the pleasure of working with him, I want to thank Prof. Dr. Carsten Gräser for always sharing his endless knowledge in a patient and profound way. I thank Max Kahnt and Dr. Uli Sack for all the fruitful discussions and good times we had, making our office feel like a second home. Ana Djurdjevac and Dr. Elias Pipping I thank for their help on functional analysis. For all the assistance and advices on the visualisation I thank Max Kahnt, Anthony Youett and especially Dr. Stefan Zachow, who also played a big role in the identification of clinical applications. I want to express my gratitude to my family who are always caring and believing in me. Last but not least I thank Evgenia for all her support, mathematically and humanly, for always being on my side and making my life beautiful.

God save the queen.

1. Contact Mechanics

In this chapter we introduce the dynamic large deformation contact problem of general non-linear hyperelastic materials. First, we derive the governing equations of motion which arise from momentum conservation in non-linear elasticity and then show how non-penetration between time-dependent contact boundaries can be modelled. Afterwards, we derive the variational formulation of the problem and prove conservation of momenta and energy under suitable conditions. Finally, based on a non-smooth Hamilton principle we will establish a non-standard variational inclusion which generalises the weak formulation to more abstract contact models.

1.1. Non-linear Elasticity

Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$ denote the reference configuration of a body which is assumed to be stress-free. Assume further that the boundary of the domain is smooth enough such that the outer unit normal field $\mathbf{n} : \partial\Omega \longrightarrow \mathbb{R}^d$ exists everywhere and is continuously differentiable. Let the boundary be decomposed into disjoint open sets $\partial\Omega = \bar{\Gamma}_D \cup \bar{\Gamma}_N$ which can possibly be empty. Further denote by $[0, T]$ the time interval of interest.

We call $\boldsymbol{\varphi} : \bar{\Omega} \times [0, T] \longrightarrow \mathbb{R}^d$ a *configuration* of Ω if for any $t \in [0, T]$ the mapping $\boldsymbol{\varphi}_t := \boldsymbol{\varphi}(\cdot, t) : \Omega \longrightarrow \mathbb{R}^d$ is a *deformation*, i.e. it is sufficiently smooth, locally injective and orientation-preserving

$$\det \nabla \boldsymbol{\varphi}_t(x) > 0 \quad \forall\, x \in \Omega. \tag{1.1}$$

By Newton's second law, i.e. balance of momentum, and Cauchy's stress principle [96, 142], the equations of motion and boundary conditions of a continuum are given by

$$\begin{aligned} \operatorname{div} \boldsymbol{\sigma}_t + \mathbf{f}(t) &= \rho\, \ddot{\boldsymbol{\varphi}}_t && \text{in } \boldsymbol{\varphi}_t(\Omega), \\ \boldsymbol{\sigma}_t \mathbf{n}_t &= \boldsymbol{\pi}(t) && \text{on } \boldsymbol{\varphi}_t(\Gamma_N), \\ \boldsymbol{\varphi}_t &= \boldsymbol{\varphi}_D(t) && \text{on } \boldsymbol{\varphi}_t(\Gamma_D), \end{aligned} \tag{1.2}$$

for each $t \in (0, T)$ and supplemented with initial deformation and velocity conditions

$$\boldsymbol{\varphi}_0 = \hat{\boldsymbol{\varphi}}_0, \quad \dot{\boldsymbol{\varphi}}_0 = \hat{\mathbf{v}}_0 \quad \text{on } \Omega.$$

The functions $\mathbf{f}(t) \in L^2(\Omega)^d$ and $\boldsymbol{\pi}(t) \in L^2(\Gamma_N)^d$ are external volume and traction forces and ρ is the mass density of the body. In this thesis we will assume that the mass density is constant over time and the external forces are dead loads, i.e. independent of the deformation. The function $\boldsymbol{\varphi}_D(t) \in C(\Gamma_D)^d$ denotes possible prescribed Dirichlet boundary conditions, e.g. clamping of the body.

For the sake of simplicity in what follows we will sometimes omit the index $_t$ and simply write $\boldsymbol{\varphi}$. The symmetric Cauchy stress tensor field $\boldsymbol{\sigma} : \boldsymbol{\varphi}(\Omega) \longrightarrow S(d)$, where $S(d)$ is the set of symmetric $(d \times d)$-matrices, represents the internal force that arises due to external loading or boundary conditions. Its existence is assured by Cauchy's stress principle which also gives an interpretation of the tensor as follows. For each sub-volume $V \subset \boldsymbol{\varphi}(\Omega)$ the traction forces acting on its boundary ∂V, with normal $\mathbf{n}$, by the surrounding tissue are given by

$$\mathbf{t}_C := \boldsymbol{\sigma}\mathbf{n} \quad \text{on } \partial V. \tag{1.3}$$

The traction force (1.3) is called *Cauchy traction* or *Cauchy stress vector*. A material is called *elastic* if the Cauchy stress tensor is depending on the deformation only through the deformation gradient and the deformed position $\boldsymbol{\sigma} = \boldsymbol{\sigma}(\boldsymbol{\varphi}(x), \nabla\boldsymbol{\varphi}(x))$.

The partial differential equations (1.2) are complicated by the fact that they are formulated on the deformed domain $\boldsymbol{\varphi}(\Omega)$ which is a priori unknown. The common approach in continuum mechanics is to reformulate the problem in reference coordinates using the *first Piola–Kirchhoff stress tensor*

$$\mathbf{P}(x) := \det(\nabla\boldsymbol{\varphi}(x))\boldsymbol{\sigma}(\boldsymbol{\varphi}(x), \nabla\boldsymbol{\varphi}(x))\nabla\boldsymbol{\varphi}(x)^{-T}. \tag{1.4}$$

This transformation is called *Piola transform* and it has the advantage that the equations of motion in reference coordinates are still in divergence form as it holds

$$\operatorname{div}\mathbf{P}(x) = \det(\nabla\boldsymbol{\varphi}(x)) \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{\varphi}(x), \nabla\boldsymbol{\varphi}(x)), \tag{1.5}$$

cf. [26, Theorem 1.7-1]. In reference coordinates the system (1.2) transforms as follows

$$\begin{aligned} \operatorname{div}\mathbf{P}_t + \mathbf{f}_R(t) &= \rho_R\, \ddot{\boldsymbol{\varphi}}_{R,t} \ \text{ in } \Omega, \\ \mathbf{P}_t\mathbf{n} &= \boldsymbol{\pi}_R(t) \quad \text{on } \Gamma_N, \\ \boldsymbol{\varphi}_t &= \boldsymbol{\varphi}_D(t) \quad \text{on } \Gamma_D, \end{aligned} \tag{1.6}$$

for each $t \in (0, T)$, where the forces and density transform according to

$$\mathbf{f}_R = \det(\nabla\boldsymbol{\varphi}_t)\mathbf{f}, \tag{1.7}$$

$$\boldsymbol{\pi}_R = \det(\nabla\boldsymbol{\varphi}_t)\|\nabla\boldsymbol{\varphi}_t^{-T}\mathbf{n}_t\|\boldsymbol{\pi}, \tag{1.8}$$

$$\rho_R = \det(\nabla\boldsymbol{\varphi}_t)\rho, \tag{1.9}$$

see e.g. [26, Section 2.6] and $\ddot{\boldsymbol{\varphi}}_{R,t}$ denotes the material acceleration, see [96, Chaper 1]. For simplicity in what follows, we will use the same notion $\mathbf{f}, \boldsymbol{\pi}, \rho, \ddot{\boldsymbol{\varphi}}_t$ for both the

reference and the deformed quantities. The Piola stress tensor has the disadvantage that it is in general not symmetric. This is why often the equations of motion are formulated in terms of the symmetric *second Piola–Kirchhoff stress tensor*

$$\mathbf{\Sigma}(\nabla\boldsymbol{\varphi}) := \nabla\boldsymbol{\varphi}^{-1}\mathbf{P}(\nabla\boldsymbol{\varphi}). \tag{1.10}$$

In this thesis we will only consider *hyperelastic* continua. A continuum is called *hyperelastic* if there exists a stored energy functional $W(x, F) : \Omega \times \mathrm{M}^+(d) \longrightarrow \mathbb{R}$ such that

$$\frac{\partial W}{\partial F}(x, \nabla\boldsymbol{\varphi}(x)) = \mathbf{P}(x, \nabla\boldsymbol{\varphi}(x)), \tag{1.11}$$

where $\mathrm{M}^+(d)$ is the space of $(d \times d)$-matrices with positive determinant. The relation (1.11) is called *constitutive* or *material law* of the hyperelastic continuum. It links the stresses to the deformation of the body and thus determines the main properties of the considered material.

Remark 1.1.1. *Usually, instead of enforcing the orientation preservation condition* (1.1) *as a hard constraint, it is built in the hyperelastic energy functional such that*

$$W(x, \nabla\boldsymbol{\varphi}(x)) = \infty \quad \text{if } \det\nabla\boldsymbol{\varphi}(x) \leq 0.$$

An important axiom in continuum mechanics is the *objectivity* of the material. It postulates that the stress is invariant under change of coordinates

$$\boldsymbol{\sigma}(QF, x) = Q\boldsymbol{\sigma}(F, x)Q^T \quad \forall\, Q \in \mathrm{SO}(d)\ \ \forall\, F \in \mathrm{M}^+(d), \tag{1.12}$$

where $\mathrm{SO}(d)$ denotes the group of rotations. If the material is objective, which from now on we will assume to hold, the stored energy functional W only depends on the deformation through the *right Cauchy-Green strain tensor*

$$C(\boldsymbol{\varphi}) := \nabla\boldsymbol{\varphi}^T\nabla\boldsymbol{\varphi}, \tag{1.13}$$

see [21, Theorem 1.6]. This tensor is a measure for the distortion of the body which can be seen using the shifted *Green–St.Venant strain tensor*

$$E(\boldsymbol{\varphi}) := \frac{1}{2}(C(\boldsymbol{\varphi}) - \mathrm{Id}). \tag{1.14}$$

Proposition 1.1.2 (Strain Measure)**.** *A deformation $\boldsymbol{\varphi}$ is a rigid body motion if and only if $E(\boldsymbol{\varphi}) = 0$.*

Proof. [26, Theorem 1.8-2]

Some important examples of hyperelastic materials are the *St.Venant–Kirchhoff material*

$$W(x, \nabla\boldsymbol{\varphi}) = \frac{\lambda(x)}{2}\operatorname{tr}(E(\boldsymbol{\varphi}))^2 + \mu(x)\|E(\boldsymbol{\varphi})\|^2, \tag{1.15}$$

which is derived by Taylor approximation and assuming a linear isotropic material behaviour [21, Theorem 1.11]; and the *Mooney–Rivlin material*

$$W(x, \nabla\boldsymbol{\varphi}) = a(x)\,\mathrm{tr}(E(\boldsymbol{\varphi})) + b(x)\,\mathrm{tr}(E(\boldsymbol{\varphi}))^2 + c(x)||E(\boldsymbol{\varphi})||^2 + \gamma(\det\nabla\boldsymbol{\varphi}), \tag{1.16}$$

which is a general model for a non-linear stress-strain relation. The function γ is penalising the violation of the local orientation-preservation of the material, cf. Remark 1.1.1. The norm of the strain tensor $\|E(\boldsymbol{\varphi})\|$ is induced by the tensor contraction which we will denote by : in the following. These energy functionals are in general non-convex whenever large deformations and rotations are considered.

Remark 1.1.3. *In the linearised framework when the deformation of the body is assumed to be small, the Green strain tensor can be approximated by*

$$E(\boldsymbol{\varphi}) \approx \varepsilon(\mathbf{u}) := \frac{1}{2}(\nabla\mathbf{u}^T + \nabla\mathbf{u}), \tag{1.17}$$

which is the linearisation of (1.14) *evaluated at the reference configuration* $\boldsymbol{\varphi} = Id$*. The displacement* $\mathbf{u}$ *is defined by*

$$\mathbf{u}(x) := \boldsymbol{\varphi}(x) - x.$$

In combination with the linear St. Venant–Kirchhoff material law (1.15) *this results in the quadratic convex strain energy functional*

$$W(\nabla\boldsymbol{\varphi}) = \frac{1}{2}\varepsilon(\mathbf{u}) : \mathcal{C} : \varepsilon(\mathbf{u}), \tag{1.18}$$

where $\mathcal{C}$ *denotes the 4th order Hooke tensor.*

In the quasi-static case, i.e. when inertia terms are neglected, the equilibrium states of hyperelastic materials are characterised as minimisers of the strain energy functional

$$\mathcal{J}(\boldsymbol{\varphi}) := \int_\Omega W(\boldsymbol{\varphi})\,dx - \text{ext. forces}, \tag{1.19}$$

cf. [26, Theorem 4.1-2]. To prove the existence of minimisers one typically assumes some kind of convexity of the functional $\mathcal{J}$. In [26] Ciarlet proved that any strain energy functional that explicitly depends on $\det(\boldsymbol{\varphi})$ cannot be convex [26, Theorem 4.8-1]. However, John Ball was able to prove existence of minimisers requiring only the weaker poly-convexity of the functional (and additional properties such as coerciveness) [26, Theorem 7.7-1]. Thus, for general non-convex hyperelastic energy functionals the coefficients, e.g. a, b, c in (1.16), are typically chosen such that these properties hold.

1.2. Frictionless large deformation contact

Next we describe how non-penetration of two bodies can be modelled in the presence of large deformations and rotations. There exist many approaches of how to enforce

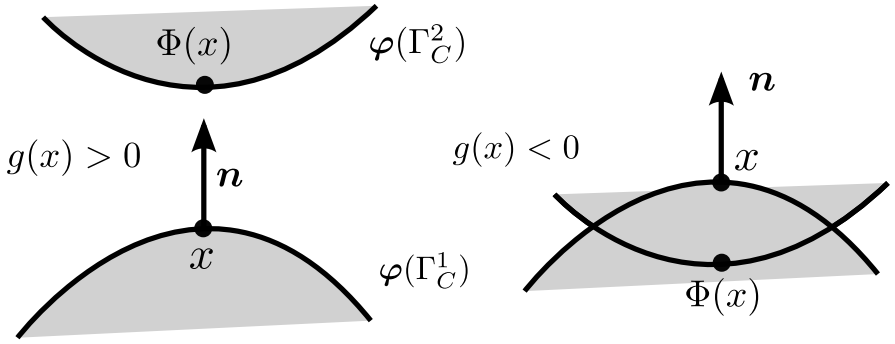


Figure 1.1.: *Left: Feasible configuration. Right: Unfeasible configuration.*

non-penetration of two bodies. In [70] Kane et al. proposed a contact description that is avoiding outer normal fields. While this approach can also be applied for domains with non-smooth boundary, it comes with the cost of having n^{seg} many constraints for each face on the contact boundary, where n^{seg} denotes the total number of faces on the contact boundary. Further, this approach, like most of the earlier contact models [69, 14], can only be formulated for discrete problems.

We will use the first and widely applied continuum-based model proposed by Laursen and Simo [90], where the existence of smooth outer normal fields on the contact boundaries is assumed. Let us consider two hyperelastic bodies $\Omega^i \subset \mathbb{R}^d$, $i = 1, 2$ with continuous differentiable boundaries that can be decomposed into open disjoint sets $\partial\Omega^i = \bar{\Gamma}_N^i \cup \bar{\Gamma}_D^i \cup \bar{\Gamma}_C^i$. We assume that the contact boundaries Γ_C^i are compactly contained within $\partial\Omega^i \setminus \Gamma_D^i$ and that in the reference configuration these bodies are not exerting any traction forces on each other through possible initial contact. As before the equations of motion in reference coordinates are given by for each $t \in (0, T)$

$$\begin{aligned} \operatorname{div} \mathbf{P}_t^i + \mathbf{f}(t) &= \rho^i \ddot{\boldsymbol{\varphi}}_t^i \quad && \text{in } \Omega^i, \\ \mathbf{P}_t^i \mathbf{n}^i &= \boldsymbol{\pi}(t) && \text{on } \Gamma_N^i, \\ \boldsymbol{\varphi}_t^i &= \boldsymbol{\varphi}_D^i(t) && \text{on } \Gamma_D^i, \end{aligned} \tag{1.20}$$

supplemented with initial deformation and velocity conditions

$$\boldsymbol{\varphi}_0^i = \hat{\boldsymbol{\varphi}}_0^i, \quad \dot{\boldsymbol{\varphi}}_0^i = \hat{\mathbf{v}}_0{}^i \quad \text{on } \Omega^i,$$

cf. (1.6). The subsets Γ_C^i denote the parts of the boundaries where possible contact may occur. Now the modelling of non-penetration can be done in several ways. In the first papers about continuous contact formulations [90, 87, 139] these surfaces were identified with each other by the *closest point projection* $\Phi^{\text{cp}} : \Gamma_C^1 \times [0, T] \longrightarrow \Gamma_C^2$ parametrised over Γ_C^1

$$\Phi^{\text{cp}}(x, t) := \operatorname*{arg\,min}_{y \in \Gamma_C^2} ||\boldsymbol{\varphi}_t^1(x) - \boldsymbol{\varphi}_t^2(y)||. \tag{1.21}$$

This projection defines a unit vector field on $\boldsymbol{\varphi}_t^1(\Gamma_C^1)$ called *contact normal* by

$$\boldsymbol{\nu}_\Phi^{\text{cp}}(x, t) := \mathbf{n}_t^2(\boldsymbol{\varphi}_t^2(\Phi^{\text{cp}}(x, t))), \tag{1.22}$$

where $\mathbf{n}_t^2$ denotes the unit outer normal on $\boldsymbol{\varphi}_t^2(\Gamma_C^2)$. Note that as Φ^{cp} is a best approximation, the vector field

$$\mathbf{V}_t(x) := \boldsymbol{\varphi}_t^1(x) - \boldsymbol{\varphi}_t^2 \circ \Phi^{\text{cp}}(x,t)$$

is itself normal to $\boldsymbol{\varphi}_t^2(\Gamma_C^2)$. Now the signed distance function of the two bodies is defined by

$$g^{\text{cp}}(x,t) := \boldsymbol{\nu}_\Phi^{\text{cp}}(x,t) \cdot (\boldsymbol{\varphi}_t^1(x) - \boldsymbol{\varphi}_t^2 \circ \Phi^{\text{cp}}(x,t)). \tag{1.23}$$

Recently using the normal projection along $\mathbf{n}_t^1$ has become more and more popular [93, 106, 66, 128]. In this case the contact boundaries are identified with each other through the projection

$$\Phi(x,t) = \min_{\mu \in \mathbb{R}} \left\{ y \in \Gamma_C^2 : \boldsymbol{\varphi}^2(y,t) = \boldsymbol{\varphi}^1(x,t) + \mu \mathbf{n}_t^1(x,t) \right\}. \tag{1.24}$$

The corresponding distance function is given by

$$g(x,t) := -\mathbf{n}_t^1(x,t) \cdot (\boldsymbol{\varphi}_t^1(x) - \boldsymbol{\varphi}_t^2 \circ \Phi(x,t)). \tag{1.25}$$

The function $g : \Gamma_C^1 \times [0,T] \to \mathbb{R}$ (and g^{cp}) is called *gap function*. Both approaches have their advantages and disadvantages. While the closest point projection does not have to be unique (cf. Curnier et al. [64]) and is more difficult to compute, the normal projection is lacking orthogonality to Γ_C^2 which is complicating the linearisation of the gap function that is needed within the solution of the discretised problems [93]. In the following we only consider the normal projection approach, although also the closest point gap can be used within the discretisation and solution techniques that we will develop in the subsequent chapters.

Remark 1.2.1. *We remark here that by* (1.25) *the gap function clearly also depends on the deformation*

$$g = g(x, \boldsymbol{\varphi}_t^1, \boldsymbol{\varphi}_t^2).$$

To keep the notation low we will omit this dependence in the strong formulation of the problem as the focus rather lies on the point wise evaluation at x. When deriving and analysing the weak integral (mortar) formulation of the constraint we will switch to considering the gap function as a mapping between function spaces (1.3 and following).

To enforce non-penetration of the bodies one has to ensure that the gap stays non-negative at all times

$$g(x,t) \geq 0 \quad \forall\, (x,t) \in \Gamma_C^1 \times [0,T], \tag{1.26}$$

which is illustrated in Figure 1.1.

Remark 1.2.2. *In the linearised framework where the deformation is assumed to be small, cf. Remark 1.1.3, a first-order Taylor approximation of the gap function at the reference configuration is used which results in the following linear constraints*

$$g^{lin}(x,t) := g(x,0) - \left(\mathbf{u}^1(x,t) - \mathbf{u}^2(x,t) \circ \Phi(x,0)\right) \cdot \mathbf{n}^1(x,0) \geq 0, \tag{1.27}$$

see [87, Section 4.4.1]. The standard linearised non-penetration constraint is then given by enforcing that the relative displacement in normal direction stays smaller than the initial gap

$$\big(\mathbf{u}^1(x,t) - \mathbf{u}^2(x,t) \circ \Phi(x,0)\big) \cdot \mathbf{n}^1(x) \leq g(x,0). \tag{1.28}$$

It is also assumed that $\mathbf{n}^1(x) \approx -\mathbf{n}^2(\Phi(x))$ *so that both versions of the gap function coincide in this framework.*

Remark 1.2.3. *The gap function can also be defined on the deformed configuration directly* $g : \boldsymbol{\varphi}_t^1(\Gamma_C^1) \longrightarrow \mathbb{R}$

$$g(x) := -\mathbf{n}^1(x) \cdot (x - \Phi(x)), \tag{1.29}$$

where $\Phi : \boldsymbol{\varphi}_t^1(\Gamma_C^1) \longrightarrow \boldsymbol{\varphi}_t^2(\Gamma_C^2)$ *is the normal projection defined on the deformed contact boundaries.*

So far the non-penetration constraint was derived only from a kinematical point of view. To investigate the effect of these constraints on the elastic systems we examine the *Piola traction*

$$\mathbf{T}_C := \mathbf{P}(\nabla \boldsymbol{\varphi}^1)\mathbf{n}^1, \tag{1.30}$$

cf. (1.3), on the contact boundary Γ_C^1 in more detail. First, the Piola traction can be decomposed into a normal and a tangential part

$$\mathbf{T}_C = T_N \mathbf{n}^1 + \mathbf{T}_T, \tag{1.31}$$

where

$$T_N := \mathbf{n}^1 \cdot (\mathbf{P}(\nabla \boldsymbol{\varphi}^1)\mathbf{n}^1), \tag{1.32}$$

$$\mathbf{T}_T := \mathbf{T}_C - T_N \mathbf{n}^1. \tag{1.33}$$

In this thesis we are considering *frictionless* contact only so the tangential traction $\mathbf{T}_T$ at the contact boundary vanishes

$$\mathbf{T}_C = T_N \mathbf{n}^1. \tag{1.34}$$

Moreover the contact forces must be compressive, i.e. directed towards the inside of Ω^1. This leads to the necessary condition

$$T_N \leq 0. \tag{1.35}$$

The *complementary condition* states that the normal stress can only be non-zero if the bodies are in contact with each other

$$g(x) \cdot T_N(x) = 0 \quad \forall\, x \in \Gamma_C^1. \tag{1.36}$$

Altogether these conditions are of the form of the classical *Karush-Kuhn-Tucker conditions* known from constrained optimization [101]

$$T_N \leq 0, \quad g \geq 0, \quad g \cdot T_N = 0. \tag{1.37}$$

The normal stress T_N can therefore be interpreted as a Lagrangian multiplier (or strictly speaking the negative multiplier).

1.3. Weak equations of motion

In this section we derive the variational formulation of the dynamic large deformation contact problem. In what follows we will use Ω to denote product spaces defined on (Ω^1, Ω^2). Let $\mathbf{L}^2(\Omega) := \prod_{i=1}^2 L^2(\Omega^i)^d$ denote the vector-valued Lebesgue space of square integrable functions and $\mathbf{H}^1(\Omega) := \prod_{i=1}^2 H^1(\Omega^i)^d$ be the first order Sobolev space with weak derivatives in $\mathbf{L}^2(\Omega)$. Further, denote by $\mathbf{H}^1_D(\Omega)$ and $\mathbf{H}^1_0(\Omega)$ the corresponding Sobolev spaces with inhomogeneous resp. zero Dirichlet boundary conditions on Γ^i_D in the sense of traces [3].

The weak equation of motion for the large deformation contact problems is obtained by multiplication with test functions $\mathbf{v} \in \mathbf{H}^1_0(\Omega)$, integration over the domains Ω^i and application of Green's formula for each $t \in (0, T)$

$$
\begin{aligned}
0 = \int_{\Omega^i} \rho^i \ddot{\boldsymbol{\varphi}}^i_t \, \mathbf{v} - \operatorname{div} \mathbf{P}^i_t \, \mathbf{v} \, dx - \int_{\Omega^i} \mathbf{f}_t \, \mathbf{v} \, dx = \\
\int_{\Omega^i} \rho^i \ddot{\boldsymbol{\varphi}}^i_t \, \mathbf{v} + \mathbf{P}^i_t : \nabla \mathbf{v} \, dx - \int_{\Omega^i} \mathbf{f}_t \, \mathbf{v} \, dx - \int_{\Gamma^i_N} \boldsymbol{\pi}_t \, \mathbf{v} \, ds - \int_{\Gamma^i_C} \mathbf{T}^i_C \, \mathbf{v} \, ds, \quad i = 1, 2.
\end{aligned}
$$

Adding up both equations we arrive at the following variational equation

$$
\mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t) = \mathbf{F}^{\mathrm{ext}}(t) + \mathbf{F}^{\mathrm{con}}(\mathbf{T}_C), \tag{1.38}
$$

where the operators $\mathbf{F}^* : \mathbf{H}^1(\Omega) \longrightarrow \mathbb{R}$ representing the internal, external, inertial and contact force terms are defined according to

$$
\mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t, \mathbf{v}) := \sum_{i=1}^2 \int_{\Omega^i} \rho^i \ddot{\boldsymbol{\varphi}}^i_t \, \mathbf{v}^i \, dx, \tag{1.39a}
$$

$$
\mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t, \mathbf{v}) := \sum_{i=1}^2 \int_{\Omega^i} \mathbf{P}^i(\nabla \boldsymbol{\varphi}^i_t) : \nabla \mathbf{v}^i \, dx, \tag{1.39b}
$$

$$
\mathbf{F}^{\mathrm{ext}}(t, \mathbf{v}) := \sum_{i=1}^2 \int_{\Omega^i} \mathbf{f}_t \, \mathbf{v}^i \, dx + \int_{\Gamma^i_N} \boldsymbol{\pi}_t \, \mathbf{v}^i \, ds, \tag{1.39c}
$$

$$
\mathbf{F}^{\mathrm{con}}(\mathbf{T}_C, \mathbf{v}) := \sum_{i=1}^2 \int_{\Gamma^i_C} \mathbf{T}^i_C \, \mathbf{v}^i \, ds. \tag{1.39d}
$$

The contact forces (1.39d) arise from the non-penetration constraints and are a priori unknown in contrast to the external tractions $\boldsymbol{\pi}_t$. To simplify this term we first push-forward the contact tractions $\mathbf{T}_C$ to the deformed configuration $\boldsymbol{\varphi}(\Omega)$

$$
\mathbf{F}^{\mathrm{con}}(\mathbf{t}_C, \mathbf{v}) = \int_{\boldsymbol{\varphi}^1_t(\Gamma^1_C)} \mathbf{t}^1_C \, \mathbf{v}^1 \, ds + \int_{\boldsymbol{\varphi}^2_t(\Gamma^2_C)} \mathbf{t}^2_C \, \mathbf{v}^2 \, ds. \tag{1.40}
$$

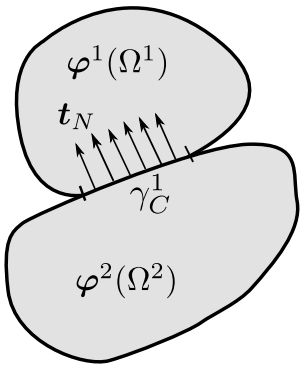


Figure 1.2.: *Contact traction forces at the active contact boundary.*

From the Karush-Kuhn-Tucker conditions (1.37) we know that the contact tractions $\mathbf{t}_C^i(x)$ at some point $x \in \varphi(\Gamma_C)$ vanish if no contact occurs. As a result, we can restrict the contact forces to the active contact boundaries, which we will denote by γ_C^i correspondingly, see Figure 1.2.

$$\mathbf{F}^{\text{con}}(\mathbf{t}_C, \mathbf{v}) = \int_{\gamma_C^1} \mathbf{t}_C^1 \, \mathbf{v}^1 \, ds + \int_{\gamma_C^2} \mathbf{t}_C^2 \, \mathbf{v}^2 \, ds. \tag{1.41}$$

Next, we identify the contact boundaries with each other through the normal projection (1.24). By the balance of forces it holds

$$\mathbf{t}_C^1 \, ds = -\mathbf{t}_C^2 \circ \Phi \, ds.$$

Therefore, we can parametrise the contact forces over γ_C^1 only:

$$\mathbf{F}^{\text{con}}(\mathbf{t}_C, \mathbf{v}) = \int_{\gamma_C^1} \mathbf{t}_C^1 (\mathbf{v}^1 - \mathbf{v}^2 \circ \Phi) \, ds, \tag{1.42}$$

which again, as in (1.34), can be decomposed into a normal and vanishing tangential part

$$\mathbf{F}^{\text{con}}(t_N, \mathbf{v}) = \int_{\gamma_C^1} t_N \mathbf{n}^1 (\mathbf{v}^1 - \mathbf{v}^2 \circ \Phi) \, ds. \tag{1.43}$$

Remark 1.3.1. *The normal contact pressure t_N acts as a Lagrangian multiplier corresponding to the non-penetration constraints, which can be seen from the following equivalent reformulation [87, Section 4.4.1] of the contact forces*

$$\mathbf{F}^{\text{con}}(t_N, \mathbf{v}) = \int_{\gamma_C^1} t_N \delta g(\boldsymbol{\varphi}_t) \, \mathbf{v} \, ds,$$

where the gap g is considered as a function

$$g : \mathbf{H}^1(\Omega) \longrightarrow H^{\frac{1}{2}}(\Gamma_C^1), \quad \boldsymbol{\varphi} \mapsto g(\boldsymbol{\varphi}),$$

cf. Remark 1.2.1, and δg denotes the corresponding directional derivative

$$\delta g(\boldsymbol{\varphi}_t)\,\mathbf{v} := \frac{d}{dh} g(\boldsymbol{\varphi}_t + h\mathbf{v})_{|h=0}.$$

Summarising, the weak equations of motion are given as follows:
Find $\boldsymbol{\varphi} \in C([0,T], \mathbf{H}^1_D(\Omega))$ with weak derivatives

$$\dot{\boldsymbol{\varphi}} \in \mathbf{H}^{\frac{1}{2}}\left((0,T), \mathbf{L}^2(\Omega)\right), \quad \ddot{\boldsymbol{\varphi}} \in L^2((0,T), \mathbf{H}^{-1}(\Omega)), \tag{1.44}$$

and $t_N \in L^2((0,T), H^{-\frac{1}{2}}(\Gamma^1_C))$, such that

$$\mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t) = \mathbf{F}^{\mathrm{ext}}(t) + \mathbf{F}^{\mathrm{con}}(t_N) \quad \text{a.e. on } (0,T), \tag{1.45}$$

where the operators $\mathbf{F}^{\mathrm{dyn}}, \mathbf{F}^{\mathrm{int}}, \mathbf{F}^{\mathrm{ext}}\mathbf{F}^{\mathrm{con}}$ are defined according to (1.39), (1.43) and it holds

$$t_N \leq 0, \quad g(\boldsymbol{\varphi}) \geq 0, \quad g(\boldsymbol{\varphi}) \cdot t_N = 0,$$
$$\boldsymbol{\varphi}_0 = \hat{\boldsymbol{\varphi}}_0, \quad \dot{\boldsymbol{\varphi}}_0 = \mathbf{v}_0 \quad \text{on } \Omega.$$

Existence and uniqueness of solutions

The existence and uniqueness of solutions of dynamic contact problems is still an open question. In the linearised framework of small displacements, cf. Remark 1.2.2, the existence and uniqueness for the weak formulation is well known [117, Lemma 3.2.1]. First results on contact problems with friction can be found in [138, 36]. Recently Schuricht derived a general existence proof for static large deformation contact problems between a poly-convex hyperelastic continuum and a rigid obstacle [122, Theorem 3.3] using the concept of generalised derivatives and non-smooth calculus [27]. Generalising this approach to an abstract two-body-contact formulation, combined with the non-smooth Hamilton principle that we will introduce in the upcoming section could lead to new insights regarding the existence of minimisers.

Conservation of momentum and energy

In this paragraph we prove conservation of linear and angular momentum of the weak formulation (1.45) in the absence of external forces. Afterwards we will show that the total energy is conserved if the persistency condition holds, cf. [88].

We assume that the external body and traction forces vanish, $\mathbf{f}_t = \boldsymbol{\pi}_t = 0$ and that there are no Dirichlet boundary conditions $\Gamma_D = \emptyset$. The total linear and angular momentum

of the system are defined by

$$K_t := \int_{\Omega^1} \rho^1 \dot{\boldsymbol{\varphi}}_t^1 \, dx + \int_{\Omega^2} \rho^2 \dot{\boldsymbol{\varphi}}_t^2 \, dx, \tag{1.46}$$

$$J_t := \int_{\Omega^1} \rho^1 \boldsymbol{\varphi}_t^1 \times \dot{\boldsymbol{\varphi}}_t^1 \, dx + \int_{\Omega^2} \rho^2 \boldsymbol{\varphi}_t^2 \times \dot{\boldsymbol{\varphi}}_t^2 \, dx, \tag{1.47}$$

where $\dot{\boldsymbol{\varphi}}_t^i$ denote the material velocities.

Proposition 1.3.2 (Momentum conservation)**.**
(i) The total linear momentum is conserved

$$\frac{d}{dt} K_t = 0.$$

(ii) The total angular momentum is conserved

$$\frac{d}{dt} J_t = 0.$$

Proof. (i) To show the conservation of linear momentum we first insert a translation, i.e. a spatially constant function $v(x) = c \in \mathbb{R}^d$ as a test function into Equation (1.45)

$$\begin{aligned} 0 &= \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t, c) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t, c) - \mathbf{F}^{\mathrm{con}}(t_N, c) \\ &= \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t, c) + \int_{\gamma_C^1} t_N \mathbf{n}^1 (c - c) \, ds \\ &= \left(\sum_{i=1}^{2} \int_{\Omega^i} \rho^i \ddot{\boldsymbol{\varphi}}_t^i \, dx \right) \cdot c \\ &= \left(\frac{d}{dt} K_t \right) \cdot c, \end{aligned}$$

which implies $\frac{d}{dt} K_t = 0$ because $c \in \mathbb{R}^d$ is arbitrary.

(ii) Similar for the conservation of angular momentum we insert the time-dependent test function $\mathbf{v} = c \times \boldsymbol{\varphi}_t, c \in \mathbb{R}^d$ into (1.45)

$$0 = \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t, \mathbf{v}) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t, \mathbf{v}) - \mathbf{F}^{\mathrm{con}}(t_N, \mathbf{v}). \tag{1.48}$$

We will now investigate these terms in more detail separately. A straight-forward calculation yields

$$\begin{aligned} \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t, \mathbf{v}) &= \int_\Omega \rho \ddot{\boldsymbol{\varphi}}_t \cdot (c \times \boldsymbol{\varphi}_t) \, dx = \int_\Omega \rho \boldsymbol{\varphi}_t \times \ddot{\boldsymbol{\varphi}}_t \, dx \cdot c \\ &= \int_\Omega \rho (\dot{\boldsymbol{\varphi}}_t \times \dot{\boldsymbol{\varphi}}_t + \boldsymbol{\varphi}_t \times \ddot{\boldsymbol{\varphi}}_t) \, dx \cdot c \\ &= \left(\frac{d}{dt} J_t \right) \cdot c. \end{aligned}$$

Before we show that the internal force vanishes, we first bring to mind that the cross product can equivalently be written as a matrix-vector multiplication (here for $d = 3$)

$$c = a \times b \Longleftrightarrow c = \hat{A}b \tag{1.49}$$

where the skew-symmetric matrix $\hat{A}$ is given by

$$\hat{A} := \begin{pmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{pmatrix}.$$

Now one can deduce

$$\begin{aligned} \mathbf{F}^{\text{int}}(\boldsymbol{\varphi}_t, \mathbf{v}) &= \int_\Omega \mathbf{P}(\nabla\boldsymbol{\varphi}_t) : \nabla(c \times \boldsymbol{\varphi}_t)\, dx = \int_\Omega \mathbf{P}(\nabla\boldsymbol{\varphi}_t) : \hat{C}\, \nabla\boldsymbol{\varphi}_t\, dx \\ &= \int_\Omega \nabla\boldsymbol{\varphi}_t \boldsymbol{\Sigma}_t : \hat{C}\, \nabla\boldsymbol{\varphi}_t\, dx = \int_\Omega \nabla\boldsymbol{\varphi}_t \boldsymbol{\Sigma}_t \nabla\boldsymbol{\varphi}_t^T :\, \hat{C}\, dx = 0, \end{aligned}$$

where $\boldsymbol{\Sigma}_t$ is the symmetric second Piola–Kirchhoff tensor (1.10). The third equality follows from basic tensor contraction rules. As $\nabla\boldsymbol{\varphi}_t \boldsymbol{\Sigma}_t \nabla\boldsymbol{\varphi}_t^T$ is symmetric and $\hat{c}$ is skew-symmetric, their contraction is zero. The contact force term also vanishes by definition of gap function

$$\begin{aligned} \mathbf{F}^{\text{con}}(t_N, \mathbf{v}) &= \int_{\gamma_C^1} t_N \mathbf{n}^1 \cdot (c \times \boldsymbol{\varphi}_t^1 - c \times \boldsymbol{\varphi}_t^2 \circ \Phi)\, ds \\ &= c \cdot \int_{\gamma_C^1} t_N \mathbf{n}^1 \times (\boldsymbol{\varphi}_t^1 - \boldsymbol{\varphi}_t^2 \circ \Phi)\, ds = 0, \end{aligned}$$

where the last equality follows from the collinearity of normal gap $(\boldsymbol{\varphi}_t^1 - \boldsymbol{\varphi}_t^2 \circ \Phi)$ and the outer normal $\mathbf{n}^1$, cf.(1.24). Combining all steps then leads to the conservation of angular momentum. □

To prove energy conservation one further has to assume that the so-called *persistency condition* holds

$$\dot{g} \cdot t_N = 0. \tag{1.50}$$

This assumption states that non-zero contact forces only occur during persistent contact. Many time discretisations for contact problem ensure energy conservation by enforcing this condition algorithmically [7, 66, 88]. Nevertheless, this assumption is considered to be invalid for many applications, see [87]. For simplicity we restrict ourselves to the case of vanishing external forces, the result can be extended to the case of dead loads, see [87, Section 7.2.1].

Proposition 1.3.3 (Energy conservation)**.** *If the persistency condition* (1.50) *holds and the external forces vanish, then the total energy*

$$E(\boldsymbol{\varphi}_t, \dot{\boldsymbol{\varphi}}_t) := E^{\text{kin}}(\dot{\boldsymbol{\varphi}}_t) + E^{\text{pot}}(\boldsymbol{\varphi}_t), \tag{1.51}$$

where

$$E^{\mathrm{kin}}(\dot{\boldsymbol{\varphi}}_t) := \frac{1}{2}\sum_{i=1}^{2}\int_{\Omega^i}\rho^i|\dot{\boldsymbol{\varphi}}_t^i|^2\,dx,$$
$$E^{\mathrm{pot}}(\boldsymbol{\varphi}_t) := \sum_{i=1}^{2}\int_{\Omega^i}W^i(\nabla\boldsymbol{\varphi}_t^i)\,dx,$$

is conserved.

Proof. We insert the material velocity $\dot{\boldsymbol{\varphi}}_t$ as a test function in (1.45)

$$\begin{aligned}0 &= \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t,\dot{\boldsymbol{\varphi}}_t) + \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t,\dot{\boldsymbol{\varphi}}_t) - \mathbf{F}^{\mathrm{con}}(t_N,\dot{\boldsymbol{\varphi}}_t)\\ &= \frac{d}{dt}E^{\mathrm{pot}}(\boldsymbol{\varphi}_t) + \frac{d}{dt}E^{\mathrm{kin}}(\dot{\boldsymbol{\varphi}}_t) - \mathbf{F}^{\mathrm{con}}(t_N,\dot{\boldsymbol{\varphi}}_t),\end{aligned}$$

where the equality follows from application of the constitutive law (1.11) and the chain rule. Due to the persistency condition the contact force term vanishes

$$\mathbf{F}^{\mathrm{con}}(t_N,\dot{\boldsymbol{\varphi}}_t) = \int_{\gamma_C^1} t_N\,\mathbf{n}^1\cdot(\dot{\boldsymbol{\varphi}}_t^1 - \dot{\boldsymbol{\varphi}}_t^2\circ\Phi)\,ds = \int_{\gamma_C^1} t_N\,\dot{g}\,ds = 0,$$

which concludes the proof. □

A variationally consistent formulation of non-penetration

We now derive an equivalent variational formulation of the non-penetration constraint (1.26), which has been done for the case of small deformations in [138, Lemma 2.1]. Let us consider the set of admissible deformations

$$\mathcal{K} := \left\{\boldsymbol{\varphi}\in\mathbf{H}_D^1(\Omega):\ g(\boldsymbol{\varphi})\geq 0 \quad \text{a.e. on } \Gamma_C^1\right\}, \tag{1.52}$$

where we assume that the normal projection $\Phi : \Gamma_C^1 \longrightarrow \Gamma_C^2$ and the normal field $\mathbf{n}^1 : \Gamma_C^1 \longrightarrow \mathbb{R}^d$ are smooth enough such that the gap function maps into the trace space $H^{\frac{1}{2}}(\Gamma_C^1)$

$$\mathbf{H}_D^1(\Omega)\ni\boldsymbol{\varphi}\mapsto g(\boldsymbol{\varphi})\in H^{\frac{1}{2}}(\Gamma_C^1). \tag{1.53}$$

Note that as we assumed the contact boundary to be compactly embedded in $\partial\Omega\setminus\Gamma_D^1$, it suffices to consider the standard trace space and its dual

$$W := H^{\frac{1}{2}}(\Gamma_C^1),$$
$$M := H^{\frac{1}{2}}(\Gamma_C^1)'.$$

The cone of positive functions and dual functionals is denoted by

$$W^+ := \{v \in W : v \geq 0\},$$
$$M^+ := \left\{\mu \in M : \langle \mu, v\rangle_{M\times W} \geq 0,\, \forall\, v \in W^+\right\},$$

where $\langle\cdot,\cdot\rangle_{M\times W}$ denotes the dual paring of the trace space. We will just write $\langle\cdot,\cdot\rangle$ if the spaces are clear from the context.

Proposition 1.3.4. *Let the non-linear function $b : M \times \mathbf{H}^1(\Omega) \longrightarrow \mathbb{R}$ be given by*

$$b(\mu, \boldsymbol{\varphi}) := \langle \mu, g(\boldsymbol{\varphi})\rangle. \tag{1.54}$$

Then, the feasible set (1.52) *is characterised by*

$$\mathcal{K} = \left\{\boldsymbol{\varphi} \in \mathbf{H}^1_D(\Omega) :\, b(\mu, \boldsymbol{\varphi}) \geq 0, \quad \forall \mu \in M^+\right\}. \tag{1.55}$$

Proof.
(i) If $g(\boldsymbol{\varphi}) \geq 0$ in the L^2 sense, then by definition $b(\mu, \boldsymbol{\varphi}) \geq 0$ for all $\mu \in M^+$.

(ii) Now assume that for some $\boldsymbol{\varphi} \in \mathbf{H}^1_D(\Omega)$ it holds that $b(\mu, \boldsymbol{\varphi}) \geq 0$ for all $\mu \in M^+$. To derive a contradiction suppose that $g(\boldsymbol{\varphi}) < 0$ on some subset $Y \subset \Gamma^1_C$ with $\mathrm{meas}(Y) > 0$. Let $\chi_Y \in L^2(\Gamma^1_C)^+ \subseteq M^+$ denote the characteristic function of the set Y

$$\chi_Y(s) = \begin{cases} 1 & s \in Y, \\ 0 & \text{else.} \end{cases} \tag{1.56}$$

Then

$$b(\chi_Y, \boldsymbol{\varphi}) = \langle \chi_Y, g(\boldsymbol{\varphi})\rangle = \int_{\Gamma^1_C} \chi_Y\, g(\boldsymbol{\varphi})\, ds = \int_Y g(\boldsymbol{\varphi})\, ds < 0 \tag{1.57}$$

which contradicts the assumption. Above we used the fact that

$$W \subseteq L^2(\Gamma^1_C) \cong (L^2(\Gamma^1_C))' \subseteq M \tag{1.58}$$

is a Gelfand triple. Hence, it holds for $f \in (L^2(\Gamma^1_C))'$ and $u \in W$

$$\langle f, u\rangle_{M\times W} = \langle f, u\rangle_{(L^2)'\times L^2} = \int_{\Gamma^1_C} f\, u\, ds,$$

see [61, Section 6.3.3]. □

1.4. A non-smooth Hamilton principle

In this paragraph we introduce a way to derive weak equations of motion for general Hamiltonian systems with inequality constraints. Therefore, we apply a non-smooth

Hamilton principle, first postulated in [70] without providing the correct non-smooth calculus to derive the resulting weak equations. Here we will introduce this principle in more detail and properly derive the weak formulation using the theory of generalised gradients and Fréchet subdifferentials [99, 100]. This approach has the advantage that the Lagrange multipliers, i.e. the contact pressure t_N, are treated completely implicitly. Consequently, only constrained problems for the primal variable have to be solved. Following Marsden et al. [70, 96] we first define the *Lagrangian* of the mechanical system as the difference of kinetic and potential energy

$$L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) := E^{\mathrm{kin}}(\dot{\boldsymbol{\varphi}}) - E^{\mathrm{pot}}(\boldsymbol{\varphi}) = \frac{1}{2}\int_\Omega \rho\,|\dot{\boldsymbol{\varphi}}|^2\,dt - \int_\Omega W(\boldsymbol{\varphi})\,dx, \tag{1.59}$$

where W denotes the hyperelastic strain energy (1.11). For mechanical systems with smooth Lagrangian function Hamilton's principle states:

Principle 1 (Hamilton's principle)**.** *The true evolution $\boldsymbol{\varphi} : \Omega \times [0,T] \longrightarrow \mathbb{R}^d$ of a mechanical system between two fixed states $\boldsymbol{\varphi}(\cdot,0) = \boldsymbol{\varphi}_0$, $\boldsymbol{\varphi}(\cdot,T) = \boldsymbol{\varphi}_T$ renders the* ***action integral*** *stationary*

$$0 = \delta\mathcal{I}(\boldsymbol{\varphi}) := \delta\int_0^T L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\,dt, \tag{1.60}$$

where δ denotes the first variation

$$\delta\mathcal{I}(\boldsymbol{\varphi})\mathbf{v} = \lim_{h\longrightarrow 0}\frac{\mathcal{I}(\boldsymbol{\varphi} + h\,\mathbf{v}) - \mathcal{I}(\boldsymbol{\varphi})}{h}. \tag{1.61}$$

If the Lagrangian is smooth the first variation of the action integral can be characterised by calculating the directional derivative along smooth perturbations $\mathbf{v} \in C_0^\infty([0,T],\mathbb{R})$ with zero boundary values $\mathbf{v}(0) = \mathbf{v}(T) = 0$,

$$\begin{aligned}\delta\mathcal{I}(\boldsymbol{\varphi})\mathbf{v} &= \int_0^T \left(\frac{\partial L}{\partial \boldsymbol{\varphi}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\,\mathbf{v} + \frac{\partial L}{\partial \dot{\boldsymbol{\varphi}}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\,\dot{\mathbf{v}}\right) dt,\\ &= \int_0^T \left(\frac{\partial L}{\partial \boldsymbol{\varphi}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{\varphi}}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\right)\mathbf{v}\,dt,\end{aligned} \tag{1.62}$$

where the second equation is obtained from Green's formula. By assumption the first variation is vanishing for all test functions $\mathbf{v} \in C_0^\infty([0,T])$ which implies the *Euler–Lagrange* equation

$$\frac{\partial L}{\partial \boldsymbol{\varphi}}(\boldsymbol{\varphi}_t, \dot{\boldsymbol{\varphi}}_t) - \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{\varphi}}}(\boldsymbol{\varphi}_t, \dot{\boldsymbol{\varphi}}_t) = 0 \quad \text{a.e. in } (0,T) \tag{1.63}$$

Remark 1.4.1. *As expected, in the case of non-linear hyperelasticity the Euler–Lagrange equation coincide with the weak equation derived in the classical way, cf. Section 1.3*

$$\mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}_t) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}_t) = \mathbf{F}^{\mathrm{ext}}(t) \quad \textit{a.e. in } (0,T). \tag{1.64}$$

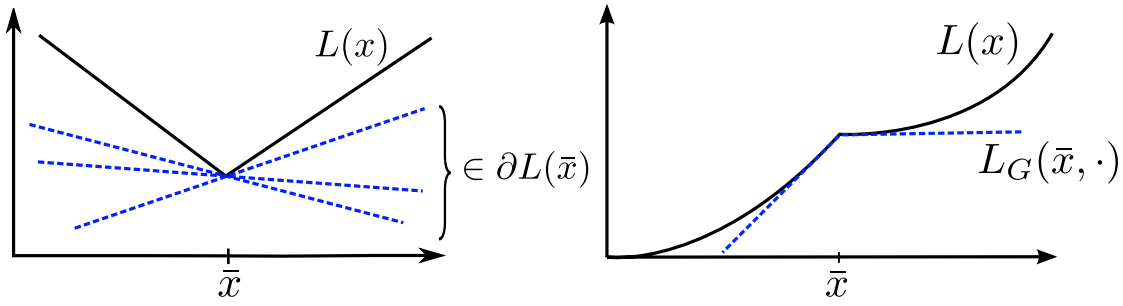


Figure 1.3.: *Left: Subgradients of a convex function. Right: Directional Gateaux derivative of a non-convex function.*

Next, we extend this principle to constrained non-convex mechanical systems like the one at hand. Following the idea of Kane et al. [70], we consider the *indicator function* of the set of admissible deformations (1.55)

$$I_{\mathcal{K}}(\boldsymbol{\varphi}) := \begin{cases} 0 & \boldsymbol{\varphi} \in \mathcal{K}, \\ \infty & \text{else}, \end{cases} \tag{1.65}$$

and incorporate it into the potential energy of the mechanical system. This results in the non-smooth Lagrangian

$$\bar{L}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) := L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) - I_{\mathcal{K}}(\boldsymbol{\varphi}). \tag{1.66}$$

This can be seen as an exact infinite penalisation to ensure that the trajectory of the mechanical system stays within the feasible set. In the following we repeat some results of non-smooth calculus and generalised derivatives which will be useful for the construction of a framework for non-differentiable Lagrangians. For convex functions a well-established extension of a differential is given by the concept of subdifferentials, see Ekeland and Temam [38].

Definition 1.4.2. Let $L : \mathbf{H}_0^1(\Omega) \longrightarrow \mathbb{R}$ be convex. A *subgradient* at $\boldsymbol{\varphi} \in \mathbf{H}_0^1(\Omega)$ is a linear functional $\mathbf{f} \in \mathbf{H}^{-1}(\Omega) := (\mathbf{H}_0^1(\Omega))'$ such that

$$\langle \mathbf{f}, \mathbf{v} \rangle \leq L_G(\boldsymbol{\varphi}, \mathbf{v}) := \lim_{h \downarrow 0} \frac{L(\boldsymbol{\varphi} + h\,\mathbf{v}) - L(\boldsymbol{\varphi})}{h} \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega), \tag{1.67}$$

where L_G is the *directional Gateaux derivative* of L.

The *convex subdifferential* is defined as the collection of all those functionals

$$\partial L(\boldsymbol{\varphi}) := \left\{ \mathbf{f} \in \mathbf{H}^{-1}(\Omega) : \langle \mathbf{f}, \mathbf{v} \rangle \leq L_G(\boldsymbol{\varphi}, \mathbf{v}) \quad \forall\, \mathbf{v} \in \mathbf{H}_0^1(\Omega) \right\}. \tag{1.68}$$

In Figure 1.3(left) this concept is illustrated. When the function is smooth the subdifferential reduces to the classical derivative $\partial L(\varphi) = \{L'(\boldsymbol{\varphi})\}$, see [121, Proposition 4.1.8]; therefore, this concept is a real extension of differentiability.
A generalised criterion to characterise minimal points is given as follows:

Proposition 1.4.3. *Let $L : \mathbf{H}_0^1(\Omega) \longrightarrow \mathbb{R}$ be convex. Then $\boldsymbol{\varphi}^* \in \mathbf{H}_0^1(\Omega)$ is a minimizer of L if and only if*

$$0 \in \partial L(\boldsymbol{\varphi}^*). \tag{1.69}$$

Proof. [111, Theorem 10.1]

For convex functions the directional Gateaux derivative always exists, even if the function itself is not differentiable ([121, Theorem 4.1.3]). In the non-convex case the subdifferential defined by (1.68) does in general not lead to a satisfactory calculus, as e.g. $\partial L(\boldsymbol{\varphi}) = \emptyset$, cf. Figure 1.3(right). A first generalisation of this concept to non-convex, local Lipschitz functions was proposed by Clarke. He replaced the directional Gateaux derivative, which does not need to exist in this case, by an upper convex sub-linear approximation ([121, Definition 7.3.1],[27]). Kruger and Mordukhovich [99] then extended this framework to arbitrary non-Lipschitzian functions, like $I_{\mathcal{K}}$, introducing the *Fréchet subdifferential* [52]

$$\partial^F L(\boldsymbol{\varphi}) := \left\{ \mathbf{f} \in \mathbf{H}^{-1}(\Omega) : \liminf_{\mathbf{v} \longrightarrow 0} \frac{L(\boldsymbol{\varphi} + \mathbf{v}) - L(\boldsymbol{\varphi}) - \langle \mathbf{f}, \mathbf{v} \rangle}{\|\mathbf{v}\|_1} \geq 0 \right\}. \tag{1.70}$$

For this subdifferential only approximate calculus rules hold in general, e.g. sum- or chain rules([121, Section 9.2]) and it is strongly linked to lower semi-continuous functions:

Lemma 1.4.4. *Let $L : \mathbf{H}_0^1(\Omega) \longrightarrow \bar{R} := \mathbb{R} \cup \{-\infty, \infty\}$ and $\boldsymbol{\varphi} \in \mathbf{H}_0^1(\Omega)$. If $\partial^F L(\boldsymbol{\varphi}) \neq \emptyset$ then L is lower semi-continuous at $\boldsymbol{\varphi}$.*

Proof. [82, Proposition 1.7]

Remark 1.4.5. *A richer and exact calculus was established recently for the Mordukhovich subdifferential which is based on generalised normals and extremal principles [99, 100]. In our case the non-smoothness arises from the indicator function $I_{\mathcal{K}}$ which is lower semi-continuous for closed $\mathcal{K}$. Consequently we restrict ourselves to Fréchet subdifferential.*

The indicator function enters in the potential energy and is subtracted from the Lagrangian. Thus it is suitable to use an upper approximation of the functional to extend the Hamilton principle. The upper construction corresponding to the Fréchet subdifferential is called *Fréchet superdifferential*.

$$\partial_F L(\boldsymbol{\varphi}) := \left\{ \mathbf{f} \in \mathbf{H}^{-1}(\Omega) : \limsup_{\mathbf{v} \longrightarrow 0} \frac{L(\boldsymbol{\varphi} + \mathbf{v}) - L(\boldsymbol{\varphi}) - \langle \mathbf{f}, \mathbf{v} \rangle}{\|\mathbf{v}\|_1} \leq 0 \right\}. \tag{1.71}$$

The relation between both differentials is given by

$$\partial_F L(\boldsymbol{\varphi}) = -\partial^F(-L), \tag{1.72}$$

see [99, Definition 1.86] and it holds

Proposition 1.4.6. *Let $L : \mathbf{H}_0^1(\Omega) \longrightarrow \overline{\mathbb{R}}$ and $\boldsymbol{\varphi}$ such that $L(\boldsymbol{\varphi})$ is finite. Then $\partial^F L(\boldsymbol{\varphi}) \neq \emptyset$ and $\partial_F L(\boldsymbol{\varphi}) \neq \emptyset$ if and only if L is Fréchet differentiable at $\boldsymbol{\varphi}$. Then $\partial^F L(\boldsymbol{\varphi}) = \partial_F L(\boldsymbol{\varphi}) = \{L'(\boldsymbol{\varphi})\}$.*

Proof. [99, Proposition 1.87].

Moreover, characterisation of critical points is given by the Fermat rule

Proposition 1.4.7 (Generalised Fermat rule)**.** *Let $L : \mathbf{H}_0^1(\Omega) \longrightarrow \overline{\mathbb{R}}$ and $\boldsymbol{\varphi}^* \in \mathbf{H}_0^1(\Omega)$ be a locally stationary point with $L(\boldsymbol{\varphi}^*)$ finite. Then*

$$0 \in \partial^F L(\boldsymbol{\varphi}^*) \cup \partial_F L(\boldsymbol{\varphi}^*). \tag{1.73}$$

Proof. [99, Proposition 1.114].

The generalised Fermat rule now allows us to extend Hamilton's principle to non-smooth mechanical systems.

Principle 2 (Non-smooth Hamilton principle)**.** *The true evolution $\boldsymbol{\varphi} : \Omega \times [0,T] \longrightarrow \mathbb{R}^d$ of a mechanical system between two fixed states $\boldsymbol{\varphi}(\cdot,0) = \boldsymbol{\varphi}_0$, $\boldsymbol{\varphi}(\cdot,T) = \boldsymbol{\varphi}_T$ renders the action integral stationary*

$$0 \in \partial^F \mathcal{I}(\boldsymbol{\varphi}) \cup \partial_F \mathcal{I}(\boldsymbol{\varphi}). \tag{1.74}$$

The corresponding equations of motion for Lagrangians of the form (1.66) can be characterised as follows:

Theorem 1.4.8. *Stationary points of the action integral, where the non-smooth Lagrangian is given by* (1.66)*, fulfil the following* ***Euler–Lagrange inclusion***

$$0 \in \frac{d}{dt}\frac{\partial L}{\partial \dot{\boldsymbol{\varphi}}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) - \frac{\partial L}{\partial \boldsymbol{\varphi}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) + \partial I_{\mathcal{K}}(\boldsymbol{\varphi}) \quad \text{a.e. in } (0,T). \tag{1.75}$$

Proof. To prove the proposition we need the following sum rule:

Proposition 1.4.9 (Sum rule)**.** *Let $L, I_{\mathcal{K}} : \mathbf{H}_0^1(\Omega) \longrightarrow \overline{\mathbb{R}}$ be finite at $\boldsymbol{\varphi} \in \mathbf{H}_0^1(\Omega)$ and L be Fréchet differentiable at $\boldsymbol{\varphi}$, then*

$$\partial^F (L + I_{\mathcal{K}})(\boldsymbol{\varphi}) = L'(\boldsymbol{\varphi}) + \partial^F I_{\mathcal{K}}(\boldsymbol{\varphi}). \tag{1.76}$$

Proof. [99, Proposition 1.107]

By Proposition 1.4.6 exactly one of $\partial^F \mathcal{I}(\boldsymbol{\varphi})$ and $\partial_F \mathcal{I}(\boldsymbol{\varphi})$ is empty when $\mathcal{I}$ is not differentiable at $\boldsymbol{\varphi}$. We will thus consider both differentials separately

$$0 \in \partial^F \mathcal{I}(\boldsymbol{\varphi}), \quad 0 \in \partial_F \mathcal{I}(\boldsymbol{\varphi}).$$

First we apply the sum rule to the principle (1.74) considering only the Fréchet subdifferential:

$$\partial^F \mathcal{I}(\boldsymbol{\varphi}) = \left(\int_0^T L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\, dt \right)' + \partial^F \int_0^T -I_{\mathcal{K}}(\boldsymbol{\varphi}(t))\, dt. \tag{1.77}$$

As $-I_{\mathcal{K}}(\cdot)$ is upper semi-continuous ($\mathcal{K}$ is closed), the function

$$\boldsymbol{\varphi} \mapsto \int_0^T -I_{\mathcal{K}}(\boldsymbol{\varphi}(t))\, dt$$

is as well upper semi-continuous by Fatou's lemma. By Lemma 1.4.4 we conclude that $\partial^F \mathcal{I}(\boldsymbol{\varphi}) = \emptyset$ whenever $\boldsymbol{\varphi}(t) \in \partial\mathcal{K}$ for some $t \in (0,T)$. If $\boldsymbol{\varphi}$ stays in the interior of the feasible set $\mathcal{K}$ at all times, then the superdifferential and subdifferential coincide. Hence, we can restrict ourselves to the superdifferential

$$0 \in \partial_F \mathcal{I}(\boldsymbol{\varphi}) = -\partial^F(-\mathcal{I})(\boldsymbol{\varphi}) \Longleftrightarrow 0 \in \partial^F(-\mathcal{I})(\boldsymbol{\varphi}), \tag{1.78}$$

Applying the sum rule to (1.78) yields

$$\partial^F(-\mathcal{I})(\boldsymbol{\varphi}) = \left(\int_0^T -L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\, dt \right)' + \partial^F \int_0^T I_{\mathcal{K}}(\boldsymbol{\varphi}(t))\, dt. \tag{1.79}$$

The derivative of the smooth part is characterised by

$$\left(\int_0^T -L(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}})\, dt \right)' \mathbf{v} = \int_0^T \left(\frac{d}{dt} \frac{\partial L}{\partial \dot{\boldsymbol{\varphi}}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) - \frac{\partial L}{\partial \boldsymbol{\varphi}}(\boldsymbol{\varphi}, \dot{\boldsymbol{\varphi}}) \right) \mathbf{v}\, dt, \tag{1.80}$$

cf. (1.63). The second term is given as

$$\int_0^T I_{\mathcal{K}}(\boldsymbol{\varphi}(t))\, dt = \begin{cases} 0 & \boldsymbol{\varphi} \in \mathcal{K} \text{ a.e. in } (0,T), \\ \infty & \text{else.} \end{cases} \tag{1.81}$$

Combining these two equations yields the result. □

For the large deformation contact problem the Euler–Lagrange inclusion is given by

$$0 \in \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}(t)) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}(t)) - \mathbf{F}^{\mathrm{ext}}(t) + \partial^F I_{\mathcal{K}}(\boldsymbol{\varphi}(t)), \tag{1.82}$$

with $\mathcal{K}$ given by (1.52) and the operators $\mathbf{F}^{\mathrm{dyn}}, \mathbf{F}^{\mathrm{int}}, \mathbf{F}^{\mathrm{ext}}$ are defined as in the classical weak formulation, cf. (1.39a) and Remark 1.4.1.
Now we will show that for smooth contact boundaries solutions of the classical weak form (1.45) are also solutions of the differential inclusion (1.82).

Theorem 1.4.10. *Let $(\bar{\boldsymbol{\varphi}}, t_N)$ be a solution of the classical weak equations of motion* (1.45) *and the contact boundaries be continuously differentiable, then $\bar{\boldsymbol{\varphi}}$ is also a solution of the inclusion* (1.82).

To prove the theorem we have to establish that the contact force term

$$\mathbf{F}^{\mathrm{con}}(t_N, \mathbf{v}) = \int_{\gamma_C^1} t_N \delta g(\boldsymbol{\varphi}) \mathbf{v} \, ds.$$

is contained in the Fréchet subdifferential $\partial^F I_{\mathcal{K}}(\bar{\varphi})$. Therefore, we first pull-back the contact forces on the reference domain

$$\mathbf{F}^{\mathrm{con}}(T_N, \mathbf{v}) = \int_{\Gamma_C^1} T_N \delta g(\boldsymbol{\varphi}) \mathbf{v} \, ds. \tag{1.83}$$

Next, we reformulate the non-penetration constraint (1.26) as a non-smooth scalar function

$$G(\boldsymbol{\varphi}) := \operatorname*{ess\,sup}_{\Gamma_C^1} -g(\boldsymbol{\varphi}, \cdot), \tag{1.84}$$

where g is the normal gap function

$$g(\boldsymbol{\varphi}, x) := -\mathbf{n}^1(x) \cdot (\boldsymbol{\varphi}^1(x) - \boldsymbol{\varphi}^2 \circ \Phi(x)),$$

cf. (1.25). By definition it holds

$$G(\boldsymbol{\varphi}) \leq 0 \Longleftrightarrow g(\boldsymbol{\varphi}, \cdot) \geq 0 \quad \text{a.e. on } \Gamma_C^1.$$

For the proof we need the following two results

Proposition 1.4.11. *Let $G : \mathbf{H}_0^1(\Omega) \longrightarrow \mathbb{R} \cup \infty$ be proper and locally Lipschitz on a neighbourhood of $\bar{\boldsymbol{\varphi}} \in \mathbf{H}^1(\Omega)$ with $G(\bar{\boldsymbol{\varphi}}) = 0$. Set*

$$\mathcal{K} := \left\{ \boldsymbol{\varphi} \in \mathbf{H}^1(\Omega) : \, G(\boldsymbol{\varphi}) \leq 0 \right\}.$$

Moreover, assume that $0 \notin \partial^C G(\bar{\boldsymbol{\varphi}})$, where ∂^C denotes the ***Clarke subdifferential***

$$\partial^C G(\boldsymbol{\varphi}) := \left\{ \mathbf{f} \in \mathbf{H}^{-1}(\Omega) : \, \langle \mathbf{f}, \mathbf{v} \rangle \leq \limsup_{\substack{h \downarrow 0 \\ \mathbf{w} \to \mathbf{v}}} \frac{1}{h} (G(\boldsymbol{\varphi} + h\mathbf{w}) - G(\boldsymbol{\varphi})) \right\}. \tag{1.85}$$

Then

$$\partial^F I_{\mathcal{K}}(\bar{\boldsymbol{\varphi}}) \subseteq \mathbb{R}^+ \partial^C G(\bar{\boldsymbol{\varphi}}). \tag{1.86}$$

If in addition G is strictly differentiable at $\bar{\varphi}$, then (1.86) *holds with equality.*

Proof. [121, Theorem 11.6.1] together with the fact that

$$\partial^F I_{\mathcal{K}}(\boldsymbol{\varphi}) \subset \partial^C I_{\mathcal{K}}(\boldsymbol{\varphi}), \tag{1.87}$$

see [99, p. 14–17]. □

The Clarke subdifferential of (1.84) can be characterised as follows

Proposition 1.4.12. *Let $\bar{\boldsymbol{\varphi}} \in \mathbf{H}^1_0(\Omega)$ and $\Gamma \subset \mathbb{R}^d$ be compact. Let $g(\bar{\boldsymbol{\varphi}}, \cdot)$ be upper semicontinuous on a convex neighbourhood U of $\bar{\boldsymbol{\varphi}}$ and $g(\cdot, x)$ be Lipschitz on U. Assume that the set $\{g(\bar{\boldsymbol{\varphi}}, x) : x \in \Gamma\}$ is essentially bounded. Then, the Clarke subdifferential of*

$$G(\boldsymbol{\varphi}) := \operatorname*{ess\,sup}_{\Gamma} g(\boldsymbol{\varphi}, \cdot),$$

at $\bar{\boldsymbol{\varphi}}$ is given by

$$\partial^C G(\bar{\boldsymbol{\varphi}}) = \left\{ \int_\Gamma \partial^C g(\bar{\boldsymbol{\varphi}}, x)\, \mu(dx) : \mu \in P[M(\bar{\boldsymbol{\varphi}})] \right\}, \tag{1.88}$$

where

$$M(\boldsymbol{\varphi}) = \{x \in \Gamma : g(\boldsymbol{\varphi}, x) = G(\boldsymbol{\varphi})\},$$

and $P[M(\boldsymbol{\varphi})]$ is the set of probability Radon measures with support $M(\boldsymbol{\varphi})$. If $g(\cdot, x)$ is strictly continuously differentiable on a neighbourhood U and $g(\cdot, \cdot)$ is continuous, then

$$\partial^C G(\bar{\boldsymbol{\varphi}}) = \left\{ \int_\Gamma \delta g(\bar{\boldsymbol{\varphi}}, x)\, \mu(dx) : \mu \in P[M(\bar{\boldsymbol{\varphi}})] \right\}, \tag{1.89}$$

where again δ denotes the first variation, cf. Remark 1.3.1.

Proof. [27, Theorem 2.8.2]

Remark 1.4.13. *The subgradients $\mathbf{f} \in \partial^C G(\boldsymbol{\varphi})$ defined above act as functionals in the following way*

$$\langle \mathbf{f}, \mathbf{v} \rangle = \int_\Gamma \delta g(\boldsymbol{\varphi})\, \mathbf{v}\, \mu_{\mathbf{f}}(dx).$$

Proof of Theorem 1.4.10.

(i) If no active contact occurred, i.e. $t_N = 0$ and $g(\bar{\boldsymbol{\varphi}}) > 0$, then $\partial^F I_{\mathcal{K}}(\bar{\boldsymbol{\varphi}}) = \{0\}$ and both formulations reduce to the problem of non-linear elasticity.

(ii) From now on let $\bar{\boldsymbol{\varphi}}$ be such that $G(\bar{\boldsymbol{\varphi}}) = 0$. If the contact boundaries are smooth enough, then the normal projection is continuously differentiable for all $x \in \Gamma^1_C$ and thus $G(\boldsymbol{\varphi})$ is locally Lipschitz. Hence, we can apply Propositions 1.4.11 and 1.4.12 and obtain an explicit representation of the Clarke subdifferential in terms of probability Radon measures

$$\partial^C G(\bar{\boldsymbol{\varphi}}) = \left\{ \int_{\Gamma^1_C} -\delta g(\bar{\boldsymbol{\varphi}}, x)\, \mu(dx) : \mu \in P[M(\bar{\boldsymbol{\varphi}})] \right\},$$

where $M(\bar{\boldsymbol{\varphi}}) \subseteq \Gamma^1_C$ is the part of contact boundary where active contact occurs $g(\bar{\boldsymbol{\varphi}}, x) = 0$. Finally, we can use the fact that in $\mathbb{R}^d$ the Lebesgue measure, which we denote by dx, is

a Radon measure itself, see e.g. [39, VIII Corollary 1.12], and for any positive weighting function $\lambda \in L^1(\Gamma_C^1, dx)$ the measure

$$m(E) := \int_E \lambda \, dx$$

is again a Radon measure, see [48, p.220 Nr.8]. We conclude that there exists a Radon measure μ_{t_N} such that

$$\int_{\Gamma_C^1} -\delta g \, d\mu_{t_N} = \int_{\Gamma_C^1} t_N \delta g \, dx, \tag{1.90}$$

and therefore $\mathbf{F}^{\mathrm{con}}(t_N) \in \partial^F I_{\mathcal{K}}(\boldsymbol{\varphi})$. □

2. Discretisation

In this chapter we describe the discretisation of the Euler–Lagrange inclusion (1.82). We employ the *method of time layers*, also called *Rothe's method*, cf. [35, Section 9.2]. In this approach first (possibly adaptive) discretisation in time is performed. The resulting semi-discrete problems are then approximated by (possibly adaptive) spatial discretisations which, from a theoretical point of view, are treated as a perturbation of the time integration scheme. Therefore, this approach combines adaptive integration in time and space. In this thesis we will not consider adaptivity but we will utilize Rothe's method to allow its incorporation in future work.
Usually the variational equation (1.38) is discretised together with the unknown contact traction forces t_N [44, 62, 66, 106, 93]. We will follow a different path consider the differential inclusion (1.82). This has the advantage that no algorithmic treatment of the contact forces (1.40) has to be done explicitly. We will show that the residual of the fully discrete inclusion corresponds to the consistent discrete mortar contact forces, which fulfil a discrete analogue of the KKT conditions (1.37). First in Section 2.1 we will discretise the inclusion (1.82) in time applying a novel contact-stabilised midpoint rule, inspired by the modified Newmark method of Kane et al. [70] and its further extension by Deuflhard, Krause and Klapproth [34, 72]. The stabilised scheme shows a dissipative energy behaviour which is investigated numerically in Chapter 4. Then in Section 2.2 we shortly describe how the spatial problems are discretised with standard linear finite elements and dual mortar elements [137, 79] for the non-penetration constraint.

2.1. Time discretisation

In this section we are concerned with the time discretisation of the differential inclusion (1.82). Most common integrators for linear dynamic systems like the Newmark schemes or the midpoint rule can lead to a blow-up of energy when applied to non-linear mechanical systems [125]. Belytschko and Schoeberle found, that this happens because the spectral stability of these schemes is only a necessary, but not sufficient condition any more for stability. They further showed that the unconditional stability of implicit schemes in non-linear mechanics is guaranteed if the total energy of the system is conserved or dissipated [18].
In the spirit of this stability criterion, time integration schemes for large deformation con-

tact can be divided into two categories: the ones that introduce artificial dissipation based on the *generalised α-method* [84] and the energy conserving schemes that algorithmically enforce energy preservation. The latter use either the *Energy–Momentum method* by Simo and Tarnow [125] or its extension, the *discrete gradient* approach by Gonzalez [51], which was originally introduced for non-linear elastodynamics. Both are based on the symplectic implicit midpoint rule [125], where in addition the discrete internal force term is modified leading to energy conservation but also to an asymmetric tangential stiffness operator. The extension of this class of integrators to large deformation contact problems is usually done by algorithmically enforcing a discrete analogue of the persistency condition (1.50), which is necessary to prove conservation of energy of the continuous system. The first extensions proposed in [88, 7] are enforcing energy conservation at the cost of modified constraints, which leads to violations of the exact non-penetration condition. This drawback is overcome by Laursen and Love through the introduction of a *discrete contact velocity* which accounts for the discontinuity of the velocity during impact [89]. Therein, the conservation of energy is recovered by computing a velocity correction in a post-processing step. Combining energy correction techniques with the discrete gradient approach, Hauret and Le Tallec developed a conserving method that avoids the contact velocity update [63].
The integrators named so far are formulated only for the point wise non-penetration constraint (1.26). In the case of variationally consistent discretisations using the mortar method, Hartmann and Ramm combined the generalised Energy–Momentum method [84] and the discrete contact velocity to establish an energy conserving integrator [62]. Due to simplifying assumptions during the constraint linearisation, this method does not conserve the total angular momentum. To overcome this problem Hesch and Betsch re-parametrised the mechanical system in terms of invariants which naturally incorporates conservation of momenta but also leads to generalised saddle point problems which have to be solved within each time step [66].
In this thesis we will take a different path. As already mentioned, Energy–Momentum methods lead to spatial problems with non-symmetric tangential stiffness operator. This in turn prevents the use of fast descent algorithms like multigrid methods for the algebraic solution [79, 117]. In the following we will introduce a contact-stabilised midpoint rule, treating the contact forces completely implicitly, as done in [70] and which reduces to the standard (possibly unstable) implicit midpoint rule in the absence of contact. We are interested in the accurate prediction of stresses arising from contact between complex geometries like the human hip joint and controlled by external (muscle) forces. Accordingly a small time step size is required in any case to resolve the full process accurately.

2.1.1. Contact-stabilised midpoint rule.

Let us split the time interval of interest $[0, T]$ into subintervals defined by $0 = t^0 < \ldots < t^M = T$, where we assume for simplicity equidistant time steps $\tau = \tau^k := t^{k+1} - t^k$. In the

following we simply write ∂ to denote the Fréchet subdifferential ∂^F and use the notation $\boldsymbol{\varphi}^n \approx \boldsymbol{\varphi}(t^n)$ to term the approximation of a quantity at the time step t^n. The midpoint rule is constructed by approximating the deformation and velocity by a truncated Taylor expansion

$$\begin{aligned} \boldsymbol{\varphi}^{n+1} &= \boldsymbol{\varphi}^n + \tau \dot{\boldsymbol{\varphi}}^n + \frac{\tau^2}{4}\left(\ddot{\boldsymbol{\varphi}}^{n+1} + \ddot{\boldsymbol{\varphi}}^n\right), \\ \dot{\boldsymbol{\varphi}}^{n+1} &= \dot{\boldsymbol{\varphi}}^n + \frac{\tau}{2}\left(\ddot{\boldsymbol{\varphi}}^{n+1} + \ddot{\boldsymbol{\varphi}}^n\right), \end{aligned} \tag{2.1}$$

and evaluating the inclusion

$$0 \in \mathbf{F}^{\mathrm{dyn}}(\boldsymbol{\varphi}(t)) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}(t)) - \mathbf{F}^{\mathrm{ext}}(t) + \partial I_{\mathcal{K}}(\boldsymbol{\varphi}(t)),$$

cf. (1.82), at the averaged midpoints

$$\boldsymbol{\varphi}(t^{n+1/2}) \approx \boldsymbol{\varphi}^{n+1/2} := \frac{\boldsymbol{\varphi}^{n+1} + \boldsymbol{\varphi}^n}{2}. \tag{2.2}$$

This leads to the following semi-discrete inclusion for the deformation

$$0 \in \frac{2}{\tau^2}\rho\left(\boldsymbol{\varphi}^{n+1} - \boldsymbol{\varphi}^n - \tau\dot{\boldsymbol{\varphi}}^n\right) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}^{n+1/2}) - \mathbf{F}^{\mathrm{ext}}(t^{n+1/2}) + \partial I_{\mathcal{K}}(\boldsymbol{\varphi}^{n+1/2}). \tag{2.3}$$

Note that in the operator formulation above the first operator acts as follows

$$\frac{2}{\tau^2}\rho\left(\boldsymbol{\varphi}^{n+1} - \boldsymbol{\varphi}^n - \tau\dot{\boldsymbol{\varphi}}^n, \mathbf{v}\right)_{\mathbf{L}^2(\Omega)} \quad \forall\, \mathbf{v} \in \mathbf{H}^1_0(\Omega). \tag{2.4}$$

Once the inclusion (2.3) is solved the new velocity is updated accordingly to

$$\dot{\boldsymbol{\varphi}}^{n+1} = \frac{2}{\tau}\left(\boldsymbol{\varphi}^{n+1} - \boldsymbol{\varphi}^n\right) - \dot{\boldsymbol{\varphi}}^n, \tag{2.5}$$

which is an immediate consequence of (2.1), obtained by subtracting the scaled first equation from the second one. In the absence of contact it is straight-forward to shown that this scheme conserves the total linear and angular momentum, by insertion of test functions $\mathbf{v} = \mathbf{c} \in \mathbb{R}^d$ and $\mathbf{v} = \mathbf{c} \times \boldsymbol{\varphi}^{n+1/2}$ correspondingly, see [125]. The proof of momenta conservation in the present contact formulation (2.3) is complicated by the Fréchet normal cone $\partial I_{\mathcal{K}}(\boldsymbol{\varphi}^{n+1/2})$. In [70] it is stated without proof, that the conservation follows directly from the translational and rotational invariance of the non-penetration constraint and *Noether's theorem* [97, Theorem 11.4.1]. A detailed extension of Noether's theorem to non-smooth Lagrangians has not been done yet and will be subject of future work. As already observed in the linearised framework of small displacements the energy behaviour of time integration schemes can be improved significantly when contact forces are computed only from the resulting $\boldsymbol{\varphi}^{n+1}$ [70, 34]. Motivated by the predictor–corrector scheme in [70] we propose to modify the implicit midpoint rule (2.3) such that the generalised gradient is evaluated at the endpoint $\boldsymbol{\varphi}^{n+1}$. This leads to

$$0 \in \frac{2}{\tau^2}\rho\left(\boldsymbol{\varphi}^{n+1} - \boldsymbol{\varphi}^n - \tau\dot{\boldsymbol{\varphi}}^n\right) + \mathbf{F}^{\mathrm{int}}(\boldsymbol{\varphi}^{n+1/2}) - \mathbf{F}^{\mathrm{ext}}(t^{n+1/2}) + \partial I_{\mathcal{K}}(\boldsymbol{\varphi}^{n+1}). \tag{2.6}$$

The resulting scheme yields an enhanced energetic behaviour and stability which is investigated in more detail in Section 4.4.3. By the generalised Fermat rule 1.4.7 all local solutions of the following constrained non-convex minimisation problem are solutions of the differential inclusion (2.6)

$$\boldsymbol{\varphi}^{n+1} \in \mathcal{K} : \quad \mathcal{J}(\boldsymbol{\varphi}^{n+1}) \leq \mathcal{J}(\mathbf{v}) \quad \forall\, \mathbf{v} \in \mathcal{K}, \tag{2.7}$$

with

$$\mathcal{J}(\mathbf{v}) := \int_\Omega \frac{1}{\tau^2}\rho|\mathbf{v}|^2 + 2W\left(\frac{\mathbf{v}+\boldsymbol{\varphi}^n}{2}\right) - l(\mathbf{v})\,dx, \tag{2.8}$$

$$l(\mathbf{v}) := \frac{2}{\tau^2}\int_\Omega \rho\left(\boldsymbol{\varphi}^n + \tau\dot{\boldsymbol{\varphi}}^n\right)\,\mathbf{v}\,dx, \tag{2.9}$$

where W denotes the hyperelastic strain energy functional (1.11) and we omitted the external forces $\mathbf{F}^{\text{ext}}$ for simplicity.

2.2. Spatial discretisation

In this section we will shortly describe the spatial discretisation of the differential inclusion (1.82) using first order finite elements. Therefore, let $\mathcal{T}_h^i = \mathcal{T}_h(\Omega^i)$ be regular simplicial grids of the bodies Ω^i with maximal mesh size $h > 0$ and denote by $\mathcal{N}^i = \mathcal{N}^i(\Omega^i)$ the set of vertices of the grid. The vector-valued space of linear finite elements is defined as

$$\mathbf{S}_h^i := S_h(\mathcal{T}_h^i)^d := \left\{\boldsymbol{\varphi}_h \in C(\Omega^i)^d : \boldsymbol{\varphi}_h \text{ linear on } T, \quad \forall\, T \in \mathcal{T}_h^i\right\}. \tag{2.10}$$

To simplify notation we introduce the product space $\mathbf{S}_h := \mathbf{S}_h^1 \times \mathbf{S}_h^2$ with $\mathcal{N} := \mathcal{N}^1 \cup \mathcal{N}^2$ and $\mathcal{T}_h := \mathcal{T}_h^1 \cup \mathcal{T}_h^2$. The total number of vertices is denoted by $N = |\mathcal{N}|$ and scalar nodal basis function $\lambda_p \in S_h$ corresponding to a node $p \in \mathcal{N}$ is characterised by

$$\lambda_p(q) = \delta_p^q := \begin{cases} 1 & p = q, \\ 0 & \text{else.} \end{cases} \tag{2.11}$$

A basis of $\mathbf{S}_h$ is given by

$$B := \left\{\lambda_p \mathbf{e}_j : p \in \mathcal{N},\, 1 \leq j \leq d\right\}, \tag{2.12}$$

where $\mathbf{e}_j$ denotes the j'th Euclidean basis vector. The deformation and velocity are discretised by replacing $\mathbf{H}_0^1(\Omega)$ by the finite dimensional subspace $\mathbf{S}_{h,0} \subset \mathbf{H}_0^1(\Omega)$. Each element of the finite element space has a unique algebraic representation

$$\boldsymbol{\varphi}_h \in \mathbf{S}_h \Longleftrightarrow \exists\, \hat{\boldsymbol{\varphi}} \in \mathbb{R}^{dN} : \boldsymbol{\varphi}_h = \sum_{p\in\mathcal{N}} \hat{\boldsymbol{\varphi}}_p \lambda_p, \tag{2.13}$$

and in the following we will not distinguish between the two representations. Exchanging the continuous operators in (2.6) by their discrete counterparts yields

$$0 \in \frac{2}{\tau^2}\left(\boldsymbol{\varphi}_h^{n+1} - \boldsymbol{\varphi}_h^n - \tau\dot{\boldsymbol{\varphi}}_h^n\right)^T \mathbf{M}_h + \mathbf{F}_h^{\text{int}}(\boldsymbol{\varphi}_h^{n+1/2}) - \mathbf{F}_h^{\text{ext}}(t^{n+1/2}) + \partial I_{\mathcal{K}}(\boldsymbol{\varphi}_h^{n+1}), \tag{2.14}$$

where $\mathbf{M}_h$ is the global mass matrix scaled with the density

$$(\mathbf{M}_h)_{pq} = \int_{\mathcal{T}_h} \rho\, \lambda_p \lambda_q \, dx, \tag{2.15}$$

and the components of the discrete internal and external forces are given by

$$\begin{aligned}\left(\mathbf{F}_h^{\text{int}}(\boldsymbol{\varphi}_h^{n+1/2})\right)_p &= \int_{\mathcal{T}_h} \mathbf{P}(\nabla\boldsymbol{\varphi}_h^{n+1/2}) : \nabla\lambda_p \, dx, \\ \left(\mathbf{F}_h^{\text{ext}}(t)\right)_p &= \int_{\mathcal{T}_h} \mathbf{f}(t)\, \lambda_p \, dx + \int_{\Gamma_{N,h}} \boldsymbol{\pi}(t)\, \lambda_p \, ds.\end{aligned} \tag{2.16}$$

What is left is the discretisation of the subdifferential $\partial I_{\mathcal{K}}(\boldsymbol{\varphi})$, which is described in the next section.

2.2.1. Dual mortar discretisation of the contact constraints

One difficulty of contact problems is the proper enforcement of the geometric non-penetration constraint (1.26). In the last decades the use of *Node-to-Segment methods* (NTS) has been very popular [7, 88, 89, 63, 65]. In this approach each node of the so-called *slave* side is identified with a segment on the opposing *master* contact surface and non-penetration is only enforced for these slave nodes. A further modification of this method additionally enforcing non-penetration of all master nodes leads to the *two pass NTS method*. When sliding occurs these methods generate artificial jumps in the discrete contact forces whenever a slave node leaves the contact area or passes to another master segment, which leads to instabilities [93].
In this context the *mortar method* could be called a *Segment-to-Segment* approach. It was first introduced by Bernardi, Maday and Patera for the coupling of spectral and finite element methods [19] and subsequently successfully applied as a domain decomposition method for non-matching grids [108, 137]. The key idea of the mortar method is to replace the point wise interface conditions, e.g. the vanishing of a jump term, by a weak integral constraint, which is variationally consistent. It is typically derived by integrating the point wise condition over the coupling interface of the slave or *nonmortar side*, the choice is arbitrary but needs to be fixed, and multiply it with "suitable" test functions $\mu \in M$. Here, "suitable" means, such that the method ensures enhanced stability.
Belgacem et al. first extended the mortar method to the linearised contact problem (1.28) using nodal Lagrange multipliers to discretise the mortar space and proved a first non-optimal a priori error estimate [17]. Krause and Wohlmuth later proposed to discretise

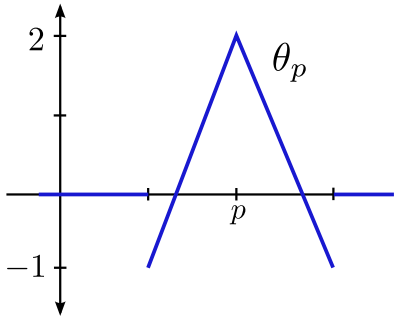


Figure 2.1.: *Dual basis function in 1D.*

the weak contact constraints using dual basis functions and derived optimal error bounds under suitable assumptions [79, 138]. Puso and Laursen [93] and Fischer&Wriggers [44] were the first who successfully applied the mortar method to static large deformation contact problems. While the first used a penalty approach, the latter combined the Lagrange multiplier method and an active-set strategy. Further extensions to e.g. frictional contact [140], quadratic elements [107] and dual mortar elements [62, 106] followed.
In the following we will consider the gap function as a mapping of $\boldsymbol{\varphi}$ only

$$\mathbf{H}^1_D(\Omega) \ni \boldsymbol{\varphi} \mapsto g(\boldsymbol{\varphi}) \in H^{\frac{1}{2}}(\gamma^1_C),$$

cf. Remark 1.2.1 and we assume from now on that g is smooth enough such that $g(\boldsymbol{\varphi}) \in H^{\frac{1}{2}}(\gamma^1_C)$. Note that we also pushed forward the gap in the deformed configuration $\gamma^1_C = \boldsymbol{\varphi}^1(\Gamma^1_C)$, cf. Remark 1.2.3, which will ease the discretisation and algorithmic treatment.
In Section 1.3 we showed that for the continuous problem the strong non-penetration constraint (1.26) (in $H^{\frac{1}{2}}$ sense) is equivalent to the variational formulation

$$b(\mu, \boldsymbol{\varphi}) \geq 0, \quad \forall\, \mu \in M^+, \tag{2.17}$$

where M^+ denotes the cone of positive functionals on $H^{\frac{1}{2}}(\gamma^1_C)^+$ and $b : M \times \mathbf{H}^1(\Omega) \longrightarrow \mathbb{R}$ is given by

$$b(\mu, \boldsymbol{\varphi}) = \langle \mu, g(\boldsymbol{\varphi}) \rangle_{M \times H^{\frac{1}{2}}},$$

cf. Proposition 1.3.4. As suggested in [137, 79] we discretise the mortar space M^+ by dual functions. For simplicial grids, on each simplex, these can be computed as a linear combination of the Lagrange basis functions

$$\theta_{p|T} = (d\lambda_p - \sum_{\substack{q \in T \\ q \neq p}} \lambda_q)_{|T} \quad \forall\, T \in \gamma^1_{C,h} \tag{2.18}$$

The dual functions form a partition of one and are in general discontinuous, see Figure 2.1. However, they fulfil the following bi-orthogonality condition

$$\int_{\gamma^1_{C,h}} \lambda_p\, \theta_q\, ds = \delta^p_q \int_{\gamma^1_{C,h}} \lambda_p\, ds, \quad \forall\, p \in \mathcal{N}(\gamma^1_{C,h}). \tag{2.19}$$

This property will be crucial in Chapter 4 to decouple the non-penetration constraints.

Remark 2.2.1. *In this thesis we will restrict ourselves to simplicial grids. In this case the construction of dual functions that fulfil* (2.19) *can be done independently of the deformed domain* $\gamma^1_{C,h}$*. For higher order finite elements or quadrilateral elements the dual functions are depending on the deformation of the domain, which complicates the linearisation of the non-penetration constraint [137, 106].*

The corresponding discrete mortar cone $M_h^+ \subset M^+$ is given by

$$M_h^+ := \left\{ \mu_h \in \Theta_h : \int_{\gamma^1_{C,h}} \mu_h \, v_h \, ds, \geq 0 \quad \forall \, v_h \in S_h\left(\gamma^1_{C,h}\right), v_h \geq 0 \right\}, \tag{2.20}$$

where Θ_h denotes the dual basis

$$\Theta_h := \operatorname{span}\left\{ \theta_p : p \in \mathcal{N}(\gamma^1_{C,h}) \right\}. \tag{2.21}$$

Note that we discretise the functionals $\mu \in M$ by functions $\mu_h \in S_h(\gamma^1_{C,h})$ in the sense

$$\langle \mu_h, v_h \rangle = \int_{\gamma^1_{C,h}} \mu_h \, v_h \, ds \quad \forall \, v_h \in S_h(\gamma^1_{C,h}). \tag{2.22}$$

This leads to the following weak non-penetration constraint

$$\int_{\gamma^1_{C,h}} -\mathbf{n}^1 \cdot (s - \Phi(s)) \; \mu_h \, ds \geq 0 \quad \forall \, \mu_h \in \Theta_h. \tag{2.23}$$

For the discretisation of the normal field $\mathbf{n}^1$ (or contact normal, cf.(1.22)) there are various approaches. In [44] at each quadrature point the mortar element normal at the closest point projection is explicitly computed. Hesch and Betsch use the discontinuous non-mortar element normal field to approximate $\mathbf{n}^1$ [66] We will use a nodally averaged non-mortar normals, i.e. for each non-mortar vertex $p \in \gamma^1_{C,h}$ with neighbouring faces $e \in E(p)$ and corresponding face normal $\mathbf{n}_e$ we set

$$\mathbf{n}_{h,p} = \frac{\sum_{e \in E(p)} \mathbf{n}_e}{\left\| \sum_{e \in E(p)} \mathbf{n}_e \right\|}. \tag{2.24}$$

This approximation in general yields a smoother behaviour when sliding occurs, cf. [93, 140]. The resulting discrete non-penetration constraint $g_h : \mathbf{S}_h \longrightarrow \mathbb{R}^m$ reads

$$g_{h,p}(\boldsymbol{\varphi}_h) := \int_{\gamma^1_{C,h}} -\mathbf{n}_{h,p} \theta_p \cdot (s - \Phi(s) \, ds, \quad p \in \mathcal{N}(\gamma^1_{C,h}), \tag{2.25}$$

where $m := |\mathcal{N}(\gamma^1_{C,h})|$. We denote the corresponding discretised feasible set by

$$\mathcal{K}_h = \left\{ \boldsymbol{\varphi}_h \in \mathbf{S}_h : \; g_h(\boldsymbol{\varphi}_h) \geq 0 \right\}. \tag{2.26}$$

More details on the implementation of the normal projection $\Phi : \gamma_{C,h}^1 \longrightarrow \gamma_{C,h}^2$ can be found in Chapter 4. We arrive at the fully discrete differential inclusion

$$0 \in \frac{2}{\tau^2}\left(\boldsymbol{\varphi}_h^{n+1} - \boldsymbol{\varphi}_h^n - \tau\dot{\boldsymbol{\varphi}}_h^n\right)^T \mathbf{M}_h + \mathbf{F}_h^{\text{int}}(\boldsymbol{\varphi}_h^{n+1/2}) - \mathbf{F}_h^{\text{ext}}(t^{n+1/2}) + \partial I_{\mathcal{K}_h}(\boldsymbol{\varphi}_h^{n+1}). \quad (2.27)$$

In the finite dimensional case we can explicitly compute the subdifferential $\partial I_{\mathcal{K}_h}(\boldsymbol{\varphi}_h)$.

Proposition 2.2.2. *Let $\boldsymbol{\varphi}_h \in \mathcal{K}_h$ and denote the active constraints by*

$$\mathbb{A}(\boldsymbol{\varphi}_h) := \left\{p \in \mathcal{N}(\gamma_{C,h}^1) : \, g_{h,p}(\boldsymbol{\varphi}_h) = 0\right\}. \quad (2.28)$$

Then

$$\partial I_{\mathcal{K}_h}(\boldsymbol{\varphi}_h) = \left\{\sum_{p\in\mathbb{A}} \xi_p \nabla g_{h,p}(\boldsymbol{\varphi}_h) : \, \xi_p \in \mathbb{R},\, \xi_p \leq 0\right\}. \quad (2.29)$$

Proof. [121, Proposition 11.6.2]

Once the differential inclusion (2.27) is solved the discrete contact forces $-\mathbf{F}_h^{\text{con}}(\boldsymbol{\varphi}_h^{n+1}) \in \partial I_{\mathcal{K}_h}(\boldsymbol{\varphi}_h^{n+1})$ corresponding to the mortar constraints (2.25) can be computed from the residual of the inclusion

$$\mathbf{F}_h^{\text{con}}(\boldsymbol{\varphi}_h^{n+1}) := \frac{2}{\tau^2}\left(\boldsymbol{\varphi}_h^{n+1} - \boldsymbol{\varphi}_h^n - \tau\dot{\boldsymbol{\varphi}}_h^n\right)^T \mathbf{M}_h + \mathbf{F}_h^{\text{int}}(\boldsymbol{\varphi}_h^{n+1/2}) - \mathbf{F}^{\text{ext}}(t^{n+1/2}). \quad (2.30)$$

Analogously to the continuous case (1.37) the following discrete KKT conditions hold

$$g_h \geq 0, \quad t_{N,h} \leq 0, \quad g_h\, t_{N,h} = 0, \quad (2.31)$$

where the components of $t_{N,h}$ are given by

$$(t_{N,h})p := \begin{cases} \xi_p & p \in \mathbb{A}(\boldsymbol{\varphi}_h), \\ 0 & \text{else}, \end{cases} \quad (2.32)$$

and

$$\mathbf{F}_h^{\text{con}}(\boldsymbol{\varphi}_h^{n+1}) = \sum_{p\in\mathbb{A}} \xi_p \nabla g_{h,p}(\boldsymbol{\varphi}_h^{n+1}). \quad (2.33)$$

Remark 2.2.3. *When considering the standard weak formulation of the problem* (1.38)*, the straight-forward finite element discretisation of the unknown contact traction forces* (1.43) *in general leads to inconsistent discrete contact forces when the mortar method is used. This results in the loss of angular momentum conservation [93]. The resulting contact forces of the weak non-penetration constraint* (2.17) *additionally include the linearisation of the deformed integration domain $\gamma_{C,h}^1$, which vanishes in the continuous case due to the KKT conditions* (1.37)*, see [44]. In the discrete case however, this is not true any more. To recover the consistent discrete contact forces* (2.29) *typically a contact potential is postulated to derive them [93, 66], which is motivated from equality*

constrained Hamilton systems, see [20]. In the non-smooth Hamilton approach presented in this thesis (1.82) *this construction is not necessary as the contact forces only enter implicitly. This allows us to discretise the continuous problem in the standard way by substituting the continuous spaces by finite-dimensional subspaces.*

Conservation of linear and angular momentum is inherited from the continuous problem if the discrete contact forces are again translational and rotational invariant. The proof is rather technical because the exact linearisation of the discrete weak non-penetration constraint (2.25) has to be analysed step by step. Therefore we will omit the proof here and refer the interested reader to [66, Section 3.3] or [106, Section 3.5]. We will return to this point in the numerical study of the time discretisation scheme in Section 4.4.3.

In the small deformation framework optimal discretisation error bounds for the dual mortar discretisation were shown recently when the normal field is constant on Γ_C^1 and additional technical assumptions

$$\|\boldsymbol{\varphi} - \boldsymbol{\varphi}_h\|_{\mathbf{H}^1(\Omega)} + \|\mu - \mu_h\|_{H^{-\frac{1}{2}}(\Gamma_C^1)} \leq c h^{s-1} |\boldsymbol{\varphi}|_{H^{s+1}(\Omega)}, \tag{2.34}$$

where $\mu \in M^+$ and $\mu_h \in M_h^+$ denote the dual Lagrange multiplier corresponding to the contact constraint, see Wohlmuth [138]. An overview of further results on quadratic finite elements and frictional problems can be found in the overview article [138]. In the large deformation framework to our knowledge no discretisation error bounds exists yet.

Like in the time-discrete case solutions of the fully discrete inclusion (2.27) can be calculated by solving the following non-convex non-linear minimisation problem.

$$\boldsymbol{\varphi}_h^{n+1} \in \mathcal{K}_h : \quad \mathcal{J}(\boldsymbol{\varphi}_h^{n+1}) \leq \mathcal{J}(\mathbf{v}_h) \quad \forall\, \mathbf{v}_h \in \mathcal{K}_h, \tag{2.35}$$

with

$$\mathcal{J}(\mathbf{v}_h) := \frac{1}{\tau^2} \mathbf{v}_h^T \mathbf{M}_h \mathbf{v}_h + 2 \int_\Omega W\left(\frac{\mathbf{v}_h + \boldsymbol{\varphi}_h^n}{2}\right) - l(\mathbf{v}_h), \tag{2.36}$$

$$l(\mathbf{v}_h) := \frac{2}{\tau^2} \mathbf{v}_h^T \mathbf{M}_h \left(\boldsymbol{\varphi}_h^n + \tau \dot{\boldsymbol{\varphi}}_h^n, \mathbf{v}_h\right), \tag{2.37}$$

cf. (2.7). Note that we again omitted the external forces, they can be incorporated if suitable external potentials exist.
The next chapter focuses on the construction of fast and globally convergent algorithms to solve the minimisation problem (2.35).

3. Filter–Trust–Region-Multigrid Methods

The aim of this chapter is to describe the construction of a fast and globally convergent algorithm for the solution of the discretised large deformation contact problem (2.35). With this goal in mind we will consider general non-linear non-convex constrained minimisation problems in the following. The simplest class of solvers for these problems are *penalty methods*. The underlying idea is to augment the energy by a penalty term that is "punishing" the violation of the constraints and then minimise this energy instead. The advantages are that only unconstrained problems have to be solved and it is very easy to implement. The critical point is the choice of the penalty parameter. If it is chosen too small the constraints are not fulfilled exactly. On the other hand the condition number of the Hessian of the energy increases with growing penalty parameter, which in turn is a necessary requirement for the convergence of the method [42, 93]. The basis of most of the remaining approaches is *sequential quadratic programming* (SQP) which is a straight-forward extension of Newton's method additionally linearising the constraints. Hence only local convergence can be expected [30].
In the *Merit function* approach SQP and penalty methods are combined. In this method the SQP corrections are accepted if the penalty function is decreased and rejected otherwise. Again, convergence only holds if the penalty parameter is sufficiently large. Moreover, if the curvature of the constraints is not taken into account, then the fast local convergence of the SQP method is in general prohibited by the *Maratos effect* [95].
If the energy functional is non-convex, the local SQP problems are possibly unbounded. The *trust–region method* can cope with these cases by imposing a supplementary constraint on the step size. While this leads to a natural and automatic damping strategy, it may also result in infeasible local problems if the trust–region and linearised SQP constraints are incompatible [30]. This can be resolved by decomposing the correction into a normal step, increasing the fulfilment of the constraints, and a tangential which is solely decreasing the objective functional. This approach is called *Composite-Step method* and it is usually used within trust–region methods for constrained optimisation [23].
A novel approach that completely avoids any penalty parameters was recently developed by Fletcher and Leyffer [45]. The *filter method* is an alternative to merit functions that provides an elegant way to compare iterates. While the decrease of energy is controlled by line-search or trust–region methods, the filter ensures the feasibility of the solution w.r.t. the non-linear constraints.
Nowadays solvers for non-convex non-linear constrained optimisation problems typically

consist of a combination of the different techniques described above [133, 23, 130, 50], where the trend is going towards using filter methods.

This chapter is organised as follows. In the first Section 3.1 we will introduce a new class of successive subset minimisation algorithms and prove global convergence of a monotone multigrid method for symmetric non-coercive quadratic obstacle problems within this framework in Section 3.2. Then, in Section 3.3 we will give a short introduction to the trust–region method and show convergence of this approach when the previously constructed multigrid method is used to solve the quadratic sub-problems. In the last Section 3.4 we will then describe an extension of the filter–trust-region method taking into account an inexact linearisation of the constraints which leads to bound constrained local problems. For this method we prove global convergence when the approximation error is controlled by the trust-region norm. In the following, if not stated otherwise, we denote by $\|\cdot\|$ the Euclidean norm for vectors and the Frobenius norm for matrices.

3.1. Successive Direction Minimisation

In this section we consider the general problem of finding a local minimum of a possibly non-convex functional $\mathcal{J} : \mathbb{R}^N \longrightarrow \mathbb{R}$ subject to bound constraints

$$\min_{v \in \mathcal{K}} \mathcal{J}(v), \tag{3.1}$$

where the compact set of feasible solutions is given by

$$\mathcal{K} := \prod_{i=1}^{N} [l_i, u_i] \subset \mathbb{R}^N. \tag{3.2}$$

For simplicity, we assume that $-\infty < l_i \leq u_i < \infty$ for all $1 \leq i \leq N$ and $\mathcal{J}$ is continuous on $\mathcal{K}$, which ensures the existence of solutions of (3.1). In the following we will introduce a general framework for iterative methods that successively minimise the energy along a set of search directions

$$\mathcal{D} := \left\{ d^1, \ldots, d^m \right\}. \tag{3.3}$$

We use the notation $\rightrightarrows$ to denote set-valued mappings, i.e. mappings that map points onto sets. Let $h : \mathbb{R}^N \times \mathbb{R}^N \rightrightarrows \mathbb{R}$ denote a mapping that for each point v and a search direction d^i returns an interval of *feasible line search parameters*, i.e.

$$\begin{aligned} h : (d^i, v) &\longmapsto [l(d^i, v), u(d^i, v)] \subset \mathbb{R}, \\ \alpha \in [l(d^i, v), u(d^i, v)] &\Longrightarrow v + \alpha d^i \in \mathcal{K}. \end{aligned} \tag{3.4}$$

Whenever it is clear from the context we will omit the dependency on v and the index of the search directions and simply write

$$[l_d, u_d] := [l(d^i, v), u(d^i, v)].$$

Note that this interval does not have to coincide with the interval of all feasible line search parameters but can also be more restrictive. This is usually the case for the coarse grid corrections within monotone multigrid methods which ensures its optimal complexity [54]. The mapping that is used within the monotone multigrid method is the *Mandel restriction* which was originally introduced by Mandel for the construction of multilevel methods for obstacle problems [94].

Example 3.1.1 (Mandel restriction)**.**

$$h(d, v) = \left[\max_{\substack{1 \le i \le N \\ d_i \neq 0}} (l - v)_i, \min_{\substack{1 \le i \le N \\ d_i \neq 0}} (u - v)_i \right],$$

where a lower subscript v_i denotes the i'th component of a vector. For each search direction d^i let $\mathcal{M}^i : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ be the set-valued mapping given by *exact* minimisation in the direction d^i within the feasible line search domain $h(d^i, \cdot)$

$$\mathcal{M}^i(w) := \left\{ w + \beta d^i : \beta \in \operatorname*{arg\,min}_{\alpha \in [l_d, u_d]} \mathcal{J}(w + \alpha d^i) \right\}. \tag{3.5}$$

Note that to reduce notation we omit the dependence of $\mathcal{M}^i$ on the feasible line search mapping h. The set-valuedness of the mappings $\mathcal{M}^i$ is a consequence of the non-convexity of the energy, as multiple minimisers might exist in this case.

We are now ready to define an abstract iterative solver for the minimisation of (3.1).

Algorithm SDM (Successive Direction Minimisation)
Let $\mathcal{D}$ be given by (3.3) *and a mapping h as in* (3.4)*.*

Given: $w^0 = v^k \in \mathcal{K}$.
For $i = 1, \ldots, m$ {
    Choose $w^{i+1} \in \mathcal{M}^i(w^i)$
}
Set $v^{k+1} = w^m$

By now this algorithm is not well-defined because the exact choice of the next iterate is left open yet. However the global convergence theorem we will present here is independent of the explicit choice. Algorithm SDM can also be written in terms of one operator

$$v^{k+1} \in \mathcal{M}^{\mathrm{sdm}}(v^k),$$

given by the composition of the directional minimisation operators $\mathcal{M}^i$

$$\mathcal{M}^{\mathrm{sdm}} := \mathcal{M}^m \circ \ldots \circ \mathcal{M}^1. \tag{3.6}$$

The composition of two set-valued mappings $S, T : \mathbb{R}^N \rightrightarrows \mathbb{R}^N$ is given by

$$S \circ T\,(v) = \{u \in\ S(w) : w \in\ T(v)\}\,.$$

We now want to analyse the mappings $\mathcal{M}^i$ in more detail and derive properties which ensure global convergence of Algorithm SDM. To cope with the non-convexity and the resulting non-uniqueness of the local solutions we will first introduce a weaker concept of continuity for set-valued mappings.

Definition 3.1.2. Let $V \subseteq \mathbb{R}^N$. A set-valued mapping $\mathcal{M} : V \rightrightarrows V$ is called *upper-hemicontinuous* (u.h.c) at $\bar{w} \in V$ if for every open neighbourhood O of $\mathcal{M}(\bar{w})$, there exists a neighbourhood K of $\bar{w}$ such that $\mathcal{M}(v) \subseteq O$ for all $v \in K$.

A slightly weaker property that we will need is given by the following:

Definition 3.1.3. Let $V \subseteq \mathbb{R}^N$. A set-valued mapping $\mathcal{M} : V \rightrightarrows V$ is called *closed* at $\bar{w} \in V$, if for every convergent sequence $w^k \longrightarrow \bar{w}$ and any sequence $v^k \in \mathcal{M}(w^k)$ with $v^k \longrightarrow \bar{v}$, it holds that $\bar{v} \in \mathcal{M}(\bar{w})$.

The relation between these two properties is shown in the next proposition

Proposition 3.1.4. *If $\mathcal{M} : V \rightrightarrows V$ is closed and the range of $\mathcal{M}$*

$$\operatorname{range}(\mathcal{M}) := \bigcup_{v \in V} \mathcal{M}(v)$$

is contained in a compact set, then $\mathcal{M}$ is u.h.c. Conversely, if $\mathcal{M}$ is u.h.c. and additionally closed valued, i.e. $\mathcal{M}(w)$ is closed for each $w \in V$, then $\mathcal{M}$ is closed.

Proof. To give an idea of the proof we will show the second part.
(i) First let $\mathcal{M}$ be u.h.c. and closed valued. Consider two convergent sequences

$$w^k \longrightarrow \bar{w}, \quad v^k \longrightarrow \bar{v},$$

with $v^k \in \mathcal{M}(w^k)$. Choose $\varepsilon > 0$ arbitrary and let O_ε be an open neighbourhood of $\mathcal{M}(\bar{w})$ such that

$$\min_{w \in \mathcal{M}(\bar{w})} \|v - w\| < \varepsilon \quad \forall\, v \in O_\varepsilon.$$

Now from the u.h.c there exists a neighbourhood $K_\varepsilon(\bar{w})$, s.t. $\mathcal{M}(w) \subset O_\varepsilon$ for all $w \in K_\varepsilon(\bar{w})$. As $w^k \longrightarrow \bar{w}$ we thus know that $w^m \in K_\varepsilon(\bar{w})$ for m large enough and

$$v^m \in \mathcal{M}(w^m) \subset O_\varepsilon.$$

From the convergence of v^k we obtain $\bar{v} \in \overline{O}_\varepsilon$. Letting $\varepsilon \longrightarrow 0$ and using the compactness of $\mathcal{M}(\bar{w})$ then yields the claim.
(ii) The other direction follows by a contradiction proof, see e.g. [4, Theorem 17.16]. □

Closed mappings are in general not u.h.c which can be seen at the following example at $w = 0$

$$\mathcal{M}(w) = \begin{cases} \{0\} & w = 0, \\ \{\frac{1}{w}\} & \text{else.} \end{cases}$$

The next convergence theorem is due to Zangwill [143] for a general class of algorithms given by closed mappings.

Theorem 3.1.5 ([143], Theorem A)**.** *Let an algorithm be given by the set-valued mapping $\mathcal{M} : V \rightrightarrows V$ and a procedure to specify the sequence of iterates $w^k \in \mathcal{M}(w^{k-1})$. Let $\mathcal{K}_{Sol} \subset V$ be a solution set with the following properties:*

- *If $w \notin \mathcal{K}_{Sol}$, then*
$$\mathcal{J}(v) < \mathcal{J}(w) \quad \forall\, v \in \mathcal{M}(w). \tag{3.7}$$

- *If $w \in \mathcal{K}_{Sol}$, then either the algorithm stops or*
$$\mathcal{J}(v) \leq \mathcal{J}(w) \quad \forall\, v \in \mathcal{M}(w). \tag{3.8}$$

Further assume that the range of $\mathcal{M}$ is contained in a compact set $\mathcal{K} \subset V$ and that $\mathcal{M}$ is closed at all $w \notin \mathcal{K}_{Sol}$.
Then either the algorithm stops at an $w \in \mathcal{K}_{Sol}$ or the limit of any convergent subsequence is in $\mathcal{K}_{Sol}$.

Proof. If the algorithm stops at some $w \in \mathcal{K}$ then by (3.7) and (3.8) it must be in $\mathcal{K}_{\text{Sol}}$. Hence let $(w^k)_k \subset \mathcal{K}$ be an infinite sequence produced by the algorithm.
From the compactness of $\mathcal{K}$ we obtain the existence of a convergent subsequence $(w^{k_l})_l$, $w^{k_l} \longrightarrow w_*$. As $\mathcal{J}$ is continuous and the sequence $\mathcal{J}(w^k)$ is monotone decreasing, we can deduce that

$$\lim_{k \longrightarrow \infty} \mathcal{J}(w^k) = \mathcal{J}(w_*). \tag{3.9}$$

Now consider the subsequence $(w^{k_l+1})_l$. By restriction to a further subsequence we get $w^{k_l+1} \longrightarrow w_{*+1}$ and thus by the same arguments as in (3.9)

$$\mathcal{J}(w_*) = \mathcal{J}(w_{*+1}). \tag{3.10}$$

To close the proof let us assume that $w_* \notin \mathcal{K}_{\text{Sol}}$. Now as $w^{k_l} \longrightarrow w_*$, $w^{k_l+1} \in \mathcal{M}(w^{k_l})$ and $w^{k_l+1} \longrightarrow w_{*+1}$, we get from the upper hemicontinuity that $w_{*+1} \in \mathcal{M}(w_*)$ and by (3.8) it holds

$$\mathcal{J}(w_{*+1}) < \mathcal{J}(w_*),$$

which contradicts (3.10). □

We now want to investigate under which assumptions the global convergence of Algorithm SDM holds.

Assumption M1. The feasible line search mappings $h(d^i, w)$ depend continuously on w for each $d^i \in \mathcal{D}$.

This assumption ensures the closedness of the mappings $\mathcal{M}^i$.

Proposition 3.1.6. *If Assumption M1 holds the mappings $\mathcal{M}^i$ are closed.*

Proof. Let the sequences $(w^k), (v^k) \subset \mathcal{K}$ be given with $v^k \in \mathcal{M}^i(w^k)$, and let $w^k \longrightarrow \bar{w}$, $v^k \longrightarrow \bar{v}$. By definition each $v^k = w^k + \alpha^k d^i$ for some $\alpha^k \in [l_d^k, u_d^k] := h(d^i, w^k)$ and further $\bar{v} = \bar{w} + \bar{\alpha} d^i$ with $\alpha^k \longrightarrow \bar{\alpha}$.
We now have to show that $\bar{v} \in \mathcal{M}^i(\bar{w})$ which by (3.5) is equivalent to

$$\mathcal{J}(\bar{v}) = \mathcal{J}(\bar{w} + \bar{\alpha} d^i) \leq \mathcal{J}(\bar{w} + \alpha d^i) \quad \forall\, \alpha \in \left[\bar{l}_d, \bar{u}_d\right]. \tag{3.11}$$

The continuity of $h(d^i, \cdot)$ yields that $l_d^k \longrightarrow \bar{l}_d$ and also $u_d^k \longrightarrow \bar{u}_d$. Thus for all $\alpha \in (\bar{l}_d, \bar{u}_d)$ and k large enough we obtain $\alpha \in [l_d^k, u_d^k]$ and

$$\mathcal{J}(w^k + \alpha^k d^i) \leq \mathcal{J}(w^k + \alpha d^i).$$

Let $k \longrightarrow \infty$, then from the continuity of $\mathcal{J}$ we deduce

$$\mathcal{J}(\bar{v}) \leq \mathcal{J}(\bar{w} + \alpha d^i) \quad \forall\, \alpha \in (\bar{l}_d, \bar{u}_d).$$

The last case that has to be shown is that (3.11) also holds for $\alpha \in \left\{\bar{l}_d, \bar{u}_d\right\}$. We only prove it for the case $\alpha = \bar{l}_d$ where $\bar{l}_d$ is contained in finitely many intervals $[l_d^k, u_d^k]$. As $l_d^k \longrightarrow \bar{l}_d$, and both w^k and α^k converge, the continuity of $\mathcal{J}$ guarantees that $\mathcal{J}(\bar{w} + \bar{\alpha} d^i) \leq \mathcal{J}(\bar{w} + \bar{l}^d d^i)$ and (3.11) holds. □

By definition of the mappings $\mathcal{M}^i$ and $h(d^i, w)$ it is assured that all intermediate iterates are feasible and

$$\text{range}(\mathcal{M}^i) \subset \mathcal{K}. \tag{3.12}$$

From this we can deduce the closedness of the composition.

Lemma 3.1.7. *Let $\mathcal{K}$ be compact and $S, T : \mathcal{K} \rightrightarrows \mathcal{K}$. If S is closed at w and B on $T(w)$, then also the composition $S \circ T$ is closed at w.*

Proof. [143, Corollary 4.2.1]

By Proposition 3.1.6 and Lemma 3.1.7 the closedness of $\mathcal{M}^{\text{sdm}}$ now follows immediately.

Corollary 3.1.8. *$\mathcal{M}^{sdm}$ is closed and* $\text{range}(\mathcal{M}^{sdm}) \subset \mathcal{K}$.

To ensure that the energy cannot increase throughout the iteration we assume that

Assumption M2. $0 \in h(d^i, w) \quad \forall\, w \in \mathcal{K}$.

Further to avoid that Algorithm SDM gets stuck at saddle-points we require:

Assumption M3. If w is no local minimum of (3.1) , then

$$\mathcal{J}(v) < \mathcal{J}(w) \quad \forall\, v \in \mathcal{M}^{\mathrm{sdm}}(w).$$

The second assumption further guarantees that the set of search directions is rich enough, i.e. in general

$$\operatorname{span} \mathcal{D} = \mathbb{R}^N. \tag{3.13}$$

Still, requiring (3.13) only is not enough, which is illustrated in Figure 3.1. Under these

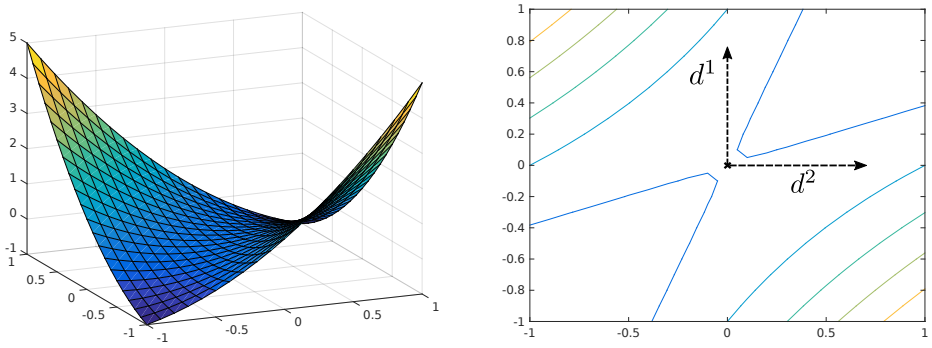


Figure 3.1.: *Example of a case where the algorithm gets stuck at a saddlepoint if Assumption M3 does not hold. Left: Graph of a non-convex quadratic functional with a saddle point at* $(0, 0)$*. Right: Contour lines of the functional. If the search directions are chosen as in the picture, then the algorithm will stay at* $(0, 0)$*.*

conditions Algorithm SDM converges globally to a local minimiser.

Theorem 3.1.9. *Assume that Assumptions M1–M3 hold. Then for any rule of choosing the next iterate* $w^{k+1} \in \mathcal{M}^{sdm}(w^k)$*, Algorithm SDM either stops at a local minimum of* (3.1) *or the limit of any convergent subsequence is a local minimiser.*

Proof. To prove this theorem we have to show that all requirements for Theorem 3.1.5 are fulfilled. First we define the solution set to be the set of all local minima

$$\mathcal{K}_{\mathrm{Sol}} := \{w \in \mathcal{K} : \mathcal{J}(w) \leq \mathcal{J}(v) \text{ locally}\}.$$

The closedness of the mapping $\mathcal{M}^{\mathrm{sdm}}$ was proven in Proposition 3.1.4. (3.8) is an immediate result of the exact minimisation property of the mappings $\mathcal{M}^i$ and Assumption M2. The condition (3.7) coincides with Assumption M3. □

3.2. Monotone Multigrid Methods

In this chapter we will introduce a monotone multigrid method for a class of non-convex obstacle problems and prove global convergence using the framework introduced in the previous section. This method will be used as a solver for the local quadratic sub-problems appearing in the filter method described in Section 3.4. Geometric multigrid methods were originally introduced for the solution of elliptic partial differential equations [40, 60]. While for elliptic obstacle problems the global convergence of these methods is well established [73, 75, 54], to our knowledge up to now no convergence proof for non-convex problems exists.
In this section we first introduce a continuous non-convex obstacle problem and then show how the discretised problem can be solved using an extension of the monotone multigrid method described in [54]. To prove global convergence of this method we will reformulate it as a successive direction minimisation SDM and apply Theorem 3.1.9. The notation is taken from the survey article by Kornhuber and Gräser [54].

Let $\Omega \subset \mathbb{R}^d$ be a bounded polygonal domain and $H \subset H^1(\Omega)$ a closed subset of the Sobolev-space $H^1(\Omega)$. We are looking for a local solution of

$$\min_{v \in \mathcal{K}} \mathcal{J}(v) := \frac{1}{2} a(v,v) - l(v), \tag{3.14}$$

where $l \in H'$ and a is a symmetric and continuous bilinearform, i.e.

$$a(w,v) \leq \gamma \|w\|_1 \, \|v\|_1.$$

Note that we do not require a to be coercive as it for example the case when a corresponds to the second derivative of a non-convex non-linear strain energy functional (1.16). The set of feasible solutions is of the form

$$\mathcal{K} := \{ v \in H : \varphi_l \leq v \leq \varphi_u \text{ a.e. on } \Omega \}, \tag{3.15}$$

with lower and upper obstacles $\varphi_l, \varphi_u \in H \cap C(\bar{\Omega})$. Next, we apply the finite element method to discretise (3.14). Let $\mathcal{T}_j$ be a triangulation of Ω and $\mathcal{S}_j = \mathcal{S}_j(\mathcal{T}_j) := \operatorname{span}\langle \lambda_1^j, \ldots, \lambda_{N_j}^j \rangle$ be the space of piecewise linear finite elements on $\mathcal{T}_j$. The number of vertices is denoted by $N_j = |\mathcal{N}_j|$ and λ_p^j is the Lagrange basis function corresponding to the node $p \in \mathcal{N}_j$, cf. (2.11). The discretised problem then reads

$$\min_{v_j \in \mathcal{K}_j} \mathcal{J}(v_j), \tag{3.16}$$

with

$$\mathcal{K}_j := \{ v_j \in \mathcal{S}_j : \varphi_l \leq v_j \leq \varphi_u \},$$

where, using the same notation, $\varphi_l, \varphi_u \in \mathcal{S}_j$ denote piecewise linear approximations of the continuous obstacles.

The idea of multigrid methods is to first apply a *smoother* like the Gauss–Seidel method, i.e. an algorithm that only removes the high-frequency parts of the error rather than reducing it and then project the smoothed iterate to a coarser finite element space. The remaining low-frequency components of the error however, correspond to high-frequencies on the coarser grid level. Hence, they can be eliminated by applying a smoother on the coarse level. Repeating this procedure on a grid hierarchy then leads to very fast algorithms [22]. Let us assume that there exists a nested hierarchy of triangulations

$$\mathcal{T}_0 \subset \ldots \subset \mathcal{T}_j,$$

and denote the corresponding linear finite element spaces by

$$\mathcal{S}_0 \subset \ldots \subset \mathcal{S}_j.$$

For each finite element space $\mathcal{S}_k$ let $\mathcal{N}_k$ denote the set of nodes with $N_k = |\mathcal{N}_k|$ and let λ_p^k be the nodal basis function corresponding to the node $p \in \mathcal{N}_k$. If no constraints are present, i.e. $\mathcal{K}_j = \mathcal{S}_j$, then the standard multigrid method with Gauss–Seidel smoothers is given by

Algorithm MG (Standard Multigrid Method)
Given: $u_j^\nu \in \mathcal{S}_j$
Initialise iterate: $v^{j+1} = u_j^\nu$
for $k = j, \ldots, 1$ {
  Gauss−Seidel smoothing:
  $w_0^k = v^{k+1}$
    for $p = 1, \ldots, N_k$ *do* {
      Choose $\alpha_p \in \arg\min \mathcal{J}(w_p^k + \alpha \lambda_p^k)$
      $w_p^k = w_{p-1}^k + \alpha_p \lambda_p^k$
    }
  Set $v^k = w_{N_k}^k$
}
Coarse grid smoothing/solution:
  $v^0 \in \underset{w \in \mathcal{S}^0}{\arg\min}\, \mathcal{J}(v^1 + w)$
Next iterate: $u_j^{\nu+1} = v^1 + v^0$

To extend this multigrid method to obstacle problems one has to decide how to incorporate the constraints into the coarse level problems. Simply choosing the exact defect obstacles $\varphi_l - v^k$ on each level will result in a suboptimal method as the fine grid $\mathcal{T}_j$ has to be revisited within each Gauss–Seidel smoothing, see [54]. The idea of the monotone multigrid method is to reduce the computational complexity by approximating the coarse grid defect obstacles using *monotone restriction operators* $R_k^{k-1} : \mathcal{S}_k \longrightarrow \mathcal{S}_{k-1}$

$$(R_k^{k-1}\psi)(p) = \max\left\{\psi(q) :\ q \in \mathcal{N}_k \cap \operatorname{int}\operatorname{supp} \lambda_p^{k-1}\right\} \quad \forall\, p \in \mathcal{N}_{k-1}, \tag{3.17}$$

see [54]. The straight-forward extension of this method to non-convex problems is given as follows:

Algorithm MMG (Monotone Multigrid Method)
Given: $u_j^\nu \in \mathcal{K}_j$
Initialise iterate: $v^{j+1} = u_j^\nu$
Initialise defect obstacles: $\psi_l^j = \varphi_l - v^{j+1}$
$\psi_u^j = \varphi_u - v^{j+1}$
For $k = j, \dots, 1$ {
Gauss−Seidel smoothing:
$w_0^k = v^{k+1}$
for $p = 1, \dots, N_k$ {
Choose $\alpha_p \in \underset{\alpha \in [\psi_l^k(p), \psi_u^k(p)]}{\arg\min} \mathcal{J}(w_{p-1}^k + \alpha \lambda_p^k)$
$w_p^k = w_{p-1}^k + \alpha_p \lambda_p^k$
}
Set $v^k = w_{N_k}^k$, $\alpha^k := (\alpha_p)_p$
Update defect obstacle: $\psi_l^k = \psi_l^k - \alpha^k, \quad \psi_u^k = \psi_u^k - \alpha^k$
Monotone restriction: $\psi_l^{k-1} = R_k^{k-1} \psi_l^k, \quad \psi_u^{k-1} = -R_k^{k-1} \psi_u^k$
}

Coarse grid smoothing/solution:
$v^0 \in \underset{w \in \mathcal{S}_0}{\arg\min} \mathcal{J}(v^1 + w)$ *s.t.* $\psi_l^0 \le w \le \psi_u^0$

Next iterate: $u_j^{\nu+1} = v^1 + v^0$

Remark 3.2.1. *Note that in practice the monotone multigrid method can be written and implemented as a V-cycle which ensures optimal complexity. In this section we are investigating the global convergence of this method when applied to non-convex problems. For this purpose the successive subset minimisation formulation proved to be more suitable. Furthermore, truncation of the coarse grid basis functions can be applied to speed-up the method, see [74]. This does not affect the convergence theory developed within this thesis and will be neglected at this point for simplicity.*

Remark 3.2.2. *The one-dimensional local problems in Algorithm MMG can be solved exactly:*

$$a(\lambda_p^k, \lambda_p^k) > 0 \Longrightarrow \alpha_p = \operatorname{med} \left\{ \psi_l^k(p), \frac{l(\lambda_p^k) - a(w_{p-1}^k, \lambda_p^k)}{a(\lambda_p^k, \lambda_p^k)}, \psi_u^k(p) \right\}. \tag{3.18}$$

$$a(\lambda_p^k, \lambda_p^k) \le 0 \Longrightarrow \alpha_p \in \underset{\alpha \in \left\{ \psi_l^k(p), \psi_u^k(p) \right\}}{\arg\min} \mathcal{J}(w_{p-1}^k + \alpha \lambda_p^k), \tag{3.19}$$

where the median is defined via

$$a = \operatorname{med} \{a, b, c\} \Longleftrightarrow b \le a \le c \text{ or } c \le a \le b.$$

If the quadratic one-dimensional problem is concave the minimal value must lie on the boundary of the feasible domain $[\psi_l^k(p), \psi_u^k(p)]$ *which corresponds to the second case* (3.19).

Remark 3.2.3. *Although, as in the previous section, we do not have to specify an explicit choice which minimal value we select in the case of multiple minima* α_p*, in practice we will choose zero, if possible, or the value on the left boundary of the interval, i.e.* $\psi_l^k(p)$.

Convergence Proof

We now will prove global convergence of Algorithm MMG by applying Theorem 3.1.5. To this end, we rewrite the algorithm as a successive direction minimisation method and show that all necessary requirements are fulfilled. First we need to define the set of search directions. As in (2.13) we directly identify finite element functions with their coefficient vectors in $\mathbb{R}^{N_j}$. The algebraic fine grid representation of quantities defined on coarser levels $k < j$ is thereby computed by interpolating them to the fine grid and considering the corresponding coefficient vector. Using the same notation for finite element functions and their algebraic representation, the set of search directions is given by the multilevel basis interpolated to the finest level j

$$\mathcal{D} = \left\{ \lambda_p^k \in \mathbb{R}^{n_j} : p \in \mathcal{N}_k,\, 0 \leq k \leq j \right\}. \tag{3.20}$$

The definition of the feasible line search mappings (3.4) is complicated by the fact that only approximate coarse obstacles are used which are restricted *after* the Gauss–Seidel smoothing on each level $\mathcal{N}_k$. A workaround is to extend the state spaces to $\mathbb{R}^{3N_j}$ and consider triples of finite element functions and the defect obstacles

$$z := (\psi_l, \psi_u, w) \in \mathbb{R}^{3N_j},$$

where again the fine grid representation of the defect obstacles is considered. Now for each multilevel basis function we define the feasible line search mapping via

$$h(\lambda_p^k, z) = h(\lambda_p^k, (\psi_l, \psi_u, w)) = [\psi_l(p), \psi_u(p)], \tag{3.21}$$

where $\psi_l(p) \in \mathbb{R}$ denotes the component of the fine grid vector corresponding to the node $p \in \mathcal{N}_k$. These coarse defect obstacles are in general more restrictive than the exact obstacles and thus guarantee the feasibility of the intermediate iterates. Furthermore, Assumptions M1 and M2 are fulfilled.

Lemma 3.2.4. *The mappings* $h(\lambda_p^k, \cdot)$ *are continuous for all* λ_p^k*. Further, for each* $w \in \mathcal{K}_j$*, the intermediate iterates* w_p^k *of Algorithm MMG are feasible and it holds*

$$0 \in h(\lambda_p^k, (\psi_l^k, \psi_u^k, w_p^k)), \tag{3.22}$$

where ψ_l^k, ψ_u^k *denote the corresponding defect obstacles.*

Proof. We only consider the lower obstacles in the following, the upper case can be obtained analogously. By construction of the defect obstacles we obtain

$$\max_{\substack{1\leq i\leq N_j\\(\lambda_p^k)_i\neq 0}} (\varphi_l - w)_i \leq \psi_l^k(p),$$

and hence the intermediate iterates are feasible. On the finest level j it holds

$$\psi_l^j = \varphi_l - w \leq 0.$$

Now (3.22) follows by observing that

$$\psi_l^k \leq 0 \Longrightarrow (R_k^{k-1}\psi_l^k)(q) \leq 0 \quad \forall\, q \in \mathcal{N}_{k-1}. \tag{3.23}$$

The continuity follows from the definition of h. □

With this at hand we now construct the directional minimisation mappings (3.5) as follows:

$$\mathcal{M}_p^k(\psi_l, \psi_u, w) = \left\{ (\psi_l - \alpha_p\lambda_p^k, \psi_u - \alpha_p\lambda_p^k, w + \alpha_p\lambda_p^k) \,:\, \alpha_p \in \operatorname*{arg\,min}_{\alpha\in[\psi_l(p),\psi_u(p)]} \mathcal{J}(w + \alpha\lambda_p^k) \right\}.$$

The mappings $\mathcal{M}_{N_k}^k$ corresponding to the last nodes of each level $N_k \in \mathcal{N}_k$ additionally perform the monotone restriction

$$(\psi_l, \psi_u, w) \mapsto (R_k^{k-1}(\psi_l - \alpha_{N_k}\lambda_{N_k}^k), -R_k^{k-1}(\psi_u - \alpha_{N_k}\lambda_{N_k}^k), w + \alpha_{n_k}\lambda_{N_k}^k).$$

The monotone multigrid can now be written in terms of these mappings

$$\mathcal{M}^{\mathrm{mmg}} = \mathcal{M}_{N_0}^0 \ldots \circ \mathcal{M}_1^0 \circ \mathcal{M}_{N_1}^1 \ldots \circ \mathcal{M}_1^j, \tag{3.24}$$

where for simplicity we assumed that the coarse grid correction is again computed using the Gauss–Seidel method.

Corollary 3.2.5. *Algorithm MMG is equivalent to*

$$u_j^{\nu+1} \in \mathcal{M}^{mmg}(\varphi_l - u_j^\nu, \varphi_u - u_j^\nu, u_j^\nu,). \tag{3.25}$$

From Proposition 3.1.6 we can deduce the closedness of the directional minimisation operators.

Lemma 3.2.6. *The mappings $\mathcal{M}_p^k$ are closed.*

Proof. The monotone restriction operator is continuous on the space of continuous functions. As a result, the closedness of $\mathcal{M}_p^k$ and $\mathcal{M}_{N_k}^k$ follows immediately from Proposition 3.1.6 and the continuity of the mapping (3.21) in the second component. □

The last requirement we have to show is Assumption M3. We consider the set of local critical points, i.e. the set of local minima and saddle-points.

$$\mathcal{K}_{\mathrm{Sol}} := \{w \in \mathcal{K}_j \,:\, w \text{ local critical point}\}\,. \tag{3.26}$$

It is shown in Figure 3.2 that there exist cases where Assumption M3 does not not hold without additional modification of the monotone multigrid method. To circumvent this

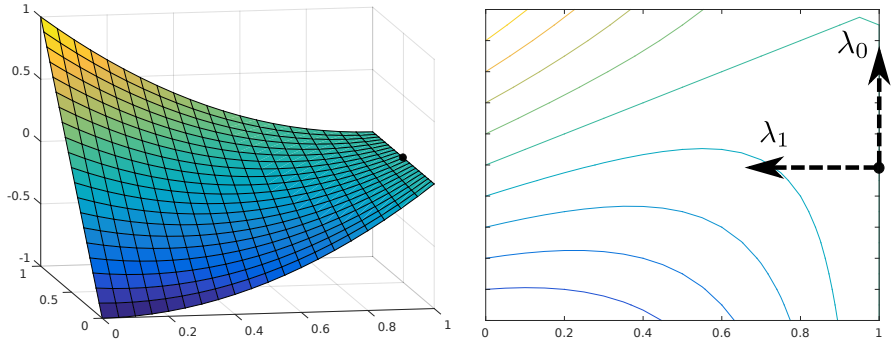


Figure 3.2.: *Example of a case where the energy is not reduced although the starting point is not critical. Left: Graph of a non-convex quadratic functional with constant zero right "boundary". Right: Contour lines of the functional. If the search directions λ_0 is chosen first, then Assumption M3 does not hold for the intermediate iterate $w_{\lambda_0} = (1, 1)$*

we prepend the leading Gauss–Seidel step by first minimising in the direction where the gradient is decreasing the most. We denote the set of the corresponding fine grid basis functions by

$$E(u_j) := \left\{ \lambda_p^j \in \mathbb{R}^{N_j} \,:\, p \in \operatorname*{arg\,min}_{q \in \mathcal{N}_j} \left\{ (-1)^r \nabla \mathcal{J}(u_j) \lambda_q^j \right\}, \quad r \text{ s.t. } (-1)^r \lambda_q^j \in T_{\mathcal{K}_j}(u_j) \right\}, \tag{3.27}$$

where $T_{\mathcal{K}_j}(u_j)$ is the contingent cone of $\mathcal{K}_j$ at u_j, see [121]. The scaling in (3.27) has to be added to sort out the directions pointing out of the feasible set $\mathcal{K}_j$ for the case where $u_j \in \partial\mathcal{K}_j$. If u_j lies in the interior of $\mathcal{K}_j$, then $T_{\mathcal{K}_j}(u_j) = \mathbb{R}^{N_j}$, cf. Figure 3.3. The corresponding directional minimisation

$$\mathcal{M}^{\mathrm{sg}}(w) := \left\{ w + \alpha_p \lambda_p^j : \alpha_p \in \operatorname*{arg\,min}_{\beta \in [\psi_l(p), \psi_u(p)]} \mathcal{J}(w + \beta\lambda_p^j),\ \lambda_p^j \in E(u_j) \right\},$$

fulfils Assumption M3.

Corollary 3.2.7. *Let $w \in \mathcal{K}_j$ not be first order critical, then*

$$\mathcal{J}(w) < \mathcal{J}(v) \quad \forall\, v \in \mathcal{M}^{\mathrm{sg}}(w). \tag{3.28}$$

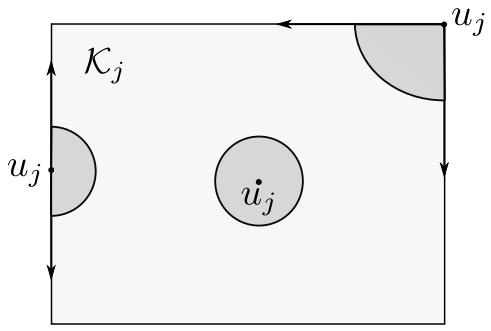


Figure 3.3.: *The contingent cone $T_{\mathcal{K}_j}$ of a convex set $\mathcal{K}_j$ at different points u_j.*

Note that in the definition of $\mathcal{M}^{\text{sg}}$ we omitted the dependency on the defect obstacles ψ_l^j, ψ_u^j for simplicity. For this mapping it further holds

Lemma 3.2.8. *The mapping $\mathcal{M}^{sg}$ is closed.*

Proof. Consider the convergent sequences

$$w^n \longrightarrow \bar{w}, \quad v^n \longrightarrow \bar{v},$$

with $v^n \in \mathcal{M}^{\text{sg}}(w^n)$. From $v^n \in \mathcal{M}^{\text{sg}}(w^n)$ we obtain that for each v^n there exists $\lambda^j_{p(n)} \in E(w^n)$, such that

$$v^n = w^n + \alpha^n \lambda^j_{p(n)},$$

for a sequence $\alpha^n \longrightarrow \bar{\alpha}$. We now have to distinguish between the case where $\bar{w} = \bar{v}$ and $\bar{w} \neq \bar{v}$.
First let $\bar{w} \neq \bar{v}$. Then, as v^n converges, for n large enough we obtain $v^n = w^n + \alpha^n \lambda^j_p$ for some fixed $\lambda^j_p \in E(\bar{w})$. Again, by definition of $\mathcal{M}^{\text{sg}}$ the following inequality holds for n large enough

$$\mathcal{J}(v^n) = \mathcal{J}(w^n + \alpha^n \lambda^j_p) \leq \mathcal{J}(w^n + \alpha \lambda^j_p) \quad \forall\, \alpha \in [\psi_l^j(p), \psi_u^j(p)].$$

Taking the limit and using the continuity of $\mathcal{J}$ we deduce

$$\mathcal{J}(\bar{v}) = \mathcal{J}(\bar{w} + \bar{\alpha} \lambda^j_p) \leq \mathcal{J}(\bar{w} + \alpha \lambda^j_p) \quad \forall\, \alpha \in [\psi_l^j(p), \psi_u^j(p)].$$

We conclude

$$\bar{v} \in \mathcal{M}^{\text{sg}}(\bar{w}) \Longleftrightarrow \lambda^j_p \in E(\bar{w}).$$

Now from the continuity of $\nabla \mathcal{J}$ it follows that for n large enough $E(w^n) \subset E(\bar{w})$ which completes the proof for the case $\bar{w} \neq \bar{v}$.
The case where $\bar{w} = \bar{v}$ can be handled analogously by choosing a subsequence w^{n_l} with $v^{n_l} = w^{n_l} + \alpha^{n_l} \lambda^j_p$. As also the subsequences converge to $\bar{w}$ and $\bar{v}$ correspondingly, the claim follows. □

Theorem 3.2.9. *The monotone multigrid method MMG prepended with the Gauss–Seidel step $\mathcal{M}^{sg}$ either stops at a first order critical point of* (3.16) *or the limit of every convergent subsequence is first order critical.*

Proof. Use Theorem 3.1.5 with the solution set (3.26) and the preceding Corollary 3.2.7 and Lemmata 3.2.4, 3.2.6 and 3.2.8. □

Remark 3.2.10. *Although this theorem only ensures convergence towards first order critical points, in practice the method does not get stuck at saddle points because the number of search directions λ_p^k exceeds the space dimension n_j. This also motivates why in practical applications prepending the monotone multigrid method with $\mathcal{M}^{sg}$ can be neglected and Assumption M3 in general still holds.*

This theory is numerically confirmed in Section 4.4.1, where the solver is tested within an obstacle problem for a non-convex non-linear hyperelastic body.

3.3. Trust–Region Methods

In preparation of the next section 3.4 we now shortly describe the trust–region method at the example of a non-linear non-convex obstacle problem. Trust–region techniques together with line search methods are typically used within non-linear optimisation schemes to ensure the sufficient decrease of energy in each iteration. A more detailed introduction and a general overview to this topic can be found in [30]. In the appendix A the interested reader can find a short introduction to non-linear optimisation and criticality measures. In this section we consider the non-linear obstacle problem

$$\min_{\varphi \in \mathcal{K}} \mathcal{J}(\varphi), \tag{3.29}$$

with $\mathcal{J} : \mathbb{R}^N \longrightarrow \mathbb{R}$ continuously differentiable on $\mathcal{K}$, which is given by

$$\mathcal{K} := \left\{ \varphi \in \mathbb{R}^N : l \leq \varphi \leq u \right\}, \quad l_i, u_i \in \bar{\mathbb{R}}, 1 \leq i \leq N. \tag{3.30}$$

To solve (3.29) efficiently, a local quadratic approximation of the energy at $\varphi_k \in \mathcal{K}$ is minimised and subsequently updated.

$$\begin{aligned} \min_{v \in \mathbb{R}^N} \; m_k(v) :=& \nabla \mathcal{J}(\varphi_k)^T v + \frac{1}{2} v^T H_k v, \\ & l \leq \varphi_k + v \leq u, \end{aligned} \tag{3.31}$$

where $H_k \approx \nabla^2 \mathcal{J}(\varphi^k)$ is a symmetric approximation of the Hessian of the energy. In the unconstrained case this scheme corresponds to applying Newton's method to the necessary

optimality condition $\nabla \mathcal{J}(\varphi) = 0$. When the energy functional (3.29) is non-convex the Hessian is not necessarily positive definite any more and thus the local problem (3.31) can be unbounded. The idea of the trust–region method is now to accept this correction step v only if (3.31) is a reasonably good approximation of (3.29). Therefore, first a constraint on the step size is imposed

$$\begin{aligned} &\min_{v \in \mathbb{R}^N} m_k(v), \\ &\qquad l \leq \varphi_k + v \leq u, \\ &\qquad \|v\|_k \leq \Delta_k. \end{aligned} \tag{3.32}$$

The choice of the norm $\|\cdot\|_k$ is arbitrary and can be adjusted to the type of problem or the actual iterate [30]. After solving (3.32) approximately, the *trust–region* Δ_k is adjusted by comparing the energy reduction predicted by the model to the actual achieved energy decrease (or increase) of the local solution v_k

$$\rho_k := \frac{\mathcal{J}(\varphi_k) - \mathcal{J}(\varphi_k + v_k)}{m_k(0) - m_k(v_k)}. \tag{3.33}$$

The general trust–region algorithm reads

Algorithm TR (Standard Trust–Region Method)

Given $\varphi_k \in \mathcal{K}, \Delta_k \in \mathbb{R}$.
Choose $0 < \mu_0 \leq \mu_1 \leq 1$, $0 < \gamma_1 < 1 \leq \gamma_2$.

Compute v_k *that approximately minimises* (3.32).

If $(\rho_k \geq \mu_1)$ // *very successful step*
$\quad \varphi_{k+1} = \varphi_k + v_k$, $\Delta_{k+1} = \gamma_2 \, \Delta_k$,
else if $(\mu_2 \leq \rho_k \leq \mu_1)$ // *successful step*
$\quad \varphi_{k+1} = \varphi_k + v_k$, $\Delta_{k+1} = \Delta_k$,
else // *unsuccessful step*
$\quad \varphi_{k+1} = \varphi_k$, $\Delta_{k+1} = \gamma_1 \, \Delta_k$.

The trust–region method can be interpreted as a Newton method with an built-in damping strategy. Close to minimisers the trust–region will become inactive and local quadratic convergence is achieved if the exact Hessians are used. Global convergence of this method can be shown under the mild assumption that the model energy m_k is reduced sufficiently within each iteration. Next, we introduce the notion of a criticality measure

Definition 3.3.1 (Criticality Measure)**.** A function $\chi : \mathbb{R}^N \longrightarrow \mathbb{R}$ is called *criticality measure* of the problem (3.29) if

- χ is non-negative and continuous.
- $\chi(\varphi) = 0 \quad \Longleftrightarrow \varphi$ is a local first order critical point.

See also Appendix A.1 for a more detailed introduction.

Theorem 3.3.2. *Let $(\varphi_k)_{k\in\mathcal{N}}$ be a sequence generated by the trust–region method TR and let χ be a criticality measure of* (3.29)*. Assume that the following assumptions hold*

- *The set*
$$\mathcal{L} := \left\{\varphi \in \mathbb{R}^N : \mathcal{J}(\varphi) \leq \mathcal{J}(\varphi_0)\right\} \tag{3.34}$$
is compact and non-empty.
- *The approximate Hessians are bounded $\|H_k\| \leq C$ indep. of k.*
- *The approximate solutions of* (3.32) *fulfil*
$$m_k(0) - m_k(v_k) \geq \kappa\chi_k \min\left\{\frac{\chi_k}{\|H_k\|}, \Delta_k\right\}, \tag{3.35}$$
for some constant $\kappa > 0$ and $\chi_k := \chi(v_k)$.

Then
$$\lim_{k\longrightarrow\infty} \chi_k = 0,$$
and thus
$$\lim_{k\longrightarrow\infty} \varphi_k = \varphi^*,$$
where φ^* is a first order critical point.

Proof. [29, Theorem 11] □

The condition (3.35) is called *sufficient Cauchy decrease*. It can be shown that the generalized Cauchy point, i.e. the first local minimum on the projected gradient path, fulfils this condition, cf. Remark A.9. In the following, to avoid incompatibility of the trust–region and the linearised constraints, cf. Section 3.4, and to enable the application of fast monotone multigrid methods, we will always choose the trust–region norm to be the infinity norm
$$\|v\|_k = \|v\|_\infty = \max_{1\leq p\leq N} \{|v_p|\}. \tag{3.36}$$
The local problems (3.32) can be reformulated as
$$\min_{v\in\mathbb{R}^N} m_k(v), \quad l_k \leq v \leq u_k, \tag{3.37}$$

with

$$l_{k,i} = \max\left\{(l-\varphi_k)_i, -\Delta_k\right\}, \quad 1 \le i \le N,$$
$$u_{k,i} = \max\left\{(u-\varphi_k)_i, \Delta_k\right\}, \quad 1 \le i \le N.$$

Corollary 3.3.3. *Let $(\varphi_k)_{k\in\mathcal{N}}$ be a sequence generated by the trust–region method TR, where (3.37) is solved using the monotone multigrid method MMG and assume that the following assumptions hold*

- *The sublevel set (3.34) is compact and non-empty.*
- *The approximate Hessians are bounded $\|H_k\| \le C$ indep. of k.*

Then φ_k converges to a first order critical point.

Proof. The corollary follows from Theorems 3.2.9 and 3.3.2 and a suitable number of iteration steps of the monotone multigrid method. □

Remark 3.3.4. *Numerical tests strongly indicate that already one monotone multigrid iteration is enough to exceed the decrease generated by the generalised Cauchy point.*

3.4. Inexact Filter–Trust-Region Methods

The filter method was first introduced by Fletcher and Leyffer in 1997 [45], as an penalty-free alternative to the widely used merit function approach. It was originally constructed as an extension of the trust–region method to non-linear constrained problems but nowadays also extensions to line search methods exist, see e.g. [132].
The standard trust–region method TR cannot be used in the presence of general constraints because the acceptance of an iterate is only decided upon by the achieved energy reduction. In most algorithms based on sequential quadratic programming (SQP) the intermediate iterates are not necessarily feasible any more. This is why the comparison of two iterates may not only involve the corresponding energy values but also their infeasibility with respect to the non-linear constraints.
In this section we start with presenting the basic first order consistent SQP method. Afterwards, we introduce a class of inexact constraint Jacobians that induce box-constrained local problems. For this class we then develop an extension of the classical filter approach that allows the use of inexact constraint Jacobians and prove global convergence towards first order critical points within the framework of composite-step methods. The inexact local problems can be solved using fast monotone multigrid methods which speeds up the overall method considerably, see Section 4.4.2. In the following we consider the general

case of minimising a smooth non-convex energy functional $\mathcal{J} : \mathbb{R}^N \longrightarrow \mathbb{R}$ subject to smooth possibly non-convex inequality constraints $g : \mathbb{R}^N \longrightarrow \mathbb{R}^m$

$$\min_{\varphi \in \mathcal{K}} \mathcal{J}(\varphi), \tag{P}$$

where the set of feasible configurations is given by

$$\mathcal{K} := \left\{ \varphi \in \mathbb{R}^N : \, g(\varphi) \geq 0 \right\}. \tag{3.38}$$

Sequential Quadratic Programming

The SQP method can be derived by applying Newton's method to the first order optimality system of the problem

$$\begin{aligned} \nabla \mathcal{J}(\varphi) + \lambda^T \nabla g(\varphi) &= 0, \\ g(\varphi) &\geq 0, \end{aligned} \tag{3.39}$$

where λ denotes the Lagrange multiplier, cf. (KKT) in the appendix. In the following we will use the following abbreviations:

$$\begin{aligned} f_k &:= \nabla \mathcal{J}(\varphi_k), & G_k &= \nabla g(\varphi_k), \\ H_k &:= \nabla^2 \mathcal{J}(\varphi_k), & L_k &= H_k + \lambda_k^T \nabla^2 g(\varphi_k). \end{aligned}$$

Now applying Newton's method to the system (3.39) results in

$$L_k u + \mu G_k = -f_k, \quad G_k \, u \geq -g(\varphi_k),$$

which has to be solved for the correction (u, μ) of the configuration and the Lagrange multiplier

$$\varphi_{k+1} = \varphi_k + u, \quad \lambda_{k+1} = \lambda_k + \mu.$$

The Newton problem can be equivalently written as a quadratic minimisation problem

$$\begin{aligned} \min_{u \in \mathbb{R}^N} m_k(u) := f_k^T u + \frac{1}{2} u^T L_k \, u, \\ G_k \, u + g(\varphi_k) \geq 0. \end{aligned} \tag{3.40}$$

Having its origin in the Newton method, successively solving (3.40) can be shown to converge locally super-linear or quadratically under suitable assumptions, see [30, Theorem 15.2.1]. In this thesis we will apply an approximate SQP method where L_k is replaced by H_k.

$$\begin{aligned} \min_{u \in \mathbb{R}^N} m_k(u) := f_k^T u + \frac{1}{2} u^T H_k \, u, \\ G_k \, u + g(\varphi_k) \geq 0. \end{aligned} \tag{3.41}$$

Remark 3.4.1. *Note that the model* (3.41) *is not second order consistent any more as the Hessian of the Lagrangian function is replaced by the Hessian of the objective functional $\mathcal{J}$ which is widely accepted [30]. Although this is preventing local quadratic convergence, it also avoids the need of updating the intermediate Lagrange multipliers. Further, local quadratic convergence in constrained optimization is usually complicated by the Maratos effect, which involves additional computational effort, e.g. by second order corrections [101].*

Inexact constraint Jacobians

In this paragraph we present a certain class of inexact constraint Jacobians $\widetilde{G}_k \approx \nabla g(\varphi_k)$ that allow to transform the local SQP problems into box-constrained ones, for which a large variety of fast solution schemes exist. We remark here that the global convergence proof of the inexact filter method does not depend on this special structure of the local problems. Hence it can be extended to a broader class of constraint approximations under additional, but standard, assumptions. From now on let $\widetilde{G}_k$ denote an approximation of the exact constraint Jacobian $\nabla g(\varphi_k)$.

Assumption J. The approximate constraint Jacobians are continuous and there exists a continuous invertible transformation $T = T(\varphi) : \mathbb{R}^N \longrightarrow \mathbb{R}^N$, such that the transformed Jacobians

$$\widetilde{G}_k T(\varphi_k)^T \tag{3.42}$$

are positive, diagonal and the inverse is bounded away from zero

$$\|(\widetilde{G}_k T(\varphi_k)^T)^{-1}\| \leq \kappa_T. \tag{3.43}$$

with $\kappa_T > 0$ independent of k.

If Assumption J holds the inexact local problems can be transformed into box-constrained ones using $T_k := T(\boldsymbol{\varphi}_k)$

$$\begin{aligned} \min_{u \in \mathbb{R}^N} m_k(u) :=& f_k^T T_k^T u + \frac{1}{2} u^T T_k \, H_k \, T_k^T u, \\ & u + \tilde{g}_k \geq 0, \end{aligned} \tag{QP}$$

where

$$\tilde{g}_k := (\widetilde{G}_k \, T_k^T)^{-1} g(\varphi_k). \tag{3.44}$$

The approximation error that we make by using an inexact constraint Jacobian will be denoted by

$$e_k(u) := \|(G_k - \tilde{G}_k)u\|.$$

To simplify notation, we will use the same notation H_k, f_k for the quantities in Euclidean coordinates and in transformed "diagonal" coordinates.

Composite-Step method

As mentioned earlier, having their origin in Newton's method, SQP can only be expected to converge locally. In the previous section 3.3 we shortly showed how the trust-region mechanism can be used as a globalisation technique, adding an automatic damping strategy. In this paragraph we will introduce the *composite-step method* that aims at incorporating the trust–region constraints into the local SQP model with approximate Jacobians (QP). Therefore, we first augment the local model as follows

$$\begin{aligned} \min_{u\in\mathbb{R}^N} \; & m_k(u), \\ & u + \tilde{g}_k \geq 0, \\ & \|u\|_\infty \leq \Delta_k, \end{aligned} \tag{TRQP}$$

where we again choose the infinity norm to restrict the step size, cf. (3.36). In general,

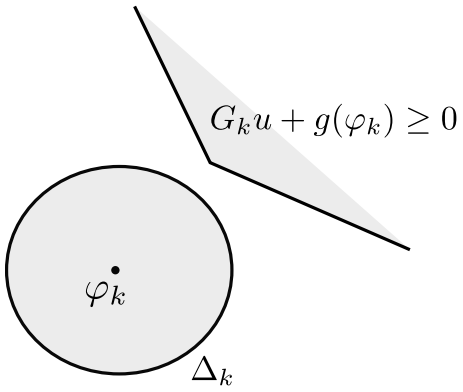


Figure 3.4.: *Incompatible linearised constraints and trust-region obstacles*

imposing the additional trust-region constraints can lead to local quadratic problems that are infeasible if the linearised constraints and the trust-region obstacles are not compatible, see Figure 3.4. To resolve this issue composite-step methods have been developed which decompose the SQP iteration into a normal and a tangential step

$$u_k = n_k + t_k. \tag{3.45}$$

The normal step n_k is aiming towards reducing the infeasibility of the iterate, e.g. by solving

$$\min_{n\in R^N} \frac{1}{2}\|n\|^2, \quad n + \tilde{g}_k \geq 0. \tag{3.46}$$

In the case of box constraints the normal step can be calculated directly by projection on the feasible set

$$n_{k,i} \begin{cases} -\tilde{g}_{k,i} & g_i(\varphi_k) < 0, \\ 0 & \text{else.} \end{cases}$$

Once a normal step is found the tangential step t_k tries to achieve a reduction of the local energy while retaining the feasibility

$$\begin{aligned} \min_{t\in\mathbb{R}^N} & m_k(n_k+t), \\ & t+n_k+\tilde{g}_k \geq 0, \\ & \|t+n_k\|_\infty \leq \Delta_k. \end{aligned} \tag{3.47}$$

If the infeasibility is too large, e.g. $\tilde{g}_k \approx -\Delta_k$, and as a result the norm of n_k is close to Δ_k, one might not be able to compute a tangential step that can sufficiently reduce the energy within the trust-region. The following definition of "good" cases is usually used within composite-step methods.

Definition 3.4.2. TRQP(φ_k, Δ_k) is called *compatible* if there are constants $\kappa_\mu > 0$, $\kappa_\Delta \in (0,1]$, and $\mu \in (0,1)$ such that

$$\|n_k\| \leq \kappa_\Delta \Delta_k \min\{1, \kappa_\mu \Delta_k^\mu\}. \tag{3.48}$$

Remark 3.4.3. *Note that in the case of diagonal constraint Jacobians the normal step n_k always exists and the compatibility of the local problems can be checked directly. In practical computations the decomposition can thus be omitted when a globally convergent solver is used for the minimisation of the local energy. Still, we will formally apply this splitting, on the one hand for the possible extension to non-diagonal inexact Jacobians and on the other hand as a theoretical tool to proof global convergence of the filter method.*

If the local problem is not compatible we have to use a more "aggressive" strategy to compute a suitable correction u_k, which will be introduced in the next section 3.4. The trust–region framework is used to ensure the sufficient decrease of energy. Accordingly, we will require the tangential step to fulfil the sufficient Cauchy decrease condition

$$m_k(n_k+t_k) - m(0) \geq \kappa_{\mathrm{sc}} \widetilde{\chi}_k \min\{\frac{\widetilde{\chi}_k}{\|H_k\|}, \Delta_k\}, \tag{3.49}$$

which was introduced in the previous section (3.35). The first order criticality measure $\widetilde{\chi}_k$ corresponds to the local tangential problem with inexact constraint Jacobians (3.47).

$$\widetilde{\chi}_k = \widetilde{\chi}(t_k) := |\min_{\substack{d+n_k+\tilde{g}_k\geq 0 \\ \|d\|=1}} \langle \nabla m(n_k+t_k), d\rangle|, \tag{3.50}$$

cf. (A.11). Analogously we define the exact criticality measure by

$$\chi(t_k) := |\min_{\substack{G_k(n_k+d)+g(\varphi_k)\geq 0 \\ \|d\|=1}} \langle \nabla m(n_k+t_k), d\rangle|, \tag{3.51}$$

Remark 3.4.4. *When the exact criticality $\chi(0) = 0$, then n_k is a critical point of the local problem TRQP. If additionally the iterate φ_k is feasible and therefore $n_k = 0$, then φ_k is itself a first order critical point of the non-linear problem* (P)*, see [32, Lemma 5].*

One advantage of this criticality measure is, that it can be computed by solving a linear convex problem. Hence, it can be used to directly monitor the convergence of a non-linear algorithm.

Filter - How to compare iterates

The second difficulty that appears when combining SQP methods with a trust-region globalisation strategy is that the intermediate iterates (3.41) are in general not feasible w.r.t the non-linear constraints. This is why the trust-region mechanism, which decides on the acceptance of a trial point only by comparing the energy decrease, cannot lead to a convergent method. The filter method resolves this issue by forcing the iterates to converge towards the feasible set. This in combination with the trust-region mechanism leads to globally convergent methods [46]. To measure the infeasibility of an iterate we use the following non-smooth function

$$\vartheta(\varphi) := \max_{i=1,\dots,m} \{0, -g_i(\varphi)\}\,. \tag{3.52}$$

In the following we use the abbreviations $\mathcal{J}_k := \mathcal{J}(\varphi_k)$ and $\vartheta_k := \vartheta(\varphi_k)$ to denote the energy and infeasibility of an iterate. If a candidate φ_{k+1} is *dominated* by an previous iterate φ_j, i.e.

$$\mathcal{J}_j \leq \mathcal{J}_{k+1} \quad \text{and} \quad \vartheta_j \leq \vartheta_{k+1}, \tag{3.53}$$

then the candidate should be rejected. The critical case is to decide for the case where

$$\mathcal{J}_{k+1} < \mathcal{J}_j, \qquad \vartheta_{k+1} > \vartheta_j, \tag{3.54}$$

whether φ_{k+1} is acceptable or not. To overcome this difficulty Fletcher and Leyffer introduced the notion of a *filter* [45]

Definition 3.4.5. A set of tuples $(\vartheta_i, \mathcal{J}_i)$ is called a *filter* $\mathcal{F}$, if

$$\mathcal{F} := \{(\vartheta_i, \mathcal{J}_i) : \text{ no pair dominates each other}\}\,. \tag{3.55}$$

Candidates are acceptable if the following criterion holds

Definition 3.4.6. A configuration φ is called *acceptable to the filter* $\mathcal{F}$, if

$$\vartheta(\varphi) < (1-\gamma_\vartheta)\vartheta_i \text{ or } \mathcal{J}(\varphi) < \mathcal{J}_i - \gamma_\vartheta\vartheta(\varphi) \quad \forall\,(\vartheta_i, \mathcal{J}_i) \in \mathcal{F}, \tag{3.56}$$

for some fixed constant $0 < \gamma_\vartheta \leq 1$, cf. Figure 3.5

Remark 3.4.7. *To require an iterate not to be dominated by the filter is not enough to ensure convergence of the iterates towards the feasible set. The use of the slightly stronger criterion* (3.56) *was proposed in [46], which guarantees a sufficient reduction of the infeasibility whenever a point is acceptable.*

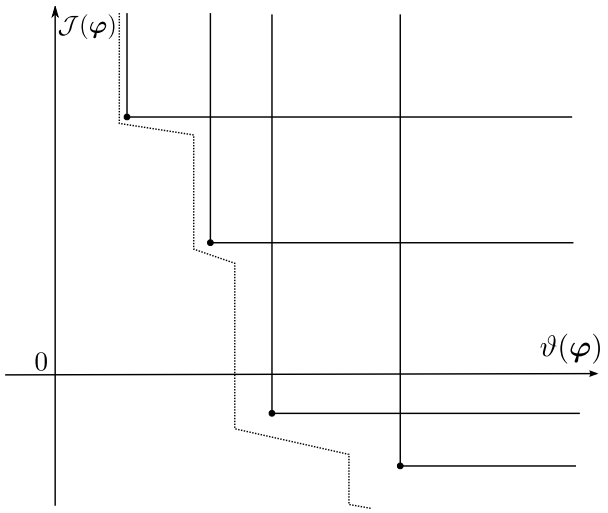


Figure 3.5.: *Illustration of a filter with four points. The area above the dotted line is not acceptable by the filter.*

Within the filter method, whenever a pair is *added to the filter* $\mathcal{F}$, all pairs that are dominated by the new pair are removed so that $\mathcal{F}$ stays a filter. From this acceptance criterion convergence of the iterates towards the feasible set can be shown

Lemma 3.4.8. *Let $(\varphi_k)_{k\in\mathcal{N}}$ be a bounded sequence which is added to the filter, then*

$$\vartheta_k \longrightarrow 0. \tag{3.57}$$

Proof. [30, Lemma 15.5.2] □

Based on the original filter–trust–region method from [47], we are now ready to formulate the main algorithm of this chapter.

Algorithm FTR (Inexact Filter Trust–Region Step)

Given φ_k, $\mathcal{F}$ (or initialize by $\emptyset$)
Choose $\varepsilon, \kappa_{\mathrm{ac}}, \kappa_\vartheta > 0$, $0 < \gamma_\vartheta \leq 1$ \\ *filter parameter*
$0 < \eta_1 < \eta_2 \leq 1$, $0 < \gamma_1 < 1 < \gamma_2$ \\ *trust−region parameter*

// Close to the optimal point, make an exact step
1. If $\widetilde{\chi}_k < \varepsilon$
Solve $TRQP(\varphi_k, \Delta_k)$ with exact G_k, go to 4.

//Check if $TRQP(\varphi_k, \Delta_k)$ is feasible
2. If $TRQP(\varphi_k, \Delta_k)$ not compatible {
add φ_k to the filter,

call restoration phase $\rightarrow (\varphi_{k+1}, \Delta_{k+1})$*, go to 1.*
}

// Check error of the inexact constraints
3. Compute $u_k = n_k + t_k$ *from* $TRQP(\varphi_k, \Delta_k)$

If

$$e_k(u_k) > \kappa_{\text{ac}} \Delta_k^2, \tag{3.58}$$

then
solve $TRQP(\varphi_k, \Delta_k)$ *with exact* G_k

// Check if iterate is acceptable
4. Evaluate $\vartheta(\varphi_k + u_k)$, $\mathcal{J}(\varphi_k + u_k)$

// Unsuccessful Iteration
If
$\varphi_k + u_k$ *not acceptable to the Filter*

or if $\rho_k < \eta_1$ *and*

$$m_k(0) - m_k(u_k) \geq \kappa_\vartheta (\vartheta_k)^2, \tag{3.59}$$

then // Reject iterate and reduce trust-region
$\varphi_{k+1} = \varphi_k$, $\Delta_{k+1} = \gamma_{\frac{1}{2}} \Delta_k$

// Accept trial point
$\varphi_{k+1} = \varphi_k + u_k$

// $\vartheta-$*type iteration to decrease infeasibility.*
If (3.59) *does not hold*
add φ_k *to filter*

// Very Successful Iteration, increase trust−region
If (3.59) *and* $\rho_k \geq \eta_2$
$\Delta_{k+1} = \gamma_2 \Delta_k$

If the local problem TRQP(φ_k, Δ_k), with exact or inexact Jacobian, is not compatible, i.e. solvable, we have to enter the so-called *feasibility restoration phase* (FRP). In this phase we try to compute a new iterate $\varphi_{k+1} = \varphi_k + r_k$ and trust-region radius $\Delta_{k+1} > 0$ such that the new iterate is acceptable to the filter and the local problem TRQP$(\varphi_{k+1}, \Delta_{k+1})$

is compatible. This is usually done by applying an iterative method that minimises the infeasibility directly

$$\min_{\varphi \in \mathbb{R}^N} \vartheta(\varphi). \tag{3.60}$$

To enable the (FRP) to find an acceptable new iterate it is thus crucial that the filter does not contain any feasible points, i.e.

$$(\vartheta_j, \mathcal{J}_j) \in \mathcal{F} \Longrightarrow \vartheta_j \neq 0.$$

This is ensured by condition (3.59) which was first proposed in [47]. If (3.59) fails, the filter iteration is merely trying to reduce the infeasibility, which is called an *ϑ-type* iteration. If the infeasibility is not dominating then the algorithm is using the standard trust-region tools to achieve a reduction in energy, an *$\mathcal{J}$-type* iteration.

The condition (3.58) ensures that we control the error that we make by using approximate constraint Jacobians. If the approximation error is larger than the bound we have to solve the quadratic problem with exact constraint Jacobian to ensure global convergence of the method towards a critical point.

Further, whenever the inexact criticality measure $\widetilde{\chi}_k$ is indicating that the iterate is close to the solution, the local problem with the exact Jacobian has to be soled, or if the exact problem is incompatible, the (FRP) is entered. In this case we assume that the solver is again generating sufficient Cauchy decrease (3.49) with respect to the exact criticality measure (3.51). This ensures that for a sequence $(\varphi_k)_{k \in \mathcal{N}}$ with

$$\lim_{k \longrightarrow \infty} \widetilde{\chi}_k = 0,$$

it also holds that

$$\lim_{k \longrightarrow \infty} \chi_k = 0.$$

Global Convergence of the filter SQP-method

Now we will prove global convergence of Algorithm FTR. To a great extend the proof goes along the lines of the proof for the filter method with exact constraint Jacobians [47]. In this section will only show the main convergence theorem and the parts of the proof that had to be adjusted to account for the inexactness. The interested reader can find the details on the whole proof in the Appendix B. For the rest of this section we make the following assumptions.

Assumption F1. The iterates φ_k of Algorithm FTR stay in a compact set $\mathcal{C} \supseteq \mathcal{K}$.

Assumption F2. The energy functional $\mathcal{J}$ and the constraints g are twice continuously differentiable on $\mathcal{C}$.

Assumption F3. The inexact Jacobians $\widetilde{G}(\cdot)$ are continuous on $\mathcal{C}$ and induce a continuous invertible transformation T. The transformed inexact Jacobians are positive, diagonal and the inverse is bounded away from zero

$$\|(\widetilde{G}(\varphi)T(\varphi)^T)^{-1}\| \leq \kappa_T. \tag{3.61}$$

Assumption F4. For a compatible sub-problem the tangential step t_k fulfils

$$m(n_k + t_k) - m(0) \geq \kappa_{\mathrm{sc}}\widetilde{\chi}_k \min\{\widetilde{\chi}_k, \Delta_k\}.$$

By Assumptions F1 and F2 the Hessians of the energy and constraints are bounded for each iterate φ_k

$$\|H_k\| \leq \kappa_H := \max_{\varphi\in\mathcal{C}} \|\nabla^2 \mathcal{J}(\varphi)\|, \tag{3.62}$$

$$\|f_k\| \leq \kappa_f := \max_{\varphi\in\mathcal{C}} \|\nabla \mathcal{J}(\varphi)\|, \tag{3.63}$$

$$\kappa_g := \max_{\varphi\in\mathcal{C}} \|\nabla^2 g(\varphi)\|. \tag{3.64}$$

Moreover, from the diagonal structure of the approximate linearised constraints we can deduce the following.

Lemma 3.4.9 (Existence + Boundedness of the normal step n_k)**.** *The normal step n_k always exists and it holds*

$$\|n_k\| \leq \overline{\kappa}_n \vartheta_k, \quad \underline{\kappa}_n \vartheta_k \leq \|n_k\|, \tag{3.65}$$

for constants $\underline{\kappa}_n, \overline{\kappa}_n > 0$ independent of k.

Proof. The existence of the normal step is trivial

$$n_{k,i} = \begin{cases} -\tilde{g}_{k,i} & \text{if } g_i(\varphi_k) < 0, \\ 0 & \text{else.} \end{cases}$$

Now assume that $n_k \neq 0$ and let i be the index of a maximal component, i.e. $n_{k,i} = \|n_k\|_\infty$. Then, from Assumption F3 it follows

$$0 \leq n_{k,i} = -\tilde{g}_{k,i} = -(\widetilde{G}_k T_k^T)^{-1}_{ii} g_i(\varphi_k) \leq \kappa_T \vartheta_k.$$

We conclude

$$\|n_k\| \leq \sqrt{N}\|n_k\|_\infty \leq \sqrt{N}\kappa_T \vartheta_k := \overline{\kappa}_n \vartheta_k.$$

The lower bound can be established as follows. Let j denote the index corresponding to the maximal constraint violation, i.e $\vartheta_k = -g_j(\varphi_k)$. Then, by definition of the normal step (3.46) and the continuity of $\widetilde{G}$ and T we obtain

$$\vartheta_k = \left(\widetilde{G}_k T_k^T n_k\right)_j \leq N\|\widetilde{G}_k T_k^T\|_\infty \|n_k\|_\infty \leq N\sqrt{N} \max_{\varphi\in\mathcal{C}} \|\tilde{G}(\varphi)T(\varphi)^T\| \|n_k\| := \underline{\kappa}_n^{-1}\|n_k\|.$$

□

The old and the new infeasibility can be bounded by the trust-region radius.

Lemma 3.4.10. *Assume that TRQP(φ_k, Δ_k) is compatible and let $u_k = n_k + t_k$ be the solution of the normal/tangential sub-problems. Then, there exists a constant $\kappa_{\text{ui}} > 0$ such that*

$$\begin{aligned}\vartheta_k &\leq \kappa_{\text{ui}} \Delta_k^{1+\mu},\\ \vartheta(\varphi_k + u_k) &\leq \kappa_{\text{ui}} \Delta_k^2.\end{aligned}$$

Proof. From the compatibility condition (3.48) and Lemma 3.4.9 we obtain

$$\underline{\kappa}_n \vartheta_k \leq \|n_k\| \leq \kappa_\Delta \Delta_k \min\{1, \kappa_\mu \Delta_k^\mu\} \leq \kappa_\Delta \kappa_\mu \Delta_k^{1+\mu}.$$

To derive a bound for the new infeasibility we have to distinguish between the case where the exact and where the inexact Jacobians are used. We will only consider the inexact case, the exact one follows similarly. Performing a first order Taylor expansion of the i'th constraint at φ_k yields

$$-g_i(\varphi_k + u_k) = -g_i(\varphi_k) - G_i(\varphi_k)u_k - \frac{1}{2} u_k^T \nabla^2 g_i(\xi_k) u_k,$$

where $\nabla^2 g(\xi_k)$ denotes the remainder term. Adding the zero $\widetilde{G}_{k,i} - \widetilde{G}_{k,i}$ and using the approximation error bound (3.58) we can deduce

$$\begin{aligned}-g_i(\varphi_k + u_k) &= -g_i(\varphi_k) - \widetilde{G}_i(\varphi_k)u_k + \left(\widetilde{G}_i(\varphi_k) - G_i(\varphi_k)\right) u_k - \frac{1}{2} u_k^T \nabla^2 g_i(\xi_k) u_k\\ &\leq 0 + \kappa_{\text{ac}} \Delta_k^2 + \kappa_N \max_{1 \leq i \leq m} \max_{\varphi \in \mathcal{C}} \|\nabla^2 g_i(\varphi)\| \|u_k\|^2\\ &\leq \kappa_{\text{ac}} \kappa_N \kappa_g \Delta_k^2.\end{aligned}$$

The first term vanishes because u_k is a solution of the local problem TRQP(φ_k, Δ_k), the second term can be bounded because of the condition on to the approximation error (3.58). The result follows with

$$\kappa_{\text{ui}} := \max\{\kappa_{\text{ac}} \kappa_N \kappa_g, \kappa_\Delta \kappa_\mu \underline{\kappa}_n^{-1}\}.$$

□

With Lemmata 3.4.9 and 3.4.10 at hand the remainder of the proof is done along the lines of the exact filter convergence proof, see [47]. For completeness we will only prove the main convergence theorem. Again, the interested reader can find the detailed steps needed on the way towards this main result with a short description of the proofs in the appendix B.

Theorem 3.4.11. *Let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR. Then either the feasibility restoration phase terminates unsuccessfully by converging to a critical point of* (3.60) *or there exists a subsequence $(\varphi_{k_l})_{l\in\mathbb{N}} \subseteq (\varphi_k)_{k\in\mathbb{N}}$ such that*

$$\lim_{l\longrightarrow\infty} \varphi_{k_l} = \varphi_*,$$

where φ_* is a first-order critical point of the non-linear problem (P).

Proof. Assume that the feasibility restoration always terminates successfully. Then by Lemmata B.6, B.7 and B.9 there exists a subsequence $(\varphi_{k_l})_{l\in\mathbb{N}}$ with

$$\lim_{l\longrightarrow\infty} \vartheta_{k_l} = \lim_{l\longrightarrow\infty} \widetilde{\chi}_{k_l} = 0.$$

For k large enough we get $\tilde{\chi}_{k_l} \leq \varepsilon$ and thus eventually only the exact local problems are solved. We conclude that it also holds

$$\lim_{l\longrightarrow\infty} \chi_{k_l} = 0. \tag{3.66}$$

By the compactness F1 we can now find a convergent subsubsequence which we denote again by $(\varphi_{k_l})_{l\in\mathbb{N}}$

$$\lim_{l\longrightarrow\infty} \varphi_{k_l} = \varphi_*.$$

From the asymptotic feasibility and Lemma 3.4.9 it follows

$$\lim_{l\longrightarrow\infty} \|n_{k_l}\| \leq \overline{\kappa}_n \lim_{l\longrightarrow\infty} \vartheta_{k_l} = 0.$$

By perturbation theory [43, Theorem 2.2.6] and (3.66) we can deduce that at φ_* the local criticality measure (3.51) at zero vanishes

$$\chi(0) = |\min_{\substack{G(\varphi_*)d+g(\varphi_*)\geq 0\\ \|d\|=1}} \langle \nabla m_{\varphi_*}(0), d\rangle| = 0. \tag{3.67}$$

From $\nabla m_{\varphi_*}(0) = \nabla\mathcal{J}(\varphi_*)$ it follows that φ_* is also a critical point of the non-linear problem (P), cf. Remark 3.4.4. □

After developing this framework for inexact constraint Jacobians and completing this part of the thesis, we got to know that recently Andrea Walther and Lorenz Biegler also proposed a filter method with inexact constraint Jacobians [134]. For this method they also proved global convergence along the lines of the exact filter convergence proof [47]. Their motivation is from optimisation problems where the assembly of the constraint Jacobians is very costly. In their approach it is assumed that the approximate Jacobians can be refined, e.g. by using quasi-Newton updates, whenever the local model is not a good approximation. As the global convergence proof is also based on [47], the model in the algorithm is refined whenever a similar condition to our approximation error bound (3.58) is violated or when the iterate is close to a locally critical point. In contrast to our framework the approximate Jacobians are not assumed to induce a diagonal structure. Consequently additional assumptions on the existence and boundedness of the normal step have to be made.

4. Application to large deformation contact problems

In this section we will apply the solver presented in Chapter 3 to the discretised large deformation contact problem (2.35). Up to now the prevailing solution strategies for large deformation contact problems are based on active-set methods. In these methods the non-penetration inequality constraints (2.25) are eliminated iteratively by estimating which constraints are active. The resulting equality constrained problems are then set-up by assembling the portion of the contact forces (2.29) that corresponds to the estimated active constraints. These are then incorporated by either a penalty approach [7, 89, 63, 140] or the Lagrange multiplier method [93, 44]. Recent approaches apply the primal-dual active set strategy, first introduced by Hübner and Wohlmuth [68], where the discrete contact forces are eliminated using dual mortar elements [62, 106]. The discrete contact forces can then be recomputed in a post-processing step, similar to our approach. In [106] the contact detection and solution of the non-linear weak equations (1.38) is further combined within one loop. Both approaches make simplifying assumptions on the contact forces, which in turn results in a loss of angular momentum conservation.
In [66] the contact forces are eliminated by applying a discrete null space method. While they retain angular momentum preservation, the method is computationally very costly because generalised saddle-point problems have to be solved. In our case the fully discrete problems (2.27) can be solved by computing local optimal points of a non-linear constrained minimisation problem (2.35). We can thus directly apply techniques from constrained optimisation without having to eliminate the inequalities first.
The chapter is organised as follows: First in Section 4.1 we specify a class of inexact constraint Jacobians and show that they fulfil all requirements for the filter method. Then, in Section 4.2 we present an extension of the monotone multigrid method MMG to contact problems, that is used for the fast solution of the inexact local filter problems. Afterwards, we shortly comment on implementational aspects. We close this chapter by showing some numerical results that illustrate the convergence theory of Chapter 3 and a numerical study of the proposed time discretisation scheme in Section 4.4.

4.1. Inexact constraint Jacobians

In what follows for simplicity we will omit all indices corresponding to the current time step. Upper indices $(\cdot)^k$ will denote the iteration index within the filter–trust–region Algorithm FTR. Recall that the algebraic problem we consider is to minimise the smooth energy functional $\mathcal{J} : \mathbb{R}^N \longrightarrow \mathbb{R}$, cf. (2.35). The problem is supplemented by the non-linear mortar discretised non-penetration constraint $g_h : \mathbb{R}^N \longrightarrow \mathbb{R}^m$, with components given by

$$g_{h,p}(\boldsymbol{\varphi}_h) := \int_{\gamma^1_{C,h}} -\mathbf{n}_{h,p}\theta_p \cdot (s - \Phi(s))\, ds \quad 1 \leq p \leq m, \tag{4.1}$$

cf. (2.25). We now apply the inexact filter method FTR to solve the algebraic problem (2.35) subject to (4.1). This method relies on the successive linearisation of the non-linear energy and the non-penetration constraint. The linearisation of the constraint can be split into the following three parts:

$$\begin{aligned} \delta g_{h,p}(\boldsymbol{\varphi}_h) = &\int_{\gamma^1_{C,h}} -\delta\mathbf{n}_{h,p}\theta_p \cdot (s - \Phi(s))\, ds + \int_{\gamma^1_{C,h}} -\mathbf{n}_{h,p}\theta_p \cdot \delta(s - \Phi(s))\, ds \\ &+ \int_{\gamma^1_{C,h}} -\mathbf{n}_{h,p}\theta_p \cdot (s - \Phi(s))\, \delta ds. \end{aligned} \tag{4.2}$$

The first part involves the linearisation of the nodally averaged normal. In the continuous case the nodally averaged normals correspond to the continuous normal field $\mathbf{n}_{h,p} \mathrel{\widehat{=}} \mathbf{n}^1(s)$. As Φ is the normal projection $s - \Phi(s) = \mu(s)\mathbf{n}^1(s)$ and $\mathbf{n}^1(s)$ are collinear. Hence the first term vanishes as it holds

$$\begin{aligned} &\mathbf{n}^1(s) \cdot \mathbf{n}^1(s) = 1, \\ \Longrightarrow\ &\delta\mathbf{n}^1(s) \cdot \mathbf{n}^1(s) = 0, \\ \Longrightarrow\ &\delta\mathbf{n}^1(s) \cdot \mu(s)\mathbf{n}^1(s) = 0. \end{aligned}$$

The second term is the linearisation of the point-wise gap function. This coincides with the non-penetration constraint in the small displacement contact framework, cf. (1.28). The third summand labels the linearisation of the deformation dependent integral domain, which we denote by δds.

Motivated by the availability of fast multigrid methods that have been developed for the small strain framework [79, 55], we will use the second term only to approximate the constraint Jacobians

$$\nabla g_{h,p}(\boldsymbol{\varphi}_h)\mathbf{v} \approx \widetilde{G}_p(\varphi)\mathbf{v} = \int_{\gamma^1_{C,h}} -\mathbf{n}_{h,p}\theta_p \cdot (\mathbf{v}^1(s) - \mathbf{v}^2 \circ \Phi(s))\, ds, \tag{4.3}$$

cf. (1.28) and Section 3.4. Note that this approximation does not completely coincide with the second term of (4.2). We additionally omit the linearisation of the contact mapping Φ.

Remark 4.1.1. *Using this approximation of the constraint Jacobians is similar to the simplifying assumptions on the contact forces that are made in [93, 106, 62]. While in other approaches this allows the condensation of the dual Lagrange multiplier, we exploit the special structure of the Jacobians to speed-up our algebraic solver using multigrid methods. Furthermore, whenever it is necessary, we switch back to the contact forces that are consistent with the mortar discretisation.*

In the following we will investigate whether the requirements introduced in Section 3.4 for the inexact Jacobians are fulfilled. The algebraic form of the approximate Jacobians applied to a function from the product space $\mathbf{v} \in \mathbf{S}_h$ is given by

$$\widetilde{G}(\boldsymbol{\varphi}_h)\mathbf{v} = -OD\mathbf{v}^1 + OM\mathbf{v}^2, \tag{4.4}$$

where the deformation-dependent non-mortar and mortar block-matrices, with diagonal $d \times d$ blocks, are given by

$$D_{pq} = \int_{\gamma^1_{C,h}} \theta_p \lambda^1_q \, ds, \tag{4.5}$$

$$M_{pq} = \int_{\gamma^1_{C,h}} \theta_p \lambda^2_q \, ds, \tag{4.6}$$

and the block-diagonal matrix $O \in \mathbb{R}^{m \times dm}$

$$O_{pp} = \mathbf{n}^T_{h,p}.$$

Note that for the assembling of the mortar matrix M, the basis functions λ^2_q have to be evaluated at the projected points $\Phi(s)$. Due to the bi-orthogonality (2.19) of the dual basis functions θ_p the non-mortar matrix is diagonal

$$D_{pp} = \int_{\gamma^1_{C,h}} \lambda^1_p \, ds. \tag{4.7}$$

The construction of a continuous transformation $T(\boldsymbol{\varphi}_h) \in R^{N \times N}$ such that $\widetilde{G}(\boldsymbol{\varphi}_h)T(\boldsymbol{\varphi}_h)^T$ is positive and diagonal, cf. Assumption J, can be done as proposed by Krause and Wohlmuth in [79]. To this end, we sort the deformation coefficient vector $\boldsymbol{\varphi}_h = (\boldsymbol{\varphi}^M_h, \boldsymbol{\varphi}^{NM}_h, \boldsymbol{\varphi}^R_h)$ into mortar $\boldsymbol{\varphi}^M_h$, non-mortar $\boldsymbol{\varphi}^{NM}_h$ and the remaining degrees of freedom $\boldsymbol{\varphi}^R_h$.
Further, as suggested by Kornhuber and Krause for Signorini's problem [75], we consider the block-diagonal matrix $\hat{O} \in \mathbb{R}^{m \times m}$, consisting of the householder reflections of the first Euclidean basis vector onto the nodally averaged normal $\mathbf{n}_{h,p}$. For the re-ordered coefficients the basis transformation takes the form

$$T(\boldsymbol{\varphi}_h) := \begin{pmatrix} \mathrm{Id} & \hat{M}^T & 0 \\ 0 & \hat{O} & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix}, \tag{4.8}$$

where $\hat{M} := D^{-1}M$. Note that the householder reflections are symmetric and invertible, hence $\hat{O}^{-1} = \hat{O}^T = \hat{O}$.

The transformed inexact Jacobian can now be computed as follows

$$\widetilde{G}(\boldsymbol{\varphi}_h)T^T(\boldsymbol{\varphi}_h) = \begin{pmatrix} OM & -OD & 0 \end{pmatrix} \begin{pmatrix} \mathrm{Id} & 0 & 0 \\ \hat{M} & \hat{O} & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix} = \begin{pmatrix} 0 & D & 0 \end{pmatrix}, \tag{4.9}$$

which are positive and diagonal.

Remark 4.1.2. *The resulting constraints only restrict the nodes of the non-mortar side, or more precisely, only the first component of each node. The corresponding degrees of freedom of the mortar contact boundary are constructed such that they always fulfil the linearised constraints, see [79, p.10].*

Moreover, the transformation is invertible with

$$T(\boldsymbol{\varphi}_h)^{-1} = \begin{pmatrix} \mathrm{Id} & -\hat{M}^T\hat{O} & 0 \\ 0 & \hat{O} & 0 \\ 0 & 0 & \mathrm{Id} \end{pmatrix}, \tag{4.10}$$

and the inverse inexact Jacobians

$$\left(\widetilde{G}(\boldsymbol{\varphi}_h)T(\boldsymbol{\varphi}_h)^T\right)^{-1} = D, \tag{4.11}$$

are bounded away from zero whenever the faces of the deformed contact boundary $\gamma^1_{C,h}$ are not degenerated. Hence, the inexact Jacobians $\widetilde{G}$ fulfil Assumption J, which is required for the global convergence of the filter method.

Remark 4.1.3. *In comparison to other recent approaches that use the dual mortar discretisation [62, 106], our inexact filter–trust–region method can also be seen as a Newton method using approximate contact forces corresponding to* (4.3)*. The difference is that, whenever it is necessary, we switch back to the fully consistent contact forces* (2.29)*.*

4.2. Multigrid methods for contact problems

In this section we describe how the monotone multigrid method MMG can be extended to the local SQP problem with inexact Jacobians (4.4) and supplemented by trust–region constraints

$$\|\mathbf{u}_h\|_\infty \leq \Delta_k. \tag{4.12}$$

The monotone multigrid method described in Section 3.2 can only handle bound constraints. An extension of this method to small deformation contact problems has been introduced in [79] by constructing a hierarchical multilevel basis, based on the transformations (4.8). While this method is very reliable and robust, it also requires the assembling of the transformation for each grid level in the hierarchy and a monotone restriction operator to obtain suitable coarse grid defect obstacles [79, p.15]. By incorporating the additional trust–region constraints and restricting them with the Mandel restriction (3.1.1), this method can be extended to the linearised filter problems and global convergence can be proven in the same way as in Section 3.2. First ideas on this method have been developed by Sander in [114].

Recently, Gräser introduced a novel multigrid method for convex quadratic problems with obstacles and more complicated non-linearities that avoids the need of constructing coarse defect obstacles [55]. The idea of the truncated non-smooth Newton multigrid method (TNNMG) is to make a non-linear pre-smoothing step only on the finest level. Then a coarse grid correction is computed by solving the *unconstrained* truncated defect problem using a linear multigrid step. To regain feasibility the resulting correction is projected back onto the feasible set $\mathcal{K}_h$. Conclusively a line search along the projected correction is performed which guarantees sufficient decrease of energy. This method is proven to be globally convergent [53].
Furthermore, as only the non-linear pre-smoothing step on the finest level has to be implemented, this algorithm is relatively easy to implement. In [117] this method is extended to small displacement contact problems using the transformation (4.8) to decouple the constraints. On the finest grid level a projected Gauss–Seidel step in the transformed coordinates is performed. The smoothed iterate is then transformed back to Euclidean coordinates and the linear multigrid method is used to compute a coarse grid correction.
Compared to the monotone multigrid method this approach has the advantage that only the canonical transfer operators are needed on the coarser levels and hence, in the case of contact problems, only the mortar matrix M_j on the finest level has to be computed. We extend this idea to large deformation contact problem with inexact constraint Jacobians (4.9) as follows. We only sketch the method here, a more detailed description can be found in [117, Section 3.4].

Algorithm TNMMG (Truncated Nonsmooth Newton Monotone Multigrid)

1. Gauss−Seidel step with inexact constraints

$$\widetilde{G}_k T_k^T \mathbf{v}_h \geq -g(\boldsymbol{\varphi}_{h,k}), \quad \|\mathbf{v}_h\|_\infty \leq \Delta_k \quad (1)$$

2. MMG step in Euclidean coordinates with trust−region constraints only

3. Projection of the correction onto the feasible set given by (1)

4. Line search in the direction of the projected correction

Proposition 4.2.1. *Under the assumptions of Theorem 3.2.9, the method TNMMG converges globally to a locally critical point.*

Proof. Follows from the monotonicity of the coarse grid correction and the global convergence of the preceding Gauss–Seidel step, see Theorem 3.2.9. □

4.3. Implementational Aspects

The normal projection Φ

For the assembling of the non-penetration constraint and the (in-)exact Jacobian the normal projection has to be implemented efficiently. Therein, each pair of faces $(\Delta^{NM}, \Delta^{M})$ needs to be checked for intersections, where $\Delta^{NM} \in \gamma_{C,h}^{1}$ corresponds to a non-mortar face and $\Delta^{M} \in \gamma_{C,h}^{2}$ to a mortar face respectively. To avoid the quadratic effort of checking all pairs of faces, we use the optimal advancing front algorithm by Gander and Japhet [49]. In this scheme, after an initial intersecting pair is found, only neighbouring information is used which leads to a linear complexity. The computation of the intersection polygon for two faces is done similarly to the method proposed in Puso and Laursen [93]:

Let $\mathbf{p}_i^{NM}$ and $\mathbf{p}_i^{M}$ denote the corners of Δ^{NM} and Δ^{M} respectively and let $\mathbf{n}_i^{NM}$ denote the nodally averaged normals (2.24) associated with the non-mortar vertices $\mathbf{p}_i^{NM}$. Then, the intersections are computed as follows:

1. Project each $\mathbf{p}_i^{NM}$ along $\mathbf{n}_i^{NM}$ onto the Δ^{M}-plane
2. Check which $\mathbf{p}_i^{NM}$ are contained in Δ^{M}.
3. Check which $\mathbf{p}_i^{M}$ are contained in the projected $\widetilde{\Delta}^{NM}$
4. Compute edge intersections and determine which neighbours intersect
5. Compute nodally averaged centre and triangulate the polygon

Exact linearisation of the mortar non-penetration constraint

The consistent linearisation of the discrete non-penetration constraint (4.2) is very complicated because the integration must be performed over the triangulated intersection

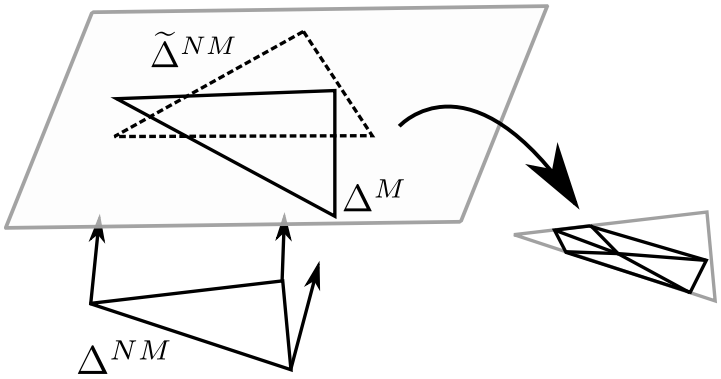


Figure 4.1.: *Computation and triangulation of an intersection polygon.*

polygons Figure 4.1. Hence, the deformation-dependent polygonal corners have to be linearised which has to be done according to the type of intersection they arise from, i.e. non-mortar/mortar corner, edge intersection or averaged centre. The corresponding formulas are very lengthy and can be adapted from the similar discretisations in [106, 93]. To validate our implementation we tested the consistent linearisation against a finite difference approximation for several grids.

4.4. Numerical Examples

We will now show some numerical examples to illustrate the performance of the multigrid method MMG and filter–trust–region method FTR presented within this thesis. Further we will investigate the behaviour of the proposed time discretisation scheme in more detail. The implementation of all the solvers and the mortar discretisation has been done within the DUNE environment which is a free C++ toolbox for the solution of partial differential equations [12, 10, 11].

The implementation of the contact mapping $\Phi : \gamma_{C,h}^1 \longrightarrow \gamma_{C,h}^2$, cf. Section 4.3, can be found within the DUNE module DUNE-GRID-GLUE [9] and originated from the PSURFACE library by Oliver Sander [118], which is also used to generate and handle the parametrised boundaries in the hip joint range of motion analysis (Chapter 5). As a grid manager we used UG [13]. For the solution of the coarsest problems within the multigrid step and for the solution of the exact local problems, cf. Section 4.3, we apply the COINIPOPT library which uses an optimised interior point algorithm [133].

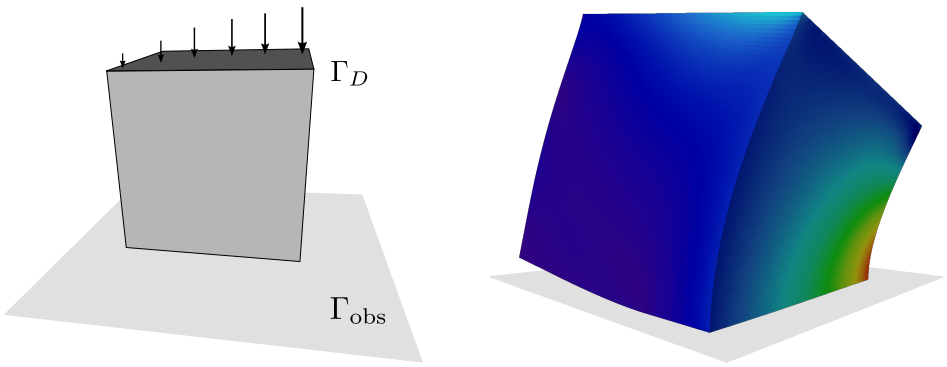


Figure 4.2.: *Left: Reference domain and boundary conditions. Right: Deformed configuration with the resulting von-Mises stresses.*

4.4.1. Obstacle problems in non-linear elasticity

In this section we want to illustrate the performance of the monotone multigrid method introduced in Section 3.2. There is a large variety of optimisation algorithms for non-linear minimisation problems subject to simple bound constraints [29, 28, 129, 56, 58]. In [28] Coleman and Li suggest a Newton approach with a non-standard *reflective linesearch* to keep the iterates feasible. Ulbrich et al. extended this approach to infinite-dimensional problems in terms of interior point methods [129]. Motivated by the convergence speed of multilevel methods Gratton et al. developed a recursive trust-region algorithm where on each grid level a local trust-region method is applied [56]. This method was refined by Krause and Gross in [58] such that the resulting method can be implemented as a V-cycle. In our approach we will go the other way round: Instead of first decomposing the system hierarchically and performing a trust–region step on each level, we apply the standard trust—region scheme TR directly to the fine level and solve the local problems with the monotone multigrid extension introduced in Section 3.2.
The following numerical example is adopted from [58]. We consider the non-linear Ogden material [103]

$$W(\varphi) = a\ \mathrm{tr}\, E(\boldsymbol{\varphi}) + b\ (\mathrm{tr}\, E(\boldsymbol{\varphi}))^2 + c\ \mathrm{tr}(E(\boldsymbol{\varphi})^2) + d\,\Gamma(\det(\nabla\boldsymbol{\varphi})), \tag{4.13}$$

where E denotes the Green strain tensor (1.14) and Γ is a penalty function

$$\Gamma(x) = -\ln\, x,$$

that "enforces" local injectivity. The constants are chosen such that

$$a = -d\,\Gamma'(1), \quad b = \frac{1}{2}(\lambda - \Gamma'(1) + \Gamma''(1)), \quad c = \mu + d\,\Gamma'(1), \quad d > 0, \tag{4.14}$$

where λ and μ denote the bulk and shear modulus. This material approximates the St.Venant–Kirchhoff material (1.15) for small deformations and reproduces it for $d = 0$.

We only consider the quasi-static problem (1.19). The initial geometry is the unit cube $\Omega = [0,1]^3$ with Dirichlet boundary $\Gamma_D = \{(x_1, x_2, 1) \in \mathbb{R}^3\}$ and skewed Dirichlet values

$$g(x) = -0.2\,(x_1 + x_2), \quad x \in \Gamma_D.$$

The cube is subject to lower bound constraints given by the affine x_1-x_2-plane through $(0, 0, -0.05)$, see Figure 4.2. We use the following material parameters: $d = 100$, $\lambda = 34$, $\mu = 136$. The local problems are solved using the monotone multigrid method MMG until the relative correction error falls beyond the tolerance $\varepsilon_{\text{mg}} = 10^{-10}$. The trust-region algorithm is performed until the first order criticality measure χ, see (A.9), falls below 10^{-9}. In the left of Figure 4.3 the reduction of the criticality measure is plotted against the

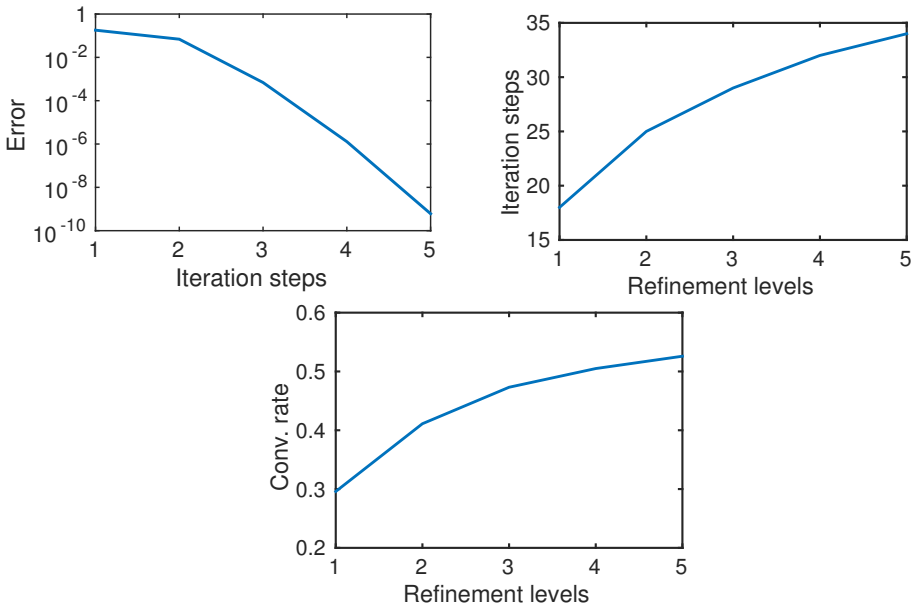


Figure 4.3.: *Left: Reduction of the criticality measure of the trust–region method in the case of 692,835 degrees of freedom. Middle and Right: Averaged convergence rates and number of iteration steps of the multigrid method for increasing number of degrees of freedom. The flattening of the curve indicates mesh-independence or midly dependent convergence.*

trust-region iterations. Already after five iterations the measure falls beyond 10^{-9} which corresponds to a convergence speed of the same order as the recursive multilevel approach in [58]. In the middle and right of Figure 4.3 we compared the multigrid iterations and convergence rates averaged over the trust-region steps for five uniform refinement steps. While for small levels the numbers still increase, they flatten for increasing degrees of freedom which indicates mesh-independent or mildly dependent convergence.

4.4.2. Convergence Analysis of the filter–trust–region

In this section we want to illustrate the convergence of the proposed filter–trust–region method at the example of the ironing problem. This problem is used to test the robustness of the mortar discretisation and the applied algebraic solver due to its difficulty [93, 106]. We adopt the initial configuration and loading set-up from [106], see Figure 4.4. In this example a block is placed under a half-spherical shell. The block is fixed at

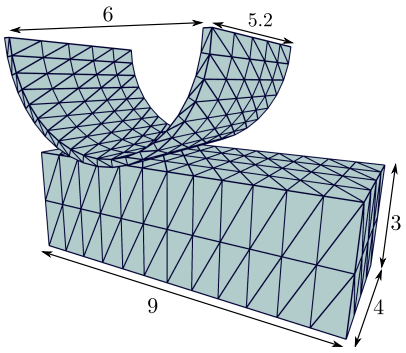


Figure 4.4.: *The initial configuration of the coarse ironing grids.*

the bottom with homogeneous Dirichlet conditions. For the shell non-homogeneous Dirichlet conditions are prescribed on the top boundary that are enforced in loading steps: First, the shell is pressed vertically into the block with a prescribed total displacement of 1.4 units. Then, in the second phase it is swiped over the block until a horizontal displacement of 2.1 is reached, see Figure 4.5. As done in [106] we apply an equidistant loading increment of 0.14. Both bodies are modelled by the Neo–Hookean material law

$$W(\varphi) = \frac{\lambda}{4}(\det(\nabla\boldsymbol{\varphi})^2 - 1) - (\frac{\lambda}{2} + \mu)\log(\det(\nabla\boldsymbol{\varphi})) + \mu \operatorname{tr} E(\boldsymbol{\varphi}), \tag{4.15}$$

with

$$\lambda_{\text{shell}} = 750, \quad \mu_{\text{shell}} = 375,$$
$$\lambda_{\text{block}} = \frac{3}{4}, \quad \mu_{\text{block}} = \frac{3}{8}.$$

In this example the stiffness ratio of the two bodies is very high with $1:1000$. In the case of a softer contact $(1:100)$ similar convergence results were obtained. The loading problems are solved with the inexact filter–trust–region method until the criticality measure χ, see (3.51), falls below 10^{-9} or the relative error in the H^1-norm is less than 10^{-15}. The linearised SQP problems with exact constraints are solved using the optimised IpOpt-Library [133], until a tolerance of 10^{-8} is reached. For the inexact problems we apply the truncated non-smooth Newton multigrid method TNMMG until the relative error in the H^1-norm falls below a tolerance of 10^{-8}. The constant in the control of the approximation error (3.58) is chosen $\kappa_{\text{ac}} = 1$. We switch from inexact to exact local problems when the approximate criticality $\widetilde{\chi}$, cf. (3.50), falls below 10^{-3}.

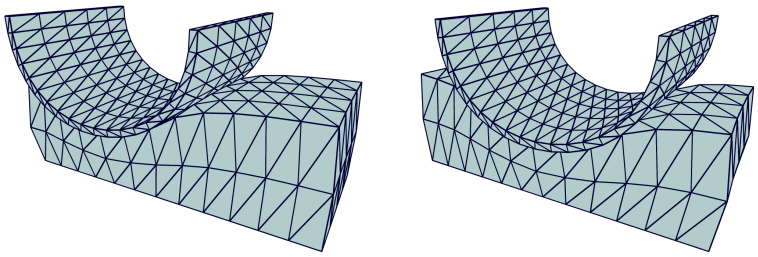


Figure 4.5.: *Left:Ironing problem after the horizontal displacement. Right: Ironing problem after the vertical displacement*

On the left side of Figure 4.6 the convergence of the inexact filter method and corresponding filter with exact Jacobians are shown for an exemplary step during the swiping phase at the increment 0.28. For this computation the block grid has been refined uniformly three times and the shell grid once. As can be seen in the picture, the decrease of the exact criticality is flattening when only the "cheap" inexact problems are solved. This is why it is important for the performance of the inexact filter method to choose the switching point, i.e. the tolerance for the inexact criticality measure, carefully or adjust it dynamically.
Already in this rather small problem with $26,550$ degrees of freedom, the computation time for the solution of the inexact local problems with the multigrid method is about ten times smaller than the time needed by the optimised solver for the exact problem ($1.5s$ resp. $15s$). In this exemplary loading step 12 inexact iterations were performed before switching to exact constraint Jacobians, which corresponds to a decrease of computation time of 15%. In total the computation time needed for this example with a three times refined block and once refined shell was reduced about 14%, see Figure 4.6 and Table 4.1
Further refining the grids ($\approx 200,000$ degrees of freedom) already leads to a difference in

	exact	inexact	difference
comp. time	18,495s	16,027s	2,648s

Table 4.1.: The total averaged computation times for the exact and inexact filter. Using inexact Jacobians results in a reduction of 14%.

computation time of $20s$ to 20 minutes per local solution.

Only linear convergence can be expected due to the approximate Hessians that are used within the filter method, cf. Remark 3.4.1. Extending this approach to second order consistent models will be done in future work.
In the right of Figure 4.7 the total inexact filter iterations are plotted for different refinement levels, where the grids are refined alternatingly. During the first phase the iteration numbers seem to remain constant for decreasing mesh size; in the second phase

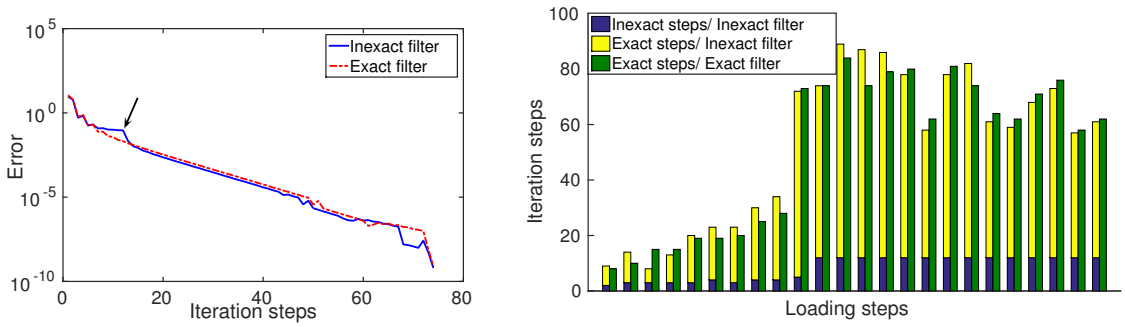


Figure 4.6.: *Left: Reduction of the criticality measure χ of the exact and inexact filter method during the 12^{th} loading step. The arrow marks the spot where the approximate criticality measure reached its tolerance. The flattening in that area shows that exact steps are necessary for the convergence. Further, the switch between exact and inexact steps has to be chosen appropriately to guarantee a high reduction of computation time. Right: Comparsion of the inexact and exact steps that were computed during all loading steps. The total iteration numbers of the inexact and exact filter are comparable. As a result, the use of cheap inexact steps results in a decrease computation time.*

this independence is lost, although the increase is still in an acceptable range.

In Figure 4.7 we compare the solutions generated by our filter method to the results that are computed by only solving the inexact local problems until $\widetilde{\chi} < 10^{-10}$. We observe that, although this error is small, convergence to the true minimisers cannot be expected in general. Finally in Figure 4.8 the L^2-norm of the Lagrange multipliers (2.32) is plotted. As expected from the mortar discretisation the contact forces evolve smoothly even in the presence of large sliding.

4.4.3. Numerical Study of modified midpoint rule

In the following example we examine the modified midpoint rule (2.6) in more detail. To this end, we consider an elastic ball bouncing on an elastic block. A coarsened representation of the initial configuration is shown in Figure 4.9. For the dynamic simulations the ball is additionally refined twice and the block once resulting in 49518 degrees of freedom. Again the bodies are modelled by the Neo–Hookean material law (4.15) with identical (soft) elastic parameters

$$\lambda = 22, \quad \mu = 11, \tag{4.16}$$

and mass density $\rho = 0.01$. The ball is supplemented with an initial velocity of $(0, 0.2, 1.5)$

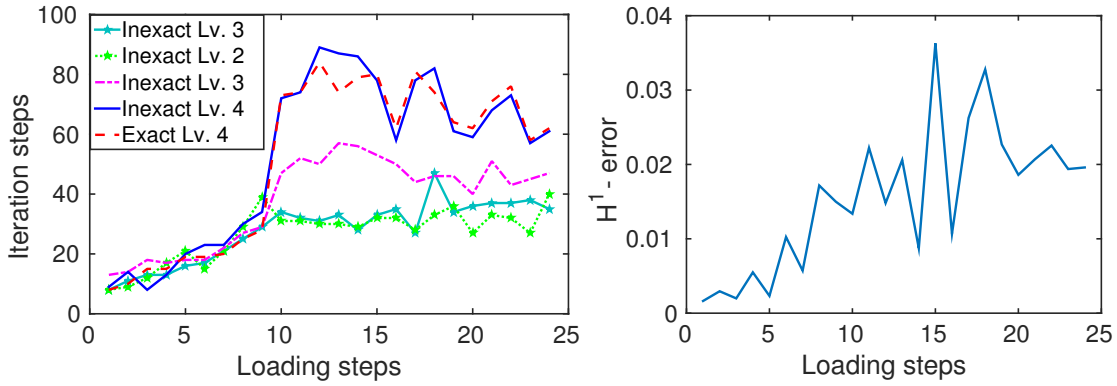


Figure 4.7.: *Left: A comparsion of the filter iteration numbers for different refinement levels and the exact filter method on the finest grid. During the swiping phase (steps* $10-24$*) a mild mesh-dependence can be observed. Right: The* H^1-*error of the true minimizers to approximate ones that are computed from the sole solution of inexact problems. While the error seems to be constant and small, convergence of a purely inexact solver cannot be expected.*

units and additional gravity forces

$$\mathbf{F}^{\text{ext}}(t, \mathbf{v}_h) = \int_{\Omega_{\text{ball}}} \rho \, \mathbf{f} \, \mathbf{v}_h(x) \, dx, \quad \mathbf{f} = (0, 0, 9.81)^T, \tag{4.17}$$

are prescribed. The block is set-up with homogeneous Dirichlet conditions prescribed at its lower boundary and a rather coarse time step size $\tau = 0.01$ is chosen. The corresponding spatial problems are solved with the inexact filter–trust–region method until the tolerance of 10^{-10} is reached. We switch from inexact to exact local problems whenever $\widetilde{\chi} \leq 10^{-5}$.
In dynamic simulations like the one at hand or the range of motion analysis that is presented in the next chapter 5 the inexact filter method shows its full potential. During time steps where no active contact occurs, the solution of inexact local problems suffices to achieve full convergence towards the solution, see Figure 4.10. In the present simulation this results in a decrease of total computation time by 30%, see Table 4.2. In Figure 4.11

	exact	inexact	difference
comp. time	212,500s	151,131s	61,369s

Table 4.2.: The total averaged computation times for the exact and inexact filter. Using inexact Jacobians results in a reduction of 30%.

the energy evolution of the midpoint rule (2.3) and our proposed modified version (2.6) are compared. While the midpoint rule is strictly generating energy during each contact phase, the modified version is almost dissipative and the energy loss is rather small

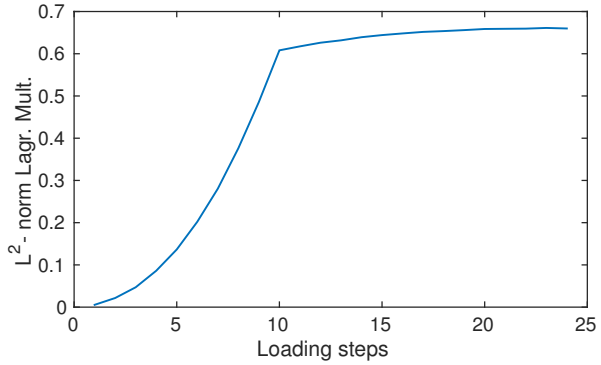


Figure 4.8.: *L^2-norm of the dual Lagrange multipliers. As we expected from the mortar discretisation the contact forces evolve smoothly during the sliding.*

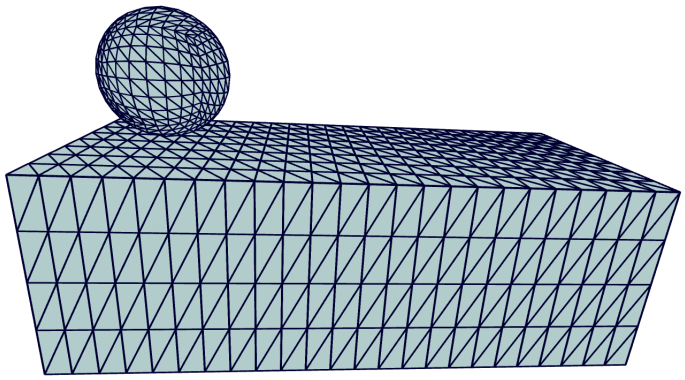


Figure 4.9.: *Initial configuration of the bouncing problem.*

compared to the increase of the unmodified scheme. In the right of Figure 4.11 the energy of the stabilised scheme is shown in more detail. One can see that the energy is lost during the attachment of the contacting bodies. In [34, 72] this phenomena was also observed in the framework of small deformations for a modified Newmark method (which coincides with the proposed scheme in the linear case). The authors suggest to add a further contact stabilisation by making a predictor-L^2-projection step. The resulting predicted contact forces enter the velocity update which improves the energetic behaviour of the scheme. This could also help to avoid the small gain in energy observed during the detachment phase.

Remark 4.4.1. *The modifications proposed in [34] and [72] also eliminate artificial oscillations at the contact boundary in the small displacement framework. In the presented numerical examples we did not observe these kind of oscillations which is possibly due to sliding and rotation. We will investigate this issue in more detail and examine the incorporation of further stabilisation techniques like in [34, 72] in future work.*

Furthermore, while improving the stability of the time discretisation significantly, the

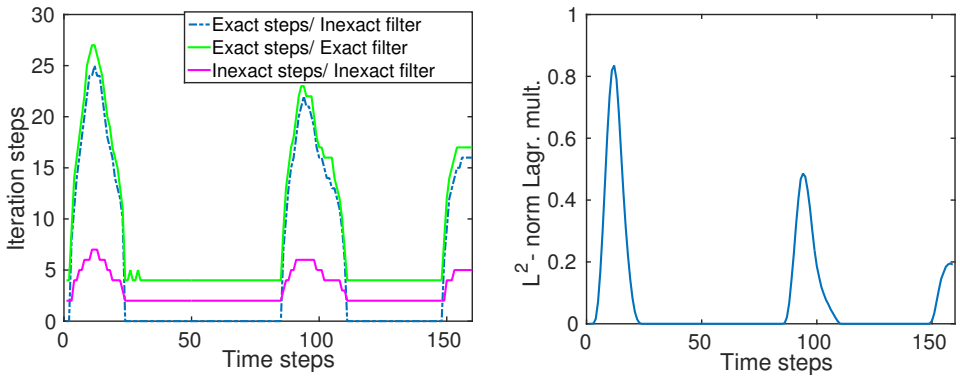


Figure 4.10.: *Left: The total number of iterations needed by the inexact and exact filter method. During the contact phases the iteration numbers increase. When no constraints are active it suffices to perform inexact steps only, which leads to a large speed-up. Right: The L^2-norm of the dual Lagrange multipliers for the stabilised scheme which again show a smooth evolution of the contact forces.*

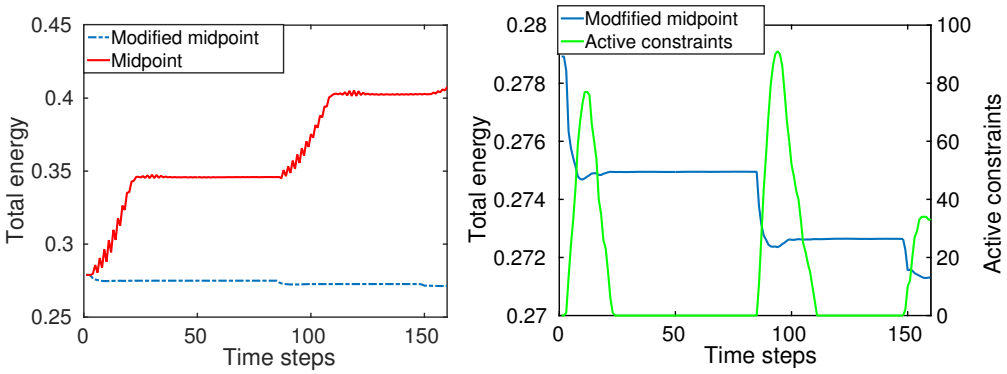


Figure 4.11.: *Left: The total energy of the midpoint rule and its stabilised pendant. While the midpoint rule is increasing the energy during each contact phase, the modified version is almost conserving. Right: The total energy of the stabilised scheme together with the number of active constraints. The main energy loss is observed during the attachment phase while it is increased a little during the detachment.*

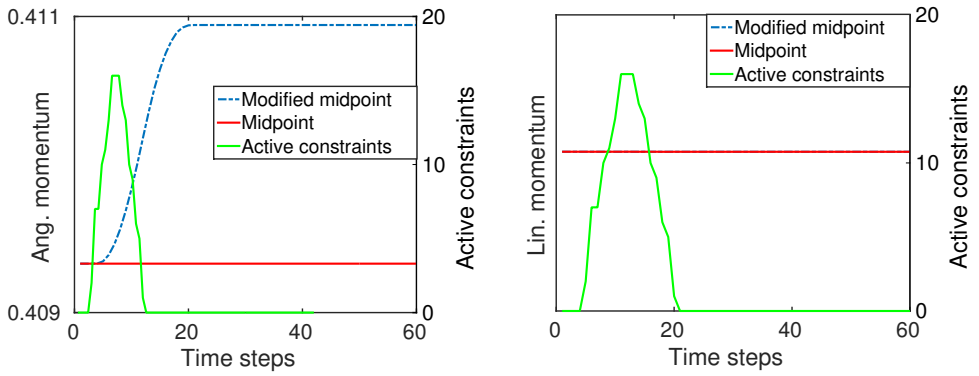


Figure 4.12.: *Linear and angular momentum of the two schemes in the absence of external forces and boundary conditions. While both schemes conserve the linear momentum, the stabilising modification of the midpoint rule leads to a loss of angular momentum preservation.*

modification leads to a loss of angular momentum conservation, see Figure 4.12. In [71] Kaufman et al. suggest to use generalised reflections to regain momentum and energy conservation which is similar to the modified velocity update in [72]. We will investigate this behaviour in more detail by carefully going through the exact linearised constraints and identify where the rotational invariance is lost.

Concluding remarks

In conclusion the monotone multigrid and the truncated non-smooth Newton multigrid method work extremely well even in the case of non-convex problems. Further, the numerical results are in accordance with the convergence theory developed in Chapter 3. Compared to its exact pendant a large reduction in computational time can be expected, especially in dynamic simulations, when the switching parameter is chosen appropriately. The proposed time discretisation scheme (2.6) for the non-standard differential inclusion showed an improved energy behaviour. Further investigations and extensions have to be made to recover energy and momentum conservation.

5. Femoroacetabular Impingement Analysis

In this chapter we present a heterogeneous joint model and a framework for the numerical simulation of patient-specific range of motion (ROM) and femoroacetabular impingement (FAI) analysis. After describing the medical background and the prevailing pre-operational hip surgery planning, we will introduce the heterogeneous joint model in Section 5.1 and a computational framework of how it can be used to estimate the range of motion and impingement zones in Section 5.2. We close this chapter by showing first promising results and compare our approach to a state-of-the-art FAI model [86].

5.0.4. The human hip-anatomy, functionality and failures

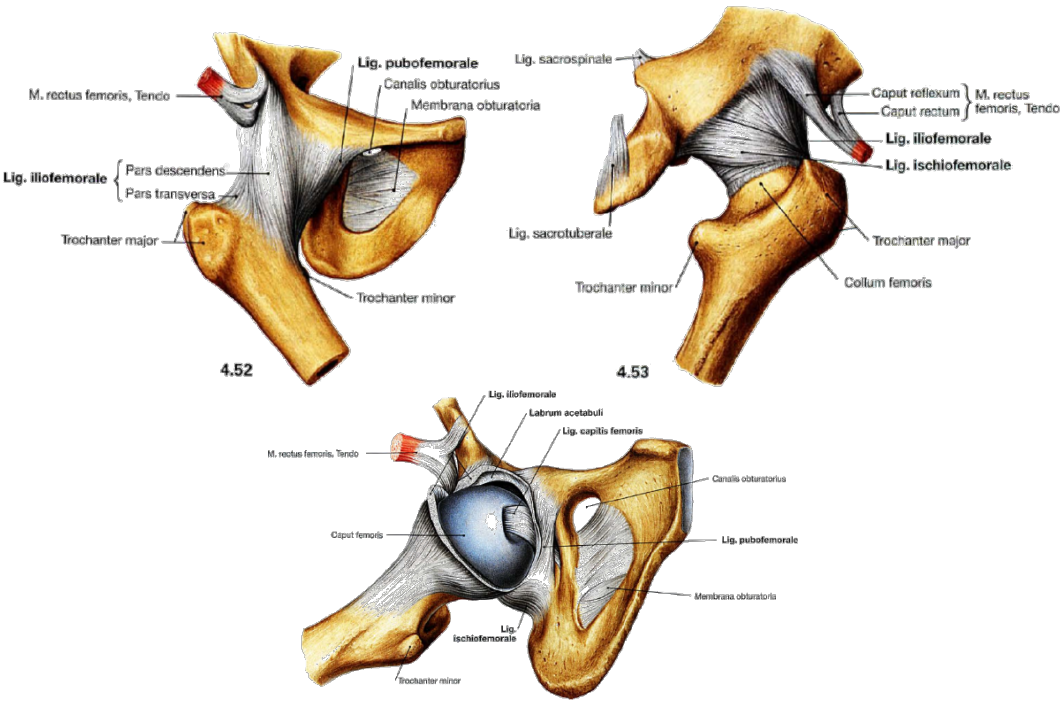


Figure 5.1.: *Anatomy of the hip joint. Picture taken from [104].*

The main task of the hip joint is to bear the body weight during movements or standing positions. It is a classical ball and socket joint which allows the movement of the leg in almost all directions. This movement is restricted by the joint capsule, the extracapsular *iliofemoral, ischiofemoral* and *pubofemoral* ligaments and surrounding muscles, in favour of enhanced stability [110]. The femoral head and the acetabulum are covered with articular hyaline cartilage which is lubricated as most of the motion involves sliding [110]. The extracapsular ligaments are reinforcing the joint and are among the strongest ligaments of the whole body [105]. The intracapsular *ligament teres* connects the acetabular notch and the femoral head. While its stabilising effect on the joint can be neglected, it can provide artery supply to the femur bone.

Range of motion (ROM)

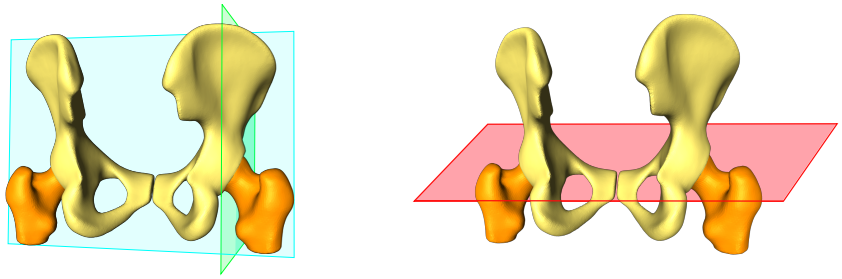


Figure 5.2.: *Sagittal (green), coronal (blue) and axial plane (red).*

The range of motion of the hip joint is quantified by measuring the maximal angles of deflection along the three orthogonal axes of motion of the ball joint. All axes pass through the centre of the femoral head and they are aligned to the *planes of motion*, the *sagittal*, *coronal* and the *axial plane*, see Figure 5.2. *Flexion* and *extension* is motion that decreases resp. increases the angle between two parts of the body. For the hip joint this corresponds to movements in the sagittal plane, cf. Figure 5.3. While the maximal extension is strongly restricted by all three ligaments, a lot of flexion is possible, only limited by contact of the thigh and the chest [102]. *Adduction and abduction* (coronal plane) describes motion that moves extremities towards resp. away from the body centre. The maximal possible angle depends on if the thigh is flexed or not. In the flexed position the tension in the iliofemoral is reduced, which allows a much higher adduction. The abduction is also reduced by the pubofemoral ligament [102]. The *internal and external rotation* denotes rotation towards resp. away from the body axis. The maximal internal rotation is restricted by the ischiofemoral and the lateral iliofemoral ligament, the external rotation is controlled by the transversal part of the iliofemoral ligament, again allowing a larger rotation when the hip is flexed [102]. Average angles of each motion in the standing position can be found in Table 5.1

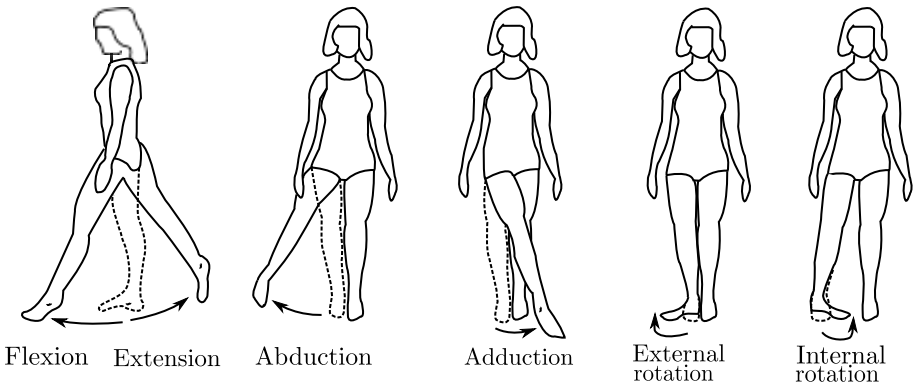


Figure 5.3.: *Movements along the axes of motion. Picture redrawn from [102][p.207].*

Flexion	$130 - 140°$
Extension	$10 - 15°$
Abduction	$30 - 40°$
Adduction	$20 - 30°$
Ext. rot.	$40°$
Int. rot.	$50°$

Table 5.1.: *Average maximal angles along the axes of motion, taken from [102, 104].*

Femoroacetabular Impingement (FAI)

It is widely accepted nowadays that FAI is one of the most common cause of osteoarthritis in hip joints [15]. FAI denotes the impingement (contact) of the femoral head and the acetabular rim resulting in unnatural stresses that cause pain. There are two types of FAI, the *pincer*-type, typically originating from an excessively deep acetabulum, and the *cam*-type impingement, caused by a diminished femoral neck offset, see Figure 5.4. Both types can be present at the same time and they in general reduce the flexion, abduction and internal rotation [16]. When the impingement type is detected, the surgeons task is to restore the femoral head neck or eliminate focal impingement lesions correspondingly. The pre-operational orthopaedic surgery planning nowadays is still mainly based on X-rays. Even in complex processes as total hip arthroplasty, simple templating, digital or by hand, cf. Figure 5.5, are used to decide on shape, size and position of the implant [76]. This gap between the 2-dimensional surgery planning and the 3-dimensional reality is one of the motivations for virtual medicine in orthopaedic surgery.

State-of-the-art FAI analysis

The prevailing toolboxes for a virtual FAI analysis are based on computerised tomography (CT) (HipOp[86],Articulis[112], HipProject[109], HipNav[126]), and due to the lack of data, rarely on magnetic resonance imaging (MRI). In a first step the data is

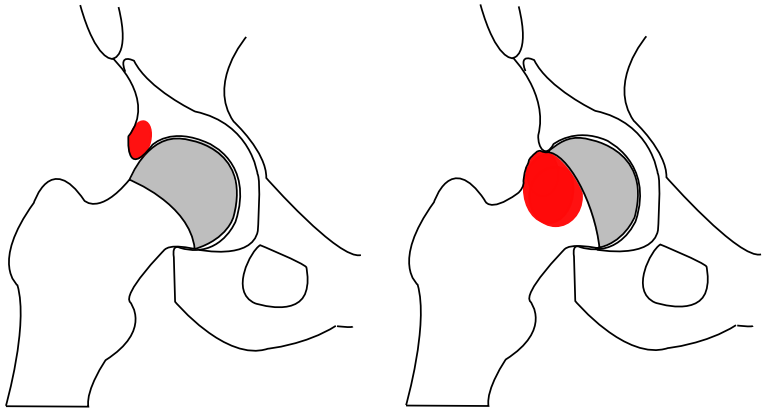


Figure 5.4.: *Impingement zones colored in red. Left: Pincer-type impingement. Right: Cam-type impingement.*

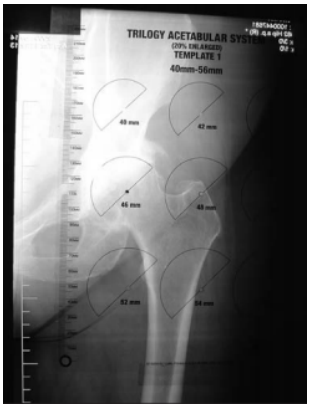


Figure 5.5.: *Acetabular templating stencil to determine the center of rotation. Picture taken from [76].*

segmented and three-dimensional geometries of the bones are extracted. The centre of rotation of the joint is computed using e.g. a sphere fitting algorithm. Then, a ROM analysis is conducted to determine the impingement zones. Practically all available FAI tools assume that the hip is a perfect ball and socket joint with fixed rotational centre. The analysis can then be accomplished by simple rigid body rotation along the axis of motion until collision is detected [86, 81, 109, 24, 83]. The maximal rotation angle is computed e.g. by doing a binary search [81]. Extensions of this approach exist where the femur, apart from rotating along the prescribed axis, additionally can translate freely. This is used to smooth out collisions that do not correspond to pincer- or cam-type impingement. In these models an impingement is detected if the joint cannot be set into a collision-free status by translation of the femur in less than e.g. 3mm [112, 16]. For the collision detection real-time algorithms can be constructed, e.g. by using a lookup table and linear transforms to speed-up the process [67], cf. Figure 5.6. The surgeon can then remodel the joint and re-evaluate the analysis to improve the ROM and, based on this,

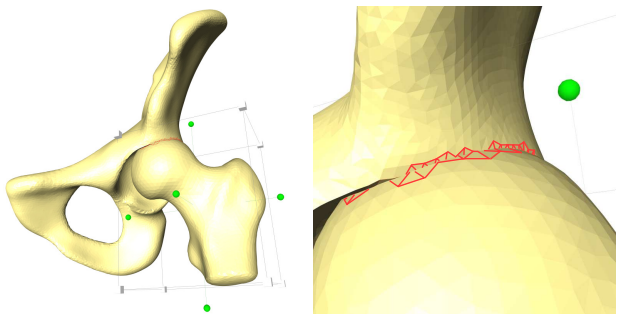


Figure 5.6.: *Real-time collision detection in the visualisation software* Amira *[41].*

develop an optimal surgery strategy [92].
Finite element models have been developed for the case where rather the joint loading than the ROM is of interest. The model in [91] is based on an idealised joint geometry that consists of rigid bones and sliding contact between linear elastic cartilage layers. Ligaments and muscles are neglected and the evolution is driven by external forces that were measured during in vivo experiments. In contrast to the previous approaches this model cannot be used for virtual testing and ROM optimisation as the evolution of the system is prescribed by the measured data. Furthermore, ligaments are not included, which are significant for an accurate estimation of the ROM, see Section 5.0.4. Similar finite element approaches can be found in [131, 25]. In [131] additionally the hip abductors are modelled by stiff cable elements. A more complex model of the knee-thigh-hip complex is proposed in [124]. The cortical bone is modelled by an orthotropic material and ligaments by one-dimensional discrete elements. The approach further contains (passive) muscles and a soft tissue mass distribution simulating the patients weight.
Finally, also near-real-time models are available using the discrete element method [6, 1]. In these approaches the bones are assumed to be rigid and cartilage between the acetabulum and femoral head is simulated by a series of springs.

5.1. A heterogeneous hip joint model

In this section we present a heterogeneous joint model that can be used for an advanced FAI analysis and further tasks in virtual medicine. The mathematical model is an extension of the knee joint model previously developed in [120]. It includes the proximal extremity of the femur, the pelvis bone and the major iliofemoral, ischiofemoral and pubofemoral ligaments, cf. Figure 5.7. We omit the joint capsule, hyaline cartilage and surrounding soft tissue for now. Although they could be incorporated in the mathematical model as passive tissues, in practical situations only CT or X-ray data is available, which makes the reconstruction of soft-tissues impossible.

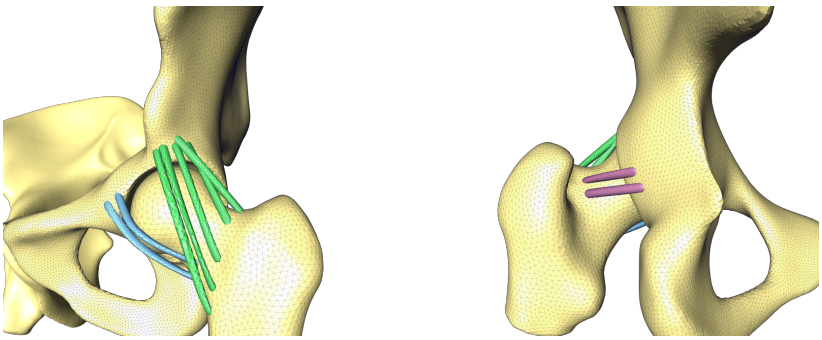


Figure 5.7.: *The heterogeneous joint model (left: AP-view, right: PA-view) with the three (re-meshed) major ligaments iliofemoral (green), ischiofemoral (purple), and pubofemoral (blue).*

Hyperelastic bones

Bone is an anisotropic multiscale material [33]. Depending on the specific application a large variety of models, ranging from biphasic models [80], to orthotropic [124], and linear elastic ones [141], exist. On the macroscopic scale that we are interested in, the cortical bone can be considered as homogeneous, isotropic elastic material [117, Chapter 2.2]. Therefore, we model the femur and the pelvis as geometrically non-linear St.Venant–Kirchhoff materials (1.15). In contrast to the small strain approach [120], we model large deformation contact between the femoral head and the acetabulum of the pelvis, see Figure 5.8, which allows for the proper simulation of rotational movements. The algorithmic treatment of these components is done as described in the previous chapters 1–4.

In [120] uniform articular cartilage layers on top of the bones are modelled by a linear

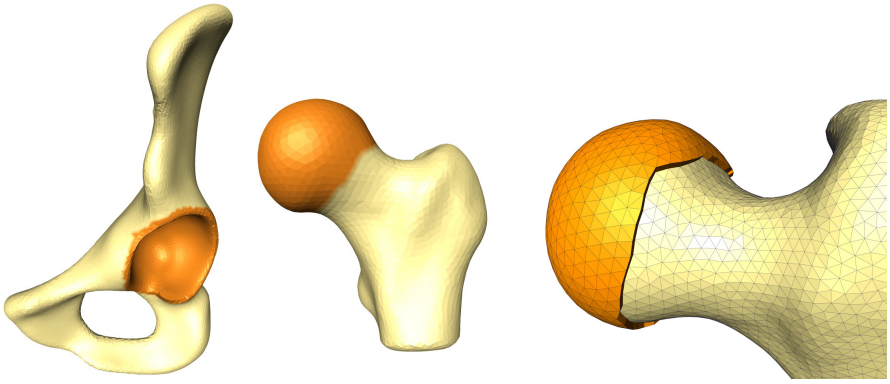


Figure 5.8.: *Left and Middle: Contact boundaries of the pelvis and the femur.Right: Articular cartilage layer.*

visco-elastic Kelvin–Voigt material. We note that the same construction can be applied to add cartilage to the present hip joint model, see Figure 5.8, although only hyperelastic

materials fit in the algorithmical framework currently.

5.1.1. Ligaments as Cosserat rods

We use Cosserat rods to model the ligaments, which was first proposed by Sander in [117]. Ligaments are long and slender objects which motivates to use a reduced model to avoid any meshing problems. The main assumption in one-dimensional Cosserat continua is that the cross-sections of a rod hardly deform and hence can be assumed to be rigid. We will only give a short introduction to Cosserat rods here, a detailed introduction, and much more, can be found in e.g. [5].
The configuration of a Cosserat rod is given by a curve in space $\mathbf{r}$ (the *centreline*) together with a set of rotations $\mathbf{q}$ describing the orientation of the cross-sections

$$\begin{aligned} \boldsymbol{\psi} : [0, L] &\longrightarrow \mathbb{R}^3 \times \mathrm{SO}(3), \\ \boldsymbol{\psi}(s) &= (\mathbf{r}(s), \mathbf{q}(s)), \end{aligned} \tag{5.1}$$

cf. Figure 5.9. The columns of the orthogonal rotation matrix $\mathbf{q}(s)$ define a local coordinate system $\{\mathbf{d}_1(s), \mathbf{d}_2(s), \mathbf{d}_3(s)\}$, where the *directors* $\mathbf{d}_1(s), \mathbf{d}_2(s)$ lie in the plane spanned by the cross-section and $\mathbf{d}_3(s)$ is orthogonal to that plane, but does not necessarily coincide with the tangent of the centreline $\mathbf{r}'(s)$, see Figure 5.9. Similar to the full-dimensional continuum case we denote by $\boldsymbol{\psi}_0$ the *reference configuration* of the rod. We assume that in the reference configuration the centreline $\mathbf{r}_0$ is parametrised by arc-length. Although the model is one-dimensional it can still represent *shearing, stretching, bending* and *twisting*, which is crucial for the proper simulation of ligaments. Suitable strain measures $\mathbf{v}, \mathbf{w}$ are defined by

$$\mathbf{v}(s) := \mathbf{r}'(s), \quad \mathbf{d}_i'(s) = \mathbf{w}(s) \times \mathbf{d}_i(s) \quad i = 1, 2, 3. \tag{5.2}$$

When represented in the local basis $\{\mathbf{d}_i\}$, the strain $\mathbf{v}$ measures shearing and stretching and $\mathbf{w}$ quantifies strain due to bending and twisting. In contrast to the previous approach [120] we will neglect the inertia terms for now and only consider quasi-static Cosserat rods. A description of how the dynamic model can be incorporated into the joint model can be found in [120]. The balance of momentum in the quasi-static case is given in terms of the resulting net forces $\mathbf{n}$ and moments $\mathbf{m}$

$$\begin{aligned} \mathbf{n}' = 0 \quad &\text{on } [0, L], \\ \mathbf{m}' + \mathbf{r}' \times \mathbf{n} = 0 \quad &\text{on } [0, L]. \end{aligned} \tag{5.3}$$

Note that we omitted external forces and momenta in (5.3) for simplicity. As usual, additional boundary conditions and the *constitutive law* linking the resulting forces to the strains have to be prescribed to close the system (5.3). Like in Chapter 1 we only consider *hyperelastic* Cosserat materials, i.e. we assume that there exists an energy functional $W(\mathbf{y}, \mathbf{z}, s)$, such that

$$\mathbf{n} = \frac{\partial W}{\partial \mathbf{y}}(\mathbf{v} - \mathbf{v}_0, \mathbf{w} - \mathbf{w}_0, s), \quad \mathbf{m} = \frac{\partial W}{\partial \mathbf{z}}(\mathbf{v} - \mathbf{v}_0, \mathbf{w} - \mathbf{w}_0, s), \tag{5.4}$$

where $\mathbf{v}_0, \mathbf{w}_0$ denote the initial strains of the reference configuration. Ligament is a multiscale material that shows elastic and visco-elastic behaviour. The prevailing models in the field of joint mechanics are simple spring models, but also more complex approaches have been used, see [117] for an overview. We are only interested in the macroscopic mechanics of the hip joint and that is why we will use a linear elastic, yet geometrically exact, material law in this model.

$$W(\mathbf{y}, \mathbf{z}, s) = \frac{1}{2} \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}^T \mathbf{C}(E, \nu) \begin{pmatrix} \mathbf{y} \\ \mathbf{z} \end{pmatrix}, \tag{5.5}$$

where the *diagonal* matrix $\mathbf{C}$ only depends on the *Poisson ratio* ν, the *elastic modulus* E and the shape of the cross-sections, see [120] for more details. We further prescribe Dirichlet boundary conditions

$$(\mathbf{r}(0), \mathbf{q}(0)) = (\boldsymbol{\varphi}_0, \mathbf{F}_0), \quad (\mathbf{r}(L), \mathbf{q}(L)) = (\boldsymbol{\varphi}_L, \mathbf{F}_L), \tag{5.6}$$

that depend on the deformation of the attached bones, as will be shown in the next section. Analogously to (1.19) the equilibrium states of a hyperelastic Cosserat rod are

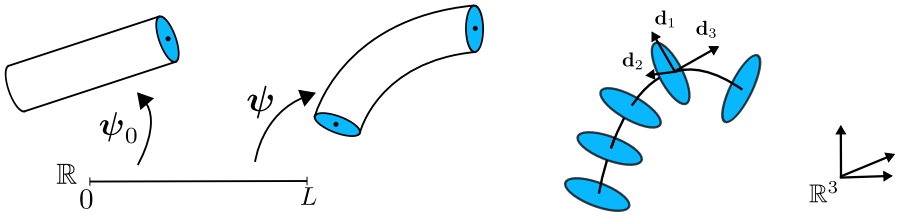


Figure 5.9.: *Left: Reference and deformed configuration of a Cosserat rod. Right: Local coordinate system defined by the directors.*

characterised as minimisers of the *non-convex* strain energy

$$\min_{\boldsymbol{\psi} \in H} \mathcal{J}(\boldsymbol{\psi}) := \int_0^L W(\mathbf{v}(\boldsymbol{\psi}) - \mathbf{v}_0, \mathbf{w}(\boldsymbol{\psi}) - \mathbf{w}_0)\, ds, \tag{5.7}$$

where the Sobolev space for manifold-valued functions is defined as

$$H := H^1_D([0, L], \mathrm{SE}(3)) = \left\{ \boldsymbol{\psi} \in H^1_D([0, L], \mathbb{R}^k) : \boldsymbol{\psi} \in \mathrm{SE}(3) \text{ a.e. on } [0, L] \right\}, \tag{5.8}$$

and $\mathrm{SE}(3) = \mathbb{R}^3 \times \mathrm{SO}(3)$ denotes the *special Euclidean group*. The integer k depends on the embedding of the orthogonal group $\mathrm{SO}(3)$. We will use *unit quaternions* which are numerically stable (in this case $k = 7$). A detailed description of how to use unit quaternions for the description of rotations can be found in [116].

Theorem 5.1.1. *There exist solutions of* (5.7) *subject to Dirichlet boundary conditions* (5.6) *and the orientation constraint*

$$\langle \mathbf{v}, \mathbf{d}_3 \rangle > 0 \tag{5.9}$$

These solutions are in general not unique.

Proof. Wolfe and Seidman [123, Theorem 3.8] □

Geodesic Finite Elements

The price we pay for the reduction in dimension is that the configuration space SE(3) is no longer a linear space. Therefore, the standard finite element method, which is based on linear interpolation, cannot be used any more. To overcome this difficulty Sander developed *geodesic finite elements* which, in the first order case, replace linear interpolation by interpolation along geodesics [116]. Let a grid $\mathcal{T}$ of $[0, L]$ be given by $0 = s_1 < \ldots < s_n = L$.

Definition 5.1.2 (Geodesic Finite Elements)**.** A function $\boldsymbol{\psi}_h : [0, L] \longrightarrow \mathrm{SE}(3)$ is a *geodesic finite element*, if it is continuous and $\boldsymbol{\psi}_h$ restricted to each element $[s_i, s_{i+1}]$ is a minimising geodesic in SE(3).

Remark 5.1.3. *In [57] the construction of higher order geodesic finite elements can be found and optimal a priori discretisation error bounds are proven.*

We set

$$\mathbf{V}_h := \{\boldsymbol{\psi}_h : [0, L] \longrightarrow \mathrm{SE}(3) : \boldsymbol{\psi}_h \text{ is a geodesic finite element}\}.$$

The discrete problem is derived by replacing the infinite-dimensional function space by the finite-element one:

$$\min_{\boldsymbol{\psi}_h \in \mathbf{V}_h} \mathcal{J}(\boldsymbol{\psi}_h), \quad \text{s.t. } \boldsymbol{\psi}_h(0) = (\boldsymbol{\varphi}_0, \mathbf{F}_0),\ \boldsymbol{\psi}_h(L) = (\boldsymbol{\varphi}_L, \mathbf{F}_L). \tag{5.10}$$

The derivation of an equivalent algebraic formulation is complicated by the fact that the identification of a geodesic finite element with its coefficient vector $\mathrm{SE}(3)^n$ is not necessarily unique. It is guaranteed if the grid $\mathcal{T}$ is sufficiently fine, see [116, Lemma 5.2]. Assuming the uniqueness of the coefficient vectors the algebraic formulation of (5.10) reads

$$\min_{\boldsymbol{\psi}_h \in \mathrm{SE}(3)^n} \mathcal{J}(\boldsymbol{\psi}_h), \quad \text{s.t. } \boldsymbol{\psi}_{h,0} = (\boldsymbol{\varphi}_0, \mathbf{F}_0),\ \boldsymbol{\psi}_{h,L} = (\boldsymbol{\varphi}_L, \mathbf{F}_L), \tag{5.11}$$

where in a slight abuse of notation we denote both the coefficient vector and the corresponding geodesic finite element function by $\boldsymbol{\psi}_h$. More details and the formulas for the evaluation of the geodesics and their derivatives using quaternions can be found in [116].

Riemannian Trust-Region Solver

In this paragraph we describe the Riemannian trust-region algorithm that can be used to solve the non-convex minimisation problems (5.11). This method was first proposed by Absil et al. [2] for matrix manifolds and then applied to Cosserat rods by Sander [116, 117], where again the details and the explicit formulas for the case $M = \mathbf{V}_h$ can be found. In Section 3.3 we introduced the standard trust-region method for non-linear optimisation problems in $\mathbb{R}^N$. The main idea was to locally approximate the energy functional by

quadratic models and then compute a correction by minimising this model instead. The extension of this algorithm to manifolds consists of locally lifting the energy to the tangential space $T_{\boldsymbol{\psi}_h}\mathrm{SE}(3)^n$ using the *exponential map* $\exp_{\boldsymbol{\psi}_h} : T_{\boldsymbol{\psi}_h}\mathrm{SE}(3)^n \longrightarrow \mathrm{SE}(3)^n$

$$\bar{\mathcal{J}}(u) = \mathcal{J}(\exp_{\boldsymbol{\psi}_h} u). \tag{5.12}$$

Now $\bar{\mathcal{J}} : T_{\boldsymbol{\psi}_h}\mathrm{SE}(3)^n \longrightarrow \mathbb{R}$ is a mapping between linear spaces which can be approximated by a quadratic model with the help of the Riemannian metric $h_{\boldsymbol{\psi}_h} : T_{\boldsymbol{\psi}_h}\mathrm{SE}(3)^n \times T_{\boldsymbol{\psi}_h}\mathrm{SE}(3)^n \longrightarrow \mathbb{R}$ in a straightforward manner

$$m_{\boldsymbol{\psi}_h}(u) := \bar{\mathcal{J}}(0) + h_{\boldsymbol{\psi}_h}(\nabla\bar{\mathcal{J}}(0), u) + \frac{1}{2}h_{\boldsymbol{\psi}_h}(\nabla^2\bar{\mathcal{J}}(0)u, u). \tag{5.13}$$

For a given iterate $\boldsymbol{\psi}_h^k$ the local problems of the Riemannian trust-region methods are given by

$$\min_{u \in T_{\boldsymbol{\psi}_h}\mathrm{SE}(3)^n} m_{\boldsymbol{\psi}_h^k}(u), \tag{5.14}$$

$$\|u\|_{\mathrm{TR}} \leq \Delta_k, \tag{5.15}$$

cf. (3.32). The trust-region norm can again be chosen arbitrarily. One option is to take the norm induced by the Riemannian metric $\|\cdot\|_{\mathrm{TR}} = \sqrt{h_{\boldsymbol{\psi}_h}(\cdot,\cdot)}$. Further selecting the infinity norm has the advantage that the trust-region constraints are bound constraints, which enables us to solve (5.14) efficiently using the monotone multigrid method introduced in Section 3.2. Once the local problem (5.14) is solved the iterate is updated using the exponential map

$$\boldsymbol{\psi}_h^{k+1} = \exp_{\boldsymbol{\psi}_h} u. \tag{5.16}$$

The acceptance criterion of the trial iterate and the update of the trust-region norm Δ_k is done analogously to Algorithm TR.

One-sided contact between ligament and bone

To predict the range of motion and impingement zones it is essential to enforce non-penetration of the bones and ligaments. We make the assumption that the contact forces which the ligaments exert on the bones are small such that we can neglect them. Hence, it is sufficient to model Signorini-type contact for the ligaments, considering the bones to be rigid moving obstacles. To our knowledge, there exist no other continuous models for contact between Cosserat rods and rigid obstacles. A discrete approach that is based on collision detection, neglecting elastic forces between the discrete degrees of freedom was proposed in [127] in the field of computer graphics.
In the following we will consider the case of one Cosserat rod and a rigid obstacle $\Gamma_{\mathrm{obs}} \subset \mathbb{R}^3$. We assume that the obstacle has a sufficiently smooth boundary, such that the outer normal field $\mathbf{n}^{\mathrm{obs}}$ exists and is continuous. Further, we only consider Cosserat rods with circular cross-section and constant radius l.

Let the boundary of a Cosserat rod with configuration $\boldsymbol{\psi}$ be decomposed into a Dirichlet part $\Gamma_D^{\boldsymbol{\psi}} = \{\mathcal{A}_0, \mathcal{A}_L\}$ given by the cross-sections at 0 and L, and the potential contact boundary $\Gamma_C^{\boldsymbol{\psi}}$. To derive a parametrisation of the contact boundary we note again that each cross-section is characterised by the position of the centreline $\mathbf{r}(s)$, the radius l and the first two directors $\mathbf{d}_1(s), \mathbf{d}_2(s)$. The contact boundary can be written as

$$\Gamma_C^{\boldsymbol{\psi}} := \{\mathbf{r}(s) + l\,(\mathbf{d}_1(s)\cos\alpha + \mathbf{d}_2(s)\sin\alpha) :\ s \in (0, L), \alpha \in [0, 2\pi)\}\,. \tag{5.17}$$

As we introduced in Section 1.2, the non-penetration of two bodies can be modelled using the closest point mapping $\Phi^{\mathrm{cp}} : \Gamma_C^{\boldsymbol{\psi}} \longrightarrow \Gamma_{\mathrm{obs}}$

$$\Phi^{\mathrm{cp}}(\mathbf{x}) := \underset{\mathbf{y}\in\Gamma_{\mathrm{obs}}}{\arg\min}\, ||\mathbf{x} - \mathbf{y}||, \tag{5.18}$$

cf. (1.21). To simplify the notation we formulate the closest point mapping on the deformed domain $\Gamma_C^{\boldsymbol{\psi}}$. Although not written explicitly, the mapping Φ^{cp} also depends on the actual rod configuration $\boldsymbol{\psi}$.

Remark 5.1.4. *Note that in contrast to the full-dimensional case in Chapter 1, we use the closest point projection rather than the normal projection, which simplifies the discretisation of the constraints significantly.*

The corresponding contact normal field on $\Gamma_C^{\boldsymbol{\psi}}$, is defined by

$$\boldsymbol{\nu}_\Phi(\mathbf{x}) := \mathbf{n}^{\mathrm{obs}}(\Phi^{\mathrm{cp}}(x)). \tag{5.19}$$

cf. (1.22). Φ^{cp} is a best approximation and that is why the vector $\mathbf{x} - \Phi^{\mathrm{cp}}(\mathbf{x})$ is itself normal to Γ_{obs} at $\Phi^{\mathrm{cp}}(\mathbf{x})$. Hence, the gap function

$$g^R(\mathbf{x}) := \boldsymbol{\nu}_\Phi(\mathbf{x}) \cdot (\mathbf{x} - \Phi^{\mathrm{cp}}(\mathbf{x})) \tag{5.20}$$

is a signed measure for the distance of the Cosserat rod to the obstacle, see (1.23). Non-penetration of the rod and the obstacle is ensured by enforcing that

$$g^R(\mathbf{x}) \geq 0 \quad \forall\, \mathbf{x} \in \Gamma^{\boldsymbol{\psi}} \tag{5.21}$$

holds. Combining (5.21) with the assumption of hyperelasticity (5.7), we arrive at the following constrained minimisation problem:

$$\begin{aligned} &\min_{\boldsymbol{\psi}\in H} \mathcal{J}(\boldsymbol{\psi}), \\ &\quad g^R(\boldsymbol{\psi}) \geq 0 \ \text{ on } \Gamma^{\boldsymbol{\psi}}. \end{aligned} \tag{5.22}$$

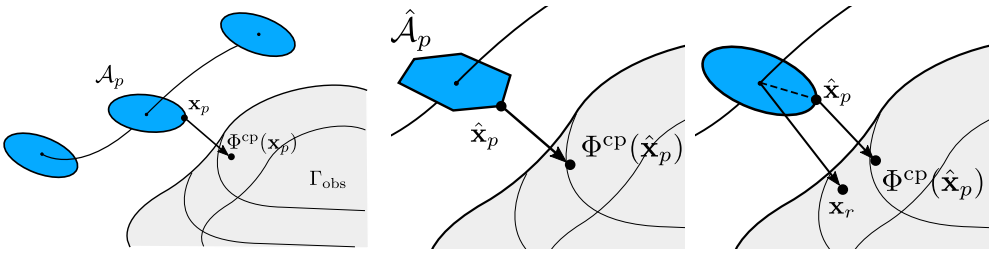


Figure 5.10.: *Left: Exact nodal non-penetration constraint. Middle: Piece wise linear discretised cross-section boundary. Right: Approximate closest point by projection.*

Discretisation of contact

We will now propose a nodal discretisation of the contact constraints. Let again $\mathcal{T}$ be a grid of $[0, L]$ and assume that the Cosserat rod is discretised using geodesic finite elements (5.10). Now consider a node $p \in \mathcal{T}$ in the grid and denote by $\mathcal{A}_p$ the corresponding cross-section within the current configuration $\boldsymbol{\psi}_h$. Let $\mathbf{x}_p \in \partial\mathcal{A}_p$ be the point closest to the obstacle, i.e.

$$\mathbf{x}_p = \operatorname*{arg\,min}_{\mathbf{x}\in\partial\mathcal{A}_p} \|\mathbf{x} - \Phi^{\mathrm{cp}}(\mathbf{x})\|. \tag{5.23}$$

A straightforward discretisation of the non-penetration constraint is to demand that

$$\bar{g}_p^R(\boldsymbol{\psi}_h) := \boldsymbol{\nu}_\Phi(\mathbf{x}_p) \cdot (\mathbf{x}_p - \Phi^{\mathrm{cp}}(\mathbf{x}_p)) \geq 0 \quad 2 \leq p \leq n-1. \tag{5.24}$$

These constraints still require the solution of a continuous minimisation problem (5.23) and thus have to be discretised further. One possibility is to discretise the cross-section boundaries $\partial\mathcal{A}_p$ by piecewise linear segments and prescribe non-penetration for all corner points or only the one that is closest to the obstacle, see the middle of Figure 5.10.
We will use a different approach: First we compute the closest point $\mathbf{x}_r$ on the obstacle to the centreline point $\mathbf{r}_p$. Then we project the distance vector $\mathbf{x}_r - \mathbf{r}_p$ onto the plane spanned by the cross-section and use the resulting point $\hat{\mathbf{x}}_p$ on the boundary $\partial\mathcal{A}_p$ as an approximation of the exact closest point $\mathbf{x}_p$, see the right of Figure 5.10. The fully-discrete nodal non-penetration constraints then read

$$g_p^R(\boldsymbol{\psi}_h) := \boldsymbol{\nu}_\Phi(\hat{\mathbf{x}}_p) \cdot (\hat{\mathbf{x}}_p - \Phi^{\mathrm{cp}}(\hat{\mathbf{x}}_p)) \geq 0 \quad 2 \leq p \leq n-1, \tag{5.25}$$

leading to the non-linear non-convex constrained minimisation problem

$$\begin{aligned} &\min_{\boldsymbol{\psi}_h\in\mathbf{V}_h} \mathcal{J}(\boldsymbol{\psi}_h), \\ &\qquad g^R(\boldsymbol{\psi}_h) \geq 0. \end{aligned} \tag{5.26}$$

Remark 5.1.5. *During the assembling of the contact constraints* (5.25) *the closest point mapping Φ^{cp} has to be evaluated. This requires the computation of the point within a face $f \in \Gamma_{obs}$, that is closest to the target point $\hat{\mathbf{x}}_p$(or $\mathbf{r}_p$). This is implemented using the polyhedral Gauss–Seidel method, where the face edges are used as search directions [59].*

Filter–Riemannian–trust–region method

Finally, we solve the constrained minimisation problem (5.26) using by a combination of the filter method Algorithm FTR and the Riemannian trust–region algorithm described in Section 5.1.1.

5.1.2. Heterogeneous coupling of bones and ligament

Motivated by classical domain decomposition theory the following coupling conditions for the Cosserat problem and a linear elastic continuum were first proposed by Sander in [115, 117]. Later Schiela and Sander extended them to the non-linear case and proved existence of solutions of the coupled system [119]. In classical domain decomposition generally two types of coupling conditions have to be prescribed to ensure the well-posedness of the system: A condition guaranteeing the continuity of the *primal* variables and a *dual* constraint, ensuring that the transmitted forces are preserved, see [108]. In the following

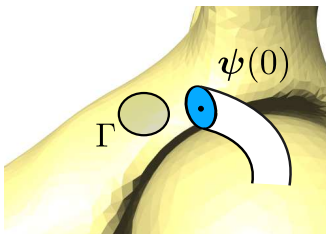


Figure 5.11.: *Coupling interface of a Cosserat rod and a continuum*

we consider a hyperelastic body Ω with deformation $\boldsymbol{\varphi}$ and a Cosserat rod $[0, L]$ with configuration $\boldsymbol{\psi}$. We denote the corresponding coupling boundaries by $\Gamma \subset \partial\Omega$ and $\mathcal{A}_o$, i.e. the cross-section corresponding to $o \in \{0, L\}$. For the primal coupling the deformation $\boldsymbol{\varphi}$ has to coincide, in some sense, with the rod configuration $\boldsymbol{\psi}(o) = (\mathbf{r}(o), \mathbf{q}(o))$. This can be achieved by prescribing that the *average deformation* equals the position of the centreline

$$\mathbf{r}(o) = \frac{1}{|\Gamma|} \int_\Gamma \boldsymbol{\varphi}(s)\, ds. \tag{5.27}$$

Furthermore, the orientation $\mathbf{q}(o)$ has to be matched. To this end we define the *average orientation* of a deformation by

$$\mathcal{F}(\boldsymbol{\varphi}) := \frac{1}{|\Gamma|} \int_\Gamma \nabla\boldsymbol{\varphi}(s)\, ds. \tag{5.28}$$

Assuming enough regularity, the *polar decomposition* can be used to decompose this matrix into a rotational part $\mathcal{F}^{\mathrm{rot}}(\boldsymbol{\varphi})$ and a stretching $H(\boldsymbol{\varphi})$

$$\mathcal{F}(\boldsymbol{\varphi}) = \mathcal{F}^{\mathrm{rot}}(\boldsymbol{\varphi})\, H(\boldsymbol{\varphi}), \tag{5.29}$$

see [117, Lemma 5.2.1]. Now let $\boldsymbol{\psi}_0$ denote the reference configuration of the rod and $\mathbf{q}_0(o)$ the corresponding orientation at $o \in \{0, L\}$. The coupling for the orientation then reads

$$\mathcal{F}^{\mathrm{rot}}(\boldsymbol{\varphi})\, \mathbf{q}_0(o) = \mathbf{q}(o). \tag{5.30}$$

In what follows we will denote the collected average deformation and orientation by

$$\mathrm{Av}_\Gamma(\boldsymbol{\varphi}) = \left(\frac{1}{|\Gamma|} \int_\Gamma \boldsymbol{\varphi}(s)\, ds, \mathcal{F}^{\mathrm{rot}}(\boldsymbol{\varphi})\, \mathbf{q}_0(o) \right). \tag{5.31}$$

The *primal coupling conditions* are then summarised by

$$\mathrm{Av}_\Gamma(\boldsymbol{\varphi}) = \boldsymbol{\psi}(o).$$

The *dual conditions*, i.e. the balance of stresses and moments, can be derived in a straightforward manner:

$$\begin{aligned} -\mathbf{n}(o)\nu^R(o) &= \int_\Gamma \mathbf{P}(\boldsymbol{\varphi})\, \mathbf{n}(s)\, ds, \\ -\mathbf{m}(o)\nu^R(o) &= \int_\Gamma (s - \mathbf{r}(o)) \times (\mathbf{P}(\boldsymbol{\varphi})\, \mathbf{n}(s))\, ds, \end{aligned} \tag{5.32}$$

where ν^R denotes the outer unit normal field of $[0, L]$, i.e.

$$\nu^R(o) = \begin{cases} -1 & o = 0, \\ 1 & o = L, \end{cases} \tag{5.33}$$

and $\mathbf{P}$ the first Piola Kirchhoff stress tensor (1.4).

A Dirichlet–Neumann algorithm

We solve the coupled system using the Dirichlet–Neumann algorithm developed by Sander [117, 115]. In this method the heterogeneous system is decoupled by iterating over the coupling-interface values and successively solving the two sub-problems. For simplicity we will consider only one Cosserat rod connecting the pelvis Ω_p and the femur Ω_f with coupling interfaces $\Gamma_0 \subset \partial\Omega_p, \Gamma_L \subset \partial\Omega_f$. Let some initial proximal and distal interface iterate $\lambda_0^0, \lambda_L^0 \in \mathrm{SE}(3)$ be given and denote by $\boldsymbol{\varphi}_p, \boldsymbol{\varphi}_f \in \mathbf{S}_h$ the deformations of the pelvis and femur at some fixed time. The method can be summarised in the following three steps:

1. Dirichlet problem for the Cosserat rods of Signorini-type

Solve the constrained Cosserat problem with rigid obstacles

$$\begin{aligned} \boldsymbol{\psi}_h^{k+1} &= \operatorname*{arg\,min}_{\boldsymbol{\psi}_h \in \mathbf{V}_h} \mathcal{J}(\boldsymbol{\psi}_h), \\ g^R(\boldsymbol{\psi}_h^{k+1}) &\geq 0, \\ \Gamma_{\mathrm{obs}} &= \boldsymbol{\varphi}_p(\Omega_p) \cup \boldsymbol{\varphi}_f(\Omega_f), \end{aligned}$$

subject to Dirichlet conditions

$$\boldsymbol{\psi}_h^{k+1}(0) = \lambda_0^k, \quad \boldsymbol{\psi}_h^{k+1}(L) = \lambda_L^k.$$

2. Large deformation contact problem with Neumann forces

From the solution of the Cosserat problem $\boldsymbol{\psi}_h^{k+1}$ we evaluate the resulting stresses and moments at the coupling interfaces $o = 0, L$

$$\boldsymbol{\psi}_h^{k+1} \longmapsto \mathbf{n}_i^{k+1}, \mathbf{m}_i^{k+1}.$$

Then a Neumann force field $\mathbf{t}^{k+1}$ for the bones is constructed such that the dual coupling conditions (5.32) hold approximately:

$$\mathbf{n}^{k+1}, \mathbf{m}^{k+1} \longmapsto \mathbf{t}^{k+1},$$

such that

$$\begin{aligned} -\mathbf{n}_i^{k+1}\nu_i^R &= \int_{\Gamma_i} \mathbf{t}^{k+1}\, ds, \\ -\mathbf{m}_i^{k+1}\nu_i^R &= \int_{\Gamma_i} (s - \mathbf{r}_i^{k+1}) \times \mathbf{t}^{k+1}\, ds. \end{aligned} \tag{5.34}$$

With this at hand the large deformation contact problem (2.35) is solved with additional external traction forces on the coupling interfaces given by $\mathbf{t}^{k+1}$

$$\mathbf{t}^{k+1} \longmapsto \boldsymbol{\varphi}_p^{k+1}, \boldsymbol{\varphi}_f^{k+1}. \tag{5.35}$$

3. Damped geodesic update

To update the interface values, first the average deformation (5.27) and orientation (5.28) are evaluated

$$\boldsymbol{\varphi}_p^{k+1}, \boldsymbol{\varphi}_f^{k+1} \longmapsto \mathrm{Av}_{\Gamma_i}(\boldsymbol{\varphi}_i^{k+1}). \tag{5.36}$$

After that, a damped correction is added to the old interface value using the exponential map

$$\lambda_i^{k+1} = \exp_{\lambda_i^k} \omega \left[\exp_{\lambda_i^k}^{-1} \mathrm{Av}_{\Gamma_i}(\boldsymbol{\varphi}_i^{k+1}) \right], \tag{5.37}$$

where $0 \leq \omega \leq 1$ is a suitable damping parameter.

Remark 5.1.6. *In [117] it is proven for the linear elastic case with small deformations, that under suitable conditions the damped Dirichlet–Neumann algorithm converges to a solution of the coupled heterogeneous system.*

Remark 5.1.7. *The Neumann force fields* $\mathbf{t}^{k+1}$ *are constructed to be "as constant as possible" to account for the rigidity of the cross-sections* $\mathcal{A}_i$*. This is achieved by minimising*

$$\min_{\substack{\mathbf{t}\in\mathbf{L}^2(\Gamma_i) \\ c\in\mathbb{R}^3}} \int_{\Gamma_s} \|\mathbf{t}(s) - c\|\, ds, \tag{5.38}$$

under the constraint that (5.34) *holds, see [117] for more details.*

Remark 5.1.8. *The damped update of the interface values* (5.37) *within Dirichlet–Neumann methods is necessary for the convergence of the scheme. An example of divergence in the undamped case is shown in [108, p.12].*

Remark 5.1.9. *An extension of the model considering two-body contact between the ligaments and the bones could be solved with the algorithm described above, by additionally updating the rigid obstacle* Γ_{obs} *in each iteration and augmenting the Neumann force field* $\mathbf{t}^{k+1}$ *by the traction forces exerted by the ligaments. The construction of suitable contact forces like in* (5.38) *is subject of future work.*

Automatic construction of the geometric model

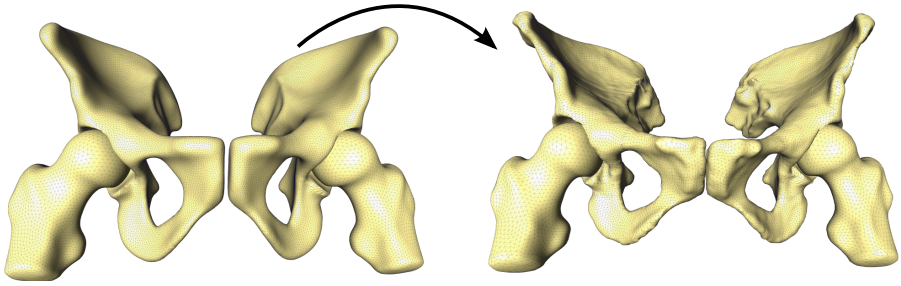


Figure 5.12.: *Morphing of the mean shape (left) into a patient-specific geometry (right).*

The joint geometries that we consider stem from the *articulated statistical shape model* developed by Lamecker et al. [85], which is making use of a principal component analysis. Starting from an averaged *mean shape* this model allows the approximate reconstruction of patient-specific hip geometries by minimising a suitable error measure [37]. The resulting *shape parameters* define the individual geometries by morphing of the mean shape, see Figure 5.12.
We constructed the insertion sites of the ligaments for the mean shape in consistency with anatomy books [104]. Hence, the shape parameters automatically define the corresponding insertion sites for the patient-specific geometry by morphing, see Figure 5.13.

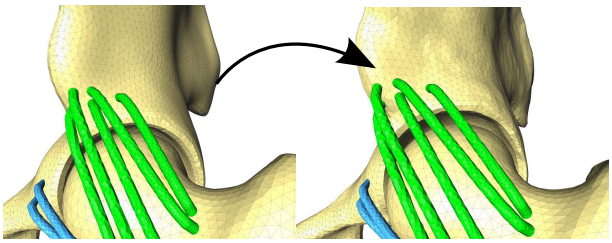


Figure 5.13.: *Movement of the insertion sites of the ligaments.*

Remark 5.1.10. *The statistical shape model [85] has another great advantage. The patient-specific geometries depend on the shape-parameter and the mean shape. Thus,*

the governing equations of motion can be formulated as a parameter-dependent system of PDEs. For these kind of problems reduced basis methods have been developed [113] that use a priori information to speed-up the costly finite element simulations drastically, which will be the focus of future work.

5.2. Finite element ROM and FAI analysis

Based on the heterogeneous joint model introduced in Section 5.1, we propose the following finite element FAI analysis. Let us assume that a 3D joint geometry was already constructed by e.g. segmentation of a CT and that the insertion sites of the ligaments are known, either from MRI data or constructed in an automatic way, cf. Section 5.1.2. Moreover, assume that the centre of rotation has been determined, e.g. using a sphere fitting algorithm, and that the corresponding axes of motion are given, see Figure 5.2. The initial configurations of the ligaments are constructed by solving a static Cosserat problem (5.26) with non-penetration constraints given by the bones and Dirichlet values given by the insertion sites.
For each of the six motions (flexion, extension, abduction, etc.) we drive the finite element joint model to maximal deflection by applying time-dependent Neumann forces to the femur while fixing the pelvis with homogeneous Dirichlet conditions, see Figure 5.14. The Neumann traction forces are chosen such that they induce a rotational motion within the corresponding plane of motion. Once impingement of the femur and the acetabular rim occurs the maximal deflection is reached and the associated angle is measured.

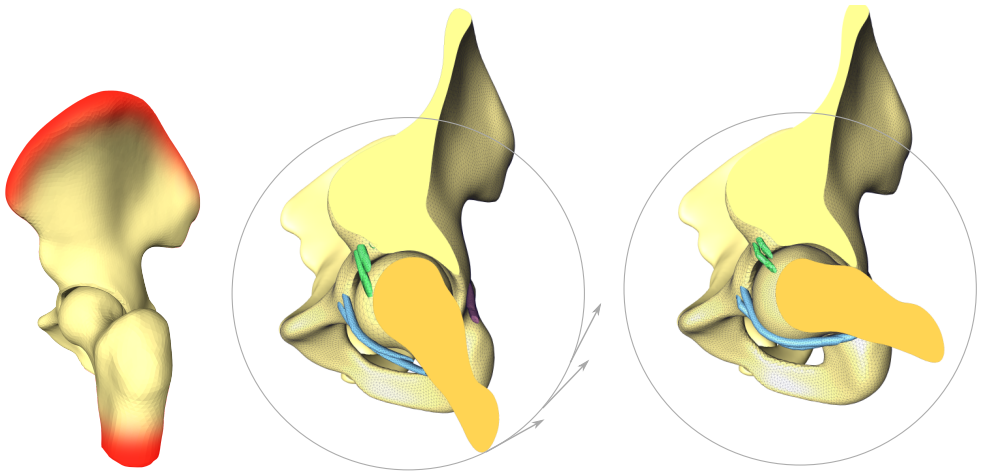


Figure 5.14.: *Left: Dirichlet patch of the pelvis and Neumann boundary of the femur. Middle: Sagittal cut of the joint illustrating the time-dependent Neumann forces in tangential direction. Right: Sagittal cut of the joint at maximal extension.*

5.2.1. Comparison with state-of-the-art models

In the following example we compare our finite element FAI model with a state-of-the-art model [86]. In this approach the centre of rotation is fixed and rigid body rotation along the axes of motion is performed until collision is detected. In this spirit we will call our FAI analysis a "flexible centre" approach hereafter. We use the 3D hip geometry of the mean shape, cf. 5.1.2, which we extended by ligaments.

In the left of Figure 5.15 the initial coarse grids of the joint are shown, that have been extracted together with a boundary parametrisation [77] from the statistical shape model [85], using the PSurface implementation [118] within Amira [41]. In the middle the three times refined grids (130,560 degrees of freedom) are shown together with the ligaments that have been visualised as triangulated surfaces. Realistic material parameters

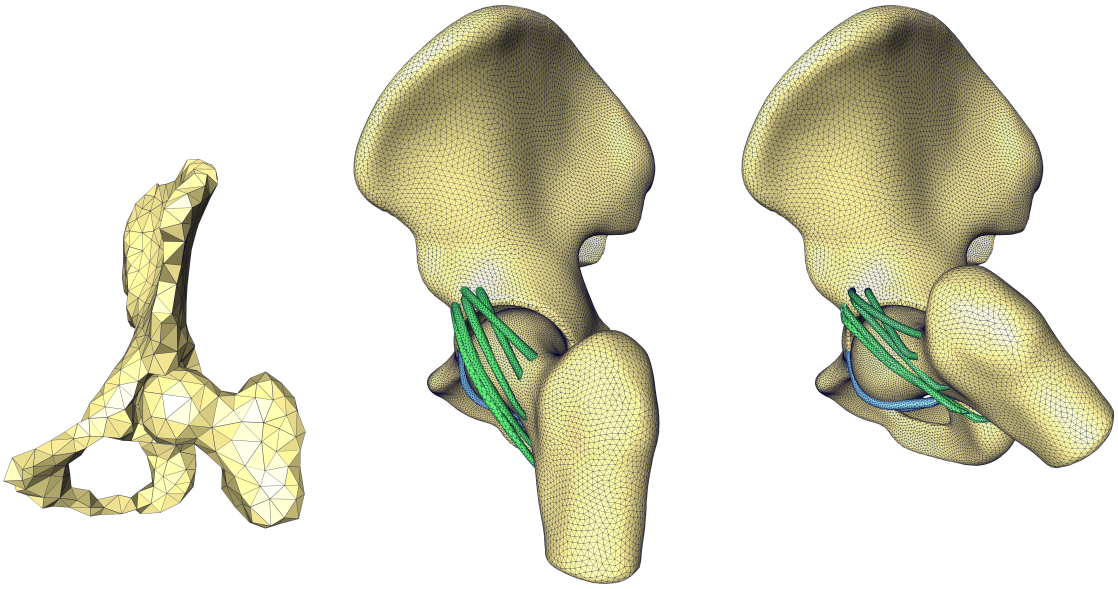


Figure 5.15.: *Left: Coarse bone grids. Middle: Fine initial joint position with remeshed ligaments. Right: Joint in the final extended position.*

for the bones and ligaments have been taken from the literature [33] and [135]. As

parameter	bone	ligament
Young's modulus	17 GPa	330 MPa
Poisson ratio	0.3	0.3
mass density	$2\,\mathrm{g/cm^3}$	$1\,\mathrm{g/cm^3}$

Table 5.2.: Material parameters for bone and ligaments.

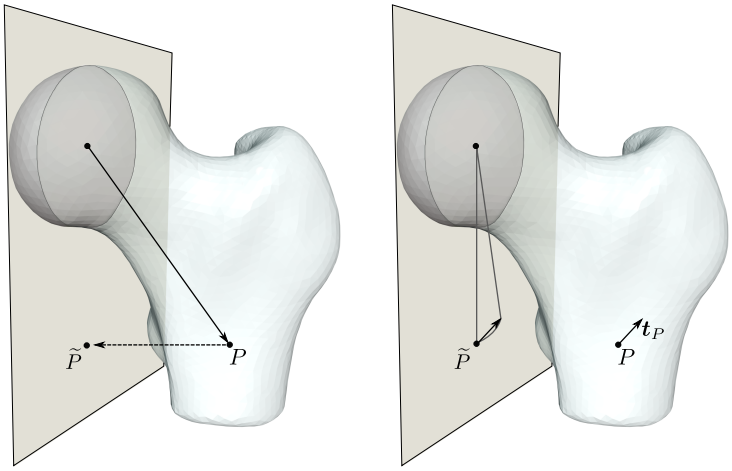


Figure 5.16.: *Construction of the Neumann traction forces. First each node P of the Neumann boundary is projected onto the plane of motion. Then the distance vector of the projected node $\widetilde{P}$ to the center is rotated about a fixed amount. The resulting displacement is used to determine the direction of the traction force $\mathbf{t}_P$ and scaled suitably.*

described in the previous Section 5.2 the pelvis grid is supplemented with homogeneous Dirichlet boundary conditions. At the lower end of the femur time-dependent rotational Neumann forces are prescribed as described in Figure 5.16. In this simulation a total number of 30 time steps with step size $\tau = 0.01$ were computed. The solution of the local problems took about $30s$ for an inexact one and $160s$ for the exact problem. Compared to the computation times obtained in the previous numerical example in Section 4.4.2, the multigrid convergence deteriorated a bit. This behaviour was observed before whenever parametrised boundaries are used [117]. In this case the refined approximation of the grid boundaries does not yield nested grid hierarchies in general.
In Figure 5.17 the configuration at the time of the first contact between the pelvis and the femur is compared to the analogous position of a rigid body rotation with fixed joint centre. Our model predicted a slightly different extension angle, but more importantly, a displacement of the joint centre about 2mm was measured, see Figure 5.18.

Concluding remarks

The model proposed in this thesis is by no means complete. Many unknowns like the patient-specific cartilage layers or ligament positions and initial strains make an exact prediction of the impingement zones very difficult. However, in contrast to prevailing approaches, a proper stability analysis of the centre under loads can be performed and even extreme movements like hyperextensions can be tested to predict possible dislocation of hip joint implants. We conclude that the proposed FAI analysis has the potential to improve present virtual medicine in pre-operational hip surgery planning.

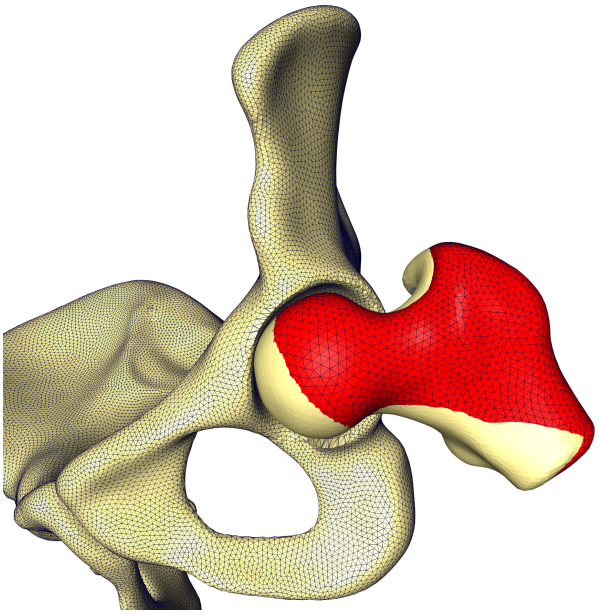


Figure 5.17.: *Left: Comparision of our finite element FAI and the pure rigid body approach. The red grid corresponds to the rotation with fixed rotational centre; the yellow shade to the flexible centre.*

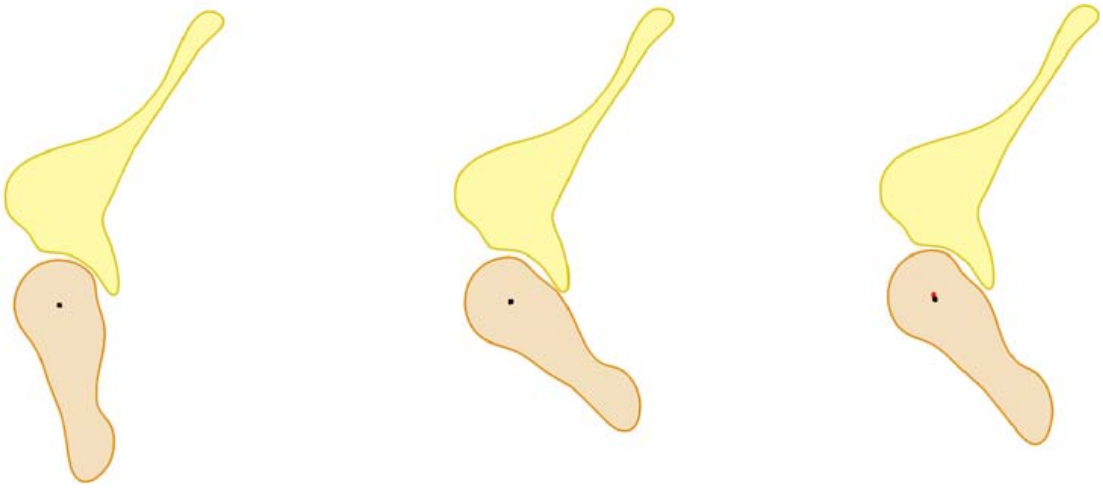


Figure 5.18.: *Range of motion analysis in the sagittal plane. The black dot denotes the initial centre and the red one the moving centre of femoral head. From left to right: a) Initial position. b) Extension position with fixed rotational centre. c) Extension position for flexible centre.*

A. Appendix

A. Introduction to Constrained Optimisation

In this section we will give a short introduction to the theory of non-linear constrained optimisation. First we start with a very general problem and repeat the first order optimality conditions and then construct a criticality measure for bound constrained problems to monitor the convergence of several descent algorithms. Let us consider the general non-linear constrained minimization problem

$$\min_{\boldsymbol{\varphi}\in\mathcal{K}} \mathcal{J}(\boldsymbol{\varphi}), \tag{P}$$

with

$$\mathcal{K} := \left\{\boldsymbol{\varphi} \in \mathbb{R}^N \,:\, g(\boldsymbol{\varphi}) \geq 0\right\}. \tag{A.1}$$

The functions $g : \mathbb{R}^N \longrightarrow \mathbb{R}^m$, $\mathcal{J} : \mathbb{R}^N \longrightarrow \mathbb{R}$ are assumed to be twice continuously differentiable on an open neighbourhood of the feasible set and can be non-convex. By adding the indicator function of (A.1), the problem (P) can be reformulated as a non-smooth unconstrained minimisation problem

$$\min_{\boldsymbol{\varphi}\in\mathbb{R}^N} \mathcal{J}(\boldsymbol{\varphi}) + I_{\mathcal{K}}(\boldsymbol{\varphi}). \tag{A.2}$$

Analogously to Theorem 1.4.8 it can be shown that for critical points $\boldsymbol{\varphi}^*$ of (A.2) it holds

$$0 \in \nabla\mathcal{J}(\boldsymbol{\varphi}^*) + \partial^F I_{\mathcal{K}}(\boldsymbol{\varphi}^*). \tag{A.3}$$

To derive the classical Karush-Kuhn-Tucker (KKT) optimality conditions one has to make a further assumption on the constraint Jacobian to exclude degenerate cases. Therefore, let the set of *active* constraints be denoted by

$$\mathcal{A}(\boldsymbol{\varphi}) := \{i \in \mathbb{N} \,:\, g_i(\boldsymbol{\varphi}) = 0\}. \tag{A.4}$$

Definition A.1 (Linear Independence Constraint Qualification)**.** Let $\boldsymbol{\varphi} \in \mathcal{K}$, then *(LICQ)* holds at $\boldsymbol{\varphi}$ if

$$\nabla g_i(\boldsymbol{\varphi}), \quad i \in \mathcal{A} \text{ are linear independent.} \tag{LICQ}$$

Remark A.2. *The condition* (LICQ) *can be weakened further to derive the (KKT) conditions [121, Chapter 12]. We will restrict ourselves to this assumption because it is equivalent to the uniqueness of the Lagrange multipliers [31].*

When this constraint qualification holds, the Fréchet subdifferential $\partial^F I_{\mathcal{K}}(\boldsymbol{\varphi})$ can be characterised as follows:

Proposition A.3. *Let $\boldsymbol{\varphi} \in \mathcal{K}$ and let* (LICQ) *hold at $\boldsymbol{\varphi}$. Then*

$$\partial^F I_{\mathcal{K}}(\boldsymbol{\varphi}) = \left\{ \sum_{i \in \mathcal{A}} \lambda_i \nabla g_i(\boldsymbol{\varphi}) \,:\, \lambda_i \in \mathbb{R},\, \lambda_i \leq 0 \right\}. \tag{A.5}$$

Proof. [121, Corollary 11.6.2] □

We are now ready to state the classical first order optimality conditions.

Theorem A.4 (Karush-Kuhn-Tucker Optimality Conditions)**.** *Let $\boldsymbol{\varphi}^* \in \mathcal{K}$ be a locally critical point of* (P) *and assume that* (LICQ) *holds at $\boldsymbol{\varphi}^*$. Then there exists a unique Lagrange multiplier $\lambda \in \mathbb{R}^m$, such that*

$$\begin{aligned} &\nabla \mathcal{J}(\boldsymbol{\varphi}^*) + \lambda^T \nabla g(\boldsymbol{\varphi}^*) = 0, \\ &g(\boldsymbol{\varphi}^*) \geq 0, \quad g(\boldsymbol{\varphi}^*)\lambda = 0, \quad \lambda \leq 0. \end{aligned} \tag{KKT}$$

Proof. Follows from Proposition A.3 and extending the multiplier by zero for all inactive components. The uniqueness of λ is proven in [31]. □

A.1. Criticality measures

In this section, following [30, Chapter 12], we will introduce a measure for the first order criticality of an iterate to monitor the convergence of an algorithm.

Definition A.5 (Criticality Measure)**.** A function $\chi : \mathbb{R}^N \longrightarrow \mathbb{R}$ is called *criticality measure* of (P) if

- χ is non-negative and continuous.
- $\chi(\boldsymbol{\varphi}) = 0 \quad \Longleftrightarrow \boldsymbol{\varphi}$ is a local first order critical point.

We now want to construct a computable criticality measure for the case of convex constraints. This measure will be used to identify solutions of the local quadratic problems within SQP methods. In the global method convergence of the sequence

towards critical points of (P) can be shown using this local criticality measure and perturbation theory, see e.g. [43, Theorem 2.2.6]. Let us assume for the rest of this section:

Assumption A. The constraints g_i are convex.

We denote the closest point projection onto the convex set $\mathcal{K}$ by $P_{\mathcal{K}} : \mathbb{R}^N \longrightarrow \mathcal{K}$. Note that for convex sets the projection $P_{\mathcal{K}}(\boldsymbol{\varphi})$ is unique, see e.g.[136]. For the case of box-constraints it can be computed explicitly

$$\mathcal{K} = \prod_{i=1}^{N} [l_i, u_i] \Longrightarrow P_{\mathcal{K}}(\boldsymbol{\varphi}) = \begin{cases} \varphi_i & l_i \leq \varphi_i \leq u_i, \\ l_i & \varphi_i < l_i, \\ u_i & u_i < \varphi_i. \end{cases} \tag{A.6}$$

The criticality measure we will construct is based on the *projected gradient path*.

Definition A.6. The path $p : \mathbb{R}^N \times [0, \infty) \longrightarrow \mathbb{R}^N$

$$p(\boldsymbol{\varphi}, t) := P_{\mathcal{K}} \left(\boldsymbol{\varphi} - t \nabla \mathcal{J}(\boldsymbol{\varphi}) \right), \tag{A.7}$$

is called *projected gradient path*.

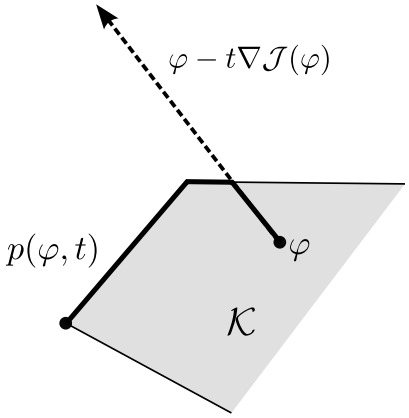


Figure A.1.: *Example of a projected gradient path.*

Similar to the gradient being zero in the unconstrained case, the projected gradient path can be used to characterise local critical points in the constrained case.

Proposition A.7. *Let Assumption A hold. Then, $\boldsymbol{\varphi}^*$ is a locally critical point of Problem* (P) *if and only if*

$$p(\boldsymbol{\varphi}^*, t) = \boldsymbol{\varphi}^* \quad \forall\, t \geq 0. \tag{A.8}$$

Proof. The proof is based on the property that if for some $t^* \geq 0$

$$0 \in \nabla J(\boldsymbol{\varphi}) + \partial I_{\mathcal{K}}\left(p(\boldsymbol{\varphi}, t^*)\right),$$

then it holds

$$p(\boldsymbol{\varphi}, t^*) = p(\boldsymbol{\varphi}, t) \quad \forall\, t \geq t^*.$$

See [30, Theorem 12.1.2] □

Proposition A.8. *Let Assumption A hold. Then the function*

$$\chi(\boldsymbol{\varphi}) := \|p(\boldsymbol{\varphi}, 1) - \boldsymbol{\varphi}\|, \tag{A.9}$$

is a criticality measure of (P).

Proof. [30, Theorem 12.1.6]

Remark A.9. *Consider the local trust–region problem* (3.32) *with* $m_k, \Delta_k, \boldsymbol{\varphi}_k \in \mathcal{K}$ *given. Then the first local minimum* $t^* \geq 0$ *on the corresponding projected gradient path* $p(\boldsymbol{\varphi}_k, \cdot)$ *fulfils*

$$m_k(0) - m_k(p(\boldsymbol{\varphi}_k, t^*) - \boldsymbol{\varphi}_k) \geq \kappa\, \chi(\boldsymbol{\varphi}_k)\, \min\left\{ \frac{\chi(\boldsymbol{\varphi}_k)}{\|\nabla^2 m_k(0)\|}, \Delta_k \right\}, \tag{A.10}$$

for some constant $\kappa > 0$, *see e.g. [30, Theorem 12.2.2].*

Remark A.10. *An equivalent definition of the criticality measure* χ *as the solution of a convex linear minimisation problem is given by*

$$\chi(\boldsymbol{\varphi}) = |\min_{\substack{g(\boldsymbol{\varphi}+d)\geq 0 \\ \|d\|=1}} \langle \nabla \mathcal{J}(\boldsymbol{\varphi}), d\rangle|. \tag{A.11}$$

See e.g. [30, Theorem 12.1.6].

B. Detailed steps towards convergence of the filter–trust–region method.

In this section we assume the following:

Assumption F1. The iterates φ_k of Algorithm FTR stay in a compact set $\mathcal{C} \supseteq \mathcal{K}$.

Assumption F2. The energy functional $\mathcal{J}$ and the constraints g are twice continuously differentiable on $\mathcal{C}$.

Assumption F3. The inexact Jacobians $\widetilde{G}(\cdot)$ are continuous on $\mathcal{C}$ and induce a continuous invertible transformation T. The transformed inexact Jacobians are positive, diagonal and the inverse is bounded away from zero

$$\|(\widetilde{G}(\varphi)T(\varphi)^T)^{-1}\| \leq \kappa_T. \tag{B.12}$$

Assumption F4. For a compatible sub-problem the tangential step t_k fulfils

$$m(n_k + t_k) - m(0) \geq \kappa_{\text{sc}}\tilde{\chi}_k \min\{\tilde{\chi}_k, \Delta_k\}.$$

With these assumptions and Lemmata 3.4.9 and 3.4.10 the single steps towards the main convergence theorem can be carried out identically to the proof of the original filter method [47]. In the following we will state the needed lemmata and give a short description of how to proof them. First one can show that the model decrease of the full step can be bounded from below by the trust-region if the radius is sufficiently small and the inexact critically measure is not zero.

Lemma B.1. *Assume that TRQP*(φ_k, Δ_k) *is compatible, that*

$$\widetilde{\chi}_k \geq \varepsilon > 0, \tag{B.13}$$

and that

$$\Delta_k \leq \min\left\{\frac{\varepsilon}{\kappa_H}, 2\left(\frac{\kappa_f}{\kappa_H\kappa_\Delta\kappa_\mu}\right)^{\frac{1}{1+\mu}}, \left(\frac{\kappa_{\text{sc}}\varepsilon}{4\kappa_f\kappa_\Delta\kappa_\mu}\right)^{1+\mu}\right\} =: \delta_m.$$

Then

$$m_k(0) - m_k(u_k) \geq \frac{1}{2}\kappa_{\text{sc}}\varepsilon\Delta_k.$$

Proof. The proof consists of using the Cauchy decrease F4 and bounding the model decrease of the normal step using Assumption F1 and the compatibility bound 3.48, see [47, Lemma 3.5]. □

By restricting the trust–region further the local model becomes a better approximation of the non-linear functional and very successful iterations can be guaranteed

Lemma B.2. *Assume that TRQP*(φ_k, Δ_k) *is compatible, that*

$$\widetilde{\chi}_k \geq \varepsilon > 0,$$

and that

$$\Delta_k \leq \min\left\{\delta_m, \frac{(1-\eta_2)\kappa_{\text{sc}}\varepsilon}{2\kappa_m}\right\} =: \delta_\rho.$$

Then

$$\rho_k \leq \eta_2.$$

Proof. Use Lemma B.1 and the fact that

$$|\mathcal{J}(\varphi_k + u_k) - m_k(u_k)| \leq \kappa_m \Delta_k^2,$$

[47, Lemma 3.6]. □

Next it can be shown that for a small trust-region the *ϑ-type* condition holds

Lemma B.3. *Assume that TRQP*(φ_k, Δ_k) *is compatible, that*

$$\widetilde{\chi}_k \geq \varepsilon > 0,$$

and that

$$\Delta_k \leq \min\left\{\delta_m, \left(\frac{\kappa_{\mathrm{sc}}\varepsilon}{2\kappa_\vartheta \kappa_{\mathrm{ui}}^2}\right)^{\frac{1}{1+2\mu}}\right\} =: \delta_f.$$

Then

$$m_k(0) - m_k(u_k) \geq \kappa_\vartheta \vartheta_k^2.$$

Proof. Follows from the compatibility 3.48 and Lemmata 3.4.10 and B.1 □

If the infeasibility is small one can further deduce that enough energy decrease is generated so that the iterate is acceptable to the filter.

Lemma B.4. *Assume that TRQP*(φ_k, Δ_k) *is compatible, that*

$$\widetilde{\chi}_k \geq \varepsilon > 0,$$

and that

$$\vartheta_k \leq \kappa_{\mathrm{ui}}^{-\frac{1}{\mu}} \left(\frac{\eta_2 \kappa_{\mathrm{sc}}\varepsilon}{2\gamma_\vartheta}\right)^{\frac{1+\mu}{\mu}} =: \delta_\vartheta.$$

Then

$$\mathcal{J}(\varphi_k + u_k) \leq \mathcal{J}(\varphi_k) - \gamma_\vartheta \vartheta_k.$$

Proof. Lemmata 3.4.10, B.1 and B.2 □

The compatibility condition of the local problem TRQP can be proven if the infeasibility is bounded by the trust-region radius. To conclude that this holds when both the radius and the infeasibility are small is the result of the next lemma.

Lemma B.5. *Assume that*

$$\widetilde{\chi}_k \geq \varepsilon > 0,$$

that $k > 0$ *with*

$$\vartheta_k \leq \delta_\vartheta,$$

and

$$\Delta_k \leq \min \left\{ \gamma_{\frac{1}{2}} \delta_\rho, \left(\frac{\gamma_{\frac{1}{2}}^2 (1 - \gamma_\vartheta) \kappa_\Delta \kappa_\mu}{\kappa_{\text{ui}} \overline{\kappa}_n} \right)^{\frac{1}{1-\mu}} \right\} =: \delta_R.$$

Then $TRQP(\varphi_k, \Delta_k)$ *is compatible.*

Proof. Proof by contradiction, using the previous lemmata, see [47, Lemma 3.9]. □

With these tools at hand it can be concluded that if during the filter method an infinite number of iterates is added to the filter, then there exist a convergent subsequence. To this end we define

$$\mathcal{Z} := \{k \in \mathbb{N} | \, \varphi_k \text{ is added to the filter}\}.$$

Lemma B.6. *Assume that the feasibility restoration phase* (FPR) *always terminates successfully and let* $(\varphi_k)_k$ *be a sequence generated Algorithm FTR such that* $|\mathcal{Z}| = \infty$*. Then there exists a subsequence* $(k_l)_l \subseteq \mathcal{Z}$ *such that*

$$\lim_{l \longrightarrow \infty} \vartheta_{k_l} = 0,$$

and

$$\lim_{l \longrightarrow \infty} \widetilde{\chi}_{k_l} = 0.$$

Proof. This lemma is again proven by contradiction. Assume that the criticality measure is bounded away from zero

$$\widetilde{\chi}_{k_l} \geq \varepsilon_1 > 0,$$

and that in this case the trust-region is not converging to zero

$$\Delta_{k_l} \geq \varepsilon_2 > 0.$$

The feasibility of the sub-sequence follows immediately from Lemma 3.4.8. See [47, Lemma 3.10] for more details. □

Finally we prove Lemma B.6 for the case when only finitely many iterates were added to the filter.

Lemma B.7. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR such that $|\mathcal{Z}| < \infty$, then*

$$\lim_{k\longrightarrow\infty} \vartheta_k = 0.$$

Proof. From $|\mathcal{Z}| < \infty$ it follows that for k large enough condition (3.59)

$$m_k(0) - m_k(u_k) \geq \kappa_\vartheta \vartheta_k^2 \geq 0,$$

never fails. Thus using Assumptions F1 and F2 and the mechanism of the algorithm yields the result, [47, Lemma 3.11]. □

With this lemma at hand one can deduce that for non-critical points the trust-region radius cannot become arbitrarily small.

Lemma B.8. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR such that $|\mathcal{Z}| < \infty$, and*

$$\widetilde{\chi}_k \geq \varepsilon > 0,$$

for all k large enough. Then there exists $\Delta_{\min} > 0$ such that

$$\Delta_k \geq \Delta_{\min},$$

for all $k \in \mathbb{N}$.

Proof. Proof by contradiction using the mechanism of the algorithm, see [47, Lemma 3.12]. □

This then yields the desired result for the case of a finite filter.

Lemma B.9. *Assume that the feasibility restoration phase (FPR) always terminates successfully and let $(\varphi_k)_{k\in\mathbb{N}}$ be a sequence generated by Algorithm FTR such that $|\mathcal{Z}| < \infty$. Then*

$$\liminf_{k\longrightarrow\infty} \widetilde{\chi}_k = 0.$$

Proof. Proof by contradiction, assuming that

$$\widetilde{\chi}_k \geq \varepsilon > 0,$$

and using the previous lemmata, see [47, Lemma 3.13]. □

Bibliography

[1] C. L. Abraham et al. "A new discrete element analysis method for predicting hip joint contact stresses". In: *Journal of biomechanics* 46.6 (2013), pp. 1121–1127.

[2] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[3] R. Adams and J. Fournier. *Sobolev spaces*. Vol. 140. Academic press, 2003.

[4] C. Aliprantis and K. Border. *Infinite dimensional analysis*. Springer, 2006. ISBN: 978-3-540-29587-7.

[5] S. Antman. *Nonlinear Problems of Elasticity*. Springer, 1995.

[6] M. Armand et al. "Computer-aided orthopedic surgery with near-real-time biomechanical feedback". In: *Johns Hopkins APL technical digest* 25.3 (2004), pp. 242–252.

[7] F. Armero and E. Petőcz. "Formulation and analysis of conserving algorithms for frictionless dynamic contact/impact problems". In: *Comp. Meth. Appl. Mech. and Eng.* 158.3 (1998), pp. 269–300. DOI: 10.1016/S0045-7825(97)00256-9.

[8] G. Aurand. "Die sphärische Pressfit-Pfanne Plasmacup als Revisionsimplantat beim Hüftendoprothesenwechsel. Mittelfristige klinische und radiologische Ergebnisse." PhD thesis. Philipps-Universität Marburg Medizin, 2008.

[9] P. Bastian, G. Buse, and O. Sander. "Infrastructure for the coupling of Dune grids". In: *Numerical Mathematics and Advanced Applications 2009*. Springer, 2010, pp. 107–114.

[10] P. Bastian et al. "A Generic Grid Interface for Parallel and Adaptive Scientific Computing. Part I: Abstract Framework". In: *Computing* 82.2–3 (2008), pp. 103–119. DOI: http://www.springerlink.com/content/4v77662363u41534/.

[11] P. Bastian et al. "A Generic Grid Interface for Parallel and Adaptive Scientific Computing. Part II: Implementation and Tests in DUNE". In: *Computing* 82.2–3 (2008), pp. 121–138. DOI: http://www.springerlink.com/content/gn177r643q2168g7/.

[12] P. Bastian et al. *DUNE Web page*. http://www.dune-project.org. 2011.

[13] P. Bastian et al. "UG–a flexible software toolbox for solving partial differential equations". In: *Computing and Visualization in Science* 1.1 (1997), pp. 27–40.

[14] K. Bathe and A. Chaudhary. "A solution method for planar and axisymmetric contact problems". In: *Int. J. Num. Meth. in Eng.* 21.1 (1985), pp. 65–88. DOI: 10.1002/nme.1620210107.

[15] M. Beck et al. "Hip morphology influences the pattern of damage to the acetabular cartilage femoroacetabular impingement as a cause of early osteoarthritis of the hip". In: *Journal of Bone & Joint Surgery, British Volume* 87.7 (2005), pp. 1012–1018.

[16] A. Bedi et al. "Surgical treatment of femoroacetabular impingement improves hip kinematics a computer-assisted model". In: *The American journal of sports medicine* 39.1 suppl (2011), 43S–49S.

[17] F. B. Belgacem, P. Hild, and P. Laborde. "Approximation of the unilateral contact problem by the mortar finite element method". In: *Comptes Rendus de l'Academie des Sciences Series I Mathematics* 324.1 (1997), pp. 123–127. DOI: 10.1016/S0764-4442(97)80024-9.

[18] T. Belytschko and D. F. Schoeberle. "On the Unconditional Stability of an Implicit Algorithm for Nonlinear Structural Dynamics". In: *J. of Appl. Mech.* 42.4 (1975), pp. 865–869. DOI: 10.1115/1.3423721.

[19] C. Bernardi, Y. Maday, and A.T. Patera. "A new nonconforming approach to domain decomposition: the mortar element method". In: *Nonliner Partial Differential Equations and Their Applications* 9 (1994), pp. 13–51. URL: http://ci.nii.ac.jp/naid/10018481740/en/.

[20] P. Betsch and C. Hesch. "Energy-momentum conserving schemes for frictionless dynamic contact problems". In: *IUTAM symposium on computational methods in contact mechanics*. Springer. 2007, pp. 77–96.

[21] D. Braess. *Finite elements: Theory, fast solvers, and applications in solid mechanics*. Cambridge University Press, 2007.

[22] A. Brandt. "Rigorous Quantitative Analysis of Multigrid, I. Constant Coefficients Two-Level Cycle with L_2-Norm". In: *SIAM Journal on Numerical Analysis* 31.6 (1994), pp. 1695–1730.

[23] R. H. Byrd, M. E. Hribar, and J. Nocedal. "An interior point algorithm for large-scale nonlinear programming". In: *SIAM Journal on Optimization* 9.4 (1999), pp. 877–900.

[24] T. Chang et al. "A pre-operative approach of range of motion simulation and verification for femoroacetabular impingement". In: *The International Journal of Medical Robotics and Computer Assisted Surgery* 7.3 (2011), pp. 318–326.

[25] G. Chen et al. "A three-dimensional finite element model for biomechanical analysis of the hip". In: *Cell biochemistry and biophysics* 67.2 (2013), pp. 803–808.

[26] P. Ciarlet. *Mathematial Elasticity: Volume I Three-Dimensional Elasticity*. Elsevier, 1988.

[27] F. Clarke. *Optimization and Nonsmooth Analysis*. Vol. 5. Siam, 1990. DOI: 10.1137/1.9781611971309.

[28] T. F. Coleman and Y. Li. "On the convergence of interior-reflective Newton methods for nonlinear minimization subject to bounds". In: *Mathematical programming* 67.1-3 (1994), pp. 189–224.

[29] A. R. Conn, N. I. M. Gould, and P. L. Toint. "Global convergence of a class of trust region algorithms for optimization with simple bounds". In: *J. of Appl. Mech.* MPS-SIAM Series on Optimization 25.2 (1988). DOI: 10.1137/0725029.

[30] A. R. Conn, N. I. M. Gould, and P. L. Toint. *Trust-Region Methods*. MPS-SIAM Series on Optimization. Siam, 2000. DOI: 10.1137/1.9780898719857.

[31] A. R. Conn and J. W. Tolle. "A necessary and sufficient qualification for constrained optimization". In: *J. on Appl. Math.* 20.2 (1971), pp. 164–172. DOI: 10.1137/0120021.

[32] A. R. Conn et al. "Global convergence of a class of trust region algorithms for optimization using inexact projections on convex constraints". In: *SIAM Journal on Optimization* 3.1 (1993), pp. 164–221.

[33] J. D. Currey. *Bones: structure and mechanics*. Princeton University Press, 2002.

[34] P. Deuflhard, R. Krause, and S. Ertel. "A contact-stabilized Newmark method for dynamical contact problems". In: *Int. J. Numer. Methods Eng.* 73 (2008), pp. 1274–1290. DOI: 10.1002/nme.2119.

[35] P. Deuflhard and M. Weiser. *Adaptive numerical solution of PDEs*. Walter de Gruyter, 2012.

[36] C. Eck, J. Jarusek, and M. Krbec. *Unilateral contact problems: variational methods and existence theorems*. Vol. 270. CRC Press, 2005.

[37] M. Ehlke et al. "Towards Robust Measurement of Pelvic Parameters from AP Radiographs using Articulated 3D Models". In: *Int. J. Comp. Ass. Rad. and Surg.* (2015).

[38] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. Siam, 1987. DOI: 10.1137/1020024.

[39] J. Elstrodt. *Maß-und Integrationstheorie*. Springer-Verlag, 2006. ISBN: 3540267034.

[40] R. P. Fedorenko. "A relaxation method for solving elliptic difference equations". In: *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki* 1.5 (1961), pp. 922–927.

[41] FEI. *Amira Web page*. http://www.amira.com.

[42] A. V. Fiacco and G. P. McCormick. *Nonlinear programming: sequential unconstrained minimization techniques*. Vol. 4. Siam, 1990.

[43] A.V. Fiacco. *Introduction to Sensitivity and Stability Analysis in Nonlinear Programming*. Mathematics in science and engineering. Academic Press, 1983. ISBN: 9780122544507.

[44] K. Fischer and P. Wriggers. "Frictionless 2D Contact formulations for finite deformations based on the mortar method". In: *Comp. Mech.* 36 (2005), pp. 226–244. DOI: 10.1007/s00466-005-0660-y.

[45] R. Fletcher and S. Leyffer. "Nonlinear programming without a penalty function." In: *Math. Programming* 91 (2002), pp. 239–269. DOI: 10.1007/s101070100244.

[46] R. Fletcher, S. Leyffer, and P. Toint. "On the Global Convergence of a Filter–SQP Algorithm". In: *SIAM Journal on Optimization* 13.1 (2002), pp. 44–59. DOI: 10.1137/S105262340038081X.

[47] R. Fletcher et al. "Global Convergence of a Trust-Region SQP-Filter Algorithm for general nonlinear programming". In: *SIAM Journal on Optimization* 13.3 (2002), pp. 635–659. DOI: 10.1137/S1052623499357258.

[48] G. B. Folland. *Real analysis: modern techniques and their applications*. John Wiley & Sons, 2013. ISBN: 1118626397.

[49] M. Gander and C. Japhet. "An algorithm for non-matching grid projections with linear complexity". In: *Domain Decomposition Methods in Science and Engineering XVIII*. Springer, 2009, pp. 185–192.

[50] P. E. Gill, W. Murray, and M. A. Saunders. "SNOPT: An SQP algorithm for large-scale constrained optimization". In: *SIAM journal on optimization* 12.4 (2002), pp. 979–1006.

[51] O. Gonzalez. "Exact energy and momentum conserving algorithms for general models in nonlinear elasticity". In: *Comp. Meth. Appl. Mech. Eng.* 190 (1999), pp. 1763–1783. DOI: 10.1016/S0045-7825(00)00189-4.

[52] M.S. Bazaraa J. Goode and M. Z. Nashed. "On the cones of tangents with applications to mathematical programming". In: *Journal of Optimization Theory and Applications* 13.4 (1974), pp. 389–426.

[53] C. Gräser. "Convex minimization and phase field models". PhD thesis. Freie Universität Berlin, 2011.

[54] C. Gräser and R. Kornhuber. "Multigrid methods for obstacle problems". In: *Journal of Computational Mathematics* 27.1 (2009), pp. 1–44.

[55] C. Gräser, U. Sack, and O. Sander. "Truncated Nonsmooth Newton Multigrid Methods for Convex Minimization Problems". In: *Proc. of DD18*. 2009, pp. 129–136. DOI: 10.1007/978-3-642-02677-5.

[56] S. Gratton, A. Sartenaer, and P. L. Toint. "Recursive trust-region methods for multiscale nonlinear optimization". In: *SIAM Journal on Optimization* 19.1 (2008), pp. 414–444.

[57] P. Grohs, H. Hardering, and O. Sander. "Optimal a priori discretization error bounds for geodesic finite elements". In: *Foundations of Computational Mathematics* (2013), pp. 1–55.

[58] C. Groß and R. Krause. "A recursive trust-region method for non-convex constrained minimization". In: *Lecture Notes in Computational Science and Engineering* 70 LNCSE.409 (2009), pp. 137–144. ISSN: 14397358. DOI: 10.1007/978-3-642-02677-5_13.

[59] C. Gräser and O. Sander. "Polyhedral Gauss-Seidel Converges". In: *J. Numer. Math* 22.3 (2011), pp. 221–254.

[60] W. Hackbusch. "A fast iterative method for solving poisson's equation in a general region". In: *Numerical treatment of differential equations*. Springer, 1978, pp. 51–62.

[61] W. Hackbusch. *Theorie und Numerik elliptischer Differentialgleichungen*. Springer-Verlag, 2013.

[62] S. Hartmann and E. Ramm. "A mortar based contact formulation for non-linear dynamics using dual Lagrange multipliers". In: *Finite Elements in Analysis and Design* 44 (2008), pp. 245–258. DOI: 10.1016/j.finel.2007.11.018.

[63] P. Hauret and P. Le Tallec. "Energy-controlling time integration methods for nonlinear elastodynamics and low-velocity impact". In: *Computer methods in applied mechanics and engineering* 195.37 (2006), pp. 4890–4916.

[64] Q. He, J. Telega, and A. Curnier. "Unilateral contact of two solids subject to large deformations: formulation and existence results". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*. Vol. 452. 1955. The Royal Society. 1996, pp. 2691–2717. DOI: 10.1098/rspa.1996.0143.

[65] C. Hesch. "Mechanische Integratoren für Kontaktvorgänge deformierbarer Körper unter großen Verzerrungen". PhD thesis. Universität Siegen, 2007. URL: http://dokumentix.ub.uni-siegen.de/opus/volltexte/2008/315/.

[66] C. Hesch and P. Betsch. "A mortar method for energy–momentum conserving schemes in frictionless dynamic contact problems". In: *Int. J. Numer. Methods Eng.* 77 (2009), pp. 1468–1500. DOI: 10.1002/nme.2466.

[67] Q. Hu et al. "A fast impingement detection algorithm for computer-aided orthopedic surgery". In: *Computer Aided Surgery* 6.2 (2001), pp. 104–110.

[68] S. Hüeber and B. Wohlmuth. "A primal–dual active set strategy for non-linear multibody contact problems". In: *Computer Methods in Applied Mechanics and Engineering* 194.27 (2005), pp. 3147–3166.

[69] T. Hughes et al. "A finite element method for a class of contact-impact problems". In: *Comp. Meth. Appl. Mech. and Eng.* 8.3 (1976), pp. 249–276. DOI: 10.1016/0045-7825(76)90018-9.

[70] C. Kane et al. "Finite element analysis of nonsmooth contact". In: *Comput. Methods Appl. Mech. Engrg.* 180 (1999), pp. 1–26. DOI: 10.1016/S0045-7825(99)00034-1.

[71] D. M. Kaufman and D. K. Pah. "Geometric numerical integration of inequality constrained, nonsmooth Hamiltonian systems". In: *Siam* (2012). DOI: 10.1137/100800105.

[72] C. Klapproth. "Adaptive numerical integration for dynamical contact problems". PhD thesis. 2012".

[73] R. Kornhuber. *Adaptive monotone multigrid methods for nonlinear variational problems*. BG Teubner, 1997.

[74] R. Kornhuber. "Monotone multigrid methods for elliptic variational inequalities I". In: *Numerische Mathematik* 69.2 (1994), pp. 167–184. ISSN: 0029-599X.

[75] R. Kornhuber and R. Krause. "Adaptive multigrid methods for Signorini's problem in linear elasticity". In: *Computing and Visualization in Science* 4.1 (2001), pp. 9–20. DOI: 10.1007/s007910100052.

[76] Y. Kosashvili et al. "Digital versus conventional templating techniques in preoperative planning for total hip arthroplasty". In: *Canadian Journal of Surgery* 52.1 (2009), p. 6.

[77] R. Krause and O. Sander. "Automatic Construction of Boundary Parametrizations for Geometric Multigrid Solvers". In: *Comp. Vis. Sci.* 9 (2006), pp. 11–22.

[78] R. Krause and M. Walloth. "Presentation and comparison of selected algorithms for dynamic contact based on the Newmark scheme". In: *Applied Numerical Mathematics* 62.10 (2012), pp. 1393–1410.

[79] R. Krause and B. Wohlmuth. "Monotone Methods on Nonmatching Grids for Nonlinear Contact Problems". In: *SIAM J. Sci. Comput.* 25.1 (2003), pp. 324–347. DOI: 10.1137/S1064827502405318.

[80] R. Krause et al. "Bone remodelling: A combined biomechanical and systems-biological challenge". In: *PAMM* 11.1 (2011), pp. 99–100.

[81] P. R. Krekel et al. "Interactive simulation and comparative visualisation of the bone-determined range of motion of the human shoulder." In: *SimVis*. 2006, pp. 275–288.

[82] A. Kruger. "On Fréchet subdifferentials". In: *Journal of Mathematical Sciences* 116.3 (2003), pp. 3325–3358.

[83] M. Kubiak-Langer et al. "Range of motion in anterior femoroacetabular impingement." In: *Clinical orthopaedics and related research* 458 (2007), pp. 117–124.

[84] D. Kuhl and M. A. Crisfield. "Energy–Conserving and Decaying Algorithms in Non-linear Structural Dynamics". In: *Int. J. Numer. Methods Eng.* 45 (1999), pp. 569–599. DOI: 10.1002/(SICI)1097-0207(19990620)45:5<569::AID-NME595>3.0.CO;2-A.

[85] H. Lamecker. "Variational and statistical shape modeling for 3D geometry reconstruction". PhD thesis. 2008. URL: http://opus4.kobv.de/opus4-zib/frontdoor/deliver/index/docId/1110/file/DissLamecker.pdf.

[86] R. Lattanzi et al. "Applications of 3d medical imaging in orthopaedic surgery: Introducing the hip-op system". In: *First International Symposium on 3D Data Processing Visualization and Transmission*. IEEE. 2002, p. 808.

[87] T. A. Laursen. *Computational Contact and Impact Mechanics*. Springer, 2003. DOI: 10.1007/978-3-211-77298-0.

[88] T. A. Laursen and V. Chawla. "Design of Energy Conserving Algorithms for Frictionless Dynamic Contact Problems". In: *Int. J. Numer. Methods Eng.* 40.5 (1997), pp. 863–886. DOI: 10.1002/(SICI)1097-0207(19970315)40:5<863::AID-NME92>3.0.CO;2-V.

[89] T. A. Laursen and G. R. Love. "Improved implicit integrators for transient impact problems–geometric admissibility within the conserving framework". In: *Int. J. Numer. Methods Eng.* 53 (2002), pp. 245–274. DOI: 10.1002/nme.264.

[90] T. A. Laursen and J. C. Simo. "A Continuum-based finite element formulation for the implicit solution of multibody, large deformation frictional contact problems". In: *Int. J. Num. Meth. in Eng.* 36.20 (1993), pp. 3451–3485. DOI: 10.1002/nme.1620362005.

[91] E. F. Liechti, S. J. Ferguson, and M. Tannast. "Protrusio acetabuli: Joint loading with severe pincer impingement and its theoretical implications for surgical therapy". In: *Journal of Orthopaedic Research* 33.1 (2015), pp. 106–113.

[92] L. Liu et al. "Computer assisted planning and navigation of periacetabular osteotomy with range of motion optimization". In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014*. Springer, 2014, pp. 643–650.

[93] T. A. Laursen M. A. Puso. "A mortar segment-to-segment contact method for large deformation solid mechanics". In: *Comput. Methods Appl. Mech. Engrg.* 193 (2004), pp. 601–629. DOI: 10.1016/j.cma.2003.10.010.

[94] J. Mandel. "A multilevel iterative method for symmetric, positive definite linear complementarity problems". In: *Applied Mathematics and Optimization* 11.1 (1984), pp. 77–95.

[95] N. Maratos. "Exact penalty function algorithms for finite dimensional and control optimization problems". PhD thesis. Imperial College London (University of London), 1978.

[96] J. Marsden and T. Hughes. *Mathematical foundations of elasticity*. Courier Corporation, 1994.

[97] J. E. Marsden and T. Ratiu. *Introduction to mechanics and symmetry*. Springer Science & Business Media, 1999.

[98] M. Michalíková et al. "The digital pre-operative planning of total hip arthroplasty". In: *Acta Polytechnica Hungarica* 7.3 (2010).

[99] B. Mordukhovich. *Variational analysis and generalized differentiation I: Basic theory*. Springer, 2006.

[100] B. Mordukhovich. *Variational Analysis and Generalized Differentiation, II: Applications*. Springer, 2006.

[101] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, 2000. DOI: 10.1007/978-0-387-40065-5.

[102] M. Nordin and V. H. Frankel. *Basic biomechanics of the musculoskeletal system*. Lippincott Williams & Wilkins, 2001.

[103] R. W. Ogden. *Non-linear elastic deformations*. Courier Corporation, 1997.

[104] F. Paulsen and J. Wattschke. *Sobotta: Anatomie des Menschen: Allgemeiner Bewegungsapparat*. Elsevier, 2010.

[105] W. Platzer. *Color Atlas of Human Anatomy, Vol. 1. Locomotor System*. Thieme, 2004.

[106] A. Popp, M. W. Gee, and W. Wall. "Finite Deformation Contact Based on 3D Dual Mortar and Semi–Smooth Newton Approach". In: *Trends in Computational Contact Mechanics*. Springer, 2011, pp. 57–77. ISBN: 978-3-642-22167-5.

[107] M. Puso, T. Laursen, and J. Solberg. "A segment-to-segment mortar contact method for quadratic elements and large deformations". In: *Computer Methods in Applied Mechanics and Engineering* 197.6 (2008), pp. 555–566.

[108] A. Quarteroni and A. Valli. *Domain decomposition methods for partial differential equations*. CMCS-BOOK-2009-019. Oxford University Press, 1999.

[109] F. Radetzki et al. "Three-dimensional virtual simulation and evaluation of the femoroacetabular impingement based on "black bone" MRA". In: *Archives of orthopaedic and trauma surgery* 135.5 (2015), pp. 667–671.

[110] N. B. Reese and W. D. Bandy. *Joint range of motion and muscle length testing*. Elsevier Health Sciences, 2013.

[111] R. T. Rockafellar and R. J. Wets. *Variational analysis*. Vol. 317. Springer Science & Business Media, 2009.

[112] M. A. Röling et al. "A quantitative non-invasive assessment of femoroacetabular impingement with CT-based dynamic simulation-cadaveric validation study". In: *BMC musculoskeletal disorders* 16.1 (2015), p. 50.

[113] G. Rozza, D. Huynh, and A. T. Patera. "Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations". In: *Archives of Computational Methods in Engineering* 15.3 (2008), pp. 229–275.

[114] O. Sander. "A Fast Solver for Finite Deformation Contact Problems". preprint. 2006.

[115] O. Sander. "Coupling geometrically exact Cosserat rods and linear elastic continua". In: *Domain Decomposition Methods in Science and Engineering XX*. Springer, 2013, pp. 443–450.

[116] O. Sander. "Geodesic finite elements for Cosserat rods". In: *International journal for numerical methods in engineering* 82.13 (2010), pp. 1645–1670.

[117] O. Sander. "Multidimensional Coupling in a Human Knee Model". PhD thesis. Freie Universität Berlin, 2008.

[118] O. Sander. "The PSurface Library". In: *Computing and visualization in science* (2012). DOI: 10.1007/s00791-013-0193-4.

[119] O. Sander and A. Schiela. "Energy minimizers of the coupling of a Cosserat rod to an elastic continuum". In: *Math. Mod* (2012).

[120] O. Sander et al. "Towards an efficient numerical simulation of complex 3D knee joint motion". In: *Computing and Visualization in Science* 16.3 (2012), pp. 119–138.

[121] W. Schirotzek. *Nonsmooth Analysis*. Springer, 2006. DOI: 10.1007/978-3-540-71333-3.

[122] F. Schuricht. "Variational approach to contact problems in nonlinear elasticity". In: *Calculus of Variations and Partial Differential Equations* 15.4 (2002), pp. 433–449.

[123] T. Seidman and P. Wolfe. "Equilibrium states of an elastic conducting rod in a magnetic field". In: *Archive for Rational Mechanics and Analysis* 102.4 (1988), pp. 307–329.

[124] C. Silvestri and M. H. Ray. "Development of a finite element model of the knee-thigh-hip of a 50th percentile male including ligaments and muscles". In: *International Journal of crashworthiness* 14.2 (2009), pp. 215–229.

[125] J. C. Simo and N. Tarnow. "The discrete energy–momentum method. Conserving algorithms for nonlinear elastodynamics". In: *J. Appl. Math. Phys.* 43 (1992), pp. 757–792. DOI: 10.1007/BF00913408.

[126] D. A. Simon et al. "Development and validation of a navigational guidance system for acetabular implant placement". In: *CVRMed-MRCAS'97*. Springer. 1997, pp. 583–592.

[127] J. Spillmann. "CORDE: Cosserat rod elements for the animation of interacting elastic rods". PhD thesis. Universitätsbibliothek Freiburg, 2008.

[128] M. Tur, F. J. Fuenmayor, and P. Wriggers. "A mortar-based frictional contact formulation for large deformations using Lagrange multipliers". In: *Computer Methods in Applied Mechanics and Engineering* 198.37 (2009), pp. 2860–2873.

[129] M. Ulbrich, S. Ulbrich, and M. Heinkenschloss. "Global Convergence of Trust-region Interior-point Algorithms for Infinite-dimensional Nonconvex Minimization Subject to Pointwise Bounds". In: *SIAM Journal on Control and Optimization* 37.3 (1999), pp. 731–764. DOI: 10.1137/S0363012997319541. URL: http://dx.doi.org/10.1137/S0363012997319541.

[130] S. Ulbrich. "On the superlinear local convergence of a filter-SQP method". In: *Mathematical Programming* 100.1 (2004), pp. 217–245.

[131] M. Vaverka et al. "Stress and strain analysis of the hip joint using FEM." In: *Technology and health care: official journal of the European Society for Engineering and Medicine* 14.4-5 (2005), pp. 271–279.

[132] A. Wächter and L. Biegler. "Line Search Methods for Nonlinear Programming: Motivation and Global Convergence". In: *SIAM J. Optim.* 16.1 (2005), pp. 1–31. DOI: 10.1137/S1052623403426556.

[133] A. Wächter and L. T. Biegler. "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". In: *Mathematical programming* 106.1 (2006), pp. 25–57.

[134] A. Walther and L. Biegler. "On an inexact trust-region SQP-filter method for constrained nonlinear optimization". In: *Computational Optimization and Applications* (2015), pp. 1–26.

[135] J. A. Weiss and J. C. Gardiner. "Computational modeling of ligament mechanics". In: *Critical Reviews^{TM} in Biomedical Engineering* 29.3 (2001).

[136] D. Werner. *Funktionalanalysis*. Springer, 2011. ISBN: 978-3-642-21017-4.

[137] B. Wohlmuth. *Discretization Methods and Iterative Solvers based on Domain Decomposition*. LNCSE vol. 17. Springer Verlag, 2001. DOI: 10.1007/978-3-642-56767-4.

[138] B. Wohlmuth. "Variationally consistent discretization schemes and numerical algorithms for contact problems". In: *Acta Numerica* 20 (2011), pp. 569–734. DOI: 10.1017/S0962492911000079.

[139] P. Wriggers and T. A. Laursen. *Computational contact mechanics*. Vol. 30167. Springer, 2006.

[140] B. Yang, T. A. Laursen, and X. Meng. "Two dimensional mortar contact methods for large deformation frictional sliding". In: *Int. J. Numer. Methods Eng.* 62 (2005), pp. 1183–1225. DOI: 10.1002/nme.1222.

[141] Z. Yosibash, N. Trabelsi, and C. Milgrom. "Reliable simulations of the human proximal femur by high-order finite element analysis validated by experimental observations". In: *Journal of Biomechanics* 40.16 (2007), pp. 3688–3699.

[142] H. Yserentant. *Einführung in die Kontinuumsmechanik*. University Lecture. URL: http://www3.math.tu-berlin.de/Vorlesungen/SoSe13/Kontinuumsmechanik/.

[143] W. Zangwill. *Nonlinear programming : a unified approach*. Prentice-Hall Englewood Cliffs, NJ, 1969.

Zusammenfassung

Die Hüftoperationsplanung heutzutage basiert trotz modernster Technologien immer noch hauptsächlich auf Röntgenbildern. Das Gelenkzentrum, die Größe und die Position eines Implantats werden dabei nur mit Hilfe von simplen Schablonen bestimmt [76]. Daraus resultierend lag die Rate an sekundären Hüftoperation im Jahre 2006 bei fast 15% [8]. Ziel der vorliegenden Arbeit ist die Entwicklung eines patienten-spezifischen Finite-Element Hüftmodells, welches für die Schätzung der Spannungen während alltäglicher Bewegungen, wie auch einer Bewegunsraumanalyse, verwendet werden kann. Der Vorteil solch eines Finite-Element Modells gegenüber den fachüblichen Starrkörper-Simulationspaketen, ist die Möglichkeit der Stabilitätsanalyse des Gelenks selbst unter extremen Belastungen.
Der Kern eines solchen Hüftmodells liegt in der akkuraten Modellierung und Simulation von dynamischen Kontaktproblemen mit großen Deformationen.
In dem ersten Teil dieser Arbeit leiten wir auf der Vorarbeit von [70] eine neuartige Formulierung von Kontaktproblemen her, welche auf einer Erweiterung des Hamiltonschen Prinzips basieren. Diese Formulierung hat den Vorteil, dass die Kontaktkräfte nur implizit auftreten und variationell konsistent aus dem Residuum der gelösten Gleichung wiederhergeleitet werden können. Für diese schwache Formulierung entwickeln wir eine Stabilisierung der implizierten Mittelpunktsregel, welche das energetische Verhalten des Zeit-Integrators deutlich verbessert. Wegen der erschwerten theoretischen Analyse des Verfahrens, wird dieses nur numerisch getestet und Schritte zur weiteren Verbesserung vorgeschlagen. Das Hauptresultat dieser Arbeit liegt in der Entwicklung eines inexakten Filter Verfahrens für die Lösung der diskretisierten Probleme. Zur Beschleunigung des Lösers werden solange wie möglich approximierte Linearisierungen der Nicht-Durchdringungs-Bedingung verwendet, welche mit schnellen monotonen Mehrgitter Verfahren gelöst werden können. Die Erweiterung der monotonen Mehrgitter für nicht-konvexe Probleme, genauso wie ein Beweis der globalen Konvergenz wird hergeleitet und numerisch bestätigt. Ebenso wird die Konvergenz des inexakten Filter Verfahrens bewiesen, solange der Approximationsfehler der inexakten Linearisierung kontrolliert wird.
Im zweiten Teil der Dissertation wird eine heterogenes Finite-Element Hüftmodell bestehend aus dem Femur und Pelvis Knochen und den drei Hauptbändern vorgestellt. Dies ist eine direkte Erweiterung des Knie-Modells [120] auf nichtlineare Materialien und große Deformationen. Im Zuge dessen wird eine neue Kontaktmodellierung, Diskretisierung und Lösungsverfahren für Cosserat Stäbe mit starren Hindernissen entwickelt, welche zur Darstellung der Bänder benutzt werden. Erste Ergebnisse einer Bewegungsraumanalyse und ein Vergleich mit einem Starrkörper-Ansatz werden durchgeführt.