Adaptive numerical simulation of contact problems: Resolving local effects at the contact boundary in space and time

Dissertation

zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von
Mirjam Walloth
aus
Bonn

Bonn, März 2012

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Rolf Krause

2. Gutachter: Prof. Dr. Sören Bartels

3. Gutachter: Prof. Dr. Helmut Harbrecht

Tag der Promotion: 3. September 2012

Erscheinungsjahr: 2012

Diese Arbeit ist mit Unterstützung der von der Deutschen Forschungsgemeinschaft getragenen Bonn International Graduate School (BIGS) entstanden.

Zusammenfassung

Die numerische Simulation von Kontaktproblemen ermöglicht Vorhersagen komplexer Prozesse in vielen Bereichen der Mechanik und Biomechanik. Die Genauigkeit der numerisch bestimmten Approximation hängt dabei wesentlich von der Wahl des Lösungsverfahrens und der Diskretisierung ab. In dieser Arbeit beschäftigen wir uns mit dem Aspekt der Diskretisierung in Ort und Zeit für statische und dynamische Kontaktprobleme, bei denen sich die deformierbaren Materialien linear elastisch verhalten. Insbesondere leiten wir einen neuen effizienten residuenbasierten a posteriori Fehlerschätzer für die Diskretisierung im Ort her und stellen eine neue Zeitdiskretisierungsmethode vor.

Für die effiziente Simulation statischer Kontaktprobleme ist eine adaptive Finite-Elemente-Diskretisierung von großem Vorteil. Bei adaptiven Diskretisierungsmethoden werden mit Hilfe sogenannter a posteriori Fehlerschätzer Bereiche lokalisiert, in denen eine Verfeinerung des Gitters für die Fehlerreduktion erforderlich ist. Um sicherzustellen, dass ein a posteriori Fehlerschätzer den Fehler nicht überschätzt, sollte er nicht nur zuverlässig sein, das heißt den Fehler nach oben beschränken, sondern auch eine untere Schranke des Fehlers liefern. Ein solcher Fehlerschätzer heißt effizient.

Betrachtet man lineare Probleme, bei denen keine Nebenbedingungen gefordert werden, so gibt es eine direkte Beziehung zwischen dem Fehler und dem linearen Residuum, aus der sich effiziente Fehlerschätzer ergeben. Fordert man hingegen Nebenbedingungen, wie zum Beispiel beim Kontaktproblem oder dem verwandten Hindernisproblem, ist die Beziehung zwischen dem Fehler und dem linearen Residuum gestört. Dadurch überschätzt das lineare Residuum den Fehler und der Fehlerschätzer ist nicht mehr effizient. Da Fehlerschätzer basierend auf dem linearen Residuum numerisch leicht zu berechnen sind, ist es wünschenswert, das Konzept der residuenbasierten Fehlerschätzer auf Kontakt- und Hindernisprobleme zu erweitern. Bei der Analyse der wenigen in der Literatur bekannten residuenbasierten Fehlerschätzer für Kontakt- und Hindernisprobleme stellt sich heraus, dass eine Anpassung des Fehlermaßes unter Berücksichtigung der zweiten Unbekannten des Systems, der Zwangskraft, von zentraler Bedeutung für den Beweis der Effizienz des Fehlerschätzers ist.

In dieser Arbeit gelingt es uns, einen neuen effizienten residuenbasierten Fehlerschätzer für Kontaktprobleme herzuleiten. Wir beweisen Zuverlässigkeit und Effizienz in $2D$ und $3D$ für lineare finite Elemente auf Gittern aus Tetraedern beziehungsweise Dreiecken. Dabei nehmen wir an, dass die Abstandsfunktion zum Hindernis eine lineare Finite-Elemente-Funktion ist. Im allgemeineren Fall, in dem wir eine beliebige Abstandsfunktion im $H^{\frac{1}{2}}$ auf dem Kontaktrand haben, können wir die Zuverlässigkeit des Fehlerschätzers beweisen.

Wir untersuchen das Verhalten unseres residuenbasierten Fehlerschätzers anhand zahlreicher numerischer Experimente. Für spezielle Beispiele in $2D$ und $3D$ vergleichen wir die auf adaptiven Gittern numerisch berechnete Kontaktspannung mit der analytischen Lösung. Auf Gittern, die mit Hilfe unseres residuenbasierten Fehlerschätzer generiert sind, ist der relative Fehler in den Kontaktspannungen schon bei einer sehr geringen

Anzahl Knoten am tatsächlichen Kontaktrand stark reduziert. Dies gilt nicht nur für Gitter aus Simplizes, für die wir Effizienz und Zuverlässigkeit beweisen, sondern auch für Gitter aus Vierecken und Hexaedern. Für beliebige Startgitter aus Hexaedern, Tetraedern, Prismen und Pyramiden und unterschiedliche Abstandsfunktionen vergleichen wir die experimentelle Konvergenzrate des Fehlerschätzers für adaptive und uniforme Verfeinerung. Des Weiteren untersuchen wir die Struktur der adaptiv generierten Gitter und den Einfluss der verschiedenen lokalen Fehlerschätzerbeiträge.

Im zweiten Teil der Arbeit befassen wir uns mit der Wahl einer geeigneten Zeitdiskretisierung für dynamische Kontaktprobleme. Da weder eine analytische Lösung des dynamischen Kontaktproblems in $3D$ noch eine Existenztheorie bekannt sind, spielen strukturerhaltende Eigenschaften wie zum Beispiel die Energieerhaltung eine wesentliche Rolle bei der qualitativen Bewertung von Zeitdiskretisierungen. Eine Zeitdiskretisierung, die für ihre strukturerhaltenden Eigenschaften bei Problemen ohne Nebenbedingungen bekannt ist, ist das Newmarkverfahren. Wendet man das Newmarkverfahren auf dynamische Kontaktprobleme an, so kommt es allerdings zu Oszillationen in den Geschwindigkeiten und Kontaktspannungen und zu einem Energieanstieg. Daher ist eine Anpassung des Newmarkverfahrens an dynamische Kontaktprobleme erforderlich. Wir analysieren die Gründe, die zu den genannten unphysikalischen Effekten des Newmarkverfahrens führen. Des Weiteren vergleichen wir verschiedene Modifikationen des Newmarkverfahrens, die für dynamische Kontaktprobleme in der Literatur zu finden sind.

Wir stellen eine neue Methode für die Zeitdiskretisierung dynamischer Kontaktprobleme vor, die basierend auf dem Newmarkverfahren die Oszillationen in den Kontaktspannungen und den Energieanstieg verhindert und zusätzlich ein physikalisch motiviertes Update der Geschwindigkeiten ermöglicht. Der zentrale Bestandteil unserer neuen Methode ist die implizite Berechnung der Aufprallzeiten für jeden Kontaktknoten des ortsdiskreten Systems mit Hilfe eines Prädiktorschritts. Mittels dieser Aufprallzeiten kann eine Änderung der Geschwindigkeiten im Moment des Aufpralls berücksichtigt werden, was mit einer klassischen Zeitdiskretisierung nicht möglich ist. Diese neue Zeitdiskretisierung, wie auch einige andere in der Literatur vorgestellte Methoden, sind Elemente einer Familie von modifizierten Newmarkverfahren abhängig von matrixwertigen Parametern. In unserer Analyse zeigen wir, welchen Einfluss die Wahl der matrixwertigen Parameter auf das Verhalten der Energie, der Kontaktspannungen und der Geschwindigkeiten hat. Numerische Beispiele in $3D$ ergänzen die theoretische Analyse der verschiedenen Zeitdiskretisierungen.

Danksagung

An dieser Stelle möchte ich mich bei Professor Rolf Krause bedanken, dem es gelungen ist, mich für das Gebiet des wissenschaftlichen Rechnens zu begeistern. Für seine wertvollen Ratschläge und Ideen als auch sein großes Vertrauen in meine Fähigkeiten bin ich sehr dankbar. Darüber hinaus möchte ich mich bei ihm für die Unterstützung zur Teilnahme an Konferenzen, für die Einladungen ans Institute of Computational Science der Università della Svizzera italiana und für die Initiierung der Kooperation mit Professor Andreas Veeser von der Università degli Studi di Milano bedanken. Ich danke Professor Sören Bartels für die Fortführung der Betreuung in Bonn und seine wertvollen Anregungen während dieser Zeit. Des Weiteren bedanke ich mich bei ihm für die Übernahme des Zweitgutachtens. Professor Andreas Veeser danke ich für die Einladungen an die Università degli Studi di Milano, die es mir ermöglicht haben, von seinen zahlreichen Ideen und seinem großen Erfahrungsschatz im Bereich der a posteriori Fehlerschätzer zu profitieren.
Der Bonn International Graduate School in Mathematics danke ich für ein großzügiges Promotionsstipendium sowie für die Finanzierung der Teilnahme an Konferenzen. Für die Bereitstellung einer hervorragenden Infrastruktur und weiterer finanzieller Mittel danke ich dem Institut für Numerische Simulation der Rheinischen Friedrich-Wilhelms-Universität und dem Institute of Computational Science der Università della Svizzera italiana. Diese Arbeit wurde weiter unterstützt durch das BMBF-Projekt ASIL (advanced solvers integrated library) und durch die Firma Ford im Rahmen des URP-Projektes „Advanced numerical algorithms to improve highly nonlinear crash simulation with multi-body contacts and friction“.
Ein großer Dank gilt allen meinen jetzigen und ehemaligen Kollegen am Institut für Numerische Simulation der Rheinischen Friedrich-Wilhelms-Universität und dem Institute of Computational Science der Università della Svizzera italiana für die angenehme Arbeitsatmosphäre. Herzlich danke ich Thomas Dickopf, Johannes Steiner, Dorian Krause und Christina Mohr für die Freundschaftlichkeit und gute Zusammenarbeit in der Arbeitsgruppe von Professor Rolf Krause trotz der großen Distanz. Vor allem danke ich Thomas Dickopf für fruchtbare Diskussionen während unserer gemeinsamen Zeit in Bonn und für das intensive Korrekturlesen großer Teile meiner Arbeit. Christina Mohr danke ich ebenfalls für das sorgfältige Korrekturlesen einiger Teile meiner Arbeit. Ein herzlicher Dank gilt auch meinem Bürokollegen Alexander Raisch, der mir bei einer Vielzahl von Details bereitwillig zugehört hat und mir oft weiterhelfen konnte. Darüber hinaus hat er einen Teil der Arbeit korrekturgelesen. Des Weiteren möchte ich mich herzlich bei Jutta Adelsberger bedanken für ihre stete und freundschaftliche Hilfsbereitschaft und das Korrekturlesen einzelner Abschnitte meiner Arbeit.
Schließlich bedanke ich mich herzlichst bei meiner Familie und meinen Freunden für ihre großartige Unterstützung während der Zeit der Promotion.

Contents

Introduction

The simulation of contact problems provides insight into many complex processes in, e.g., mechanics or biomechanics. The accuracy of the numerical approximation depends strongly on the choice of the solver and the discretization methods. In this work we deal with the aspects of discretization in space and time of static and dynamic contact problems. In particular, we derive a new efficient residual-type a posteriori error estimator for static contact problems and a new space-time connecting discretization scheme for dynamic contact problems in linear elasticity.

Residual-type a posteriori error estimators for contact problems

We consider a discretization in space by means of linear finite elements. In order to reduce the computational costs without sacrificing the accuracy of the numerical solution, an adaptive mesh generation is highly advantageous. For the detection of the critical regions, where the mesh has to be refined in order to improve the accuracy of the numerical solution, a posteriori error estimators are used. A posteriori error estimators should be reliable, i.e., give an upper bound of the error. If the error estimator also constitutes a lower bound of the error, it is called efficient. The efficiency of the error estimator ensures that the accuracy is achieved with an almost minimal amount of grid points. Due to these reasons, a reliable and efficient a posteriori error estimator for contact is desirable.

For linear elliptic problems, where no constraints are imposed, the standard residual a posteriori error estimator is reliable and efficient. This estimator is very common due to its easy computation. The proofs of upper and lower bound are based on the equivalence $\|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1} \lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_1 \lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1}$ between the linear residual and the error. This relation is disturbed if constraints are imposed as, e.g., in contact problems or the closely related obstacle problems. Therefore, the linear residual overestimates the error and the standard residual error estimator is not efficient. However, in the literature one can find residual-type a posteriori error estimators which extend the concept of the standard residual estimator to contact and obstacle problems.

We present the basic ideas of selected residual-type estimators for contact and obstacle problems found in the literature and we explain the difficulties arising in their construction. For the Signorini problem and the simplified Signorini problem residual-type a posteriori error estimators can be found in [HN05, HN07]. These error estimators have been derived for the two-dimensional case and the gap function $g = 0$. Unfortunately, upper and lower bounds are non-optimal for the a posteriori error estimator proposed in [HN05]. For the closely related obstacle problem residual-type a posteriori error estimators are given in [CN00, Vee01, FV03, NSV03, NSV05, MNvPZ07]. The first efficient residual-type a posteriori error estimator for obstacle problems with optimal upper and lower bounds is presented in [Vee01]. Therein the error measure is given in terms of the solution and the second unknown of the system, the constraining force which turns the inequality into an equality. A so-called Galerkin functional, which considers the error in the constraining force as well as in the solution, takes the role of the linear residual. This ansatz has been applied and further developed in, e.g., [FV03, NSV03, NSV05, MNvPZ07]. In these works the definition of a suitable constraining force depending on the discrete solution is of great importance.

In this thesis we derive a new efficient residual-type a posteriori error estimator for contact problems. Inspired by [FV03, MNvPZ07] we define a suitable constraining force depending on the discrete solution. It is a functional which mimics the properties of the continuous constraining force and is computable from the discrete solution and given data. The resulting Galerkin functional enables localization of error estimator contributions to the free-boundary zone and to the area where no actual contact occurs. We note that in contrast to obstacle problems the constraints are imposed in a subset of the boundary instead of in the whole domain. Therefore, the a posteriori analysis for contact problems is quite different to the case of obstacle problems. The contributions of our new residual-type a posteriori error estimator are motivated by inherent properties of the solution as, e.g., the fulfillment of the complementarity condition. If no actual contact occurs the error estimator coincides with the standard residual error estimator for linear elliptic problems. We prove reliability and efficiency in $2D$ and $3D$ if the gap function is discrete. Even for arbitrary, non-discrete gap functions the reliability is proven.

The proofs of upper and lower bound are given for meshes of simplices. However, numerical examples show the performance of the new residual-type a posteriori error estimator for meshes of hexahedra, prisms, and pyramids, too. We compare the rate of convergence of the error estimator on adaptively and uniformly refined grids. For special examples in $2D$ and $3D$ we compare the contact stresses computed on adaptively refined grids with the exact contact stresses. Furthermore, the structure of the adaptively refined grids and the relevance of the different error estimator contributions are analyzed by means of different numerical experiments in $3D$.

Time discretization schemes for dynamic contact problems

In the numerical simulation of dynamic contact problems the construction of a suitable time discretization scheme is of crucial importance. Due to the non-smooth character of dynamic contact problems, classical time discretization schemes cannot be applied in a straightforward way. Therefore, several methods have been proposed which adapt classical time discretization schemes to the case of contact problems; see, e.g., [AB08, Mor99, HTS$^+$76, TP93, LC97, KROM99, LL02, DKE08, KLR08, HHW08, KW09a, DEP11]. Since up to now no existence results for the hyperbolic system of dynamic contact problems are available, the quality of time discretization schemes of contact problems is measured by means of physical properties of the time-discretized system. Besides the displacements, velocities and contact stresses, the energy and momentum conservation and the persistency condition are relevant.

In this work we focus on modifications of the classical Newmark scheme. The classical Newmark scheme is very common in continuum mechanics as it is of second order consistency and conserves the energy in the unconstrained case. Unfortunately, in the case of contact constraints, the classical Newmark scheme evokes oscillations in the contact stresses, the displacements and the velocities at the contact boundary and even energy blow-ups may occur which spoil the accuracy of the solution. We give a deeper insight into the causes of these instabilities. Further, we present and compare selected modifications of the Newmark scheme which can be found in the literature, see [LC97, KROM99, DKE08, KLR08, HHW08]. Each of this methods has its advantage concerning the stability of contact stresses or the course of energy. We illustrate these properties by means of numerical examples in $3D$.

We present a new space-time connecting discretization scheme for dynamic contact problems which is based on the Newmark scheme. It avoids oscillations in the contact stresses, is provably dissipative and allows for a physically motivated update of the velocities. We find out that by means of the predictor step used in [DKE08] the impact times of the single nodes can be computed implicitly. By means of these impact times, the change in the velocity in the moment of impact can be taken into account. This is not possible in classical time discretization schemes. Further, we show that this new time discretization scheme as well as other methods proposed in, e.g., [KROM99, DKE08] are elements of a family of modified Newmark schemes depending on matrix-valued parameters. We discuss the influence of different choices of the matrix-valued parameters on the course of energy and the behavior of the contact stresses and velocities. Numerical examples in $3D$ complement the theoretical analysis.

Outline

This thesis is structured as follows. The first chapter deals with the strong and weak formulation of the Signorini problem. The equations of linear elasticity are deduced from the general equations of motion and constitutive laws of elastic materials. Further, we derive the linearized non-penetration condition. With regard to the validation of numerical experiments we shortly comment on analytic solutions of contact problems. Besides the Signorini contact problem, the simplified Signorini problem and the closely related obstacle problem are stated. At the end of Chapter 1 we give the weak formulation of the Signorini problem and we recall existence and uniqueness results.

The aim of Chapter 2 is to explain the difficulties arising in the construction of residual-type a posteriori error estimators for contact and obstacle problems. Therefore, we review the standard residual error estimator for linear elliptic problems without constraints and present the basic ideas of residual-type a posteriori error estimators for obstacle and contact problems which can be found in the literature.

Chapter 3 is devoted to our new residual-type a posteriori error estimator. The definition of the quasi-discrete contact force density and the corresponding Galerkin functional is introduced. We give the proofs of reliability and efficiency.

In Chapter 4 several numerical examples illustrate the performance of our new residual-type a posteriori error estimator for contact. For special examples the analytical contact stresses and the radius of the contact zone are compared with the corresponding quantities computed on adaptively refined grids. The structure of the adaptively refined grids, the relevance of the local error estimator contributions and the experimental order of convergence are investigated for different examples in $3D$.
In the final chapter we deal with the discretization in time of dynamic contact problems. After introducing the strong and weak formulation of dynamic contact problems we explain the difficulties arising in the construction of suitable time discretization schemes. Chapter 5 provides a detailed comparison of time discretization schemes for dynamic contact problems which can be found in the literature. We present our new space-time connecting discretization scheme for contact problems.

Publications

Finally, we note that several results of this thesis have already been published as article or preprint. The presentation and comparison of the modified Newmark schemes in Section 5.2 has been accepted for publication in *Applied Numerical Mathematics*. It can be found in the preprint [KW09a]. Our new space-time connecting discretization scheme has been published in *Computer Methods in Applied Mechanics and Engineering*, see [KW11]. Just recently, the results concerning our new residual-type a posteriori error estimator have been presented in a shortened version in the preprint [KVW12].

1 Static contact problems in linear elasticity

This chapter deals with the strong and weak formulation of static contact problems. We mainly focus on one-body contact problems in linear elasticity. Nevertheless, in Section 1.1.1 we derive the equations of motion and the constitutive laws of elastic materials in general. The special case of linearized elasticity is considered in Section 1.1.2.
The condition of non-penetration of solid bodies which is enforced in the case of contact problems is derived in Section 1.2.1. In Section 1.2.2 the Signorini contact problem is stated along with some examples for which the contact stresses can be computed analytically. We make use of these analytic solutions in Chapter 4 for the evaluation of the numerical results. In Section 1.2.3 we present the closely related simplified Signorini problem and the obstacle problem. As we will see in Section 2.2.3 and 2.2.2 similar difficulties arise in the derivation of residual-type a posteriori error estimators for the Signorini problem, the simplified Signorini problem and the obstacle problem.
Finally, in Section 1.3 we give the weak formulation of the Signorini problem with several useful definitions. Notably, the constraining force density is introduced in Section 1.3.1 which is of interest for the following results.

1.1 Equations of equilibrium and constitutive laws

In continuum mechanics the materials are modeled as continuous mass. Although the microscopic structure is ignored this models are highly accurate as long as the length scales are large enough comparative to interatomic distances. The two main areas of continuum mechanics are solid and fluid mechanics. The area of solid mechanics is further subdivided into elasticity and plasticity. The field of fluid mechanics splits up in Newtonian and Non-Newtonian fluids. Materials which have solid and fluid characteristics are described by so-called rheological models which combine elastic, plastic and viscous phenomena. Elastic materials as, e.g., rubber return quickly to their initial state when applied stresses are released whereas plastic materials as, e.g., plasticine stay deformed after release of applied stresses. Often materials behave elastically up to a special amount of applied stresses and afterwards the reaction is plastic. In this work we are interested in the elastic reaction of materials. Our focus lies on the special case of linear elastic materials. However, we briefly introduce nonlinear elasticity because numerical methods for linear elasticity might be inspiring for the related area of nonlinear elasticity.

1.1.1 Elasticity

If we bear force on a deformable body as, e.g., stretching an elastic band, the body undergoes a certain deformation. We aim to determine this deformation field caused by the reaction of the elastic body to given forces. In a first step we give a precise definition of the deformation and the closely related displacement. We motivate the definition of a strain tensor which measures the deformation of a body resulting in a change of form and size compared to rigid body motions. In the deformed configuration the important stress principle of Euler and Cauchy (Axiom 1.1.1) states the existence of interior surface tractions which fulfill the laws of momentum balance. Based on this axiom Cauchy's theorem (Theorem 1.1.1) introduces a stress tensor field and derives a partial differential equation relating the stress tensor field to the applied forces in the deformed configuration. As the deformed configuration is unknown we transform the equilibrium equation to the reference configuration. Finally, we derive constitutive equations in elasticity which relate stress and strain tensor. The equilibrium equation together with the constitutive law gives rise to a boundary value problem where the applied forces are given and the deformations are unknown. In this section we follow mainly the lines of [Cia88, Sän05, Bro95]. For further presentations we refer the reader to, e.g., [Gur81, EGK08, Ogd84, FdV79, Ant05].

The undeformed body is identified with its reference configuration, represented by a domain $\Omega \subset \mathbb{R}^d$. We assume the domain to be a Lipschitz domain, i.e., an open, bounded, connected subset of the Euclidean space with Lipschitz boundary $\Gamma := \partial\Omega$. Each material particle in the closure $\bar{\Omega}$ is identified with a point $\boldsymbol{x} = (x_1, ..., x_d)^T$. Throughout this work we denote all quantities which refer to tensors of order ≥ 1 by bold symbols as, e.g., the displacements $\boldsymbol{u}$ which are vector-valued. Their components are printed in normal type and are indicated by subindices, e.g., u_i. The Einstein summation convention is used for all repeated indices and $\boldsymbol{e}_i$ denotes the Cartesian basis vectors of $\mathbb{R}^d$ such that, e.g., $\boldsymbol{u} = u_i \boldsymbol{e}_i$. For a function depending on the variable $\boldsymbol{x} \in \mathbb{R}^d$, the partial derivative with respect to x_j, $j = 1, .., d$ is abbreviated with ∂_j. The symbol for the identity mapping and its matrix representation is $\boldsymbol{id}$. In this section we consider the full-dimensional case $d = 3$.

A deformation of the reference configuration $\bar{\Omega}$ is a vector field

$$\boldsymbol{\varphi} : \bar{\Omega} \longrightarrow \mathbb{R}^d, \qquad \boldsymbol{x} \mapsto \boldsymbol{\varphi}(\boldsymbol{x}),$$

that is smooth enough, injective except possibly on the boundary of the set Ω, and orientation preserving. The gradient of the deformation $\boldsymbol{\varphi} = \varphi_i \boldsymbol{e}_i$ is given by the matrix

$$\boldsymbol{\nabla\varphi} := (\partial_j \varphi_i)_{1 \leq i,j \leq d}.$$

It follows from the orientation preserving property that $\boldsymbol{\nabla\varphi}(\boldsymbol{x}) \in \mathbb{R}_+^{d\times d} := \{\boldsymbol{M} \in \mathbb{R}^{d\times d} \mid \det \boldsymbol{M} > 0\}$. The deformation $\boldsymbol{\varphi}$ measures each kind of movement even rigid deformations like rotations and translations which do not lead to a change in form or size. Rigid deformations are defined by

$$\boldsymbol{\varphi}(\boldsymbol{x}) := \boldsymbol{Q}\boldsymbol{x} + \boldsymbol{a},$$

where $\boldsymbol{a} \in \mathbb{R}^d$ and $\boldsymbol{Q} \in \mathbb{O}_+$ is an orthogonal matrix with $\det \boldsymbol{Q} = 1$. Let us consider two points $\boldsymbol{x}$ and $\boldsymbol{x} + \boldsymbol{\delta x}$ where $|\boldsymbol{\delta x}|$ is an infinitesimal distance, changing their positions to

$\boldsymbol{\varphi}(\boldsymbol{x})$ and $\boldsymbol{\varphi}(\boldsymbol{x}+\boldsymbol{\delta x})$. The distance between the two points in the deformed configuration is given by

$$\boldsymbol{\varphi}(\boldsymbol{x}+\boldsymbol{\delta x})-\boldsymbol{\varphi}(\boldsymbol{x})=\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})\boldsymbol{\delta x}+o(|\boldsymbol{\delta x}|)$$

and the square of the euclidean distance $|\cdot|$ is

$$|\boldsymbol{\varphi}(\boldsymbol{x}+\boldsymbol{\delta x})-\boldsymbol{\varphi}(\boldsymbol{x})|^2=\boldsymbol{\delta x}^T\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^T\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})\boldsymbol{\delta x}+o(|\boldsymbol{\delta x}|^2).$$

If $\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})=\boldsymbol{Q}$ the deformation causes no change in the distance between $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{\delta x}$. Thus, the (right) Cauchy–Green strain tensor $\boldsymbol{C}:=\boldsymbol{\nabla}\boldsymbol{\varphi}^T\boldsymbol{\nabla}\boldsymbol{\varphi}$ is a measure of the strain evoked by the deformation. Another measure of the strain is the Green–St. Venant strain tensor $\boldsymbol{E}:=\frac{1}{2}(\boldsymbol{C}-\boldsymbol{id})$, quantifying the local deviation of the deformation $\boldsymbol{\varphi}$ from a rigid body motion.

The displacement field is given by the mapping $\boldsymbol{u}:\bar{\Omega}\longrightarrow\mathbb{R}^d$ related to the deformation by $\boldsymbol{u}:=\boldsymbol{\varphi}-\boldsymbol{id}$. Inserting the definition of the displacements in the Green-St. Venant strain tensor leads to the following strain-displacements relation

$$\boldsymbol{E}(\boldsymbol{u})=\frac{1}{2}\left(\boldsymbol{\nabla}(\boldsymbol{u})^T+\boldsymbol{\nabla}(\boldsymbol{u})+\boldsymbol{\nabla}(\boldsymbol{u})^T\boldsymbol{\nabla}(\boldsymbol{u})\right). \tag{1.1}$$

In the deformed body the stress principle of Euler and Cauchy holds, which is the foundation of continuum mechanics. We set $\bar{\Omega}^\varphi:=\boldsymbol{\varphi}(\bar{\Omega})$ and $\boldsymbol{x}^\varphi:=\boldsymbol{\varphi}(\boldsymbol{x})$. The later are called Euler variables. The corresponding volume and area elements are denoted by dx^φ and da^φ. Further, we assume that the deformation of the body is due to volume forces and surface tractions on the Neumann boundary $\Gamma_N^\varphi\subset\partial\Omega^\varphi$ which is an open measurable subset of the boundary. They are represented by the densities $\boldsymbol{f}^\varphi:\Omega^\varphi\to\mathbb{R}^d$ and $\boldsymbol{\pi}^\varphi:\Gamma_N^\varphi\to\mathbb{R}^d$. The principle of Euler and Cauchy, see [Cia88, Axiom 2.2-1], states the existence of surface tractions in the interior of the body such that equilibrium of momentum and angular momentum is obtained locally. The vector field of interior surface tractions is called Cauchy stress vector.

Axiom 1.1.1. *Principle of Euler and Cauchy*
Consider a body in its deformed configuration $\bar{\Omega}^\varphi$, subjected to applied forces represented by the densities $\boldsymbol{f}^\varphi:\Omega^\varphi\to\mathbb{R}^d$ and $\boldsymbol{\pi}^\varphi:\Gamma_N^\varphi\to\mathbb{R}^d$. Then there exists a vector field

$$\boldsymbol{t}^\varphi:\bar{\Omega}^\varphi\times\mathbb{S}^{d-1}\to\mathbb{R}^d,\qquad(\boldsymbol{x}^\varphi,\boldsymbol{n})\mapsto\boldsymbol{t}^\varphi(\boldsymbol{x}^\varphi,\boldsymbol{n}),$$

where $\mathbb{S}^{d-1}=\{\boldsymbol{v}\in\mathbb{R}^d\mid|\boldsymbol{v}|=1\}$ such that

- *For any subdomain $\mathcal{A}^\varphi\subset\bar{\Omega}^\varphi$, and at any point $\boldsymbol{x}^\varphi\in\Gamma_N^\varphi\cap\partial\mathcal{A}^\varphi$ where the unit outward normal vector $\boldsymbol{n}^\varphi$ to $\Gamma_N^\varphi\cap\partial\mathcal{A}^\varphi$ exists,*

  $$\boldsymbol{t}^\varphi(\boldsymbol{x}^\varphi,\boldsymbol{n}^\varphi)=\boldsymbol{\pi}^\varphi(\boldsymbol{x}^\varphi).$$

- *Balance of linear momentum: For any subdomain $\mathcal{A}^\varphi\subset\bar{\Omega}^\varphi$,*

  $$\int_{\mathcal{A}^\varphi}\boldsymbol{f}^\varphi(\boldsymbol{x}^\varphi)dx^\varphi+\int_{\partial\mathcal{A}^\varphi}\boldsymbol{t}^\varphi(\boldsymbol{x}^\varphi,\boldsymbol{n}^\varphi)da^\varphi=\boldsymbol{0},$$

  where $\boldsymbol{n}^\varphi$ denotes the unit outward normal vector along $\partial\mathcal{A}^\varphi$.

- *Balance of angular momentum: For any subdomain* $\mathcal{A}^\varphi \subset \bar{\Omega}^\varphi$,

$$\int_{\mathcal{A}^\varphi} \boldsymbol{x}^\varphi \times \boldsymbol{f}^\varphi(\boldsymbol{x}^\varphi) dx^\varphi + \int_{\partial\mathcal{A}^\varphi} \boldsymbol{x}^\varphi \times \boldsymbol{t}^\varphi(\boldsymbol{x}^\varphi, \boldsymbol{n}^\varphi) da^\varphi = \boldsymbol{0}.$$

The following theorem asserts a linear dependence of the Cauchy stress vector on its second argument $\boldsymbol{n} \in \mathbb{S}^{d-1}$, i.e., at each point $\boldsymbol{x}^\varphi \in \bar{\Omega}^\varphi$ there exists a symmetric tensor $\boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi) \in \mathbb{R}^{d\times d}_{\text{sym}}$, called Cauchy stress tensor, such that $\boldsymbol{t}^\varphi(\boldsymbol{x}^\varphi, \boldsymbol{n}) = \boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi)\boldsymbol{n}$ for all $\boldsymbol{n}$. Further, the tensor field $\boldsymbol{T}^\varphi$ and the given volume forces and surface tractions are related by a partial differential equation in Ω^φ and by a boundary condition on Γ_N^φ.

Theorem 1.1.1. *Cauchy's theorem*
Assume that the applied body force density $\boldsymbol{f}^\varphi : \bar{\Omega}^\varphi \to \mathbb{R}^d$ *is continuous, and that the Cauchy stress vector field* $\boldsymbol{t}^\varphi : \bar{\Omega}^\varphi \times \mathbb{S}^{d-1} \to \mathbb{R}^d$ *is continuously differentiable with respect to the variable* $\boldsymbol{x}^\varphi \in \bar{\Omega}^\varphi$ *for each* $\boldsymbol{n} \in \mathbb{S}^{d-1}$ *and continuous with respect to the variable* $\boldsymbol{n} \in \mathbb{S}^{d-1}$ *for each* $\boldsymbol{x}^\varphi \in \bar{\Omega}^\varphi$. *Then the principle of Euler and Cauchy implies the existence of a continuously differentiable symmetric tensor field*

$$\boldsymbol{T}^\varphi : \bar{\Omega}^\varphi \to \mathbb{R}^{d\times d}_{\text{sym}}, \qquad \boldsymbol{x}^\varphi \mapsto \boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi),$$

such that the Cauchy stress vector satisfies

$$\boldsymbol{t}^\varphi(\boldsymbol{x}^\varphi, \boldsymbol{n}) = \boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi)\boldsymbol{n}, \quad \forall \boldsymbol{x}^\varphi \in \bar{\Omega}^\varphi, \boldsymbol{n} \in \mathbb{S}^{d-1},$$

and such that

$$\begin{aligned} -\text{div}^\varphi \boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi) &= \boldsymbol{f}^\varphi(\boldsymbol{x}^\varphi), \quad \forall \boldsymbol{x}^\varphi \in \Omega^\varphi \\ \boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi)\boldsymbol{n}^\varphi &= \boldsymbol{\pi}^\varphi(\boldsymbol{x}^\varphi), \quad \forall \boldsymbol{x}^\varphi \in \Gamma_N^\varphi \end{aligned} \tag{1.2}$$

where $\boldsymbol{n}^\varphi$ *is the unit outward normal vector along* Γ_N^φ *and* $\text{div}^\varphi := \partial_j T_{ij}^\varphi \boldsymbol{e}_i$ *is the divergence of the vector field with respect to* $\boldsymbol{x}^\varphi$.

Proof. See [Cia88, Theorem 2.3-1]. qed.

Cauchy's theorem gives rise to the equations of equilibrium (1.2) in the deformed configuration with respect to the Euler variables $\boldsymbol{x}^\varphi = \boldsymbol{\varphi}(\boldsymbol{x})$. Unfortunately, the deformation is the sought-after quantity of our system. Thus, we need to express the equations in terms of the Lagrange variables $\boldsymbol{x}$. The relations between the force densities in reference and deformed configuration are given by $\boldsymbol{f}(\boldsymbol{x}) = (\det \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x}))\boldsymbol{f}^\varphi(\boldsymbol{x}^\varphi)$ and $\boldsymbol{\pi}(\boldsymbol{x}) = (\det \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x}))|\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^{-T}\boldsymbol{n}|\boldsymbol{\pi}^\varphi(\boldsymbol{x}^\varphi)$. For the transformation of the stress tensor $\boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi)$ depending on Euler variables into a stress tensor $\boldsymbol{T}(\boldsymbol{x})$ depending on Lagrange variables we use the Piola transform

$$\boldsymbol{T}(\boldsymbol{x}) = (\det \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x}))\boldsymbol{T}^\varphi(\boldsymbol{x}^\varphi)\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^{-T}, \quad \boldsymbol{x}^\varphi = \boldsymbol{\varphi}(\boldsymbol{x}).$$

The resulting stress tensor in the reference configuration is called first Piola-Kirchhoff stress tensor. The advantage of the Piola transform is the relation between the divergences of $\boldsymbol{T}^\varphi$ and $\boldsymbol{T}$ and between the area elements da^φ and da; see, e.g., [Cia88, Theorem

1.7-1]. Thus, leading to the following equations of equilibrium in the reference configuration

$$\begin{aligned} -\operatorname{div} \boldsymbol{T}(\boldsymbol{x}) &= \boldsymbol{f}(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \\ \boldsymbol{T}(\boldsymbol{x})\boldsymbol{n} &= \boldsymbol{\pi}(\boldsymbol{x}), \quad \boldsymbol{x} \in \Gamma_N; \end{aligned} \tag{1.3}$$

see, e.g., [Cia88, Theorem 2.6-1]. Unfortunately, the first Piola-Kirchhoff stress tensor is not symmetric. Thus, it is common to use the so-called second Piola-Kirchhoff stress tensor

$$\boldsymbol{\Sigma}(\boldsymbol{x}) := \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^{-1}\boldsymbol{T}(\boldsymbol{x}) = (\det \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x}))\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^{-1}\boldsymbol{T}^{\varphi}(\boldsymbol{x}^{\varphi})\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^{-T}, \quad \boldsymbol{x}^{\varphi} = \boldsymbol{\varphi}(\boldsymbol{x}). \tag{1.4}$$

As we will see in Theorem 1.1.2, the symmetry of the second Piola-Kirchhoff stress tensor simplifies the relation between the stress and strain tensors. In the reference configuration in terms of the second Piola-Kirchhoff stress tensor the equations of equilibrium are given by

$$\begin{aligned} -\operatorname{div}(\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})\boldsymbol{\Sigma}(\boldsymbol{x})) &= \boldsymbol{f}(\boldsymbol{x}), \quad \boldsymbol{x} \in \Omega \\ \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})\boldsymbol{\Sigma}(\boldsymbol{x})\boldsymbol{n} &= \boldsymbol{\pi}(\boldsymbol{x}), \quad \boldsymbol{x} \in \Gamma^N; \end{aligned} \tag{1.5}$$

see, e.g., [Cia88, Theorem 2.6-2].

In order to determine the deformation or displacement from (1.5) we need to express the stress tensor in terms of deformation. This is done by the constitutive equation which relates stress and strain tensor depending on the material properties. A material is elastic if the Cauchy stress tensor solely depends on $\boldsymbol{x}$ and on the gradient of the deformation $\boldsymbol{\nabla}\boldsymbol{\varphi}$. Equivalently, a material is elastic if each of the Piola-Kirchhoff stress tensors solely depends on $\boldsymbol{x}$ and $\boldsymbol{\nabla}\boldsymbol{\varphi}$. Thus, for all $\boldsymbol{x} \in \bar{\Omega}$ there exists a so-called response function $\hat{\boldsymbol{\Sigma}} : \bar{\Omega} \times \mathbb{R}^{d\times d}_{+} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$, $(\boldsymbol{x}, \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})) \mapsto \hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x}))$ such that

$$\boldsymbol{\Sigma}(\boldsymbol{x}) := \hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})), \quad \boldsymbol{x} \in \bar{\Omega}.$$

The objectivity, also called material frame indifference, is an axiom which states the invariance of a physical quantity under a change of observer. The response function $\hat{\boldsymbol{\Sigma}}$ fulfills the axiom of material frame indifference if and only if there exists a mapping $\tilde{\boldsymbol{\Sigma}} : \bar{\Omega} \times \mathbb{R}^{d\times d}_{\mathrm{spd}} \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$, where $\mathbb{R}^{d\times d}_{\mathrm{spd}}$ is the space of symmetric positive definite matrices, so that

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{F}) = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{F}^T\boldsymbol{F}), \quad \forall \boldsymbol{F} \in \mathbb{R}^{d\times d}_{+}, \boldsymbol{x} \in \bar{\Omega};$$

see, e.g., [Cia88, Theorem 3.3.-1]. Thus, we get a stress-strain relation between the second Piola-Kirchhoff stress tensor and the right Cauchy-Green strain tensor

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})) = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})^T\boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})) = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{C}(\boldsymbol{x})).$$

Further, we make the assumption of isotropy. A material is isotropic if the stress response does not depend on a direction, i.e., the material has no preferential direction. Expressed in terms of the response function to the second Piola-Kirchhoff stress tensor at a point $\boldsymbol{x} \in \bar{\Omega}$ we get the relation

$$\hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{F}\boldsymbol{Q}) = \boldsymbol{Q}^T \hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{F})\boldsymbol{Q}, \quad \forall \boldsymbol{F} \in \mathbb{R}_+^{d \times d}, \boldsymbol{Q} \in \mathbb{O}_+,$$

cf., [Cia88, Section 3.4]. Under this assumption we get a more precise form for the response function of a general isotropic elastic material as formulated in the following theorem.

Theorem 1.1.2. *Constitutive equation for an isotropic material*
Let the material be elastic, frame-indifferent and isotropic. Given an arbitrary deformation $\boldsymbol{\varphi} : \bar{\Omega} \to \mathbb{R}^d$*, then the second Piola-Kirchhoff stress tensor at a point* $\boldsymbol{x}$ *is given by*

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \hat{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{\nabla}\boldsymbol{\varphi}(\boldsymbol{x})) = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{C}(\boldsymbol{x})),$$

where the response function $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \cdot) : \mathbb{R}_{\mathrm{spd}}^{d \times d} \to \mathbb{R}_{\mathrm{sym}}^{d \times d}$ *is of the form*

$$\tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{C}) = \sum_{i=0}^{2} \gamma_i(\boldsymbol{x}, \operatorname{tr} \boldsymbol{C}, \operatorname{tr} \operatorname{Cof}(\boldsymbol{C}), \det \boldsymbol{C}) \boldsymbol{C}^i, \tag{1.6}$$

where $\gamma_i(\boldsymbol{C})$ *are scalar-valued functions of* $\boldsymbol{x}$ *and of the principal invariants of the Cauchy-Green strain tensor* $\boldsymbol{C}$*.*

Proof. See, e.g., [Cia88, Theorem 3.6.-2]. qed.

The theorem is a consequence of the Rivlin-Erickson representation theorem. The stress-strain relation (1.6) is in general highly nonlinear. From now on we assume the reference configuration to be in a natural state, i.e., $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{id}) = \boldsymbol{0}$. In the following theorem the stress-strain relation close to the reference configuration $\boldsymbol{C} = \boldsymbol{id}$ is investigated.

Theorem 1.1.3. *Let* Ω *be a reference configuration in natural state of an elastic, frame-indifferent and isotropic material. We assume the coefficients* γ_i*, appearing in (1.6), to be continuously differentiable with respect to the principal invariants of the Cauchy-Green strain tensor. Then there exist functions* $\lambda(\boldsymbol{x}), \mu(\boldsymbol{x}) : \bar{\Omega} \to \mathbb{R}$ *so that*

$$\boldsymbol{\Sigma}(\boldsymbol{x}) = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{x}, \boldsymbol{C}(\boldsymbol{x})) = \lambda(\boldsymbol{x}) \mathrm{tr} \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{id} + 2\mu(\boldsymbol{x}) \boldsymbol{E}(\boldsymbol{x}) + o(\boldsymbol{x}, \boldsymbol{E}(\boldsymbol{x}))$$

where $\boldsymbol{E} = \frac{1}{2}(\boldsymbol{C} - \boldsymbol{id})$ *is the Green-St. Venant strain tensor.*

Proof. See, e.g., [Cia88, Theorem 3.7-1]. qed.

If we further assume the material to be homogeneous, which means that the response function is independent of the point $\boldsymbol{x} \in \bar{\Omega}$, we get $\lambda(\boldsymbol{x}) = \lambda$ and $\mu(\boldsymbol{x}) = \mu$. These constants are called Lamé constants. Ignoring the terms of higher order we get the stress-strain relation of St. Venant-Kirchhoff materials

$$\tilde{\boldsymbol{\Sigma}}(\boldsymbol{E}) = \lambda \mathrm{tr} \boldsymbol{E}(\boldsymbol{x}) \boldsymbol{id} + 2\mu \boldsymbol{E}(\boldsymbol{x}). \tag{1.7}$$

Inserting (1.7) in (1.5) and exploiting $\boldsymbol{E}(\boldsymbol{x}) = \boldsymbol{E}(\boldsymbol{u}(\boldsymbol{x}))$, we get in terms of displacement field

$$\begin{aligned} -\operatorname{div}((\boldsymbol{id} + \boldsymbol{\nabla u})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{E}(\boldsymbol{u}))) &= \boldsymbol{f} \quad \text{in } \Omega \\ (\boldsymbol{id} + \boldsymbol{\nabla u})\tilde{\boldsymbol{\Sigma}}(\boldsymbol{E}(\boldsymbol{u}))\boldsymbol{n} &= \boldsymbol{\pi} \quad \text{on } \Gamma_N. \end{aligned} \tag{1.8}$$

By definition, St. Venant-Kirchhoff materials can only be used for small strains. Thus, it is common to linearize the Green-St. Venant strain tensor as well which leads to the linear elastic constitutive equation, see Section 1.1.2.
The deformation should be invertible and orientation preserving $\det \boldsymbol{\nabla\varphi}(\boldsymbol{x}) > 0$. But in the definition of the stress tensor this has not been considered. However, under the assumption of small deformations, $\det \boldsymbol{\nabla\varphi}(\boldsymbol{x})$ should be close to $\det \boldsymbol{id} = 1$.
We would expect an infinite stress if large deformations cause $\det \boldsymbol{\varphi}(\boldsymbol{x}) \to 0^+$. This property is usually expressed in terms of the so-called stored energy function which exists for hyperelastic materials.

Definition 1.1.1. *A material is called hyperelastic if there exists a stored energy function* $W : \bar{\Omega} \times \mathbb{R}^{d\times d}_+ \to \mathbb{R}$ *such that for the first Piola-Kirchhoff stress tensor the following relation is valid*

$$\boldsymbol{T}(\boldsymbol{x}) = \hat{\boldsymbol{T}}(\boldsymbol{x}, \boldsymbol{F}) = \frac{\partial W}{\partial \boldsymbol{F}}(\boldsymbol{x}, \boldsymbol{F}), \quad \forall \boldsymbol{x} \in \bar{\Omega}, \boldsymbol{F} \in \mathbb{R}^{d\times d}_+.$$

The functional $\mathcal{J}_{\text{elast}}(\boldsymbol{\psi}) := \int_\Omega W(\boldsymbol{x}, \boldsymbol{\nabla\psi}(\boldsymbol{x}))\, dx$ *defined for a smooth mapping* $\boldsymbol{\psi}$ *is called elastic energy and the functional*

$$\mathcal{J}(\boldsymbol{\psi}) := \int_\Omega W(\boldsymbol{x}, \boldsymbol{\nabla\psi}(\boldsymbol{x}))\, dx - \left(\int_\Omega \boldsymbol{f} \cdot \boldsymbol{\psi}\, dx + \int_{\Gamma_N} \boldsymbol{\pi} \cdot \boldsymbol{\psi}\, da \right) \tag{1.9}$$

is the total energy.

We note that a stationary point of the total energy is the solution of (1.3), cf., [Cia88, Theorem 4.1-1]. The relation between the second Piola-Kirchhoff stress tensor and the stored energy is given in [Cia88, Theorem 4.2-2] and the stored energy function for the special St. Venant Kirchhoff material can be found in [Cia88, p. 190].
For hyperelastic materials we assume an infinite energy caused by a deformation $\boldsymbol{\varphi}$ in order to annihilate volume

$$W(\boldsymbol{x}, \boldsymbol{\nabla\varphi}(\boldsymbol{x})) \to +\infty \quad \text{if} \quad \det \boldsymbol{\nabla\varphi} \to 0^+.$$

A stored energy function fulfilling this assumption is given for so-called Ogden materials

$$W(\boldsymbol{\nabla\varphi}) := \sum_{i=1}^{I} a_i (\operatorname{tr} \boldsymbol{\nabla\varphi}^T \boldsymbol{\nabla\varphi})^{\gamma_i/2} + \sum_{j=1}^{J} b_j \operatorname{tr}\operatorname{Cof}(\boldsymbol{\nabla\varphi}^T \boldsymbol{\nabla\varphi})^{\delta_j/2} + \bar{\Gamma}(\det \boldsymbol{\nabla\varphi})$$

with $a_i, b_j > 0, \gamma_i, \delta_j \geq 1$ and $\bar{\Gamma} : (0, \infty) \to \mathbb{R}$ is a convex function satisfying $\bar{\Gamma}(\delta) \to \infty$ for $\delta \to 0^+$. In [Cia88, pp. 189 ff.] further examples of stored energy functions are listed.

1.1.2 Linearized elasticity in $3D$ and $2D$

In linear elasticity small displacements and small strains are assumed allowing for linearization of the strain-displacement relation as well as of the stress-strain relation, thus simplifying the equations of motion. The magnitude of strains and displacements depends on the applied forces as well as on material properties. Iron steel which has a high elastic modulus ($E = 210kN/mm^2$) is typically modeled as linear elastic material.
In Section 1.1.1 the Green-St. Venant strain tensor $\boldsymbol{E}$ is given in terms of the displacements $\boldsymbol{u}$. For small displacements we may neglect the quadratic terms leading to the linearized strain tensor

$$\boldsymbol{\epsilon}(\boldsymbol{u}) := \frac{1}{2}\left(\boldsymbol{\nabla u}^T + \boldsymbol{\nabla u}\right). \tag{1.10}$$

Furthermore, in Section 1.1.1 we derived the stress-strain relation of St. Venant Kirchhoff materials which is designed for small strains. Inserting (1.10) in (1.7) we get the constitutive equation in linearized elasticity

$$\boldsymbol{\sigma}(\boldsymbol{\epsilon}(\boldsymbol{u})) = \lambda\, \mathrm{tr}\boldsymbol{\epsilon}(\boldsymbol{u})\, \boldsymbol{id} + 2\mu\boldsymbol{\epsilon}(\boldsymbol{u}) \tag{1.11}$$

which is called Hooke's law. We note that the strain and the stress tensor are both symmetric. The Lamé constants λ and μ are expressed in terms of the material constants Young's modulus E and Poisson ratio ν as follows

$$\lambda = \frac{E\nu}{(1+\nu)(1-2\nu)}, \qquad \mu = \frac{E}{2(1+\nu)}.$$

The Young's modulus also called tensile modulus or elastic modulus is the ratio of the uniaxial stress over the uniaxial strain. The Poisson ratio is the ratio of the lateral contraction to the longitudinal strain of a bar under pure extension. We refer to [Gur81, pp. 202 ff.] and [Cia88, pp. 123 ff.] for examples about the determination of material constants. By means of a fourth order material tensor $\boldsymbol{H}$, called Hooke's tensor with the entries

$$H_{ijkl} = \frac{E}{2(1+\nu)}(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + \frac{E\nu}{(1+\nu)(1-2\nu)}\delta_{ij}\delta_{kl} \tag{1.12}$$

we may reformulate (1.11) to

$$\sigma_{ij}(\boldsymbol{u}) = H_{ijkl}\epsilon_{kl}(\boldsymbol{u}), \quad 1 \le i,j,k,l \le 3. \tag{1.13}$$

Hooke's tensor is symmetric

$$H_{ijkl} = H_{jilk} = H_{lkij},$$

elliptic, i.e., there exists a constant a_0 such that

$$H_{ijkl}\zeta_{ij}\zeta_{kl} \ge a_0\zeta_{ij}\zeta_{kl}$$

for all symmetric $\{\zeta_{ij}\}_{i,j=1}^3$ and it is bounded, i.e., there exists a constant A_0

$$H_{ijkl}\zeta_{ij}\zeta_{kl} \le A_0\zeta_{ij}\zeta_{kl}.$$

for all symmetric $\{\zeta_{ij}\}_{i,j=1}^{3}$.
Replacing $\tilde{\boldsymbol{\Sigma}}$ by $\boldsymbol{\sigma}$ and approximating $\boldsymbol{\nabla\varphi} = \boldsymbol{id} + \boldsymbol{\nabla u}$ by $\boldsymbol{id}$ in (1.8) we get the equations of linearized elasticity

$$\begin{aligned} -\mathrm{div}\boldsymbol{\sigma}(\boldsymbol{u}) &= \boldsymbol{f} \quad \text{on } \Omega \\ \boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n} &= \boldsymbol{\pi} \quad \text{on } \Gamma_N. \end{aligned} \tag{1.14}$$

For a formal derivation of (1.14) by means of a linearization of the operator of nonlinear elasticity defined as the left hand sides of (1.5), we refer to [Cia88, Theorem 6.2-1].
The elastic energy of the linearized elastic problem caused by the displacements $\boldsymbol{u}$ is given by

$$\mathcal{J}_{\mathrm{elast}}(\boldsymbol{u}) = \frac{1}{2}\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{u})\,dx$$

where we use the Frobenius inner product $\boldsymbol{A} : \boldsymbol{B} := A_{ij}B_{ij}$. We note that the elastic energy defines a symmetric bilinear form $a(\boldsymbol{\psi}, \boldsymbol{\psi}) := \int_\Omega \boldsymbol{\sigma}(\boldsymbol{\psi}) : \boldsymbol{\epsilon}(\boldsymbol{\psi})\,dx$. The total energy corresponding to the displacements $\boldsymbol{u}$ is given by

$$\mathcal{J}(\boldsymbol{u}) = \frac{1}{2}a(\boldsymbol{u}, \boldsymbol{u}) - \left(\int_\Omega \boldsymbol{f}\cdot\boldsymbol{u}\,dx + \int_{\Gamma_N} \boldsymbol{\pi}\cdot\boldsymbol{u}\,da\right). \tag{1.15}$$

The displacement $\boldsymbol{u}$ fulfilling the equations (1.14) is a stationary point of the total energy (1.15).
Due to the symmetry of the linearized strain tensor (1.10) and the linearized stress tensor (1.13) we have only six different components in each of these tensors. Thus, it is more convenient to express the linearized strain and stress tensor as vectors with the entries

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{22} \\ \epsilon_{33} \\ 2\epsilon_{12} \\ 2\epsilon_{23} \\ 2\epsilon_{13} \end{pmatrix}, \qquad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{33} \\ \sigma_{12} \\ \sigma_{23} \\ \sigma_{13} \end{pmatrix}$$

and accordingly Hooke's tensor may be represented by a symmetric 6 × 6 matrix

$$\boldsymbol{H} = \begin{pmatrix} \lambda+2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda+2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda+2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}$$

such that

$$\boldsymbol{\sigma} = \boldsymbol{H}\boldsymbol{\epsilon}.$$

Each body can be considered as a three-dimensional object. However, for special problems like, e.g., plane strain or plane stress it is advantageous or even necessary to reduce the dimensionality. For a schedule of these special cases we refer to [Bat86, pp.160-161]. In the present work when dealing with $2D$ linear elasticity we refer to plane strain. The plane strain assumption is fulfilled if the three-dimensional body has a large extension in one direction, is fixed at the ends, and is exposed to forces which are perpendicular to this direction and do not vary along the length. Typical examples are retaining dams, culverts and tunnels. There is no displacement in the third coordinate direction and the strain components $\epsilon_{33}, \epsilon_{13}, \epsilon_{23}$ vanish. Thus, it is sufficient to consider an arbitrary cross section. In this two-dimensional case the strain and stress vectors and Hooke's tensor reduce to

$$\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_{11} \\ \epsilon_{22} \\ 2\epsilon_{12} \end{pmatrix}, \qquad \boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{pmatrix}, \qquad \boldsymbol{H} = \begin{pmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & \mu \end{pmatrix}. \tag{1.16}$$

1.2 Contact problems

The aim of this section is to state the Signorini problem and to give some examples for which the exact contact stresses are known. In Chapters 2, 3 and 4 we solely deal with the Signorini problem. However, as the one-body contact problem is a simplification we have to consider two-body contact problems when analyzing the physical properties of the dynamic system in Chapter 5. Therefore, we derive the linearized non-penetration condition based on the general formulation as two-body contact problem.

In Section 1.2.3 we present the closely related simplified Signorini problem and the obstacle problem for which we discuss existing residual-type a posteriori error estimators in Sections 2.2.2 and 2.2.3.

1.2.1 Contact constraints

When two-solid bodies Ω^S and Ω^M come into contact they deform, transmit forces but do not penetrate each other. As the actual contact zone depends on the deformations of the bodies, which are unknowns of the system we do not know in advance which points on Γ_C^S and Γ_C^M are coming into contact. In this section we primarily derive an approximation of the exact non-penetration condition. Further, we comment on frictional effects which may be considered additionally.

We assume a bijective mapping $\boldsymbol{\Phi} : \Gamma_C^S \to \Gamma_C^M$ to be given, which relates every point on Γ_C^S to a potential contact point on Γ_C^M. Such a mapping $\boldsymbol{\Phi}$ may be obtained by means of a closest point projection $\boldsymbol{\Phi}(\boldsymbol{x}) := \operatorname{argmin}_{y \in \Gamma_C^M} |\boldsymbol{x} - \boldsymbol{y}|$, see, e.g., [Lau02, Wri02, Chapter 3]. The distance between points $\boldsymbol{x}$ and $\boldsymbol{\Phi}(\boldsymbol{x})$ is given by

$$g(\boldsymbol{x}) := |\boldsymbol{x} - \boldsymbol{\Phi}(\boldsymbol{x})|$$

and the direction vector is

$$\boldsymbol{\nu}(\boldsymbol{x}) := \frac{\boldsymbol{x} - \boldsymbol{\Phi}(\boldsymbol{x})}{|\boldsymbol{x} - \boldsymbol{\Phi}(\boldsymbol{x})|}.$$

The relative displacement in normal direction is the difference in displacements on Γ_C^S and Γ_C^M

$$[\boldsymbol{u}] \cdot \boldsymbol{\nu} := \left(\boldsymbol{u}^S - \boldsymbol{u}^M \circ \boldsymbol{\Phi}\right) \cdot \boldsymbol{\nu}.$$

Here, we adopted the notation from the theory of mortar finite elements, where S and M stand for slave and master sides. With the above definitions we get a non-penetration condition

$$[u_\nu] \leq g, \tag{1.17}$$

where we used the abbreviation $u_\nu := \boldsymbol{u} \cdot \boldsymbol{\nu}$. The question under which conditions $\boldsymbol{\Phi}$ is a good choice or (1.17) is a good approximation of the exact non-penetration condition, respectively, has been answered in [Eck96, Lemma 1.3]. Under the following conditions

- the displacements and strains are small in both bodies, $|u_i(\boldsymbol{x})|, |\epsilon_{ij}(\boldsymbol{x})| < \varepsilon$
- the contact boundaries are close to each other, i.e., $|g(\boldsymbol{x})| < 2\varepsilon$
- the curvature of the potential boundaries is bounded
- $\boldsymbol{\nu}(\boldsymbol{x}) \approx \boldsymbol{\nu}^S(\boldsymbol{x}) \approx \boldsymbol{\nu}^M(\boldsymbol{x})$, where $\boldsymbol{\nu}^S(\boldsymbol{x})$ and $\boldsymbol{\nu}^M(\boldsymbol{x})$ are the unit outward normals of Γ_C^S and Γ_C^M

for all $\boldsymbol{x} \in \Gamma_C^S$, the non-penetration condition (1.17) defined by the bijective mapping $\boldsymbol{\Phi}$ is a good approximation in the sense that the error to the exact non-penetration condition is of order $\varepsilon^{3/2}$. Due to these assumptions (1.17) is called linearized non-penetration condition. We refer also to [KO88, Chapter 2] for the linearization of the exact non-penetration condition.

For theoretical and numerical investigation of multibody contact problems in linear elasticity we refer to [Wri02, Lau02, BGK87, BBHL99, WK03, HMW05, HW05, Kra09, DK09]. In this work we mostly deal with one-body contact problems, describing the contact of a deformable body with a rigid obstacle. The relative displacement $[\boldsymbol{u}]$ is replaced by $\boldsymbol{u}$ because the rigid obstacle does not move. Usually, the linearized non-penetration condition

$$u_\nu \leq g \tag{1.18}$$

is used if the material of the deformable body is assumed to be linear elastic because the assumption of small displacements and strains is made in linear elasticity as well as for the linearization of the non-penetration condition.

In the case of non-linear elasticity where the assumption of small displacements is not fulfilled, the linearized non-penetration condition is replaced by an incremental contact condition; see, e.g., [KO88, Chapter 2]. In each iteration step of the solution of the underlying non-linear equation the incremental contact condition is a linearized non-penetration condition like (1.18) with respect to the deformation of the previous iteration. In [Lau02] a closest point projection is used in each iteration step. Another possibility is to compute the signed distance function which is characterized by the Eikonal equation in a preprocess. Thus, in each iteration step the non-penetration condition is obtained by the evaluation of the signed distance function with respect to the

deformation of the previous iteration, [KM11]. Even in linear elasticity an incremental contact condition should be applied for complex geometries.
The non-penetration condition is a constraint, hence evoking constraining forces which we call contact forces. In terms of the boundary stresses $\hat{\boldsymbol{\sigma}}(\boldsymbol{u}) := \boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n}$ the linearized non-penetration condition evokes compressive stresses $\hat{\sigma}_\nu := \hat{\boldsymbol{\sigma}} \cdot \boldsymbol{\nu}$ in the direction $\boldsymbol{\nu}$ of the constraint along the actual contact boundary. We note that in the case of two-body contact the contact forces on both sides have to be equal, opposite and collinear due to Newton's axiom of action and reaction. The boundary stresses $\hat{\boldsymbol{\sigma}}_T := \hat{\boldsymbol{\sigma}} - \hat{\sigma}_\nu \cdot \boldsymbol{\nu}$ orthogonal to $\hat{\sigma}_\nu \cdot \boldsymbol{\nu}$ are assumed to be zero in an idealized contact problem where frictional effects are neglected. We refer to $\hat{\boldsymbol{\sigma}}_T$ as frictional or tangential contact stresses because $\boldsymbol{\nu}$ is assumed to be an approximation of the unit outward normal on Γ_C. In a more realistic contact situation frictional effects should be incorporated; see, e.g., [Lau02, Wri02, Eck96, KO88]. The most common models of friction are the Coulomb friction law

$$\begin{aligned} \boldsymbol{u}_T = 0 &\Rightarrow |\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u})| \leq \mathcal{F}|\hat{\sigma}_\nu(\boldsymbol{u})| \\ \boldsymbol{u}_T \neq 0 &\Rightarrow \hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}) = -\mathcal{F}|\hat{\sigma}_\nu(\boldsymbol{u})|\frac{\boldsymbol{u}_T}{|\boldsymbol{u}_T|} \end{aligned} \tag{1.19}$$

and the Tresca friction law where $\hat{\sigma}_\nu$ in (1.19) is replaced by a given function $\kappa(\boldsymbol{x})$. We call $\boldsymbol{u}_T$ tangential displacement corresponding to the tangential stresses. The coefficient of friction $\mathcal{F}$ has to be measured experimentally. It may depend on several factors, e.g., the material, the roughness of the surface, the local pressure, the temperature and the humidity. We refer to, e.g., [Wri02, Chapter 4.2.8] where friction laws are computed from the roughness on the microscale by means of homogenization techniques.

1.2.2 Signorini contact

In this section we state the strong formulation of the one-body contact problem in linear elasticity, called Signorini contact problem. Furthermore, we give some exact solutions which we use later for the evaluation of our numerical results in Chapter 4.
The Signorini contact problem describes the contact of a linear elastic body with a rigid obstacle. The non-penetration is modeled by the linearized non-penetration condition (1.18) and frictional effects are ignored. Thus, we get the boundary value problem from the equations of linear elasticity (1.14) and the primal and dual constraints presented in Section 1.2.1 and we add a Dirichlet boundary condition on Γ_D.

Problem 1.2.1. *Strong formulation of the Signorini contact problem*

$$-\operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}) = \boldsymbol{f} \quad \text{in } \Omega \tag{1.20}$$

$$\hat{\boldsymbol{\sigma}}(\boldsymbol{u}) = \boldsymbol{\pi} \quad \text{on } \Gamma_N \tag{1.21}$$

$$\boldsymbol{u} = \boldsymbol{u}_D \quad \text{on } \Gamma_D \tag{1.22}$$

$$u_\nu \leq g \quad \text{on } \Gamma_C \tag{1.23}$$

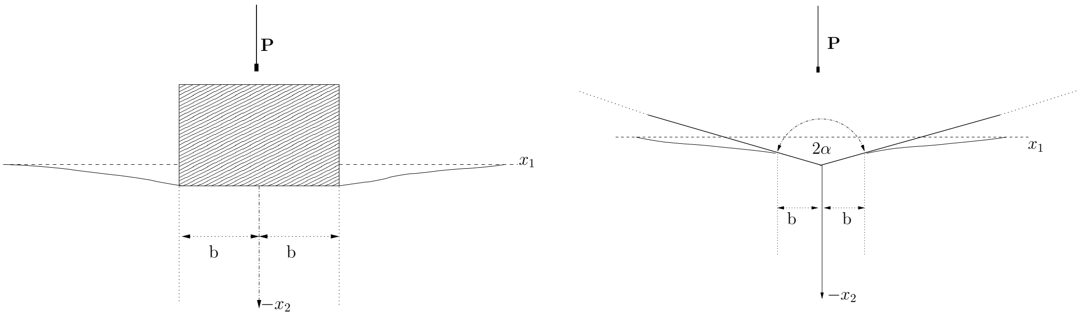
$$\hat{\sigma}_\nu(\boldsymbol{u}) \leq 0 \quad \text{on } \Gamma_C \tag{1.24}$$

$$(u_\nu - g) \cdot \hat{\sigma}_\nu(\boldsymbol{u}) = 0 \quad \text{on } \Gamma_C \tag{1.25}$$

$$\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}) = \boldsymbol{0} \quad \text{on } \Gamma_C \tag{1.26}$$

The boundary is subdivided in three pairwise disjoint parts, the Neumann boundary Γ_N which is an open subset of Γ, the Dirichlet boundary Γ_D which is a closed subset of Γ and the potential contact boundary Γ_C which is also a closed subset. As the boundary segments are disjoint, $\Gamma_D \cap \Gamma_C = \emptyset$. As usual, we assume that the actual contact boundary, where $u_\nu = g$, is a strict subset of the potential contact boundary. The condition (1.25) in Problem 1.2.1 is called complementarity condition and ensures that the compressive contact stresses (1.24) are zero if no contact occurs. We note that although we are dealing with linear elasticity and a linearized non-penetration condition Problem 1.2.1 is nonlinear and non-differentiable. This is due to the a priori unknown contact boundary.

Consequently, it is far from trivial to find an analytic solution of contact problems. For a special problem Hertz [Her81] derived the solution of contact stresses $\hat{\sigma}_\nu$ and normal displacements u_ν in the actual contact zone. In the engineering literature, cf., [Joh85], some more academic examples of contact problems with exact solutions of displacements on Γ_C or stresses may be found. In the following we present the Hertzian contact problem and two examples of contact between an elastic and a rigid body, taken from [Joh85]. In these examples the contact between rigid and elastic bodies is enforced by the application of a vertical load P on the top of the rigid body. In contrast, our model problem (Problem 1.2.1) stated above is a one-body contact problem where the rigid obstacle is motionless. However, the contact stresses in the contact zone are the same and are used for the evaluation of our numerical solutions in Sections 4.2.1 and 4.2.2.



(a) Rigid punch

(b) Rigid wedge

Figure 1.1: Indentation of elastic half-space

For the following two-dimensional examples we choose $\boldsymbol{\nu} = \boldsymbol{e}_2$, so that $\hat{\sigma}_2 = \hat{\sigma}_\nu$ and $x_T = x_1$. As first example we consider the indentation of a linear elastic two-dimensional half-space by a rigid flat punch, see Figure 1.1(a). In x_1-direction the punch ranges from $-b$ to b so that the distribution of the contact stresses is given by

$$\hat{\sigma}_2(\boldsymbol{x}) = \frac{P}{\pi(b^2 - x_1^2)^{\frac{1}{2}}}. \tag{1.27}$$

Theoretically, the pressure is infinite where the edges of the punch $x_1 = \pm b$ indent. For details about the derivation of (1.27) and formulas describing the motion of the elastic contact surface, we refer to [Joh85, Chapter 2.8]. In the second example a two-dimensional wedge indents a linear elastic half space, see [Joh85, Chapter 5.2]. The semi-angle α of the wedge has to be close to 90°, see Figure 1.1(b), such that the theory of linear elasticity is valid. The half of the width of the contact strip is

$$b = \frac{P(1-\nu^2)}{E\cot\alpha}, \tag{1.28}$$

where E and ν are the Young's modulus and the Poisson ratio, cf. Section 1.1.2, and the distribution of the contact stresses is given by

$$\hat{\sigma}_2(\boldsymbol{x}) = \frac{E\cot\alpha}{(1-\nu^2)2\pi}\ln\left(\frac{b+\left(b^2-x_1^2\right)^{\frac{1}{2}}}{b-\left(b^2-x_1^2\right)^{\frac{1}{2}}}\right). \tag{1.29}$$

The value of (1.29) is infinite at the apex of the wedge. For the three-dimensional indentation by a blunt cone, we refer to [Lov39, Sne48, HS45].
Originally, the Hertzian contact problem [Her81] describes the contact between two linear elastic spherical bodies of dimension $d = 3$ with radii R^S, R^M, Young's moduli E^S, E^M and Poisson ratio ν^S, ν^M . Here, we consider the extreme case, the contact of a linear elastic cube ($R^S = \infty$) and a rigid ball ($E^M = \infty$), see Figure 1.2 for a schematic view of a cross-section. We choose $\boldsymbol{\nu} = \boldsymbol{e}_3$, so that $\hat{\sigma}_3 = \hat{\sigma}_\nu$ and $\boldsymbol{x}_T = \boldsymbol{x} - x_3\boldsymbol{e}_3$. If the rigid ball

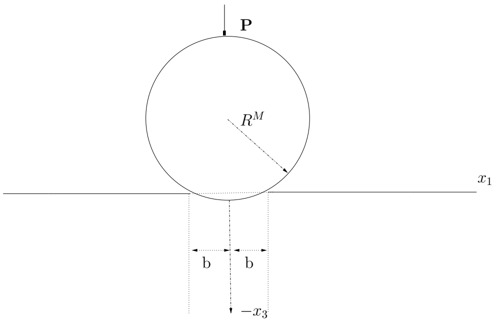


Figure 1.2: Cross-section of Hertzian contact

is pressed against the linear elastic cube by a load P, the radius of the actual contact zone is

$$b = \left(\frac{3PR^M\left(1-\left(\nu^S\right)^2\right)}{4E^S}\right)^{\frac{1}{3}}, \tag{1.30}$$

and the pressure distribution is

$$\hat{\sigma}_3(\boldsymbol{x}) = \frac{3P}{2\pi b^3}\left(b^2 - |\boldsymbol{x}_T|^2\right)^{\frac{1}{2}}, \tag{1.31}$$

such that the largest contact stress in the origin is

$$\hat{\sigma}_3(\mathbf{0}) = \frac{3P}{2\pi b^2}. \tag{1.32}$$

For more details about the Hertzian contact problem we refer to [Joh85, Chapter 3.4] and [Hub04, SH83, Yun91].

1.2.3 Simplified Signorini contact and obstacle problem

In this section we present two problems which are closely related to the Signorini problem, the simplified Signorini problem and the obstacle problem. The simplified Signorini problem is often used in the numerical analysis of contact problems, e.g., in the development of a posteriori error estimators, instead of the Signorini problem. The underlying partial differential equation of the simplified Signorini problem and the obstacle problem is the Poisson equation. Furthermore, the constraints imposed in the obstacle problem are valid in the whole domain. However, numerical methods applied to the obstacle problem may be used for contact problems, too.

Problem 1.2.2. *Simplified Signorini problem*

$$\begin{aligned} -\Delta u &= f && \text{in } \Omega \\ \boldsymbol{\nabla} u \cdot \boldsymbol{n} &= \pi && \text{on } \Gamma_N \\ u &= u_D && \text{on } \Gamma_D \\ u &\leq g && \text{on } \Gamma_C \\ \boldsymbol{\nabla} u \cdot \boldsymbol{n} &\leq 0 && \text{on } \Gamma_C \\ (u-g)(\boldsymbol{\nabla} u \cdot \boldsymbol{n}) &= 0 && \text{on } \Gamma_C \end{aligned}$$

Problem 1.2.3. *Obstacle problem*

$$\begin{aligned} -\Delta u &\geq f && \text{in } \Omega \\ u &\geq g && \text{in } \Omega \\ (u-g)(-\Delta u - f) &= 0 && \text{in } \Omega \\ u &= u_D && \text{on } \partial\Omega \end{aligned}$$

In both problems a solution $u : \Omega \to \mathbb{R}$ is sought-after. We note that for obstacle problems the Dirichlet value has to be compatible with the constraints, i.e., $u_D \geq g$. In the obstacle problem, cf., [Rod87, EGK08], the domain is divided in two regions where either the Poisson equation is solved or $u = g$. This is in contrast to the Signorini problem where the a priori unknown contact stresses influence the solution in the interior like Neumann boundary conditions. Thus, academic examples of exact solution can be constructed more easily; see, e.g., the examples in [NSV03, BC04].

The most common application of an obstacle problem is a membrane which is constrained to lie above an obstacle. However, problems of this kind appear in many other applications as, e.g., the pricing of options.

1.3 Weak formulation, existence and uniqueness

In this section we derive the weak formulation of the Signorini problem and introduce several important definitions. Moreover, we comment on existence and uniqueness results.

1.3.1 Weak formulation

This section introduces the different weak formulations of the Signorini problem as variational inequality, variational inclusion, minimization problem and saddle point formulation.

Let $\boldsymbol{u}$ be the solution of Problem 1.2.1 and $\boldsymbol{v}$ a trial function which is assumed to be sufficiently smooth and fulfilling the Dirichlet condition as well as the non-penetration condition $v_\nu \leq g$. Further, we assume $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega)$ and $\boldsymbol{\pi} \in \boldsymbol{L}^2(\Gamma_N)$. We multiply (1.20) with $\boldsymbol{v} - \boldsymbol{u}$ and integrate over Ω

$$-\int_\Omega \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}) \cdot (\boldsymbol{v} - \boldsymbol{u})\, dx = \int_\Omega \boldsymbol{f} \cdot (\boldsymbol{v} - \boldsymbol{u})\, dx.$$

Then, we apply Green's formula, exploit the Neumann boundary condition (1.21) and $\boldsymbol{v} - \boldsymbol{u} = \boldsymbol{0}$ on Γ_D

$$\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{v} - \boldsymbol{u})\, dx = \int_\Omega \boldsymbol{f} \cdot (\boldsymbol{v} - \boldsymbol{u})\, dx + \int_{\Gamma_C} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}) \cdot (\boldsymbol{v} - \boldsymbol{u})\, da + \int_{\Gamma_N} \boldsymbol{\pi} \cdot (\boldsymbol{v} - \boldsymbol{u})\, da.$$

The trial function is assumed to fulfill the non-penetration condition, so that

$$\hat{\sigma}_\nu(\boldsymbol{u})(v_\nu - u_\nu) = \hat{\sigma}_\nu(\boldsymbol{u})(v_\nu - g + g - u_\nu) \overset{(1.25)}{=} \hat{\sigma}_\nu(\boldsymbol{u})(v_\nu - g) \geq 0.$$

As we consider the frictionless case $\hat{\boldsymbol{\sigma}}_T = \boldsymbol{0}$, the integral over the contact boundary is greater or equal zero ($\int_{\Gamma_C} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}) \cdot (\boldsymbol{v} - \boldsymbol{u})\, da \geq 0$), so that the solution $\boldsymbol{u}$ fulfills the following variational inequality

$$\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{v} - \boldsymbol{u})\, dx \geq \int_\Omega \boldsymbol{f} \cdot (\boldsymbol{v} - \boldsymbol{u})\, dx + \int_{\Gamma_N} \boldsymbol{\pi} \cdot (\boldsymbol{v} - \boldsymbol{u})\, da. \tag{1.33}$$

The requirements of regularity of a solution of (1.33) are less than for the strong formulation where $\boldsymbol{u} \in \mathcal{C}^2(\bar{\Omega})$. In order to define the weak solution space we need some basic definitions and results, which can be found, e.g., in [KO88, Eck96]. We define for $p \in [1, \infty]$, $m \in \mathbb{N}_0$ the Sobolev space $H^{m,p}(\Omega)$ as the set of all functions $v \in L^p(\Omega)$ such that for every multi-index $\alpha = (\alpha_1, ..., \alpha_d) \in \mathbb{N}_0^d$ with $|\alpha| = \alpha_1 + ... + \alpha_d \leq m$, the weak partial derivative $D^\alpha v$ belongs to $L^p(\Omega)$

$$\begin{aligned} H^{m,p}(\Omega) :=& \{v \in L^p(\Omega) \mid \forall \alpha \in \mathbb{N}_0^d, |\alpha| \leq m \\ & \qquad \exists D^\alpha v \in L^p(\Omega)\ \forall \zeta \in \mathcal{C}_0^\infty(\Omega) \int_\Omega v D^\alpha \zeta \, dx = (-1)^{|\alpha|} \int_\Omega D^\alpha v \zeta \, dx \ \} \end{aligned}$$

where $D^\alpha \zeta := \partial_1^{\alpha_1} ... \partial_n^{\alpha_n} \zeta$.
We only use the case $p = 2$, hence we abbreviate $H^{m,2}$ with H^m. The space H^m equipped with the scalar product

$$\langle v, \varphi \rangle_{H^m(\Omega)} := \sum_{|\alpha| \leq m} \langle D^\alpha v, D^\alpha \varphi \rangle_{L^2(\Omega)} = \sum_{|\alpha| \leq m} \int_\Omega D^\alpha v \cdot D^\alpha \varphi \, dx \tag{1.34}$$

is a Hilbert space. The corresponding norm is

$$\|v\|_{H^m(\Omega)} = \left(\langle v, v \rangle_{L^2(\Omega)} + \sum_{0 < |\alpha| \leq m} \langle D^\alpha v, D^\alpha v \rangle_{L^2(\Omega)} \right)^{1/2} . \tag{1.35}$$

It is also possible to define Sobolev spaces H^s of fractional order $s \in \mathbb{R}_0^+$, $s = [m] + l$ with $0 < l < 1$ called Sobolev-Slobodětskij space; see, e.g., [Wlo82]. Sobolev spaces of fractional order often occur for functions defined on the boundary Γ of Ω. As we are mainly concerned with $H^1(\Omega)$ we give the definition of $H^s(\Gamma)$ for $0 < s < 1$

$$H^s(\Gamma) = \{v \in L^2(\Gamma) \mid \|v\|_{H^s(\Gamma)} < \infty\}$$

by means of the Slobodětskij norm

$$\|v\|_{H^s(\Gamma)} = \left(\|v\|_{L^2(\Gamma)}^2 + \int_\Gamma \int_\Gamma \frac{|v(x) - v(y)|^2}{|x - y|^{(d-1)+2s}} \, dx \, dy \right)^{1/2} . \tag{1.36}$$

The following theorem justifies the definition of weak boundary values as, e.g., Dirichlet values.

Theorem 1.3.1. *Trace theorem*
Let $\Omega \subset \mathbb{R}^d$ be a Lipschitz domain. The mapping

$$\begin{aligned} \mathcal{C}^{0,1}(\bar{\Omega}) &\to \mathcal{C}^{0,1}(\Gamma), \\ u &\mapsto u|_\Gamma, \end{aligned}$$

defined by $u|_\Gamma(\boldsymbol{x}) = u(\boldsymbol{x})$ for $\boldsymbol{x} \in \Gamma$ can be extended to

$$\begin{aligned} H^1(\Omega) &\to H^{\frac{1}{2}}(\Gamma), \\ u &\mapsto \Upsilon(u). \end{aligned}$$

Thus, it is

$$\|\Upsilon(u)\|_{H^{\frac{1}{2}}(\Gamma)} \leq C \|u\|_{H^1(\Omega)} ,$$

where the constant C depends on the domain.

Proof. See, e.g., [Ada75, Wlo82, Eva10]. qed.

As our solution spaces are d-dimensional we write them in bold characters $\boldsymbol{H}^1(\Omega) := (H^1(\Omega))^d$, $\boldsymbol{L}^2(\Omega) := (L^2(\Omega))^d$ and $\boldsymbol{H}^{1/2}(\Gamma) = (H^{1/2}(\Gamma))^d$. Let $\Upsilon|_{\Gamma_K}$ for $K \in \{D, N, C\}$ be the restrictions of Υ on Γ_K, then the subspace of all functions in $\boldsymbol{H}^1(\Omega)$ fulfilling the Dirichlet boundary condition is given by

$$\boldsymbol{\mathcal{H}} := \{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \mid \Upsilon|_{\Gamma_D}(\boldsymbol{v}) = \boldsymbol{u}_D\}.$$

Further, we need the subspace

$$\boldsymbol{\mathcal{H}}_0 := \{\boldsymbol{\varphi} \in \boldsymbol{H}^1(\Omega) \mid \Upsilon|_{\Gamma_D}(\boldsymbol{\varphi}) = \boldsymbol{0}\}$$

and its dual $\boldsymbol{\mathcal{H}}^*$. We emphasize that $\boldsymbol{\mathcal{H}}_0$ is the space of all functions $\boldsymbol{H}^1(\Omega)$ with $\Upsilon|_{\Gamma_D}(\boldsymbol{\varphi}) = \boldsymbol{0}$ on the subset Γ_D of Γ, not on the whole boundary Γ. Whenever it is clear from the context that the restriction to the boundary requires the trace operator we omit the special notation of the trace Υ.

The non-penetration condition is given in direction $\boldsymbol{\nu}$. As the absolute value of the direction of constraints $\boldsymbol{\nu}(\boldsymbol{x})$ is assumed to be $|\boldsymbol{\nu}(\boldsymbol{x})| = 1$, it follows from Hölder's inequality that $\boldsymbol{u}|_{\Gamma_C} \cdot \boldsymbol{\nu} \in H^{\frac{1}{2}}(\Gamma_C)$. Especially for frictional contact problems it is convenient to decompose $\boldsymbol{H}^{\frac{1}{2}}(\Gamma_C)$ in $H^{\frac{1}{2}}(\Gamma_C) \times \boldsymbol{H}_T^{\frac{1}{2}}(\Gamma_C)$ with $u_\nu = \boldsymbol{u}|_{\Gamma_C} \cdot \boldsymbol{\nu} \in H^{\frac{1}{2}}(\Gamma_C)$ and $\boldsymbol{u}_T = \boldsymbol{u}|_{\Gamma_C} - u_\nu \boldsymbol{\nu} \in \boldsymbol{H}_T^{\frac{1}{2}}(\Gamma_C)$. For a gap function $g \in H^{\frac{1}{2}}(\Gamma_C)$ we define the admissible set

$$\boldsymbol{\mathcal{K}} := \{\boldsymbol{v} \in \boldsymbol{\mathcal{H}} \mid v_\nu \leq g \text{ on } \Gamma_C\}$$

where the order relation $\leq$ has to be understood in the sense of almost everywhere, see [KO88, Chapter 5] for a precise definition.

The following theorem claims the existence of contact stresses in the dual $H^{-\frac{1}{2}}(\Gamma)$ of $H^{\frac{1}{2}}(\Gamma)$ fulfilling a generalization of Green's formula.

Theorem 1.3.2. *Trace theorem for stresses*
Let Ω be a Lipschitz domain. We define $\mathcal{S} := \{\boldsymbol{\sigma} \in \mathbb{R}^{3\times 3}_{\text{sym}} \mid \sigma_{ij} \in L^2(\Omega)$ and $\partial_j \sigma_{ij} \in L^2(\Omega)\}$. Then there exists a uniquely determined linear continuous mapping $\boldsymbol{\Pi}$ from $\mathcal{S}$ into $\boldsymbol{H}^{-\frac{1}{2}}(\Gamma)$ such that

$$\boldsymbol{\Pi}(\boldsymbol{\sigma}) = \hat{\boldsymbol{\sigma}} \quad \text{if } \boldsymbol{\sigma} \in \mathcal{C}^1(\bar{\Omega})$$

and such that the following generalized Green's formula holds for every $\boldsymbol{\sigma} \in \mathcal{S}$ and every $\boldsymbol{v} \in \boldsymbol{H}^1(\Omega)$

$$\int_\Omega \boldsymbol{\sigma} : \boldsymbol{\nabla} \boldsymbol{v} \, dx + \int_\Omega \operatorname{div} \boldsymbol{\sigma} \boldsymbol{v} \, dx = \langle \boldsymbol{\Pi}(\boldsymbol{\sigma}), \Upsilon(\boldsymbol{v}) \rangle_{-\frac{1}{2}, \frac{1}{2}, \Gamma}$$

where $\langle \cdot, \cdot \rangle_{-\frac{1}{2}, \frac{1}{2}, \Gamma}$ denotes the duality pairing on $\boldsymbol{H}^{-\frac{1}{2}}(\Gamma) \times \boldsymbol{H}^{\frac{1}{2}}(\Gamma)$.

Proof. See, e.g., [KO88]. qed.

We write $\hat{\boldsymbol{\sigma}}$ instead of $\boldsymbol{\Pi}(\boldsymbol{\sigma})$ if it is clear from the context that $\hat{\boldsymbol{\sigma}} \in \boldsymbol{H}^{-\frac{1}{2}}(\Gamma)$. Similar to the decomposition of $H^{\frac{1}{2}} \times \boldsymbol{H}_T^{\frac{1}{2}}$ it is possible to decompose the dual space $H^{-\frac{1}{2}} \times \boldsymbol{H}_T^{-\frac{1}{2}}$.

From now on, we avoid the intricate notation of norms and scalar products with the spaces in the subindex. The L^2-norm and its scalar product are denoted by $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ without any subindex. The duality pairing between H^1 and its dual H^{-1} is given by $\langle\cdot,\cdot\rangle_{-1,1}$ and the corresponding norms are $\|\cdot\|_1$ and $\|\cdot\|_{-1}$. The duality pairing between $H^{\frac{1}{2}}$ and its dual $H^{-\frac{1}{2}}$ is denoted by $\langle\cdot,\cdot\rangle_{-\frac{1}{2},\frac{1}{2}}$ and the corresponding norms are $\|\cdot\|_{\frac{1}{2}}$ and $\|\cdot\|_{-\frac{1}{2}}$. Restrictions to subdomains are indicated by a further subindex, e.g., $\|\cdot\|_{1,\omega}$ for $\omega \subset \Omega$. Further, we recall the definition of the symmetric bilinear form

$$a(\cdot,\cdot) := \int_\Omega \boldsymbol{\sigma}(\cdot) : \boldsymbol{\epsilon}(\cdot)\, dx \tag{1.37}$$

which corresponds to the elastic energy, compare Section 1.1.2. Now we have defined everything to state the weak formulation of the Signorini problem.

Problem 1.3.1. *Variational inequality of the Signorini problem*
Let $\boldsymbol{f} \in \boldsymbol{L}^2(\Omega)$, $\boldsymbol{\pi} \in \boldsymbol{L}^2(\Gamma_N)$*. Then we seek a solution* $\boldsymbol{u} \in \boldsymbol{\mathcal{K}}$*, such that*

$$a(\boldsymbol{u}, \boldsymbol{v}-\boldsymbol{u}) \geq \langle \boldsymbol{f}, \boldsymbol{v}-\boldsymbol{u}\rangle + \langle \boldsymbol{\pi}, \boldsymbol{v}-\boldsymbol{u}\rangle_{\Gamma_N} \quad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{K}}. \tag{1.38}$$

It follows from Korn's inequality (Lemma 1.3.1) and from the properties of Hooke's tensor, see Section 1.1.2, that $a(\cdot,\cdot)$ is elliptic.

Lemma 1.3.1. *Korn's inequality*
Let Ω *be a Lipschitz domain and* $\operatorname{meas}(\Gamma_D) > 0$*, then there exists a positive constant* $c > 0$ *which depends on the domain, so that*

$$\int_\Omega \boldsymbol{\epsilon}(\boldsymbol{v}) : \boldsymbol{\epsilon}(\boldsymbol{v})\, dx \geq c\|\boldsymbol{v}\|_1^2.$$

Due to the ellipticity of $a(\cdot,\cdot)$ it is possible to write the weak formulation 1.3.1 as constrained minimization problem.

Problem 1.3.2. *Minimization problem of Signorini contact*
Let $\mathcal{J}(\cdot)$ *be the energy functional, defined in (1.15). Then we seek a solution* $\boldsymbol{u} \in \boldsymbol{\mathcal{K}}$ *such that*

$$\mathcal{J}(\boldsymbol{u}) = \min_{\boldsymbol{v}\in\boldsymbol{\mathcal{K}}} \mathcal{J}(\boldsymbol{v}). \tag{1.39}$$

We define the indicator functional

$$\mathcal{I}_{\mathcal{K}}(\boldsymbol{v}) = \mathcal{I}_{\mathcal{K}}(v_\nu) := \begin{cases} 0\,, & \text{if } \boldsymbol{v} \in \boldsymbol{\mathcal{K}}, \\ \infty\,, & \text{else,} \end{cases} \tag{1.40}$$

which is lower semi-continuous and convex because $\boldsymbol{\mathcal{K}}$ is a convex set. It leads to a further characterization of the solution of Problem 1.3.1 as the minimum of

$$(\mathcal{J} + \mathcal{I}_{\mathcal{K}})(\boldsymbol{u}) = \min_{\boldsymbol{v}\in\boldsymbol{\mathcal{H}}} (\mathcal{J} + \mathcal{I}_{\mathcal{K}})(\boldsymbol{v}) \tag{1.41}$$

over the whole space $\boldsymbol{\mathcal{H}}$. In [ET76, Proposition 2.2] it is proven that the solution of (1.39) solves the variational inequality

$$a(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) + \mathcal{I}_{\mathcal{K}}(\boldsymbol{v}) - \mathcal{I}_{\mathcal{K}}(\boldsymbol{u}) \geq \langle \boldsymbol{f}, \boldsymbol{v} - \boldsymbol{u} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{v} - \boldsymbol{u} \rangle_{\Gamma_N} \quad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{H}}. \tag{1.42}$$

The indicator functional is not differentiable but we may define the subdifferential of $\mathcal{I}_{\mathcal{K}}$.

Definition 1.3.1. *Subdifferential*
Let H be a Hilbert space and $\vartheta : H \to \mathbb{R} \cup \{-\infty, \infty\}$. Further, let $u \in H$ with $\vartheta(u) \neq \pm\infty$. Then the subdifferential of ϑ in u is given by

$$\partial^{\text{sub}} \vartheta(u) = \{u^* \in H^{-1} \mid \vartheta(v) - \vartheta(u) \geq u^*(v - u) \; \forall v \in H\}.$$

Exploiting the definition of the subdifferential we can reformulate the variational inequality (1.42) as variational inclusion

$$0 \in a(\boldsymbol{u}, \boldsymbol{\varphi}) - \langle \boldsymbol{f}, \boldsymbol{\varphi} \rangle - \langle \boldsymbol{\pi}, \boldsymbol{\varphi} \rangle_{\Gamma_N} + \partial^{\text{sub}} \mathcal{I}_{\mathcal{K}}(\boldsymbol{u})(\boldsymbol{\varphi}) \quad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0 \tag{1.43}$$

where

$$\partial^{\text{sub}} \mathcal{I}_{\mathcal{K}}(u_\nu) = \begin{cases} 0\,, & \text{if } u_\nu < g\,, \\ [0, \infty)\,, & \text{if } u_\nu = g. \end{cases}$$

We define $\boldsymbol{\lambda} \in \partial^{\text{sub}} \mathcal{I}_{\mathcal{K}}(\boldsymbol{u})$ as the element of the subdifferential fulfilling the variational inclusion

$$\langle \boldsymbol{f}, \boldsymbol{\varphi} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi} \rangle_{\Gamma_N} - a(\boldsymbol{u}, \boldsymbol{\varphi}) = \langle \boldsymbol{\lambda}, \boldsymbol{\varphi} \rangle_{-1,1} \quad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0. \tag{1.44}$$

From a physical point of view $\boldsymbol{\lambda}$ has the meaning of a constraining force density which we call contact force density. Comparing (1.44) with the generalization of Green's formula (Theorem 1.3.2) we get

$$\langle \boldsymbol{\lambda}, \boldsymbol{\varphi} \rangle_{-1,1} = - \langle \hat{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{\varphi} \rangle_{-\frac{1}{2},\frac{1}{2}}\,.$$

For details about the equivalent formulations as variational inequalities (1.38), (1.42), variational inclusion (1.43) or minimization problems (1.39), (1.41) we refer to [ET76, Chapter 2].

It is also possible to give a saddle point formulation of contact problems. Therefore, we define the space of positive Lagrange multipliers

$$\mathcal{M}^+ := \{\psi \in H^{-\frac{1}{2}}(\Gamma_C) \mid \langle \psi, v_\nu \rangle_{-\frac{1}{2},\frac{1}{2}} \geq 0 \quad \forall v_\nu \in H^{\frac{1}{2}}(\Gamma_C),\; v_\nu \geq 0\}$$

and the bilinear form

$$c(\psi, \boldsymbol{v}) := \langle \psi, v_\nu \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C}\,.$$

The saddle point

$$(\boldsymbol{u}, \mu) \in \boldsymbol{\mathcal{H}} \times \mathcal{M}^+, \quad \mathcal{L}(\boldsymbol{u}, \psi) \leq \mathcal{L}(\boldsymbol{u}, \mu) \leq \mathcal{L}(\boldsymbol{v}, \mu) \quad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{H}},\; \forall \psi \in \mathcal{M}^+$$

of the Lagrangian

$$\mathcal{L} : \boldsymbol{\mathcal{H}} \times \mathcal{M}^+ \to \mathbb{R}, \quad \mathcal{L}(\boldsymbol{v}, \psi) := \mathcal{J}(\boldsymbol{v}, \boldsymbol{v}) + c(\psi, \boldsymbol{v}) - \langle \psi, g \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C}$$

is the solution of the following system of equations

Problem 1.3.3. *Saddle point formulation of the Signorini contact*
Find $\boldsymbol{u} \in \boldsymbol{\mathcal{H}}$ *and* $\mu \in \mathcal{M}^+$*, such that*

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{\varphi}) + c(\mu, \boldsymbol{\varphi}) = &\quad \langle \boldsymbol{f}, \boldsymbol{\varphi} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi} \rangle_{\Gamma_N} \qquad && \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0 \\ c(\psi, \boldsymbol{u}) \leq &\quad \langle \psi, g \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C} && \forall \psi \in \mathcal{M}^+. \end{aligned}$$

In [HHN96, Theorem 9.4] is proven that the solutions $\boldsymbol{u}$ of Problems 1.3.3 and 1.3.1 are equivalent and that $\mu = -\hat{\sigma}_\nu(\boldsymbol{u})$.

1.3.2 Existence and uniqueness

The existence and uniqueness of the solution of the weak formulation of the Signorini problem follows from the Theorem of Lions and Stampacchia [LS67].

Theorem 1.3.3. *Theorem of Lions and Stampacchia*
Let $\boldsymbol{X}$ *be a Hilbert space and* $\boldsymbol{Y} \subset \boldsymbol{X}$ *non-empty, closed and convex. If the bilinear form* $b : \boldsymbol{X} \times \boldsymbol{X} \to \mathbb{R}$ *is continuous and elliptic and the linear form* $\boldsymbol{F} : \boldsymbol{X} \to \mathbb{R}$ *is continuous, then the variational inequality*

$$b(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) \geq \boldsymbol{F}(\boldsymbol{v} - \boldsymbol{u}) \quad \forall \boldsymbol{v} \in \boldsymbol{Y} \tag{1.45}$$

has a unique solution which depends continuously on $\boldsymbol{F}$.

Proof. See, e.g., [KS80, Theorem 2.1]. qed.

In the case of the Signorini problem the Hilbert space is $\boldsymbol{\mathcal{H}}$ and the closed and convex subset is the admissible set $\boldsymbol{\mathcal{K}}$. The bilinear form $b(\cdot,\cdot)$ in (1.45) is replaced by $a(\cdot,\cdot)$, defined in (1.37), and we set $\boldsymbol{F}(\boldsymbol{v} - \boldsymbol{u}) = \langle \boldsymbol{f}, \boldsymbol{v} - \boldsymbol{u} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{v} - \boldsymbol{u} \rangle_{\Gamma_N}$. From Korn's inequality follows that the bilinear form $a(\cdot,\cdot)$ used in (1.38) is elliptic. Further, $a(\cdot,\cdot)$ is continuous due to the boundedness of Hooke's tensor, see Section 1.1.2. Thus, it follows from Theorem 1.3.3 that the Problem 1.3.1 has a unique solution. Regularity results have been proven in [Sch89, Kin81] under stronger assumptions on the smoothness of the domain and the data. For the case of two-body contact problems we refer to, e.g., [BGK87].

The weak formulations of the simplified Signorini problem and the obstacle problem give rise to the same type of variational inequalities, hence the existence and uniqueness follows from Theorem 1.3.3, too; see, e.g., [Rod87].

Variational inequalities like (1.45) are called elliptic variational inequalities of the first kind whereas (1.42) is an elliptic variational inequality of the second kind. The weak formulation of the Signorini problem with Tresca friction, see Section 1.2 gives rise to an elliptic variational inequality of the second kind. The indicator functional in (1.42) is replaced by the friction potential. The existence and uniqueness follow from another Theorem of Lions and Stampachia; see, e.g., [LS67, Glo84]. If the frictional effects are described by Coulomb friction instead of Tresca friction we get a quasi-variational inequality. Under the assumption of a sufficiently small coefficient of friction $\mathcal{F}$ the existence of a solution has been proven in [Eck96, NJH80]. In [NJH80] the quasi-variational inequality has been perceived as fixed point mapping which is of special interest for numerical solution methods. In each iteration step a problem with Tresca friction has to be solved and under further assumptions on the mesh size even the uniqueness of the discrete solution can be proven, see [Has83, LPR91].

We note that the case of non-linear elastic materials is more involved even without constraints because the energy (1.9) is usually non-convex. An existence result of a one-body contact problem in non-linear elasticity with exact non-penetration condition can be found in [Sch02].

2 Finite element approximation and error estimation

This chapter deals with the discrete approximation by means of linear finite elements. The accuracy of the discrete solution depends highly on the underlying mesh. Thus, the construction of a posteriori error estimators enabling an appropriate choice of a mesh is essential for the numerical approximation. We here focus on residual-type a posteriori error estimators which are favorite due to the easy computation of the error estimator contributions. Unfortunately, the residual is no adequate measure of the error in the case of variational inequalities in contrast to linear equations. In Sections 2.2.2 and 2.2.3 we discuss existing residual-type error estimators for obstacle and contact problems.

2.1 Discrete formulation of the Signorini problem

We assume a polygonal domain resolved by a mesh $\mathfrak{m}$, consisting of elements $\mathfrak{e} \in \mathfrak{m}$ which are triangles or quadrilaterals in $2D$ and tetrahedra, hexahedra, prism or pyramids in $3D$. Any two elements are either disjoint or share a node, an edge or a face. The boundary segments $\Gamma_D, \Gamma_C, \Gamma_N$ are resolved by the mesh, meaning that their boundaries $\partial\Gamma_C, \partial\Gamma_N, \partial\Gamma_D$ are either nodes or edges. The set of nodes $\boldsymbol{p}$ is given by $N_\mathfrak{m}$ and we distinguish between the set $N_\mathfrak{m}^D$ of nodes on the Dirichlet boundary, the set $N_\mathfrak{m}^N$ of nodes at the Neumann boundary, the set $N_\mathfrak{m}^C$ of nodes at the potential contact boundary and the set of interior nodes $N_\mathfrak{m}^I$ where $\boldsymbol{p} \in \Omega$. Let $h_\mathfrak{e} := \operatorname{diam}(\mathfrak{e})$ be the diameter of an element and $\varrho_\mathfrak{e}$ the radius of the incircle of an element. The minimal angle of an element should be bounded away from zero. Thus, there exists a constant $\varsigma := \max_{\mathfrak{e}\in\mathfrak{m}} \frac{h_\mathfrak{e}}{\varrho_\mathfrak{e}} \ll \infty$ which is a measure of the shape-regularity. A family of meshes $\{\mathfrak{m}_j\}_{j=0}^{\infty}$ is called shape-regular if the constants of shape-regularity ς_j for all meshes $\mathfrak{m}_j$ are uniformly bounded, i.e., $c := \sup_j \varsigma_j \ll \infty$. We call $h := \max_{\mathfrak{e}\in\mathfrak{m}} h_\mathfrak{e}$ the meshsize of the mesh $\mathfrak{m}$.
For the finite element approximation of $\boldsymbol{\mathcal{H}}$ one uses continuous and piecewise polynomial functions with respect to the underlying mesh $\mathfrak{m}$. In this work we consider approximations by means of linear finite elements. The space with incorporated Dirichlet values $\boldsymbol{u}_D$ is

$$\boldsymbol{\mathcal{H}}_\mathfrak{m} := \{\boldsymbol{v}_\mathfrak{m} \in \boldsymbol{\mathcal{C}}^0(\bar{\Omega}) \mid \forall\mathfrak{e} \in \mathfrak{m},\ \boldsymbol{v}_\mathfrak{m}|_\mathfrak{e} \in \boldsymbol{P}_1(\mathfrak{e}) \text{ and } \boldsymbol{v}_\mathfrak{m}(\boldsymbol{p}) = \boldsymbol{u}_D(\boldsymbol{p})\ \forall \boldsymbol{p} \in N_\mathfrak{m}^D\}.$$

Further, we need the space of all linear finite elements which are zero on the Dirichlet boundary

$$\boldsymbol{\mathcal{H}}_{\mathfrak{m},0} := \{\boldsymbol{\varphi}_\mathfrak{m} \in \boldsymbol{\mathcal{C}}^0(\bar{\Omega}) \mid \forall\mathfrak{e} \in \mathfrak{m},\ \boldsymbol{\varphi}_\mathfrak{m}|_\mathfrak{e} \in \boldsymbol{P}_1(\mathfrak{e}) \text{ and } \boldsymbol{\varphi}_\mathfrak{m}(\boldsymbol{p}) = \boldsymbol{0}\ \forall \boldsymbol{p} \in N_\mathfrak{m}^D\}$$

and its dual $\boldsymbol{\mathcal{H}}_{\mathfrak{m}}^*$. The nodal basis functions of the finite element spaces are denoted by ϕ_p hence a discrete vector quantity has the representation

$$\boldsymbol{v}_{\mathfrak{m}} = \sum_{p \in N_{\mathfrak{m}}} v_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p \boldsymbol{e}_i$$

where we used the Einstein summation convention for repeated indices. We define the discrete admissible set by pointwise constraints

$$\boldsymbol{\mathcal{K}}_{\mathfrak{m}} := \{\boldsymbol{v}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m}} \mid v_{\mathfrak{m},\nu}(\boldsymbol{p}) \leq g_{\mathfrak{m}}(\boldsymbol{p}) \; \forall \boldsymbol{p} \in N_{\mathfrak{m}}^C\} \tag{2.1}$$

where $v_{\mathfrak{m},\nu}(\boldsymbol{p}) := \boldsymbol{v}_{\mathfrak{m}}(\boldsymbol{p}) \cdot \boldsymbol{\nu}(\boldsymbol{p})$ and the discrete gap function $g_{\mathfrak{m}}$ is a finite element approximation of g on Γ_C.

Problem 2.1.1. *Discrete variational inequality of the Signorini problem*
Find $\boldsymbol{u}_{\mathfrak{m}} \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}$ fulfilling the variational inequality

$$a(\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}) \geq \langle \boldsymbol{f}, \boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}} \rangle_{\Gamma_N} \quad \forall \boldsymbol{v}_{\mathfrak{m}} \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}.$$

The discrete problem formulation of the obstacle problem follows in the same way, except the constraints are enforced at all nodes in the domain. The unique solvability of the discrete Signorini problem (Problem 2.1.1) as well as of the discrete obstacle problem follows just as in the continuous case from Theorem 1.3.3 with the spaces $\boldsymbol{X} = \boldsymbol{\mathcal{H}}_{\mathfrak{m}}$ and $\boldsymbol{Y} = \boldsymbol{\mathcal{K}}_{\mathfrak{m}}$. For the discretization of variational inequalities and convergence results $\boldsymbol{u} \overset{h \to 0}{\longrightarrow} \boldsymbol{u}_{\mathfrak{m}}$ we refer to, e.g., [Glo84, KO88, HHN96]. A priori error estimates have been proven in, e.g., [BHR77, HH81, BBHL99, Hil00, CHLS02, HL02, HW05] for a sufficiently regular continuous solution $\boldsymbol{u}$. Further assumptions concern the regularity of the domain and data, the uniformity of the family of meshes and the choice of the Lagrange multiplier space. Most of these results are for two-body contact problems and are based on the saddle point formulation giving the discretization error in terms of both unknowns, the displacement and the Lagrange multiplier.

Problem 2.1.2. *Discrete saddle point formulation of the Signorini problem*
Find $\boldsymbol{u}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m}}$ and $\mu_{\mathfrak{m}} \in \mathcal{M}_{\mathfrak{m}}^+$, such that

$$\begin{aligned} a(\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}}) + c(\mu_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}}) = & \quad \langle \boldsymbol{f}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{\Gamma_N} && \forall \boldsymbol{\varphi}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m},0} \\ c(\psi_{\mathfrak{m}}, \boldsymbol{u}_{\mathfrak{m}}) \leq & \quad \langle \psi_{\mathfrak{m}}, g_{\mathfrak{m}} \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C} && \forall \psi_{\mathfrak{m}} \in \mathcal{M}_{\mathfrak{m}}^+ \end{aligned}$$

where $c(\psi_{\mathfrak{m}}, \boldsymbol{u}_{\mathfrak{m}}) := \langle \psi_{\mathfrak{m}}, u_{\mathfrak{m},\nu} \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C}$.

There are different choices for the discretization of the Lagrange multiplier $\mu_{\mathfrak{m}}$, by means of piecewise constant or linear functions or by so called dual basis functions ξ_p, fulfilling the biorthogonality property

$$\int_{\Gamma_C} \xi_p \phi_q \, da = \delta_{pq} \int_{\Gamma_C} \phi_q \, da \quad \forall \boldsymbol{p}, \boldsymbol{q} \in N_{\mathfrak{m}}^C.$$

The dual Lagrange multipliers have been introduced in [Woh00] for mortar finite elements and applied to two body contact in [WK03]. In [HW05] an a priori error estimate is given for two-body contact problems based on dual Lagrange multipliers. Therein $\boldsymbol{\nu}$ is assumed to be constant on Γ_C and the space of suitable Lagrange multipliers is

$$\mathcal{M}_{\mathfrak{m}}^{+} := \left\{ \psi_{\mathfrak{m}} = \sum_{p \in N_{\mathfrak{m}}^{C}} \psi_{\mathfrak{m}}(\boldsymbol{p}) \xi_p \mid \langle \psi_{\mathfrak{m}}, \varphi_{\mathfrak{m},\nu} \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C} \geq 0 \ \text{ if } \varphi_{\mathfrak{m},\nu}|_{\Gamma_C} \geq 0 \right\}.$$

We note that due to the biorthogonality property the weak constraint $c(\psi_{\mathfrak{m}}, \boldsymbol{u}_{\mathfrak{m}}) \leq \langle \psi_{\mathfrak{m}}, g_{\mathfrak{m}} \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C}$ equals the pointwise non-penetration condition (2.1) for $g_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m}}$. The domain is assumed to be polyhedral and the number of points on Γ_C where a change between actual and non-actual contact occurs is assumed to be finite. Then the following optimal a priori error estimate holds for $\boldsymbol{u} \in \boldsymbol{H}^{\frac{3}{2}+\beta}(\Omega)$, $0 < \beta \leq \frac{1}{2}$

$$\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}_j}\|_1 + \|\mu - \mu_{\mathfrak{m}_j}\|_{-\frac{1}{2},\Gamma_C} \leq C h_j^{\frac{1}{2}+\beta} |\boldsymbol{u}|_{\frac{3}{2}+\beta}$$

where $|\boldsymbol{u}|_{\frac{3}{2}+\beta}$ is the semi-norm. For an a priori error estimate for contact problems with Coulomb friction we refer to [HR07].

2.2 Residual-type a posteriori error estimators

Usually, in the numerical simulation of physical problems no information about the regularity of the solution is given a priori. The solution may be more or less regular in different regions of the domain and even singularities may occur. Increasing the number of grid points near the critical region improves the accuracy. The detection of those critical regions is feasible due to a posteriori error estimators which do not require regularity assumptions on the solution as a priori error estimators.

A posteriori error estimators should be reliable and efficient. Reliable means that the error estimator gives an upper bound of the error. This upper bound is global which is due to the fact that upper bounds involve the inverse of the differential operator which is a global operator. In an adaptive mesh-refinement process the accuracy of the solution can be estimated by a reliable a posteriori error estimator. In order to ensure that the error estimator does not overestimate the error so that the accuracy is achieved with an almost minimal amount of grid points, the error estimator has to be a lower bound of the error, too. Such error estimators are called efficient. The lower bound should be local in order to reproduce the spatial distribution of the error.

We concentrate on the topic of a posteriori error estimators. For theoretical details about the adaptive strategy, i.e., marking strategy, mesh refinement techniques and convergence results for linear elliptic PDE's we refer to [Ver96, SS05, NSV09] and the references cited therein. Convergence results for obstacle problems can be found in, e.g., [BCH07, PP10].

An overview over different types of error estimators for linear elliptic boundary value problems can be found in [Ver96, Ver98, AO00]. The most common error estimator is the residual error estimator which is obtained by an L^2-representation of the residual. It is an explicit error estimator which can be easily computed from the given data and the discrete solution. In Section 2.2.1 we present the proof of upper and lower bound of the residual error estimator for the equation of linear elasticity. The extension of residual error estimators to variational inequalities is not straightforward because for it the linear residual is no measure of the error. We will explain this difficulty in detail in Section 2.2.2. However, error estimators of residual type have been derived for obstacle problems in [CN00, Vee01, NSV03, NSV05, MNvPZ07] and for contact problems in [HN05, HN07] which are presented in Sections 2.2.2 and 2.2.3.

Another explicit error estimator, often used in engineering application, is given by the difference between $\nabla \boldsymbol{u}_{\mathfrak{m}}$ and a better approximation of $\nabla \boldsymbol{u}$ which is usually obtained by averaging techniques. Such error estimators have been analyzed for obstacle problems in [BC04].

A posteriori error estimators are called implicit if they require the solution of auxiliary discrete problems which should be simpler to evaluate than the original problem; see, e.g., [AO00, Chapter 6]. Examples of obstacle problems can be found in [WW10, BHS08] and for contact problems in [Woh07, WW09].

Hierarchical error estimators evaluate the residual of the computed finite element solution with respect to another finite element space corresponding to higher order elements or to a refined grid. In order to reduce the computational costs a hierarchical splitting of the enhanced finite element space is used. Hierarchical error estimators have been developed for obstacle problems in [Kor96, KZ11, ZVKG11], see also [SV07].

We here focus on residual-type a posteriori error estimators.

2.2.1 Residual error estimator for linear elliptic problems

The aim of this section is to explain the idea of residual error estimators for linear elliptic problems and to get familiar with the notation and the technical details of the proofs of upper and lower bound. In view of the new residual-type a posteriori error estimator for contact problems which is presented in Chapter 3 we consider the equation of linear elasticity. The continuous solution solves the variational equation

$$a(\boldsymbol{u}, \boldsymbol{\varphi}) = \langle \boldsymbol{f}, \boldsymbol{\varphi} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi} \rangle_{\Gamma_N} \quad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0 \tag{2.2}$$

and in the discrete setting the equation

$$a(\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}}) = \langle \boldsymbol{f}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{\Gamma_N} \quad \forall \boldsymbol{\varphi}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m},0} \tag{2.3}$$

has to be solved. We define $\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}} \in \boldsymbol{\mathcal{H}}^*$ as the residual

$$\begin{aligned} \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}, \boldsymbol{\varphi} \right\rangle_{-1,1} &:= \langle \boldsymbol{f}, \boldsymbol{\varphi} \rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi} \rangle_{\Gamma_N} - a(\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}) \\ &= a(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}) \end{aligned} \tag{2.4}$$

of the linear equation and call it linear residual. The corresponding norm of $\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}$ is given by

$$\|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}\|_{-1} := \sup_{\boldsymbol{\varphi}\in\boldsymbol{\mathcal{H}}_0} \frac{\left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}, \boldsymbol{\varphi}\right\rangle_{-1,1}}{\|\boldsymbol{\varphi}\|_1}.$$

It follows from the positive definiteness of Hooke's tensor and from Korn's inequality that the error in the H^1-norm $\|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_1$ is bounded by the norm of the residual

$$\|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_1 \lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}\|_{-1}. \tag{2.5}$$

Here and in the following, we neglect the constants in the error bounds; we write $|\cdot| \lesssim |\cdot|$ for a relation $|\cdot| \leq C|\cdot|$. Conversely, it follows from the boundedness of the bilinear form $a(\cdot,\cdot)$ that the norm of the residual is bounded by the error

$$\|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}\|_{-1,\omega} \lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega} \tag{2.6}$$

for each open subset ω of Ω. Unfortunately, $\|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}\|_{-1}$ is not a computable quantity. By means of its L^2-representation and due to the Galerkin orthogonality we find a computable upper bound of $\|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}\|_{-1}$. For the lower bound the properties of bubble functions are employed.

The discrete and continuous equations (2.2) and (2.3) imply the Galerkin orthogonality

$$a(\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}}) = 0 \quad \forall \boldsymbol{\varphi}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m}}. \tag{2.7}$$

because of $\boldsymbol{\mathcal{H}}_{\mathfrak{m},0} \subset \boldsymbol{\mathcal{H}}_0$. Thus, we have $\left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}, \boldsymbol{\varphi}\right\rangle_{-1,1} = \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}, \boldsymbol{\varphi} - \boldsymbol{I}_{\mathfrak{m}}\boldsymbol{\varphi}\right\rangle_{-1,1}$ for $\boldsymbol{I}_{\mathfrak{m}}\boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m},0}$. For the upper bound it would be desirable to choose a quasi-interpolation operator

$$I_{\mathfrak{m}}\varphi := \sum_{p\in N_{\mathfrak{m}}\setminus N_{\mathfrak{m}}^D} c_p(\varphi)\phi_p \tag{2.8}$$

which fulfills an L^2-approximation property (Lemma 2.2.1). Usually, the Clément type quasi-interpolation operator with $c_p(\varphi) := \frac{\int_{\omega_p} \varphi\, dx}{\int_{\omega_p} 1\, dx}$ is used. Here ω_p is the union of all elements sharing the node $\boldsymbol{p}$ and is called patch. In the following sections different choices of quasi-interpolation operator are used. Thus, the definition of $c_p(\varphi)$ varies in this work. For details about the Clément interpolation we refer to [Clé75, Car06]. In Section 2.1 we introduced the notation $\mathfrak{e}$ for an element of the mesh $\mathfrak{m}$ and $h_{\mathfrak{e}}$ for the diameter of an element. Accordingly, we denote a side of an element by $\mathfrak{s}$ and the diameter of the side is $h_{\mathfrak{s}}$. Here and in the following, elements $\mathfrak{e}$ and sides $\mathfrak{s}$ are understood as elements of the mesh $\mathfrak{m}$ and as subdomains of $\bar{\Omega}$, too. We set

$$\omega_{\mathfrak{s}} := \bigcup_{\mathfrak{s}\subset\partial\mathfrak{e}} \mathfrak{e}$$

$$\tilde{\omega}_{\mathfrak{s}} := \bigcup_{\mathfrak{s}\cap\mathfrak{e}\neq\emptyset} \mathfrak{e}$$

$$\tilde{\omega}_{\mathfrak{e}} := \bigcup_{\mathfrak{e}\cap\tilde{\mathfrak{e}}\neq\emptyset} \tilde{\mathfrak{e}}.$$

Lemma 2.2.1. *L^2-approximation property*
Let $I_\mathfrak{m}\varphi$ be defined as in (2.8), then for all $\varphi \in \mathcal{H}_0$ and all elements $\mathfrak{e}$ and sides $\mathfrak{s}$ of the mesh $\mathfrak{m}$ the following L^2-approximation properties hold

$$\begin{aligned} \|\varphi - I_\mathfrak{m}\varphi\|_\mathfrak{e} &\lesssim h\|\varphi\|_{1,\tilde{\omega}_\mathfrak{e}} \\ \|\varphi - I_\mathfrak{m}\varphi\|_\mathfrak{s} &\lesssim h^{\frac{1}{2}}\|\varphi\|_{1,\tilde{\omega}_\mathfrak{s}} \end{aligned}$$

where the constants depend on the shape parameter ς.

Proof. The proof follows from the Poincaré inequality and the trace inequality, see e. g. [Ver99, Ver11b]. qed.

For vector-valued functions $\boldsymbol{I}_\mathfrak{m}$ is defined by applying $I_\mathfrak{m}$ to the components of the function.

Next, for a clear presentation we define the so-called jump terms which are either the difference between the stresses $\boldsymbol{\sigma}|_\mathfrak{e}(\boldsymbol{u}_\mathfrak{m}) \cdot \boldsymbol{n}$ and $\boldsymbol{\sigma}|_{\tilde{\mathfrak{e}}}(\boldsymbol{u}_\mathfrak{m}) \cdot \boldsymbol{n}$ of two neighboring elements $\mathfrak{e}$ and $\tilde{\mathfrak{e}}$ where $\boldsymbol{n}$ is the unit outward normal to the common side $\mathfrak{s}$ in the interior of Ω or the difference between the given Neumann data $\boldsymbol{\pi}$ and the boundary stress $\hat{\boldsymbol{\sigma}}|_\mathfrak{e}(\boldsymbol{u}_\mathfrak{m})$ at a Neumann boundary side $\mathfrak{s}$

$$\begin{aligned} \boldsymbol{J}^I(\boldsymbol{u}_\mathfrak{m}) &:= (\boldsymbol{\sigma}|_{\tilde{\mathfrak{e}}}(\boldsymbol{u}_\mathfrak{m}) - \boldsymbol{\sigma}|_\mathfrak{e}(\boldsymbol{u}_\mathfrak{m}))\boldsymbol{n} \\ \boldsymbol{J}^N(\boldsymbol{u}_\mathfrak{m}) &:= \boldsymbol{\pi} - \hat{\boldsymbol{\sigma}}|_\mathfrak{e}(\boldsymbol{u}_\mathfrak{m}). \end{aligned} \tag{2.9}$$

If the equation of linear elasticity is replaced by the Laplace equation the jump terms $\boldsymbol{J}$ have to be replaced by the difference in gradients which are denoted by

$$\begin{aligned} [[\boldsymbol{\nabla} u_\mathfrak{m}]]^I &:= (\boldsymbol{\nabla}|_{\tilde{\mathfrak{e}}} u_\mathfrak{m} - \boldsymbol{\nabla}|_\mathfrak{e} u_\mathfrak{m}) \cdot \boldsymbol{n} \\ [[\boldsymbol{\nabla} u_\mathfrak{m}]]^N &:= \pi - \boldsymbol{\nabla}|_\mathfrak{e} u_\mathfrak{m} \cdot \boldsymbol{n}. \end{aligned} \tag{2.10}$$

We assume the faces of pyramids, prisms and hexahedra to be planar such that for all kinds of elements the unit outward normals $\boldsymbol{n}$ are constant. Further, we abbreviate $\boldsymbol{r}(\boldsymbol{u}_\mathfrak{m}) := \mathrm{div}\boldsymbol{\sigma}(\boldsymbol{u}_\mathfrak{m}) + \boldsymbol{f}$ which is sometimes called element or interior residual. With these notations we derive the residual a posteriori error estimator

$$\begin{aligned} \left\langle \boldsymbol{\mathcal{R}}^{\mathrm{lin}}_\mathfrak{m}, \boldsymbol{\varphi} \right\rangle_{-1,1} &= \left\langle \boldsymbol{\mathcal{R}}^{\mathrm{lin}}_\mathfrak{m}, \boldsymbol{\varphi} - \boldsymbol{I}_\mathfrak{m}\boldsymbol{\varphi} \right\rangle_{-1,1} \\ &= \sum_{\mathfrak{e}\in\mathfrak{m}} \int_\mathfrak{e} \boldsymbol{r}(\boldsymbol{u}_\mathfrak{m}) \cdot (\boldsymbol{\varphi} - \boldsymbol{I}_\mathfrak{m}\boldsymbol{\varphi})\, dx \\ &\quad + \sum_{\mathfrak{s}\in\mathfrak{m}} \left(\int_{\mathfrak{s}\cap\Omega} \boldsymbol{J}^I(\boldsymbol{u}_\mathfrak{m}) \cdot (\boldsymbol{\varphi} - \boldsymbol{I}_\mathfrak{m}\boldsymbol{\varphi})\, dx + \int_{\mathfrak{s}\cap\Gamma_N} \boldsymbol{J}^N(\boldsymbol{u}_\mathfrak{m}) \cdot (\boldsymbol{\varphi} - \boldsymbol{I}_\mathfrak{m}\boldsymbol{\varphi})\, dx \right) \\ &\lesssim \left(\sum_{\mathfrak{e}\in\mathfrak{m}} h_\mathfrak{e}^2 \|\boldsymbol{r}(\boldsymbol{u}_\mathfrak{m})\|_\mathfrak{e}^2 + \sum_{\mathfrak{s}\subset\Omega} h_\mathfrak{s} \|\boldsymbol{J}^I(\boldsymbol{u}_\mathfrak{m})\|_\mathfrak{s}^2 + \sum_{\mathfrak{s}\subset\Gamma_N} h_\mathfrak{s} \|\boldsymbol{J}^N(\boldsymbol{u}_\mathfrak{m})\|_\mathfrak{s}^2 \right) \|\boldsymbol{\varphi}\|_1 \end{aligned}$$

where we used the Cauchy-Schwarz inequality and the L^2-approximation property. Thus, it is proven that the error estimator $\eta := \left(\sum_{\mathfrak{e}\in\mathfrak{m}} \eta_{\mathfrak{e}}^2\right)^{\frac{1}{2}}$ with the local contributions

$$\eta_{\mathfrak{e}} := \left(h_{\mathfrak{e}}^2 \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}^2 + \sum_{\mathfrak{s}\subset\partial\mathfrak{e}\cap\Omega} h_{\mathfrak{s}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 + \sum_{\mathfrak{s}\subset\partial\mathfrak{e}\cap\Gamma_N} h_{\mathfrak{s}} \|\boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 \right)^{\frac{1}{2}} \tag{2.11}$$

bounds the error.

For the converse proof that the error estimator is bounded by the residual the bubble functions are a very helpful device. We define the element bubble functions for an element $\mathfrak{e}$ by

$$\Psi_{\mathfrak{e}} := c \prod_{p\in\mathfrak{e}} \phi_p \tag{2.12}$$

where we assume $c := \left(\max_{x\in\mathfrak{e}} \prod_{p\in\mathfrak{e}} \phi_p\right)^{-1}$ such that $0 \le \Psi_{\mathfrak{e}}(\boldsymbol{x}) \le 1$ and the side bubble functions for a side $\mathfrak{s}$ by

$$\Psi_{\mathfrak{s}} := c \prod_{p\in\mathfrak{s}} \phi_p \tag{2.13}$$

where we assume $c := \left(\max_{x\in\mathfrak{e}} \prod_{p\in\mathfrak{s}} \phi_p\right)^{-1}$ such that $0 \le \Psi_{\mathfrak{s}}(\boldsymbol{x}) \le 1$, see e.g. [Ver11a].

Basic properties of element bubble functions:

- $\operatorname{supp}\Psi_{\mathfrak{e}} \subset \mathfrak{e}$
- $\int_{\mathfrak{e}} \Psi_{\mathfrak{e}} \sim \int_{\mathfrak{e}} 1 \sim h_{\mathfrak{e}}^d$
- $\|\Psi_{\mathfrak{e}}\|_{\mathfrak{e}} \sim h_{\mathfrak{e}}^{\frac{d}{2}}$
- $\|\nabla\Psi_{\mathfrak{e}}\|_{\mathfrak{e}} \sim h_{\mathfrak{e}}^{-1}\|\Psi_{\mathfrak{e}}\|_{\mathfrak{e}} \sim h_{\mathfrak{e}}^{\frac{d-2}{2}}$

Basic properties of side bubble functions:

- $\operatorname{supp}\Psi_{\mathfrak{s}} \subset \omega_{\mathfrak{s}}$
- $\int_{\mathfrak{s}} \Psi_{\mathfrak{s}} \sim \int_{\mathfrak{s}} 1 \sim h_{\mathfrak{s}}^{d-1}$
- $\|\Psi_{\mathfrak{s}}\|_{\mathfrak{s}} \sim h_{\mathfrak{s}}^{\frac{d-1}{2}}$
- $\|\Psi_{\mathfrak{s}}\|_{\omega_{\mathfrak{s}}} \sim h_{\mathfrak{s}}^{\frac{d}{2}}$
- $\|\nabla\Psi_{\mathfrak{s}}\|_{\omega_{\mathfrak{s}}} \sim h_{\mathfrak{s}}^{-1}\|\Psi_{\mathfrak{s}}\|_{\omega_{\mathfrak{s}}} \sim h_{\mathfrak{s}}^{-\frac{1}{2}}\|\Psi_{\mathfrak{s}}\|_{\mathfrak{s}}$

These properties follow from explicit computation and scaling arguments and will be sufficient for the subsequent calculation as long as we consider linear finite elements on triangles or tetrahedra. Otherwise we need some more general properties of the bubble functions.

Lemma 2.2.2. *Properties of bubble functions*
For all polynomials w defined on $\mathfrak{e}$ and $\mathfrak{s}$, respectively, the following inequalities hold

$$
\begin{aligned}
\|w\|_{\mathfrak{e}}^2 &\lesssim \int_{\mathfrak{e}} \Psi_{\mathfrak{e}} w^2 \lesssim \|w\|_{\mathfrak{e}}^2 \\
\|\Psi_{\mathfrak{e}} w\|_{1,\mathfrak{e}} &\lesssim h_{\mathfrak{e}}^{-1} \|w\|_{\mathfrak{e}} \\
\|w\|_{\mathfrak{s}}^2 &\lesssim \int_{\mathfrak{s}} \Psi_{\mathfrak{s}} w^2 \lesssim \|w\|_{\mathfrak{s}}^2 \\
\|\Psi_{\mathfrak{s}} w\|_{1,\omega_{\mathfrak{s}}} &\lesssim h_{\mathfrak{s}}^{-\frac{1}{2}} \|w\|_{\mathfrak{s}} \\
\|\Psi_{\mathfrak{s}} w\|_{\omega_{\mathfrak{s}}} &\lesssim h_{\mathfrak{s}}^{\frac{1}{2}} \|w\|_{\mathfrak{s}}.
\end{aligned}
$$

The constants depend on the reference elements, on the bubble functions on the reference elements and on the shape parameter ς but they are independent of w and $h_{\mathfrak{s}}, h_{\mathfrak{e}}$.

Proof. See, e.g., [Ver96, Lemma 3.3] or [AO00, Section 2.3.1]. The fact that $\int_{\mathfrak{e}} \Psi_{\mathfrak{e}} w^2$ defines a norm on the finite-dimensional space of polynomials is an essential ingredient of the proof. qed.

In the vector-valued case, each component of the bubble functions $\boldsymbol{\Psi}_{\mathfrak{e}}, \boldsymbol{\Psi}_{\mathfrak{s}}$ is defined as in (2.12), (2.13). Let $\bar{\boldsymbol{f}}$ be a discrete approximation of $\boldsymbol{f}$ whose components $\bar{f}_i$ are piecewise constant on each $\mathfrak{e} \in \mathfrak{m}$. The corresponding element residual $\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}) := \mathrm{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}}) + \bar{\boldsymbol{f}}$ can be bounded as follows

$$
\begin{aligned}
\|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}^2 \lesssim \int_{\mathfrak{e}} (\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}))^2 \boldsymbol{\Psi}_{\mathfrak{e}} &= \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}, \bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{e}} \right\rangle_{-1,1,\mathfrak{e}} + \int_{\mathfrak{e}} (\bar{\boldsymbol{f}} - \boldsymbol{f}) \bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{e}} \\
&\leq \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\mathfrak{e}} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{e}}\|_{1,\mathfrak{e}} + \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{e}}\|_{\mathfrak{e}} \\
&\lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\mathfrak{e}} h_{\mathfrak{e}}^{-1} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} + \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}.
\end{aligned}
\tag{2.14}
$$

Dividing the last expression by $h_{\mathfrak{e}}^{-1} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}$ and using the triangle inequality $\|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} \leq \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} + \|\boldsymbol{f} - \bar{\boldsymbol{f}}\|_{\mathfrak{e}}$ we get the desired result on each element

$$
h_{\mathfrak{e}} \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} \lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\mathfrak{e}} + h_{\mathfrak{e}} \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}}. \tag{2.15}
$$

As $\bar{f}_i$ has been defined as arbitrary piecewise constant functions we may replace them by the best approximation $\inf_{\bar{f}_i} \|\bar{f}_i - f_i\|$. This term is formally assumed to be of higher order $o(h)$.
The part of the error estimator involving the jump terms of interior sides $\mathfrak{s}$ can be estimated similar to (2.14)

$$
\begin{aligned}
\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 &\lesssim \int_{\mathfrak{s}} \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{s}} \\
&\lesssim \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}, \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{s}} \right\rangle_{-1,1,\omega_{\mathfrak{s}}} - \int_{\omega_{\mathfrak{s}}} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{s}} \\
&\lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\omega_{\mathfrak{s}}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{s}}\|_{1,\omega_{\mathfrak{s}}} + \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \boldsymbol{\Psi}_{\mathfrak{s}}\|_{\omega_{\mathfrak{s}}} \\
&\lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\omega_{\mathfrak{s}}} h_{\mathfrak{s}}^{-\frac{1}{2}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} + \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}} h_{\mathfrak{s}}^{\frac{1}{2}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}.
\end{aligned}
\tag{2.16}
$$

Dividing by $h_{\mathfrak{s}}^{-\frac{1}{2}}\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}$ and applying (2.15), we get the desired estimate

$$h_{\mathfrak{s}}^{\frac{1}{2}}\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\omega_{\mathfrak{s}}} + h_{\mathfrak{s}}\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_{\mathfrak{s}}}.$$

Let $\bar{\boldsymbol{\pi}}$ be a piecewise constant approximation of $\boldsymbol{\pi}$ one can deduce in the same way

$$h_{\mathfrak{s}}^{\frac{1}{2}}\|\boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \lesssim \|\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}\|_{-1,\omega_{\mathfrak{s}}} + h_{\mathfrak{s}}\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_{\mathfrak{s}}} + h_{\mathfrak{s}}^{\frac{1}{2}}\|\bar{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\mathfrak{s}}.$$

Thus, the efficiency of the error estimator follows from (2.6).

2.2.2 Residual-type error estimators for obstacle problems

In this section we present residual-type error estimators for obstacle problems and discuss the difficulties arising in the construction.

In Section 1.2.3 we presented the strong formulation of obstacle problems and we already mentioned in Section 1.3.2 that the weak formulation gives rise to a variational inequality whose existence and uniqueness follows from the Theorem of Lions and Stampacchia. For the ease of presentation and to be conform to the presented literature we assume in this section $u_D = 0$ on $\Gamma_D = \partial\Omega$. Thus, the variational inequality reads as

$$\textit{Find } u \in \mathcal{K} \textit{ such that } \langle \boldsymbol{\nabla} u, \boldsymbol{\nabla}(v-u)\rangle \geq \langle f, v-u\rangle \quad \forall v \in \mathcal{K} \tag{2.17}$$

where $\mathcal{K} := \{v \in \mathcal{H}_0 \mid v \geq g\}$ and $g \leq 0$ on $\partial\Omega$. Similar to (1.44) there exists a distribution $\Lambda \in \mathcal{H}^*$ which turns the variational inequality (2.17) into an equation

$$\langle \Lambda, \varphi\rangle_{-1,1} := \langle f, \varphi\rangle - \langle \boldsymbol{\nabla} u, \boldsymbol{\nabla}\varphi\rangle \quad \forall \varphi \in \mathcal{H}_0$$

and which has the meaning of a contact force density. For obstacle problems, where the constraints hold in the domain, the contact force density coincides with the Lagrange multiplier. We refer to Λ as Lagrange multiplier and note that $\langle \Lambda, \varphi\rangle_{-1,1} \leq 0$ for all $\varphi \geq 0$. In the discrete setting the constraints are enforced at each node, i. e., $u(\boldsymbol{p}) \geq g(\boldsymbol{p})$ $\forall \boldsymbol{p} \in N_{\mathfrak{m}}$. The discrete Lagrange multiplier $\Lambda_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m}}^*$ is defined by

$$\langle \Lambda_{\mathfrak{m}}, \varphi_{\mathfrak{m}}\rangle_{-1,1} := \langle f, \varphi_{\mathfrak{m}}\rangle - \langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla}\varphi_{\mathfrak{m}}\rangle \quad \forall \varphi_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m},0} \tag{2.18}$$

with $\langle \Lambda_{\mathfrak{m}}, \varphi_{\mathfrak{m}}\rangle_{-1,1} \leq 0$ for all $\varphi_{\mathfrak{m}} \geq 0$. We note that the extension of $\Lambda_{\mathfrak{m}}$ to a functional in $\mathcal{H}^*$ is not unique. A straightforward choice could be

$$\langle \Lambda_{\mathfrak{m}}, \varphi\rangle_{-1,1} := \langle f, \varphi\rangle - \langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla}\varphi\rangle \quad \forall \varphi \in \mathcal{H}_0. \tag{2.19}$$

Comparing (2.19) with the definition of the linear residual (2.4) we see that $\langle \Lambda_{\mathfrak{m}}, \varphi\rangle_{-1,1} = \left\langle \mathcal{R}_{\mathfrak{m}}^{\mathrm{lin}}, \varphi\right\rangle_{-1,1}$. If the discrete solutions $u_{\mathfrak{m}}$ and $\Lambda_{\mathfrak{m}}$ approximate the continuous solutions u and Λ, the linear residual will not tend to zero as $\Lambda \neq 0$. Therefore, the linear residual is no appropriate error measure as in the linear case, see Section 2.2.1. It overestimates the error. The relation between $\left\langle \mathcal{R}_{\mathfrak{m}}^{\mathrm{lin}}, \varphi\right\rangle_{-1,1}$ and $\langle \boldsymbol{\nabla}(u - u_{\mathfrak{m}}), \varphi\rangle$ is disturbed

$$\langle \boldsymbol{\nabla}(u - u_{\mathfrak{m}}), \varphi\rangle = \left\langle \mathcal{R}_{\mathfrak{m}}^{\mathrm{lin}}, \varphi\right\rangle_{-1,1} - \langle \Lambda, \varphi\rangle_{-1,1} \quad \forall \varphi \in \mathcal{H}_0. \tag{2.20}$$

However, in the first residual-type error estimator of $\|u - u_{\mathfrak{m}}\|_1$ for obstacle problems [CN00] the same contributions $\eta_{\mathfrak{e}}$ as in (2.11) occur for all elements $\mathfrak{e}$ regardless of whether the solution coincides with the obstacle in the element or not. Unfortunately, the disturbed relation between the linear residual and the error precludes to give a lower bound of the error exclusively in terms of $\eta_{\mathfrak{e}}$. The remaining error estimator contributions in [CN00] account for the differences between the discrete and continuous obstacles and the interelement jumps $[[\boldsymbol{\nabla} g_{\mathfrak{m}}]]^I$ of the discrete obstacle. The latter contribution is restricted to the coincidence set where $u_{\mathfrak{m}} = g_{\mathfrak{m}}$ avoiding overestimation in areas where the obstacle is inactive. This first attempt to distinguish between coincidence and non-coincidence sets has been further developed and significantly improved in [NSV03, FV03, NSV05, MNvPZ07]. For the proof of the upper bound suitable test functions have been inserted in the variational inequality. For this purpose a special positivity preserving interpolation operator has been introduced which is further investigated in [NW02]. This interpolation operator gives rise to higher order stability estimates and has been employed for residual-type error estimates for obstacle problems in the L^∞-norm, see [NSV03] and [NSV05].

The Lagrange multiplier Λ as well as the displacement field u are both unknowns of the obstacle problem. Further, the Lagrange multiplier is directly related to the coincidence set. In the region where $u > g$ a linear equation is solved and the Lagrange multiplier is zero, so that we might expect a residual error estimator as in the linear case. In the coincidence set where $u = g$ the displacement is fixed and thus changing the data f affects the Lagrange multiplier which is given by $\Lambda = f + \Delta g$ in a distributional sense. The first residual-type error estimator considering this structure of obstacle problems can be found in [Vee01]. Therein the error is measured in both unknowns $\|u - u_{\mathfrak{m}}\|_1 + \|\Lambda - \Lambda_{\mathfrak{m}}\|_{-1}$. It is worth to mention that it is the first residual-type a posteriori error estimator for obstacle problems for which the efficiency has been proven. Accordingly, the linear residual is replaced by a Galerkin functional whose abstract definition is given by

$$\begin{aligned} \langle \mathcal{G}_{\mathfrak{m}}, \varphi \rangle_{-1,1} :=& \ \langle \boldsymbol{\nabla}(u - u_{\mathfrak{m}}), \boldsymbol{\nabla}\varphi \rangle + \langle \Lambda - \Lambda_{\mathfrak{m}}, \varphi \rangle_{-1,1} && (2.21)\\ =& \ \langle f - \Lambda_{\mathfrak{m}}, \varphi \rangle_{-1,1} - \langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla}\varphi \rangle \qquad \forall \varphi \in \mathcal{H}_0. && (2.22) \end{aligned}$$

In practice $\langle \Lambda_{\mathfrak{m}}, \varphi \rangle_{-1,1}$ still needs to be specified as there is no unique extension of $\Lambda_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m}}^*$ defined in (2.18) to a functional in $\mathcal{H}^*$. The choice (2.19) of $\Lambda_{\mathfrak{m}}$ as functional on $\mathcal{H}_0$ would lead to $\langle \mathcal{G}_{\mathfrak{m}}, \varphi \rangle_{-1,1} = 0$ and is thus inappropriate. In [Vee01] the bilinear form

$$\langle \psi_{\mathfrak{m}}, \varphi_{\mathfrak{m}} \rangle_{\mathfrak{m}} := \sum_{p \in N_{\mathfrak{m}}} \psi_{\mathfrak{m}}(\boldsymbol{p}) \varphi_{\mathfrak{m}}(\boldsymbol{p}) \int_\Omega \phi_p$$

corresponding to the lumped mass matrix and thus sometimes called lumped L^2-scalar product is used. The discrete variational inequality

$$\langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla}(v_{\mathfrak{m}} - u_{\mathfrak{m}}) \rangle \geq \langle f_{\mathfrak{m}}, (v_{\mathfrak{m}} - u_{\mathfrak{m}}) \rangle_{\mathfrak{m}} \quad \forall v_{\mathfrak{m}} \in \mathcal{K}_{\mathfrak{m}} \tag{2.23}$$

where $f_{\mathfrak{m}}$ is a discrete approximation to f is solved. From the linear residual of this variational inequality (2.23) by means of the lumped L^2-scalar product

$$\langle \Lambda_{\mathfrak{m}}, \varphi_{\mathfrak{m}} \rangle_{\mathfrak{m}} := \langle f_{\mathfrak{m}}, \varphi_{\mathfrak{m}} \rangle_{\mathfrak{m}} - \langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla}\varphi_{\mathfrak{m}} \rangle \quad \forall \varphi_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m},0} \tag{2.24}$$

a function $\Lambda_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m},0}$ is defined with node values $\Lambda_{\mathfrak{m}}(\boldsymbol{p}) := \frac{\langle \Lambda_{\mathfrak{m}}, \phi_p \rangle_{\mathfrak{m}}}{\int_\Omega \phi_p}$. This function is in L^2 and hence a functional on $\mathcal{H}_0$ by means of the standard L^2 scalar product, i.e., $\langle \Lambda_{\mathfrak{m}}, \varphi \rangle_{-1,1} := \int_\Omega \Lambda_{\mathfrak{m}} \varphi$. The fact that $\Lambda_{\mathfrak{m}}$ is a discrete function with $\Lambda_{\mathfrak{m}}(\boldsymbol{p}) \leq 0$ is advantageous for the proof of upper and lower bounds. The error measure is bounded by the dual norm of the Galerkin functional plus an additional term

$$\|u - u_{\mathfrak{m}}\|_1^2 + \|\Lambda - \Lambda_{\mathfrak{m}}\|_{-1}^2 \lesssim \|\mathcal{G}_{\mathfrak{m}}\|_{-1}^2 + \langle \Lambda_{\mathfrak{m}} - \Lambda, u - u_{\mathfrak{m}} \rangle_{-1,1} . \tag{2.25}$$

Thus, upper bounds of $\|\mathcal{G}_{\mathfrak{m}}\|_{-1}$ and $\langle \Lambda_{\mathfrak{m}} - \Lambda, u - u_{\mathfrak{m}} \rangle_{-1,1}$ are derived. The proof of the upper bound of the Galerkin functional is similar to the proof of the upper bound of the linear residual in Section 2.2.1. This is due to the fact that the Galerkin functional is the residual of the linear equation

$$\langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla} \varphi \rangle = \langle f - \Lambda_{\mathfrak{m}}, \varphi \rangle_{-1,1} \quad \forall \varphi \in \mathcal{H}_0, \tag{2.26}$$

with given $\Lambda_{\mathfrak{m}}$ of (2.19). We refer to [Bra05] for an abstract analysis of residual-type error estimators for obstacle problems based on this perception (2.26). We note that in the presented works linear finite elements on simplices are taken such that $\Delta u_{\mathfrak{m}} = 0$. Due to (2.26) the error estimator contribution considering the interior residual changes to $h_{\mathfrak{s}} \|f - \Lambda_{\mathfrak{m}}\|_{\omega_{\mathfrak{s}}}$. In a region where $u_{\mathfrak{m}} \neq g_{\mathfrak{m}}$ this contribution coincides with the one for linear elliptic equations because $\Lambda_{\mathfrak{m}}(\boldsymbol{p}) = 0$ if $u_{\mathfrak{m}}(\boldsymbol{p}) \neq g_{\mathfrak{m}}(\boldsymbol{p})$. If the obstacle is an affine function the discrete Lagrange multiplier should tend to the given force density in the coincidence set provided f is well approximated by $f_{\mathfrak{m}}$. In this sense this contribution is localized. The error estimator $\eta := \sum_{\mathfrak{s} \in \mathfrak{m}} \eta_{\mathfrak{s}}$ is side-based with the contributions

$$\eta_{\mathfrak{s}}^2 := h_{\mathfrak{s}}^{\frac{1}{2}} \|[[\boldsymbol{\nabla} u_{\mathfrak{m}}]]^I\|_{\mathfrak{s}}^2 + h_{\mathfrak{s}} \|f - \Lambda_{\mathfrak{m}}\|_{\omega_{\mathfrak{s}}}^2 + h_{\mathfrak{s}}^2 \|\boldsymbol{\nabla} \Lambda_{\mathfrak{m}}\|_{\omega_{\mathfrak{s}}}^2 \tag{2.27}$$

for affine obstacles. The additional term $h_{\mathfrak{s}}^2 \|\boldsymbol{\nabla} \Lambda_{\mathfrak{m}}\|_{\omega_{\mathfrak{s}}}$ measures the consistency error of the discrete choice of $\Lambda_{\mathfrak{m}}$ and vanishes in the non-contact set. If no contact occurs the error estimator coincides with the standard residual error estimator for linear elliptic equations. Besides the proof of the upper bound also the lower bound is proven in [Vee01] and also the case of non-affine obstacles is treated. One of the three additional error estimator contributions for general obstacles accounts for the interelement jumps of the discrete obstacle $\|[[\boldsymbol{\nabla}(u_{\mathfrak{m}} - g_{\mathfrak{m}})]]^I\|_{\mathfrak{s}}$ like in [CN00]. The other ones take care of the quality of approximation of g by $g_{\mathfrak{m}}$, the contribution $\|(g - u_{\mathfrak{m}})^+\|_{1,\omega_{\mathfrak{s}}}$ penalizes $u_{\mathfrak{m}} \not\geq g$ which may occur if $g - g_{\mathfrak{m}} > 0$ and the contribution $\int_{\omega_{\mathfrak{s}}} (-\Lambda_{\mathfrak{m}})(g - g_{\mathfrak{m}})^-$ penalizes $\Lambda_{\mathfrak{m}} \neq 0$ if $g - g_{\mathfrak{m}} < 0$.

In [NSV03] a quasi-discrete Lagrange multiplier $\tilde{\Lambda}_{\mathfrak{m}}$ replaces $\Lambda_{\mathfrak{m}}$ of (2.24) in the Galerkin functional. This quasi-discrete Lagrange multiplier $\tilde{\Lambda}_{\mathfrak{m}}$ is a functional on $\mathcal{H}_0$, computed from the discrete solution, and given data and mimics the properties of the continuous Lagrange multiplier enabling localization of error estimator contributions to the non-contact area. Therefore, the discrete Lagrange multiplier obtained by (2.24) is modified for all elements where the surrounding elements do not belong to the coincidence set. We do not go into further details because we are interested in error estimators for the error in the H^1-norm and pointwise error estimators require different techniques. We rather analyze the error estimator of [MNvPZ07] where the idea of defining a quasi-discrete Lagrange multiplier according to the different kinds of contact regions has been improved for estimating the error in the H^1-norm of parabolic obstacle problems. Therein the partition of unity is used $\left\langle \tilde{\Lambda}_{\mathfrak{m}}, \varphi \right\rangle_{-1,1} = \sum_{p \in N_{\mathfrak{m}}} \left\langle \tilde{\Lambda}_{\mathfrak{m}}, \varphi\phi_p \right\rangle_{-1,1}$ and the definition of each $\left\langle \tilde{\Lambda}_{\mathfrak{m}}, \varphi\phi_p \right\rangle_{-1,1}$ depends on the properties of the node with respect to their contact status. This ansatz for the quasi-discrete Lagrange multiplier is taken from [FV03]. The set of nodes is subdivided into full-contact nodes, non-contact nodes and the remaining free boundary nodes. For non-contact nodes $u_{\mathfrak{m}} > g_{\mathfrak{m}}$ holds in the whole patch ω_p and $\left\langle \tilde{\Lambda}_{\mathfrak{m}}, \varphi\phi_p \right\rangle_{-1,1}$ is set to zero. Full-contact nodes $\boldsymbol{p}$ fulfill $u_{\mathfrak{m}} = g_{\mathfrak{m}}$ in the patch ω_p and $\mathcal{R}_{\mathfrak{m}}^{\text{lin}} \leq 0$. The latter property means that modifying the discrete solution locally does not improve the resolution. In fact, let $\tilde{\mathcal{K}}_p := \{v + u_{\mathfrak{m}} \mid 0 \leq v \in H_0^1(\omega_p)\}$ and consider the local problem:

$$\textit{Find } U \in \tilde{\mathcal{K}}_p \textit{ such that } \langle \boldsymbol{\nabla} U, \boldsymbol{\nabla}(V-U) \rangle \geq \langle f, V-U \rangle \quad \forall V \in \tilde{\mathcal{K}}_p.$$

If $u_{\mathfrak{m}}$ is locally not improvable we have $U = u_{\mathfrak{m}}$ and

$$\left\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, v \right\rangle_{-1,1} = \langle f, v \rangle - \langle \boldsymbol{\nabla} u_{\mathfrak{m}}, \boldsymbol{\nabla} v \rangle \leq 0 \quad \forall v \in H_0^1(\omega_p).$$

In the case of full-contact nodes the quasi-discrete Lagrange multiplier is defined as $\left\langle \tilde{\Lambda}_{\mathfrak{m}}, \varphi\phi_p \right\rangle_{-1,1} := \langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, \varphi\phi_p \rangle_{-1,1}$ and otherwise set to $\left\langle \tilde{\Lambda}_{\mathfrak{m}}, \varphi\phi_p \right\rangle_{-1,1} := \int_{\omega_p} s_p \varphi\phi_p$ with $s_p := \frac{\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, \phi_p \rangle_{-1,1}}{\int_{\omega_p} \phi_p}$. We note that s_p coincides with the node value of a discrete Lagrange multiplier obtained by lumping the mass matrix and that $s_p = 0$ for non-contact nodes. Thus, we have $\tilde{\Lambda}_{\mathfrak{m}} \leq 0$. This choice of $\tilde{\Lambda}_{\mathfrak{m}}$ replaces $\Lambda_{\mathfrak{m}}$ in the abstract definition of the Galerkin functional (2.21). Thus, the following Galerkin functional

$$\begin{aligned} \langle \mathcal{G}_{\mathfrak{m}}, \varphi \rangle_{-1,1} &:= \langle \boldsymbol{\nabla}(u - u_{\mathfrak{m}}), \boldsymbol{\nabla}\varphi \rangle + \left\langle \Lambda - \tilde{\Lambda}_{\mathfrak{m}}, \varphi \right\rangle_{-1,1} \\ &= \left\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}} - \tilde{\Lambda}_{\mathfrak{m}}, \varphi \right\rangle_{-1,1} \end{aligned} \tag{2.28}$$

is used in [MNvPZ07]. Exploiting the definition of $\tilde{\Lambda}_{\mathfrak{m}}$ we deduce $\langle \mathcal{G}_{\mathfrak{m}}, \varphi\phi_p \rangle_{-1,1} = 0$ for full-contact nodes. Therefore, the error estimator contributions bounding the dual norm of the Galerkin functional have no contributions from the full-contact zone. For the nodes at the free boundary we have

$$\begin{aligned} \langle \mathcal{G}_{\mathfrak{m}}, \varphi\phi_p \rangle_{-1,1} &:= \left\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, \varphi\phi_p \right\rangle_{-1,1} - \int_{\omega_p} s_p \varphi\phi_p \\ &= \left\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, \varphi\phi_p \right\rangle_{-1,1} - \left\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, \phi_p \right\rangle_{-1,1} \frac{\int_{\omega_p} \varphi\phi_p}{\int_{\omega_p} \phi_p} \\ &= \left\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, (\varphi - c_p(\varphi))\phi_p \right\rangle_{-1,1} \end{aligned}$$

with $c_p(\varphi) := \frac{\int_{\omega_p} \varphi\phi_p}{\int_{\omega_p} \phi_p}$. Accordingly, the quasi-interpolation operator used in [FV03, MNvPZ07] is given by

$$I_{\mathfrak{m}}(\varphi) := \sum_{p \in N_{\mathfrak{m}}} c_p(\varphi)\phi_p.$$

We note that due to the partition of unity the resulting error estimator is based on patches instead of elements. However, the error estimator contributions obtained from the upper bound of $\|\mathcal{G}_{\mathfrak{m}}\|_{-1}$ are similar to $\eta_{\mathfrak{c}}$ in the linear elliptic case. The additional error estimator contribution which bounds the second term $\left\langle \tilde{\Lambda}_{\mathfrak{m}} - \Lambda, u - u_{\mathfrak{m}} \right\rangle_{-1,1}$ in (2.25) is a kind of discrete complementarity condition $s_p \int_{\omega_p} (u_{\mathfrak{m}} - g_{\mathfrak{m}})\phi_p$ for all free boundary nodes. Unfortunately, no lower bound is proven for this additional contribution. Finally, we note that the proof of the lower bound in terms of $[[\boldsymbol{\nabla} u_{\mathfrak{m}}]]^I$ is not as easy as for the linear elliptic case. A special linear combination of bubble functions on elements and sides has to be constructed such that the additional term $\langle \mathcal{R}_{\mathfrak{m}}^{\text{lin}}, c_p(\varphi)\phi_p \rangle_{-1,1}$ occurring in the Galerkin functional due to the quasi-discrete Lagrange multiplier at free boundary nodes disappears while keeping the properties of a side bubble function. This construction has already been used in [FV03]. An upper bound for general obstacles has also been shown in [MNvPZ07] with similar terms as in [Vee01]. For an abstract analysis of error estimation by means of modified Lagrange multipliers we refer to [NvPZ10].

2.2.3 Residual-type error estimators for contact problems

In contrast to obstacle problems only few a posteriori error estimators for contact problems exist. Due to the localization of the constraints to a subset of the boundary the finite element analysis of error estimators for contact problems is significantly different from the one for obstacle problems. In the following we present the residual-type a posteriori error estimators of [HN05, HN07].

In [HN05] an a posteriori error estimator for a simplified Signorini problem (Problem 1.2.2) is derived. The two-dimensional case with obstacle $g = 0$ is studied. The basic idea of [HN05] is to apply a $(d-1)$-dimensional version of the special positivity preserving interpolation operator of [CN00] at the contact boundary.

In [CN00] the constants $c_p(\varphi)$ occurring in the quasi-interpolation operator $I_{\mathfrak{m}}(\varphi) := \sum_{p \in N_{\mathfrak{m}}} c_p(\varphi)\phi_p$ are mean values on symmetric domains around interior nodes $\boldsymbol{p}$ and the boundary is assumed to be a homogeneous Dirichlet boundary. Thus, a $(d-1)$-dimensional version of the special positivity preserving interpolation operator is straightforward within segments at the contact boundary. However, at vertices and at the $(d-2)$-dimensional boundary of Γ_C modifications are required. Unfortunately, the error estimator contributions stemming from these nodes give rise to a non-optimal upper bound involving a constant $C(h) = \sqrt{-\ln(h)}$. Apart from the standard residual error estimator contributions $\eta_{\mathfrak{c}}$, contributions measuring the violation of the complementarity condition and the negativeness of $\boldsymbol{\nabla} u_{\mathfrak{m}} \cdot \boldsymbol{n}$ occur. The lower bound of the contribution accounting for the complementarity condition is non-optimal with a constant $C(h) = h^{-\frac{1}{2}}$.

Residual-based a posteriori error estimators for a Signorini problem (Problem 1.2.1) can be found in [HN07]. The two-dimensional case with gap function $g = 0$ is considered. While the starting point in [HN05] is the variational inequality, the error estimators of [HN07] are given for the saddle point formulation (Problem 1.3.3). Thus, the error is measured in primal and dual variables $\|u - u_{\mathfrak{m}}\|_1 + \|\mu - \mu_{\mathfrak{m}}\|_{-\frac{1}{2},\Gamma_C}$. The discrete Lagrange multipliers are continuous piecewise affine functions. The bilinear form $c(\cdot,\cdot)$ describing the weak formulation of the constraints in the saddle point formulation (Problem 1.3.3) is defined in two ways. One is the usual L^2-scalar product on the boundary representing the mass matrix and the other corresponds to the lumped mass matrix.

For both saddle point formulations residual-type a posteriori error estimators are proposed. Apart from the usual contributions $\eta_{\mathfrak{e}}$ as in the linear case, contributions measuring the non-positivity of the Lagrange multiplier $\|(\mu_{\mathfrak{m}})^-\|_{\mathfrak{e}}$, the violation of the complementarity condition $\int_{\mathfrak{e}} -(\mu_{\mathfrak{m}})^+ u_{\mathfrak{m},\nu}$ and the difference between the piecewise linear Lagrange multiplier and the piecewise constant contact stresses $\|\mu_{\mathfrak{m}} + \sigma(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}$ belong to the a posteriori error estimator. In the second case where the bilinear form is defined by lumping, the Lagrange multiplier is always positive such that the contributions $\|(\mu_{\mathfrak{m}})^-\|_{\mathfrak{e}}$ do not exist. For the proof of the upper bound of $\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1 + \|\mu - \mu_{\mathfrak{m}}\|_{-\frac{1}{2},\Gamma_C}$ integration by parts and the Clement-type quasi-interpolation operator (2.8) as in the linear case is used. Further, the complementarity condition and the sign condition of the Lagrange multiplier are exploited. In the proof of the lower bound $\mu \in L^2(\Gamma_C)$ is assumed.

Up to now, there exists no residual-type a posteriori error estimator for contact problems based on a suitable definition of a contact force density associated to the discrete solution and a Galerkin functional as for obstacle problems. We follow this issue and present a new a posteriori error estimator in the next section.

3 A new residual-type a posteriori error estimator for the Signorini problem

In this chapter we introduce a new residual-type a posteriori error estimator for contact problems. We prove reliability and efficiency for dimensions $d = 2, 3$ and meshes of simplices. Even for arbitrary, non-discrete gap functions the reliability is proven.
The error is measured in terms of both unknowns, the displacements and the contact force densities, compare [Vee01]. Inspired by [FV03, MNvPZ07] a quasi-discrete contact force density and a Galerkin functional is defined, see also Section 2.2.2.
The local error estimator contributions stemming from areas where no constraints are imposed or are active coincide with the local error estimator contributions of a standard residual error estimator for linear elliptic problems. In the areas where actual contact occurs we get an additional error estimator contribution measuring the fulfillment of a kind of complementarity condition. A similar contribution occurs in the upper bound of the error estimator proposed for obstacle problems in [MNvPZ07]. We achieve to prove that this contribution is a local lower bound of the error measure.
We note that just recently, we wrote a preprint [KVW12] where the results of Chapters 3 and 4 have been presented in a shortened version.

3.1 Definitions and main results

In this section we motivate the definition of the quasi-discrete contact force density which is used for the definition of the Galerkin functional and the error measure. Further the error estimator contributions are presented and we state the upper and lower a posteriori estimates but postpone the proofs to Section 3.2 and 3.3.

3.1.1 Quasi-discrete contact force density and the Galerkin functional

We derive the definition of the quasi-discrete contact force density. It mimics the properties of the continuous contact force density $\boldsymbol{\lambda}$ while computed from the discrete solution and given data. We note that in the case of contact problems we distinguish between the Lagrange multiplier which is a functional on $H^{\frac{1}{2}}(\Gamma_C)$ and the contact force density which is a functional on $\boldsymbol{\mathcal{H}}_0$.

We start with some preliminary definitions and recall the properties of the continuous contact force density. In Section 2.2.1 the patch ω_p the union of all surrounding elements, open with respect to Ω, has been defined. The corresponding diameter is abbreviated with $h_p := \operatorname{diam} \omega_p$ and the union of all sides of elements belonging to $\bar{\omega}_p$ is denoted by

γ_p. We call the union of all sides in the interior of ω_p, not including the boundary of ω_p, skeleton and denote it by $\gamma_{p,I}$. For the intersections between Γ and $\partial\omega_p$ $\forall \boldsymbol{p} \in \Gamma$ we distinguish between the three following types

$$\begin{aligned} \gamma_{p,C} &:= \Gamma_C \cap \partial\omega_p \\ \gamma_{p,N} &:= \Gamma_N \cap \partial\omega_p \\ \gamma_{p,D} &:= \Gamma_D \cap \partial\omega_p. \end{aligned}$$

We assume the unit outward normal $\boldsymbol{\nu}$ to be constant on Γ_C and choose the coordinate system such that $\boldsymbol{e}_1 = \boldsymbol{\nu}$. Then the discrete admissible set (2.1) is given by

$$\boldsymbol{\mathcal{K}}_{\mathfrak{m}} = \{\boldsymbol{v}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m}} \mid v_{\mathfrak{m},1} \leq g_{\mathfrak{m}} \text{ on } \Gamma_C\}. \tag{3.1}$$

In the following, we make use of the representation

$$\langle \boldsymbol{\lambda}, \boldsymbol{\varphi} \rangle_{-1,1} = \sum_{i=1}^{d} \langle \lambda_i, \varphi_i \rangle_{-1,1}$$

of the contact force density with the components

$$\begin{aligned} \langle \lambda_1, \varphi_1 \rangle_{-1,1} &:= -\langle \hat{\sigma}_1(\boldsymbol{u}), \varphi_1 \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C} \\ \langle \lambda_i, \varphi_i \rangle_{-1,1} &:= -\langle \hat{\sigma}_i(\boldsymbol{u}), \varphi_i \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C} = 0, \quad i \in \{2, .., d\} \end{aligned}$$

where we exploited that the tangential stresses are zero in the frictionless case. From the variational inequality (1.33) and Green's formula follows $\langle \hat{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{v} - \boldsymbol{u} \rangle_{\Gamma_C} \geq 0$ and thus

$$\langle \boldsymbol{\lambda}, \boldsymbol{v} - \boldsymbol{u} \rangle_{-1,1} \leq 0 \quad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{K}}. \tag{3.2}$$

The discrete contact force density coincides with the linear residual as functional on $\boldsymbol{\mathcal{H}}_{\mathfrak{m},0}$, compare Section 2.2.2

$$\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{-1,1} := \langle \boldsymbol{f}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle - a(\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}}) + \langle \boldsymbol{\pi}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{\Gamma_N} = \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}, \boldsymbol{\varphi}_{\mathfrak{m}} \right\rangle_{-1,1} \quad \forall \boldsymbol{\varphi}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m},0}. \tag{3.3}$$

In order to investigate $\boldsymbol{\lambda}_{\mathfrak{m}}$ further, we use integration by parts like in Section 2.2.1

$$\begin{aligned} &\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{-1,1} \\ &= \sum_{i=1}^{d} \sum_{p \in N_{\mathfrak{m}}} \langle \boldsymbol{\lambda}_{\mathfrak{m}}, \varphi_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p \boldsymbol{e}_i \rangle_{-1,1} \\ &= \sum_{i=1}^{d} \sum_{p \in N_{\mathfrak{m}}} \int_{\gamma_{p,I}} \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p \boldsymbol{e}_i + \sum_{i=1}^{d} \sum_{p \in N_{\mathfrak{m}}} \int_{\omega_p} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p \boldsymbol{e}_i \\ &\quad + \sum_{i=1}^{d} \sum_{p \in N_{\mathfrak{m}}^{\bar{N}}} \int_{\gamma_{p,N}} \boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p \boldsymbol{e}_i - \sum_{i=1}^{d} \sum_{p \in N_{\mathfrak{m}}^C} \int_{\gamma_{p,C}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p \boldsymbol{e}_i \end{aligned} \tag{3.4}$$

for all $\boldsymbol{\varphi}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}_{\mathfrak{m},0}$. Here, we use the notation $N_{\mathfrak{m}}^{\bar{N}}$ for the set of all nodes on $\bar{\Gamma}_N$.
In the support of ϕ_p for all interior and Neumann boundary nodes no constraints are imposed such that we have for all $i = 1,..,d$

$$\begin{aligned} 0 &= \int_{\gamma_{p,I}} \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i + \int_{\omega_p} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i \\ &= \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}, \phi_p \boldsymbol{e}_i \right\rangle_{-1,1} \qquad \forall \boldsymbol{p} \in N_{\mathfrak{m}}^I \end{aligned} \tag{3.5}$$

and

$$\begin{aligned} 0 &= \int_{\gamma_{p,I}} \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i + \int_{\omega_p} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i + \int_{\gamma_{p,N}} \boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i \\ &= \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}, \phi_p \boldsymbol{e}_i \right\rangle_{-1,1} \qquad \forall \boldsymbol{p} \in N_{\mathfrak{m}}^N. \end{aligned} \tag{3.6}$$

At the contact boundary in tangential direction homogeneous Neumann boundary conditions hold because the constraints are solely imposed in direction $\boldsymbol{\nu} = \boldsymbol{e}_1$. Additionally to (2.9) we define the tangential jumps at the contact boundary $\boldsymbol{J}_T^C(\boldsymbol{u}_{\mathfrak{m}}) := \hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}_{\mathfrak{m}})$ and we get a similar equation like (3.6) for $\boldsymbol{p} \in N_{\mathfrak{m}}^C$ and $i \neq 1$

$$\begin{aligned} 0 &= \int_{\gamma_{p,I}} \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i + \int_{\omega_p} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i + \int_{\gamma_{p,N}} \boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i - \int_{\gamma_{p,C}} \boldsymbol{J}_T^C(\boldsymbol{u}_{\mathfrak{m}}) \cdot \phi_p \boldsymbol{e}_i \\ &= \left\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}, \phi_p \boldsymbol{e}_i \right\rangle_{-1,1}. \end{aligned} \tag{3.7}$$

Thus, (3.4) reduces to

$$\begin{aligned} \langle \boldsymbol{\lambda}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{-1,1} = &\sum_{p \in N_{\mathfrak{m}}^C} \int_{\gamma_{p,I}} J_1^I(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},1}(\boldsymbol{p}) \phi_p + \sum_{p \in N_{\mathfrak{m}}^C} \int_{\omega_p} r_1(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},1}(\boldsymbol{p}) \phi_p \\ &+ \sum_{p \in N_{\mathfrak{m}}^C} \int_{\gamma_{p,N}} J_1^N(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},1}(\boldsymbol{p}) \phi_p - \sum_{p \in N_{\mathfrak{m}}^C} \int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \cdot \varphi_{\mathfrak{m},1}(\boldsymbol{p}) \phi_p \end{aligned} \tag{3.8}$$

which motivates the representation

$$\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{-1,1} = \sum_{i=1}^{d} \langle \lambda_{\mathfrak{m},i}, \varphi_{\mathfrak{m},i} \rangle_{-1,1}$$

with

$$\begin{aligned} \langle \lambda_{\mathfrak{m},1}, \varphi_{\mathfrak{m},1} \rangle_{-1,1} &:= \langle \boldsymbol{\lambda}_{\mathfrak{m}}, \boldsymbol{\varphi}_{\mathfrak{m}} \rangle_{-1,1} \\ \langle \lambda_{\mathfrak{m},i}, \varphi_{\mathfrak{m},i} \rangle_{-1,1} &:= 0, \quad i \in \{2,..,d\}. \end{aligned}$$

As $\langle \boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\mathrm{lin}}, \phi_p \boldsymbol{e}_1 \rangle_{-1,1} = \langle \lambda_{\mathfrak{m},1}, \phi_p \rangle_{-1,1} \geq 0$ which follows from the discrete variational inequaliy the linear residual is no adequate measure of the error, see also Section 2.2.2.

We replace the linear residual by a Galerkin functional

$$\langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \boldsymbol{\varphi} \rangle_{-1,1} := a(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}) + \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1}$$

where $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}$ is a quasi-discrete contact force density depending on the discrete solution and data and reflecting the properties of $\boldsymbol{\lambda}$ as, e.g., in [FV03, MNvPZ07]. We note that $\boldsymbol{\lambda}_{\mathfrak{m}}$ given by (3.8) has no direct relation to the contact stresses $\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})$ like the continuous contact force density $\langle \boldsymbol{\lambda}, \boldsymbol{\varphi} \rangle_{-1,1} = -\langle \hat{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{\varphi} \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C}$. This is due to the fact that a node at the contact boundary is responsible for the displacement at the boundary as well as in the interior. In other words, the reaction to the non-penetration condition is not restricted to the boundary; it is blurred into the interior. The quasi-discrete contact force density $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}^*$ should be related to a functional on $H^{\frac{1}{2}}(\Gamma_C)$. Therefore, we take the node values of a discrete contact force density obtained by lumping the boundary mass matrix $s_p := \frac{\langle \lambda_{\mathfrak{m},1}, \phi_p \rangle_{-1,1}}{\int_{\gamma_{p,C}} \phi_p} = \frac{\left\langle \mathcal{R}^{\text{lin}}_{\mathfrak{m},1}, \phi_p \right\rangle_{-1,1}}{\int_{\gamma_{p,C}} \phi_p}$ and define a quasi-discrete contact force density

$$\begin{aligned} \left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \right\rangle_{-1,1} &:= \int_{\Gamma_C} \left(\sum_{p \in N^C_{\mathfrak{m}}} s_p \phi_p \right) \varphi_1 \\ &= \int_{\Gamma_C} \left(\sum_{p \in N^C_{\mathfrak{m}}} \frac{\left\langle \mathcal{R}^{\text{lin}}_{\mathfrak{m},1}, \phi_p \right\rangle_{-1,1}}{\int_{\gamma_{p,C}} \phi_p} \phi_p \right) \varphi_1 \\ &= \sum_{p \in N^C_{\mathfrak{m}}} \left\langle \mathcal{R}^{\text{lin}}_{\mathfrak{m},1}, \phi_p \right\rangle_{-1,1} c_p(\varphi_1) \end{aligned} \tag{3.9}$$

with $c_p(\varphi_1) := \frac{\int_{\gamma_{p,C}} \varphi_1 \phi_p}{\int_{\gamma_{p,C}} \phi_p}$. Further, the sign condition $\left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \right\rangle \geq 0$, for all $\varphi_1 \geq 0$, compare (3.2), follows from $s_p \geq 0$. Another possibility to define a quasi-discrete contact force density is to exploit the fact that $\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})$ is an L^2-function on Γ_C. Therefore, we set

$$\begin{aligned} \langle \mathcal{R}_{\mathfrak{m},1}, \varphi_1 \rangle_{-1,1} :=& \left\langle \mathcal{R}^{\text{lin}}_{\mathfrak{m},1}, \varphi_1 \right\rangle_{-1,1} + \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \varphi_1 \\ =& \sum_{p \in N_{\mathfrak{m}}} \int_{\gamma_{p,I}} J^I_1(\boldsymbol{u}_{\mathfrak{m}}) \varphi_1 \phi_p + \sum_{p \in N_{\mathfrak{m}}} \int_{\omega_p} r_1(\boldsymbol{u}_{\mathfrak{m}}) \varphi_1 \phi_p + \sum_{p \in N^{\bar{N}}_{\mathfrak{m}}} \int_{\gamma_{p,N}} J^N_1(\boldsymbol{u}_{\mathfrak{m}}) \varphi_1 \phi_p \end{aligned} \tag{3.10}$$

so that especially the following relation holds

$$\langle \mathcal{R}_{\mathfrak{m},1}, \phi_p \rangle_{-1,1} = \int_{\gamma_{p,I}} J^I_1(\boldsymbol{u}_{\mathfrak{m}}) \phi_p + \int_{\omega_p} r_1(\boldsymbol{u}_{\mathfrak{m}}) \phi_p + \int_{\gamma_{p,N}} J^N_1(\boldsymbol{u}_{\mathfrak{m}}) \phi_p.$$

By means of this abbreviation we define a further alternative of a quasi-discrete contact force density

$$\begin{aligned}\left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \right\rangle_{-1,1} &:= \int_{\Gamma_C} \sum_{p \in N_{\mathfrak{m}}^C} \left(\frac{\langle \mathcal{R}_{\mathfrak{m},1}, \phi_p \rangle_{-1,1}}{\int_{\gamma_{p,C}} \phi_p} - \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \right) \phi_p \varphi_1 \\ &= \sum_{p \in N_{\mathfrak{m}}^C} \left(\langle \mathcal{R}_{\mathfrak{m},1}, \phi_p \rangle_{-1,1} c_p(\varphi_1) - \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \varphi_1 \phi_p \right)\end{aligned}$$

with the same mean value $c_p(\varphi_1) := \frac{\int_{\gamma_{p,C}} \varphi_1 \phi_p}{\int_{\gamma_{p,C}} \phi_p}$ as before. Of course, we set $\left\langle \tilde{\lambda}_{\mathfrak{m},i}, \varphi_i \right\rangle_{-1,1} = 0$ for $i \neq 1$.

In order to get the right cancellations in the Galerkin functional and localization of the error estimator contributions we combine the local contributions of both choices of quasi-discrete contact force densities depending on the contact status of the nodes. Inspired by [FV03, MNvPZ07], we use the partition of unity

$$\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1} = \sum_{p \in N_{\mathfrak{m}}^C} \left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \phi_p \right\rangle_{-1,1}. \tag{3.11}$$

The actual contact nodes with $u_{\mathfrak{m},1}(\boldsymbol{p}) = g_{\mathfrak{m}}(\boldsymbol{p})$ are classified into two types, the *full-contact* nodes $\boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{fC}}$ with

- $u_{\mathfrak{m},1} = g_{\mathfrak{m}}$ on $\gamma_{p,C}$
- $\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \leq 0$ on $\gamma_{p,C}$,

fulfilling the properties of the continuous solution u_ν and $\hat{\sigma}_\nu$ on the contact boundary, see Problem 1.2.1, and the remaining actual contact nodes which are called *semi-contact* nodes $\boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{sC}}$. For a full-contact node $\boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{fC}}$ we set

$$\begin{aligned}\left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \phi_p \right\rangle_{-1,1} &:= \langle \mathcal{R}_{\mathfrak{m},1}, \phi_p \rangle_{-1,1} c_p(\varphi_1) - \int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \varphi_1 \phi_p \\ &= s_p c_p(\varphi_1) \int_{\gamma_{p,C}} \phi_p - \int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})(\varphi_1 - c_p(\varphi_1)) \phi_p\end{aligned} \tag{3.12}$$

and for a semi-contact node $\boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{sC}}$ we define

$$\left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \phi_p \right\rangle_{-1,1} := \left\langle \mathcal{R}_{\mathfrak{m},1}^{\mathrm{lin}}, \phi_p \right\rangle_{-1,1} c_p(\varphi_1) = s_p c_p(\varphi_1) \int_{\gamma_{p,C}} \phi_p. \tag{3.13}$$

We may also use the definition (3.13) of $\left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \phi_p \right\rangle_{-1,1}$ for nodes which are no actual contact nodes because $s_p = 0$ if $u_{\mathfrak{m}}(\boldsymbol{p}) \neq g_{\mathfrak{m}}(\boldsymbol{p})$.

As we have already seen in (2.28) the Galerkin functional is the difference between the quasi-discrete contact force density and the linear residual. Inserting the definition of $\tilde{\lambda}_{\mathfrak{m},1}$ for full-contact nodes we get

$$\begin{aligned}\langle \mathcal{G}_{\mathfrak{m},1}, \varphi_1 \phi_p \rangle_{-1,1} &= \left\langle \mathcal{R}^{\text{lin}}_{\mathfrak{m},1} - \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1 \phi_p \right\rangle_{-1,1} \\ &= \langle \mathcal{R}_{\mathfrak{m},1}, (\varphi_1 - c_p(\varphi_1))\phi_p \rangle_{-1,1} .\end{aligned}$$

From the definition of $\mathcal{R}_{\mathfrak{m},1}$, compare (3.10), follows that the part $\int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_\mathfrak{m})\varphi_1\phi_p$ cancels out, avoiding that $\hat{\sigma}_1(\boldsymbol{u}_\mathfrak{m})$ contributes to the error estimator. This is justified by the fact that the discrete contact stress and discrete displacement at full-contact nodes have the same properties as the exact solution.

It will turn out that in order to prove an upper and lower bound of the error measure $\|\boldsymbol{u} - \boldsymbol{u}_\mathfrak{m}\|_1 + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_\mathfrak{m}\|_{-1}$ we need specific choices of $c_p(\varphi)$ for semi- and full-contact nodes. For full-contact nodes we use $c_p(\varphi_1) := \frac{\int_{\bar{\mathfrak{s}}} \varphi_1 \phi_p}{\int_{\bar{\mathfrak{s}}} \phi_p}$ where $\bar{\mathfrak{s}} \subset \gamma_{p,C}$ is a side in $\gamma_{p,C}$ fulfilling the following condition

$$\frac{\int_{\bar{\mathfrak{s}}} (u_1 - u_{\mathfrak{m},1})\phi_p}{\int_{\bar{\mathfrak{s}}} \phi_p} \geq \frac{\int_{\mathfrak{s}} (u_1 - u_{\mathfrak{m},1})\phi_p}{\int_{\mathfrak{s}} \phi_p} \quad \forall \mathfrak{s} \subset \gamma_{p,C}$$

if $g = g_\mathfrak{m}$ and otherwise

$$\frac{\int_{\bar{\mathfrak{s}}} ((u_1 - u_{\mathfrak{m},1}) + (u_{\mathfrak{m},1} - g)^+)\phi_p}{\int_{\bar{\mathfrak{s}}} \phi_p} \geq \frac{\int_{\mathfrak{s}} ((u_1 - u_{\mathfrak{m},1}) + (u_{\mathfrak{m},1} - g)^+)\phi_p}{\int_{\mathfrak{s}} \phi_p} \quad \forall \mathfrak{s} \subset \gamma_{p,C}.$$

For semi-contact nodes $\boldsymbol{p} \in N^{\text{sC}}_\mathfrak{m}$ we take $c_p(\varphi_1) = \frac{\int_{\tilde{\gamma}_{p,C}} \varphi_1 \phi_p}{\int_{\tilde{\gamma}_{p,C}} \phi_p}$, where $\tilde{\gamma}_{p,C}$ is a strict subset of $\gamma_{p,C}$, such that for every two different nodes $\boldsymbol{p}_1 \in \gamma_{p,C}$ and $\boldsymbol{p}_2 \in \gamma_{p,C}$, $\tilde{\gamma}_{p_1,C} \cap \tilde{\gamma}_{p_2,C} = \emptyset$. The motivation for a subset $\tilde{\gamma}_{p,C}$ can be found in the fact that at least one of the conditions $u_{\mathfrak{m},1} = g_\mathfrak{m}$ and $\hat{\sigma}_1(\boldsymbol{u}_\mathfrak{m}) \leq 0$ is not fulfilled on the entire set $\gamma_{p,C}$.

We note that we can choose any of the above examples of $c_p(\varphi)$ for nodes $\boldsymbol{p} \in N^C_\mathfrak{m}$ which are no actual contact nodes.

3.1.2 Error estimator and main results

The error estimator for which we prove efficiency and reliability in the following sections consists of the contributions

$$
\begin{aligned}
\eta_1 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}} \eta_{1,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{1,p} &:= h_p \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_p},\\
\eta_2 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}} \eta_{2,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{2,p} &:= h_p^{\frac{1}{2}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,I}},\\
\eta_3 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}^{\bar{N}}} \eta_{3,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{3,p} &:= h_p^{\frac{1}{2}} \left\|\boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}})\right\|_{\gamma_{p,N}},\\
\eta_4 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}^{C}} \eta_{4,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{4,p} &:= h_p^{\frac{1}{2}} \|\boldsymbol{J}_T^C(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,C}},\\
\eta_5 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}^{C}\setminus N_{\mathfrak{m}}^{\mathrm{fC}}} \eta_{5,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{5,p} &:= h_p^{\frac{1}{2}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,C}},\\
\eta_6 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}} \eta_{6,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{6,p} &:= (s_p d_p)^{\frac{1}{2}},\\
\eta_7 &:= \Bigg(\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}\cup N_{\mathfrak{m}}^{\mathrm{fC}}} \eta_{7,p}^2\Bigg)^{\frac{1}{2}}, & \eta_{7,p} &:= \begin{cases} \left(s_p \int_{\gamma_{p,C}} \phi_p \frac{\int_{\tilde{\gamma}_{p,C}} (g-g_{\mathfrak{m}})^+ \phi_p}{\int_{\tilde{\gamma}_{p,C}} \phi_p}\right)^{\frac{1}{2}} & \text{if } \boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{sC}}\\ \left(s_p \int_{\gamma_{p,C}} \phi_p \left(\sum_{\mathfrak{s}\subset\gamma_{p,C}} \frac{\int_{\mathfrak{s}} (g-g_{\mathfrak{m}})^+ \phi_p}{\int_{\mathfrak{s}} \phi_p}\right)\right)^{\frac{1}{2}} & \text{if } \boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{fC}} \end{cases}\\
\eta_8 &:= \|(u_{\mathfrak{m},1} - g)^+\|_{\frac{1}{2},\Gamma_C} \,.
\end{aligned}
$$

We denote the positive part of a function by $(\varphi)^+ := \max\{\varphi, 0\}$ and the negative part by $(\varphi)^- := \max\{-\varphi, 0\}$ such that $(\varphi) = (\varphi)^+ - (\varphi)^-$. The abbreviation d_p in η_6 stands for

$$
d_p := \int_{\tilde{\gamma}_{p,C}} (g_{\mathfrak{m}} - u_{\mathfrak{m},1})\phi_p. \tag{3.14}
$$

Thus, η_6 reminds of a complementarity condition. In fact, for a semi-contact node $s_p d_p$ would be a complementarity condition with respect to the quasi-discrete contact force density $\left\langle \tilde{\lambda}_{\mathfrak{m},1}, (g_{\mathfrak{m}} - u_{\mathfrak{m},1})\phi_p \right\rangle_{-1,1}$ if $\tilde{\gamma}_{p,C}$ was replaced by $\gamma_{p,C}$. We note that in contrast to the residual error estimator in Section 2.2.1 the local error contributions are defined on patches instead of elements. In the absence of any contact, we have $\eta_6 = \eta_7 = \eta_8 = 0$ and η_5 has contributions from all potential contact nodes such that η is a residual error estimator for linear elliptic boundary value problems where the potential contact boundary is replaced by a Neumann boundary with $\boldsymbol{\pi} = \boldsymbol{0}$. If contact occurs such an error estimator for linear equations would overestimate the error because the expected boundary stresses in the actual contact zone are non-zero. The contributions η_7 and η_8 deal with the case $g \neq g_{\mathfrak{m}}$ where η_7 is responsible for the case $g > g_{\mathfrak{m}}$, meaning that contact is achieved for $g_{\mathfrak{m}}$ but not for g and η_8 cares about the converse case that

$g \leq g_{\mathfrak{m}}$ where the non-penetration condition with respect to g is violated. The different structure of $\eta_{7,p}$ and $\eta_{8,p}$ is due to the fact that the difference between $u_{\mathfrak{m},1}$ and g is computable while the difference between u_1 and $g_{\mathfrak{m}}$ is not known. The contributions η_5 and η_6 are localized to semi-contact nodes and nodes which are not actually in contact. In the case that the mesh consists of simplices we prove the following Theorems in Sections 3.2 and 3.3.

Theorem 3.1.1. Reliability of the error estimator
Let $\boldsymbol{u}$ and $\boldsymbol{u}_{\mathfrak{m}}$ be the continuous and discrete solutions of Problem 1.3.1 and Problem 2.1.1, respectively, and $\boldsymbol{\lambda}$ the continuous contact force density and $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}$ the quasi-discrete contact force density defined in Section 3.1.1. Then the error estimator $\eta := \left(\sum_{k=1}^{8} \eta_k^2\right)^{\frac{1}{2}}$ defined above is reliable

$$\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1 + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1} \lesssim \eta.$$

Theorem 3.1.2. Local lower bounds
Let $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{\pi}}$ be piecewise constant approximations of $\boldsymbol{f}$ and $\boldsymbol{\pi}$, respectively, then we get the following estimate

$$\eta_{k,p} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_p} + h_p^{\frac{1}{2}}\|\bar{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\gamma_{p,N}}$$

for $k = 1, 2$ with $\boldsymbol{p} \in N_{\mathfrak{m}}$, for $k = 3$ with $\boldsymbol{p} \in N_{\mathfrak{m}}^{\bar{N}}$, for $k = 4$ with $\boldsymbol{p} \in N_{\mathfrak{m}}^{C}$ and for $k = 5$ with $\boldsymbol{p} \in N_{\mathfrak{m}}^{C} \backslash N_{\mathfrak{m}}^{\mathrm{fC}}$.
Under the assumption that each $\boldsymbol{p} \in N_{\mathfrak{m}}^{\mathrm{sC}}$ has a neighboring interior node $\boldsymbol{p} \in N_{\mathfrak{m}}^{I}$ and for a suitable extension of $g_{\mathfrak{m}}$ to a finite element function $\bar{g}_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m}}$ on Ω, the following estimate holds for $\eta_{6,p}$

$$\eta_{6,p} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_p} + h_p^{\frac{1}{2}}\|[[\boldsymbol{\nabla}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})]]^I\|_{\gamma_{p,C}}$$

where for simplicity we supposed that the actual contact boundary is a strict subset of the potential contact boundary.

We recall that $h_p\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_p}$ and $h_p^{\frac{1}{2}}\|\bar{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\gamma_{p,N}}$ are formally assumed to be of higher order. The additional error estimator parts $\eta_{7,p}$ and $\eta_{8,p}$ do not give rise to a lower bound of the error as the gap function is a given data, not related to the solution. But they cannot be neglected in the upper bound because all the other error estimator contributions might be zero whereas the real problem is not solved due to $g_{\mathfrak{m}} \neq g$, compare Examples 4.1 and 4.2 in [Vee01].

3.2 Reliability of the error estimator

In this section we give the proof of Theorem 3.1.1. The Galerkin functional $\langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \boldsymbol{\varphi} \rangle_{-1,1} := a(\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}) + \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1}$ defined by means of the quasi-discrete contact force density plays an essential role in the upper bound of the error. The error in the contact force densities $\|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1}$ is bounded by means of $\|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1}$ and $\|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_1$. This follows from the definition of the Galerkin functional, the boundedness of the bilinear form, i.e., $a(\boldsymbol{\varphi}, \tilde{\boldsymbol{\varphi}}) \leq \|\boldsymbol{\varphi}\|_1 \|\tilde{\boldsymbol{\varphi}}\|_1$ for all $\boldsymbol{\varphi}, \tilde{\boldsymbol{\varphi}} \in \boldsymbol{\mathcal{H}}$ and Young's inequality

$$\begin{aligned} \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1}^2 &= \left(\sup_{\varphi \in H_0^1(\Omega)} \frac{\Big(\langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \boldsymbol{\varphi} \rangle_{-1,1} - a(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}) \Big)}{\|\boldsymbol{\varphi}\|_1} \right)^2 \\ &\lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1}^2 + \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1^2. \end{aligned} \tag{3.15}$$

If we evaluate the Galerkin functional for $\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}$ and exploit the ellipticity of the bilinear form $a(\cdot,\cdot)$ on $\boldsymbol{\mathcal{H}}_0$, we obtain

$$\begin{aligned} \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1^2 &\lesssim a\left(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\right) \\ &= \langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \rangle_{-1,1} - \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \right\rangle_{-1,1} \\ &\leq \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1} \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1 - \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \right\rangle_{-1,1} \\ &\leq \frac{1}{2} \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1}^2 + \frac{1}{2} \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1^2 + \left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} - \boldsymbol{\lambda}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \right\rangle_{-1,1}, \end{aligned}$$

such that

$$\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_1^2 \lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1}^2 + 2 \left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} - \boldsymbol{\lambda}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \right\rangle_{-1,1}. \tag{3.16}$$

Thus, the error in the displacements is bounded by the dual norm of the Galerkin functional and a duality pairing between the contact force densities and the displacements. In Section 3.2.1 we show that $\|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1}$ is bounded by the error estimator and we deal with the second term in Sections 3.2.2 and 3.2.3 depending on the kind of gap function.

3.2.1 Upper bound of the Galerkin functional

For the proof of the upper bound of $\|\boldsymbol{\mathcal{G}}_\mathfrak{m}\|_{-1}$ we reformulate $\langle \boldsymbol{\mathcal{G}}_\mathfrak{m}, \boldsymbol{\varphi}\rangle_{-1,1}$ appropriately

$$\langle \boldsymbol{\mathcal{G}}_\mathfrak{m}, \boldsymbol{\varphi}\rangle_{-1,1} \tag{3.17}$$

$$= a(\boldsymbol{u} - \boldsymbol{u}_\mathfrak{m}, \boldsymbol{\varphi}) + \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_\mathfrak{m}, \boldsymbol{\varphi}\right\rangle_{-1,1} \tag{3.18}$$

$$= \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi}\rangle_{\Gamma_N} - \left\langle \tilde{\boldsymbol{\lambda}}_\mathfrak{m}, \boldsymbol{\varphi}\right\rangle_{-1,1} - a(\boldsymbol{u}_\mathfrak{m}, \boldsymbol{\varphi}) \tag{3.19}$$

$$= \sum_{i=1}^{d} \sum_{p\in N_\mathfrak{m}\setminus N_\mathfrak{m}^C} \left\langle \mathcal{R}^{\mathrm{lin}}_{\mathfrak{m},i}, \varphi_i\phi_p\right\rangle_{-1,1} + \sum_{i=2}^{d} \sum_{p\in N_\mathfrak{m}^C} \left\langle \mathcal{R}^{\mathrm{lin}}_{\mathfrak{m},i}, \varphi_i\phi_p\right\rangle_{-1,1} \tag{3.20}$$

$$+ \sum_{p\in N_\mathfrak{m}^C} \left(\langle \mathcal{R}_{\mathfrak{m},1}, \varphi_1\phi_p\rangle_{-1,1} - \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_\mathfrak{m})\varphi_1\phi_p - \left\langle \tilde{\lambda}_{\mathfrak{m},1}, \varphi_1\phi_p\right\rangle_{-1,1}\right) \tag{3.21}$$

$$= \sum_{i=1}^{d} \sum_{p\in N_\mathfrak{m}\setminus N_\mathfrak{m}^C} \left\langle \mathcal{R}^{\mathrm{lin}}_{\mathfrak{m},i}, (\varphi_i - c_p(\varphi_i))\phi_p\right\rangle_{-1,1} + \sum_{i=2}^{d} \sum_{p\in N_\mathfrak{m}^C} \left\langle \mathcal{R}^{\mathrm{lin}}_{\mathfrak{m},i}, (\varphi_i - c_p(\varphi_i))\phi_p\right\rangle_{-1,1} \tag{3.22}$$

$$+ \sum_{p\in N_\mathfrak{m}^C} \langle \mathcal{R}_{\mathfrak{m},1}, (\varphi_1 - c_p(\varphi_1))\phi_p\rangle_{-1,1} - \sum_{p\in N_\mathfrak{m}^C\setminus N_\mathfrak{m}^{\mathrm{fC}}} \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_\mathfrak{m})(\varphi_1 - c_p(\varphi_1))\phi_p. \tag{3.23}$$

Here we added the equations (3.5), (3.6) and (3.7) multiplied with constants $c_p(\varphi_i)$ and exploited the definition of $\tilde{\boldsymbol{\lambda}}_\mathfrak{m}$. For Dirichlet nodes we set $c_p(\varphi_i) = 0$ and for all other non-contact nodes and all contact nodes with $i \neq 1$ we choose the constants

$$c_p(\varphi_i) := \frac{\int_{\omega_p} \varphi_i\phi_p}{\int_{\omega_p} \phi_p}.$$

This mean value has been used for an a posteriori error estimator in [FV03]. It fulfills the L^2-approximation properties

$$\begin{aligned} \|\varphi - c_p(\varphi)\|_{\omega_p} &\lesssim h_p\|\boldsymbol{\nabla}\varphi\|_{\omega_p} \\ \|\varphi - c_p(\varphi)\|_{\gamma_p} &\lesssim h_p^{\frac{1}{2}}\|\boldsymbol{\nabla}\varphi\|_{\omega_p}. \end{aligned} \tag{3.24}$$

For the proof of (3.24) one defines $\psi := \varphi - c_p(\varphi)$ so that due to the definition of $c_p(\varphi)$, $\int_{\omega_p} \psi\phi_p = \int_{\omega_p}(\varphi - c_p(\varphi))\phi_p = 0$. As the local approximation $c_p(\varphi)$ preserves constants c we have $\psi = (\psi - c) - \frac{1}{\int_{\omega_p}\phi_p}\int_{\omega_p}(\psi - c)\phi_p$. Applying Poincaré's inequality with the mean value c of ψ we get the result

$$\|\varphi - c_p(\varphi)\|_{\omega_p} = \|\psi - c\|_{\omega_p} + \frac{\left(\int_{\omega_p} 1\right)^{\frac{1}{2}}}{\left(\int_{\omega_p}\phi_p\right)^{\frac{1}{2}}}\|\psi - c\|_{\omega_p} \lesssim \|\psi - c\|_{\omega_p} \lesssim h_p\|\boldsymbol{\nabla}\varphi\|_{\omega_p}.$$

The second inequality follows from the first by means of the scaled trace inequality. For Dirichlet nodes we have at least one edge or side $\mathfrak{s} \subset \gamma_{p,D}$, where the test function φ is zero, therefore, we can deduce $\|\varphi\phi_p\|_{\omega_p} \lesssim h_p\|\boldsymbol{\nabla}\varphi\|_{\omega_p}$ directly from the Poincaré-Friedrichs inequality. For the constants $c_p(\varphi)$ defined in Section 3.1.1 for semi- and full-contact nodes the L^2-approximation properties hold, too. We note that this kind of mean value over a $(d-1)$-dimensional subset has been used for a posteriori error estimators in [SV07] and we give the proof.

Lemma 3.2.1.
For an arbitrary function φ *and* $c_p(\varphi) = \frac{\int_{\bar{\gamma}_{p,C}} \varphi\phi_p}{\int_{\bar{\gamma}_{p,C}} \phi_p}$ *where* $\bar{\gamma}_{p,C}$ *is a subset of* $\gamma_{p,C}$ *as, e.g.,* $\bar{\mathfrak{s}}$ *or* $\tilde{\gamma}_{p,C}$*, the following* L^2*-approximation properties hold*

$$\|\varphi - c_p(\varphi)\|_{\omega_p} \lesssim h_p\|\boldsymbol{\nabla}\varphi\|_{\omega_p} \tag{3.25}$$

$$\|\varphi - c_p(\varphi)\|_{\bar{\gamma}_p} \lesssim h_p^{\frac{1}{2}}\|\boldsymbol{\nabla}\varphi\|_{\omega_p}. \tag{3.26}$$

Proof. The local approximation $c_p(\varphi)$ preserves constants and for $\psi := \varphi - c_p(\varphi)$ holds $\int_{\bar{\gamma}_{p,C}} \psi\phi_p = 0$. Therefore, we have $\psi = \psi - \frac{1}{\int_{\bar{\gamma}_{p,C}} \phi_p}\int_{\bar{\gamma}_{p,C}} \psi\phi_p = (\psi - c) - \frac{1}{\int_{\bar{\gamma}_{p,C}} \phi_p}\int_{\bar{\gamma}_{p,C}}(\psi - c)\phi_p$ for an arbitrary constant c. We assume $\operatorname{diam}(\bar{\gamma}_{p,C}) \simeq \operatorname{diam}(\gamma_{p,C}) \simeq \operatorname{diam}(\omega_p) = h_p$ and choose $c := \frac{\int_{\omega_p}\psi}{|\omega_p|}$. Then applying first the triangle inequality, second Hölder's inequality, third the scaled trace inequality $\left(\|\cdot\|_{\gamma_p} \lesssim h_p^{-\frac{1}{2}}\|\cdot\|_{\omega_p} + h_p^{\frac{1}{2}}\|\boldsymbol{\nabla}\cdot\|_{\omega_p}\right)$ and finally the Poincaré inequality with mean value zero, we get the desired estimate

$$\begin{aligned}
\|\psi\|_{\omega_p} &\lesssim \|\psi - c\|_{\omega_p} + \left\|\frac{1}{\int_{\bar{\gamma}_{p,C}} \phi_p}\int_{\bar{\gamma}_{p,C}}(\psi - c)\phi_p\right\|_{\omega_p} \\
&\lesssim \|\psi - c\|_{\omega_p} + \frac{\left(\int_{\omega_p} 1\right)^{\frac{1}{2}}}{\left(\int_{\bar{\gamma}_{p,C}} \phi_p\right)^{\frac{1}{2}}}\|\psi - c\|_{\bar{\gamma}_{p,C}} \\
&\lesssim \|\psi - c\|_{\omega_p} + \frac{h_p^{d/2}}{h_p^{(d-1)/2}}\left(h_p^{-\frac{1}{2}}\|\psi - c\|_{\omega_p} + h_p^{\frac{1}{2}}\|\boldsymbol{\nabla}\psi\|_{\omega_p}\right) \\
&\lesssim h_p\|\boldsymbol{\nabla}\varphi\|_{\omega_p}.
\end{aligned}$$

The proof of the second inequality (3.26) follows directly from the first one by applying the scaled trace inequality. qed.

With the help of these L^2-approximation properties we are able to derive the upper bound of the dual norm of the Galerkin functional in terms of the error estimator.

Lemma 3.2.2. *Let the Galerkin functional and the quasi-discrete contact force density be defined as in Section 3.1.1, then*

$$\|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1} \lesssim \left(\sum_{k=1}^{5}\eta_k^2\right)^{\frac{1}{2}}. \tag{3.27}$$

Proof. Inserting the definition of $\mathcal{R}^{\mathrm{lin}}_{\mathtt{m},i}$ and $\mathcal{R}_{\mathtt{m},i}$ in (3.17), applying Hölder's inequality and the L^2-approximation properties (3.24) and (3.25, 3.26) we get

$$
\begin{aligned}
&\langle \boldsymbol{\mathcal{G}}_{\mathtt{m}}, \boldsymbol{\varphi} \rangle_{-1,1} \\
&= \sum_{i=1}^{d} \sum_{p \in N_{\mathtt{m}}} \Bigg(\int_{\gamma_{p,I}} J_i^I(\boldsymbol{u}_{\mathtt{m}})(\varphi_i - c_p(\varphi_i))\phi_p + \int_{\omega_p} r_i(\boldsymbol{u}_{\mathtt{m}})(\varphi_i - c_p(\varphi_i))\phi_p \Bigg) \\
&\quad + \sum_{i=1}^{d} \sum_{p \in N_{\mathtt{m}}^{\bar{N}}} \int_{\gamma_{p,N}} J_i^N(\boldsymbol{u}_{\mathtt{m}})(\varphi_i - c_p(\varphi_i))\phi_p + \sum_{i=2}^{d} \sum_{p \in N_{\mathtt{m}}^{C}} \int_{\gamma_{p,C}} J_i^C(\boldsymbol{u}_{\mathtt{m}})(\varphi_i - c_p(\varphi_i))\phi_p \\
&\quad - \sum_{p \in N_{\mathtt{m}}^{C} \setminus N_{\mathtt{m}}^{\mathrm{fC}}} \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_{\mathtt{m}})(\varphi_1 - c_p(\varphi_1))\phi_p \\
&\leq \sum_{i=1}^{d} \sum_{p \in N_{\mathtt{m}}} \Big(\|J_i^I(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,I}} \|(\varphi_i - c_p(\varphi_i))\phi_p\|_{\gamma_{p,I}} + \|r_i(\boldsymbol{u}_{\mathtt{m}})\|_{\omega_p} \|(\varphi_i - c_p(\varphi_i))\phi_p\|_{\omega_p} \Big) \\
&\quad + \sum_{i=1}^{d} \sum_{p \in N_{\mathtt{m}}^{\bar{N}}} \Big(\left\|J_i^N(\boldsymbol{u}_{\mathtt{m}})\right\|_{\gamma_{p,N}} \|(\varphi_i - c_p(\varphi_i))\phi_p\|_{\gamma_{p,N}} \Big) \\
&\quad + \sum_{i=2}^{d} \sum_{p \in N_{\mathtt{m}}^{C}} \Big(\left\|J_i^C(\boldsymbol{u}_{\mathtt{m}})\right\|_{\gamma_{p,C}} \|(\varphi_i - c_p(\varphi_i))\phi_p\|_{\gamma_{p,C}} \Big) \\
&\quad + \sum_{p \in N_{\mathtt{m}}^{C} \setminus N_{\mathtt{m}}^{\mathrm{fC}}} \Big(\|\hat{\sigma}_1(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,C}} \|(\varphi_1 - c_p(\varphi_1))\phi_p\|_{\gamma_{p,C}} \Big) \\
&\lesssim \sum_{i=1}^{d} \sum_{p \in N_{\mathtt{m}}} \Bigg(h_p^{\frac{1}{2}} \|J_i^I(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,I}} \|\boldsymbol{\nabla}\varphi_i\|_{\gamma_{p,I}} + h_p \left\|r_i(\boldsymbol{u}_{\mathtt{m}})\right\|_{\omega_p} \|\boldsymbol{\nabla}\varphi_i\|_{\omega_p} \Bigg) \\
&\quad + \sum_{i=1}^{d} \sum_{p \in N_{\mathtt{m}}^{\bar{N}}} \Bigg(h_p^{\frac{1}{2}} \left\|J_i^N(\boldsymbol{u}_{\mathtt{m}})\right\|_{\gamma_{p,N}} \|\boldsymbol{\nabla}\varphi_i\|_{\gamma_{p,N}} \Bigg) + \sum_{i=2}^{d} \sum_{p \in N_{\mathtt{m}}^{C}} \Bigg(h_p^{\frac{1}{2}} \left\|J_i^C(\boldsymbol{u}_{\mathtt{m}})\right\|_{\gamma_{p,C}} \|\boldsymbol{\nabla}\varphi_i\|_{\gamma_{p,C}} \Bigg) \\
&\quad + \sum_{p \in N_{\mathtt{m}}^{C} \setminus N_{\mathtt{m}}^{\mathrm{fC}}} \Bigg(h_p^{\frac{1}{2}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,C}} \|\boldsymbol{\nabla}\varphi_1\|_{\gamma_{p,C}} \Bigg) \\
&\leq \Bigg(\Bigg(\sum_{p \in N_{\mathtt{m}}} \Big(\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,I}} h_p^{\frac{1}{2}} \Big)^2 \Bigg)^{\frac{1}{2}} + \Bigg(\sum_{p \in N_{\mathtt{m}}} \big(\|\boldsymbol{r}(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,I}} h_p \big)^2 \Bigg)^{\frac{1}{2}} \\
&\quad + \Bigg(\sum_{p \in N_{\mathtt{m}}^{\bar{N}}} \Big(\|\boldsymbol{J}^N(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,N}} h_p^{\frac{1}{2}} \Big)^2 \Bigg)^{\frac{1}{2}} + \Bigg(\sum_{p \in N_{\mathtt{m}}^{C}} \Big(\|\boldsymbol{J}_T^C(\boldsymbol{u}_{\mathtt{m}})\|_{\gamma_{p,C}} h_p^{\frac{1}{2}} \Big)^2 \Bigg)^{\frac{1}{2}}
\end{aligned}
$$

$$\left. + \left(\sum_{p\in N_{\mathfrak{m}}^{C}\setminus N_{\mathfrak{m}}^{\mathrm{fC}}}\left(\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,C}}h_p^{\frac{1}{2}}\right)^2\right)^{\frac{1}{2}}\right)\left(\sum_{p\in N_{\mathfrak{m}}}\|\boldsymbol{\nabla}\boldsymbol{\varphi}\|_{\omega_p}^2\right)^{\frac{1}{2}}$$

$$\lesssim\left(\sum_{k=1}^{5}\eta_k\right)\|\boldsymbol{\nabla}\boldsymbol{\varphi}\|_{\Omega}\lesssim\left(\sum_{k=1}^{5}\eta_k^2\right)^{\frac{1}{2}}\|\boldsymbol{\varphi}\|_{1,\Omega}.$$

Dividing the inequality by $\|\boldsymbol{\varphi}\|_1$ we get $\frac{\langle\boldsymbol{\mathcal{G}}_{\mathfrak{m}},\boldsymbol{\varphi}\rangle_{-1,1}}{\|\boldsymbol{\varphi}\|_1}\lesssim\left(\sum_{k=1}^{5}\eta_k^2\right)^{\frac{1}{2}}$ for each $\boldsymbol{\varphi}\in\boldsymbol{\mathcal{H}}_0$ and taking the supremum on both sides we get the desired result. qed.

3.2.2 Upper bound in the case of a discrete gap function

In this section we derive the upper bound of $\left\langle\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}-\boldsymbol{\lambda},\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\right\rangle_{-1,1}$ in the case $g_{\mathfrak{m}}=g$. We have $\langle\lambda_i,\varphi_i\rangle=\left\langle\tilde{\lambda}_{\mathfrak{m},i},\varphi_i\right\rangle=0$ for $i\neq1$ and $\langle\lambda_1,u_{\mathfrak{m},1}-u_1\rangle_{-1,1}\leq0$ because $\boldsymbol{\mathcal{K}}_{\mathfrak{m}}\subset\boldsymbol{\mathcal{K}}$ if $g_{\mathfrak{m}}=g$. Thus, it remains to bound $\left\langle\tilde{\lambda}_{\mathfrak{m},1},u_1-u_{\mathfrak{m},1}\right\rangle_{-1,1}$. From the definition of the quasi-discrete contact force density $\tilde{\lambda}_{\mathfrak{m},1}$ (3.11, 3.12, 3.13) follows

$$\begin{aligned}\left\langle\tilde{\lambda}_{\mathfrak{m},1},u_1-u_{\mathfrak{m},1}\right\rangle_{-1,1}=&\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}}\left(s_pc_p(u_1-u_{\mathfrak{m},1})\int_{\gamma_{p,C}}\phi_p\right)\\&+\sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}}\left(s_pc_p(u_1-u_{\mathfrak{m},1})\int_{\gamma_{p,C}}\phi_p\right)\\&-\sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}}\left(\int_{\gamma_{p,C}}\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\left(u_1-u_{\mathfrak{m},1}-c_p(u_1-u_{\mathfrak{m},1})\right)\phi_p\right).\end{aligned}\tag{3.28}$$

The first sum in (3.28) is bounded by η_6

$$\begin{aligned}&\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}}s_p\int_{\gamma_{p,C}}\phi_p\frac{\int_{\tilde{\gamma}_{p,C}}(u_1-u_{\mathfrak{m},1})\phi_p}{\int_{\tilde{\gamma}_{p,C}}\phi_p}\\&=\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}}\left(s_p\int_{\gamma_{p,C}}\phi_p\frac{\int_{\tilde{\gamma}_{p,C}}(u_1-g)\phi_p}{\int_{\tilde{\gamma}_{p,C}}\phi_p}+s_p\int_{\gamma_{p,C}}\phi_p\frac{\int_{\tilde{\gamma}_{p,C}}(g-u_{\mathfrak{m},1})\phi_p}{\int_{\tilde{\gamma}_{p,C}}\phi_p}\right)\\&\lesssim\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}}s_pd_p\end{aligned}$$

where we exploit $u_1\leq g$, $g=g_{\mathfrak{m}}$ and $\frac{\int_{\gamma_{p,C}}\phi_p}{\int_{\tilde{\gamma}_{p,C}}\phi_p}$ is a computable constant independent of h_p if $\tilde{\gamma}_{p,C}$ is always a fixed fraction of $\gamma_{p,C}$.

For a full-contact node $\boldsymbol{p}$, we have $u_1 \le g = g_{\mathfrak{m}} = u_{\mathfrak{m},1}$ which implies $u_1 - u_{\mathfrak{m},1} \le 0$ and, therefore, $c_p(u_1 - u_{\mathfrak{m},1}) \le 0$. As further $s_p \ge 0$ we have $s_p c_p(u_1 - u_{\mathfrak{m},1}) \le 0$ such that the second sum is bounded by zero.

In order to estimate the third sum we decompose it into a sum of integrals over sides. We exploit $\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})$ constant on each side as we assumed meshes of simplices and $c_p(u_1 - u_{\mathfrak{m},1}) = \frac{\int_{\bar{\mathfrak{s}}}(u_1 - u_{\mathfrak{m},1})\phi_p}{\int_{\bar{\mathfrak{s}}}\phi_p} \ge \frac{\int_{\mathfrak{s}}(u_1 - u_{\mathfrak{m},1})\phi_p}{\int_{\mathfrak{s}}\phi_p}$ for each side $\mathfrak{s}$ in $\gamma_{p,C}$

$$\begin{aligned} &-\sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})(u_1 - u_{\mathfrak{m},1} - c_p(u_1 - u_{\mathfrak{m},1}))\phi_p \\ &\quad = \sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}} \sum_{\mathfrak{s}\subset\gamma_{p,C}} -\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \underbrace{\int_{\mathfrak{s}} (u_1 - u_{\mathfrak{m},1} - c_p(u_1 - u_{\mathfrak{m},1}))\phi_p}_{\le 0} \le 0. \end{aligned} \tag{3.29}$$

Thus, we have

$$\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} - \boldsymbol{\lambda}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \right\rangle_{-1,1} \lesssim \eta_6^2 = \sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}} s_p d_p. \tag{3.30}$$

Putting together (3.15), (3.16), (3.27) and (3.30) we have the proof of Theorem 3.1.1 if $g = g_{\mathfrak{m}}$.

3.2.3 Upper bound for a general gap function in $H^{\frac{1}{2}}$

In the foregoing section we used the fact, that $u_{\mathfrak{m},1} \le g$ if $g_{\mathfrak{m}} = g$ in order to get an upper bound of $\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} - \boldsymbol{\lambda}, \boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}} \right\rangle_{-1,1}$. This condition may not hold for an arbitrary function $g \in H^{\frac{1}{2}}(\Gamma_C)$. However, it is possible to give an upper bound of the error measure by means of the error estimator for arbitrary gap functions. For this purpose, we define a function

$$u_{\mathfrak{m},1}^* = \min\{u_{\mathfrak{m},1}|_{\Gamma_C}, g\} \in H^{\frac{1}{2}}(\Gamma_C)$$

and a harmonic extension $\tilde{w}$ of $w := u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^* \in H^{\frac{1}{2}}(\Gamma_C)$ so that the stability estimate (see, e.g., [Ste08], pp. 70–71)

$$\|u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^*\|_1 \lesssim \|u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^*\|_{\frac{1}{2},\Gamma_C} \tag{3.31}$$

holds. We set $u_{\mathfrak{m},1}^* := u_{\mathfrak{m},1} - \tilde{w} \in \mathcal{H}$. In a first step we find an upper bound of $\langle \lambda_1, u_{\mathfrak{m},1} - u_1 \rangle$. Therefore, we use $\left\langle \lambda_1, u_{\mathfrak{m},1}^* - u_1 \right\rangle_{-1,1} \le 0$ as $u_{\mathfrak{m},1}^* \le g$, the Young's inequality and (3.31)

$$\begin{aligned}
&\left\langle \lambda_1, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} + u^*_{\mathfrak{m},1} - u_1 \right\rangle_{-1,1}\\
&\leq \left\langle \lambda_1, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} \right\rangle_{-1,1}\\
&= \left\langle \lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} \right\rangle_{-1,1} + \left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} \right\rangle_{-1,1}\\
&\leq \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1}\|^2_{1} + \left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} \right\rangle_{-1,1}\\
&\lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1}\|^2_{\frac{1}{2},\Gamma_C} + \left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} \right\rangle_{-1,1}.
\end{aligned}$$

Therefrom we deduce

$$\begin{aligned}
\left\langle \lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u_1 \right\rangle_{-1,1} \lesssim\ & \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|(u_{\mathfrak{m},1} - g)^+\|^2_{\frac{1}{2},\Gamma_C}\\
&+ \left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} + u_1 - u_{\mathfrak{m},1} \right\rangle_{-1,1}
\end{aligned} \tag{3.32}$$

exploiting the identity

$$(u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1})|_{\Gamma_C} = (u_{\mathfrak{m},1} - g)^+|_{\Gamma_C}. \tag{3.33}$$

We make use of the relation

$$\begin{aligned}
(u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} + u_1 - u_{\mathfrak{m},1}) &= (u_{\mathfrak{m},1} - g)^+ + (u_1 - u_{\mathfrak{m},1})\\
&= (u_{\mathfrak{m},1} - g)^+ + (u_1 - g) + (g - u_{\mathfrak{m},1})\\
&\leq (u_{\mathfrak{m},1} - g)^+ + (g - u_{\mathfrak{m},1})\\
&= (u_{\mathfrak{m},1} - g)^- = (g - u_{\mathfrak{m},1})^+
\end{aligned} \tag{3.34}$$

which holds on Γ_C to give an upper bound of $\left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} + u_1 - u_{\mathfrak{m},1} \right\rangle_{-1,1}$

$$\begin{aligned}
&\left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} + u_1 - u_{\mathfrak{m},1} \right\rangle_{-1,1}\\
&= \sum_{p \in N^C_{\mathfrak{m}}} \left\langle \tilde{\lambda}_{\mathfrak{m},1}, \left(u_{\mathfrak{m},1} - u^*_{\mathfrak{m},1} + u_1 - u_{\mathfrak{m},1}\right) \phi_p \right\rangle_{-1,1}\\
&\leq \sum_{p \in N^{\mathrm{sC}}_{\mathfrak{m}}} \left(s_p c_p \left((g - u_{\mathfrak{m},1})^+ \right) \int_{\gamma_{p,C}} \phi_p \right)\\
&\quad + \sum_{p \in N^{\mathrm{fC}}_{\mathfrak{m}}} \left(s_p c_p \left((g - u_{\mathfrak{m},1})^+ \right) \int_{\gamma_{p,C}} \phi_p \right)\\
&\quad - \sum_{p \in N^{\mathrm{fC}}_{\mathfrak{m}}} \left(\int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \left((u_1 - u_{\mathfrak{m},1}) + (u_{\mathfrak{m},1} - g)^+ - c_p \left((u_1 - u_{\mathfrak{m},1}) + (u_{\mathfrak{m},1} - g)^+ \right) \right) \phi_p \right)
\end{aligned} \tag{3.35}$$

In the case of semi-contact nodes we exploit that $(g_{\mathfrak{m}} - u_{\mathfrak{m},1}) = (g_{\mathfrak{m}} - u_{\mathfrak{m},1})^+$ so that

$$(g-u_{\mathfrak{m},1})^+ = (g-u_{\mathfrak{m},1}+g_{\mathfrak{m}}-g_{\mathfrak{m}})^+ \leq (g_{\mathfrak{m}}-u_{\mathfrak{m},1})^+ + (g-g_{\mathfrak{m}})^+ = (g_{\mathfrak{m}}-u_{\mathfrak{m},1}) + (g-g_{\mathfrak{m}})^+.$$

Thus, we have for the first sum in (3.35)

$$\begin{aligned}&\sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}} \left(s_p c_p \left((g-u_{\mathfrak{m},1})^+ \right) \int_{\gamma_{p,C}} \phi_p \right) \\ &\lesssim \sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}} s_p d_p + \sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}} \left(s_p c_p \left((g-g_{\mathfrak{m}})^+ \right) \int_{\gamma_{p,C}} \phi_p \right) \\ &\leq \eta_6^2 + \eta_7^2.\end{aligned} \tag{3.36}$$

In the case of full-contact nodes we have $u_{\mathfrak{m},1} = g_{\mathfrak{m}}$ on $\gamma_{p,C}$ so that $(g - u_{\mathfrak{m},1})^+$ can be replaced by $(g - g_{\mathfrak{m}})^+$ in the second sum of (3.35)

$$\begin{aligned}&\sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}} \left(s_p c_p \left((g-u_{\mathfrak{m},1})^+ \right) \int_{\gamma_{p,C}} \phi_p \right) \\ &= \sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}} \left(s_p c_p \left((g-g_{\mathfrak{m}})^+ \right) \int_{\gamma_{p,C}} \phi_p \right) \\ &\leq \eta_7^2.\end{aligned} \tag{3.37}$$

The third sum of (3.35) is bounded by zero which follows from the definition of $c_p(\varphi)$ and $\bar{\mathfrak{s}}$ for full contact nodes in the same way as in (3.29)

$$-\int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \left((u_1 - u_{\mathfrak{m},1}) + (u_{\mathfrak{m},1} - g)^+ - c_p \left((u_1 - u_{\mathfrak{m},1}) + (u_{\mathfrak{m},1} - g)^+ \right) \right) \phi_p \leq 0.$$

Thus, it follows

$$\left\langle \lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}, u_{\mathfrak{m},1} - u_1 \right\rangle_{-1,1} \lesssim \frac{1}{2} \|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|_{-1}^2 + \frac{1}{2} \underbrace{\|(u_{\mathfrak{m},1} - g)^+\|_{\frac{1}{2},\Gamma_C}^2}_{\eta_8^2} + \eta_6^2 + \eta_7^2. \tag{3.38}$$

Putting together (3.15), (3.16), (3.27) and (3.38) we have the proof of Theorem 3.1.1 if $g \in H^{\frac{1}{2}}(\Gamma_C)$.

3.3 Efficiency of the error estimator

In this section we give the proof of Theorem 3.1.2.

3.3.1 Lower bound in terms of the contributions $\eta_{1,p}, ..., \eta_{4,p}$

The proof of lower bounds in terms of the contributions $\eta_{1,p}, ..., \eta_{4,p}$ is very similar to the one given in Section 2.2.1. In fact $\boldsymbol{\mathcal{G}}_{\mathfrak{m}}$ plays the role of $\boldsymbol{\mathcal{R}}_{\mathfrak{m}}^{\text{lin}}$ and the properties of the bubble functions, compare Lemma 2.2.2, are used.
We start with the contribution of the interior residual $\eta_{1,p} = h_p \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_p}$. Let $\bar{\boldsymbol{f}}$ be defined as in Theorem 3.1.2 with constant components $\bar{f}_i$ and $\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) = \text{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}}) + \boldsymbol{f}$ the interior residual with the approximation $\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}) = \text{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}}) + \bar{\boldsymbol{f}}$. The element bubble functions $\boldsymbol{\Psi}_{\mathfrak{e}}$ as defined in Section 2.2.1 are zero on the sides of the elements so that $\int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}} = 0$. Therefrom and due to the properties of the element bubble functions we obtain

$$
\begin{aligned}
\|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}^2 &\lesssim \int_{\mathfrak{e}} (\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}}))^2 \boldsymbol{\Psi}_{\mathfrak{e}} \\
&= \int_{\mathfrak{e}} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}} + \int_{\mathfrak{e}} (\bar{\boldsymbol{f}} - \boldsymbol{f})\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}} \\
&= \langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}} \rangle_{-1,1,\omega_p} + \int_{\mathfrak{e}} (\bar{\boldsymbol{f}} - \boldsymbol{f})\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}} \\
&\leq \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}}\|_{1,\mathfrak{e}} + \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{e}}\|_{\mathfrak{e}} \\
&\lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} h_{\mathfrak{e}}^{-1} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} + \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}} \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}},
\end{aligned}
$$

for a node $\boldsymbol{p}$ and an element $\mathfrak{e} \subset \omega_p$. Dividing the last expression by $h_{\mathfrak{e}}^{-1}\|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}$, using the triangle inequality for $\|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} \leq \|\bar{\boldsymbol{r}}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} + \|\boldsymbol{f} - \bar{\boldsymbol{f}}\|_{\mathfrak{e}}$ and

$$
\|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} \leq \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p}, \tag{3.39}
$$

which follows from the definition of the Galerkin functional by means of the triangle inequality and the boundedness of the bilinear form, we get on each element

$$
\begin{aligned}
h_{\mathfrak{e}} \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} &\lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_{\mathfrak{e}} \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}} \\
&\lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_{\mathfrak{e}} \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}}.
\end{aligned}
$$

Due to the shape-regularity, which implies that the number of elements belonging to a patch is bounded, it follows

$$
h_p \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_p} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p \|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_p}, \tag{3.40}
$$

where we assume $h_p \approx h_{\mathfrak{e}}$.
Next, we consider $\eta_{2,p}$. In order to prove that the error measure is locally bounded by the inner jump terms we exploit the properties of the side bubble functions $\boldsymbol{\Psi}_{\mathfrak{s}}$ of an interior side $\mathfrak{s} \subset \gamma_{p,I}$

$$
\begin{aligned}
\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 &\lesssim \langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{s}} \rangle_{-1,1,\omega_p} - \int_{\omega_{\mathfrak{s}}} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{s}} \\
&\lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{s}}\|_{1,\omega_{\mathfrak{s}}} + \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\Psi}_{\mathfrak{s}}\|_{\omega_{\mathfrak{s}}} \\
&\lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} h_{\mathfrak{s}}^{-\frac{1}{2}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} + \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}} h_{\mathfrak{s}}^{\frac{1}{2}} \|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}.
\end{aligned} \tag{3.41}
$$

Dividing by $h_{\mathfrak{s}}^{-\frac{1}{2}}\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}$ we get for each side

$$h_{\mathfrak{s}}^{\frac{1}{2}}\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_{\mathfrak{s}}\|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}}$$

from which follows by (3.40) and the shape-regularity

$$h_p^{\frac{1}{2}}\|\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,I}} \lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}}-\boldsymbol{f}\|_{\omega_p}$$

where we assume $h_{\mathfrak{s}} \approx h_p$.
Finally, the local lower bounds in terms of $\eta_{3,p}$ and $\eta_{4,p}$ follow in the same way as in (3.41)

$$\begin{aligned} h_p^{\frac{1}{2}}\|\boldsymbol{J}^N(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,N}} &\lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}}-\boldsymbol{f}\|_{\omega_p} + h_p^{\frac{1}{2}}\|\bar{\boldsymbol{\pi}}-\boldsymbol{\pi}\|_{\gamma_{p,N}} \\ h_p^{\frac{1}{2}}\|\boldsymbol{J}_T^C(\boldsymbol{u}_m)\|_{\gamma_{p,C}} &\lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}}-\boldsymbol{f}\|_{\omega_p}. \end{aligned}$$

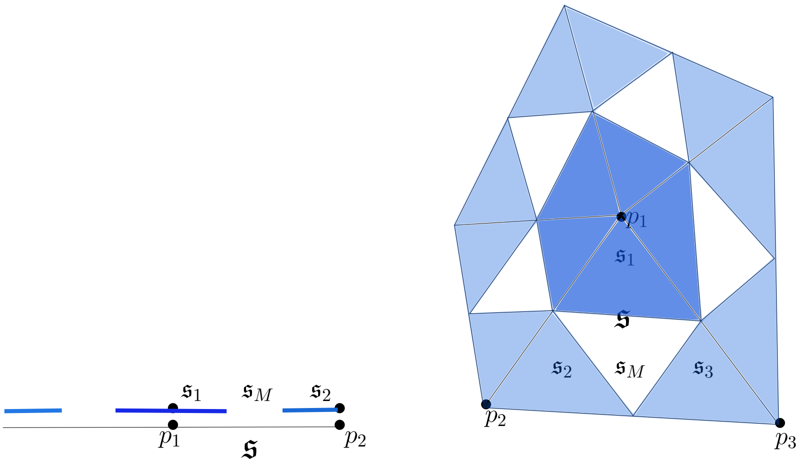
Thus, Theorem 3.1.2 is proven for $\eta_{1,p}, ..., \eta_{4,p}$.

3.3.2 Lower bound in terms of the contribution $\eta_{5,p}$

In this section we show that $\eta_{5,p}$ is a local lower bound of the error for all nodes $\boldsymbol{p} \in N_{\mathfrak{m}}^C \backslash N_{\mathfrak{m}}^{fC}$. In the foregoing section we exploited the relation between the Galerkin functional and the quantities occurring in the error estimator contributions, e.g., $\boldsymbol{J}^I(\boldsymbol{u}_{\mathfrak{m}})$. Similarly, we make use of the relation between the Galerkin functional and the boundary stresses; see, e.g., (3.17).
Let $\bar{\boldsymbol{p}} \in N_{\mathfrak{m}}^C \backslash N_{\mathfrak{m}}^{\mathrm{fC}}$ be an arbitrary but fixed node. In the following, $\mathfrak{s}$ denotes a side which belongs to $\gamma_{\bar{p},C}$. We take the side bubble function $\Psi_{\mathfrak{s}}\boldsymbol{e}_1$ as test function in (3.17)

$$\begin{aligned} &\sum_{p\in N_{\mathfrak{m}}^C\backslash N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\Psi_{\mathfrak{s}}\phi_p \\ &= -\left\langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \Psi_{\mathfrak{s}}\boldsymbol{e}_1\right\rangle_{-1,1} + \sum_{p\in N_{\mathfrak{m}}\backslash N_{\mathfrak{m}}^C} \left\langle \mathcal{R}_{\mathfrak{m},1}^{\mathrm{lin}}, \Psi_{\mathfrak{s}}\phi_p\right\rangle_{-1,1} \\ &\quad + \sum_{p\in N_{\mathfrak{m}}^C} \left\langle \mathcal{R}_{\mathfrak{m},1}, (\Psi_{\mathfrak{s}} - c_p(\Psi_{\mathfrak{s}}))\phi_p\right\rangle_{-1,1} + \sum_{p\in N_{\mathfrak{m}}^C\backslash N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\Gamma_C} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})c_p(\Psi_{\mathfrak{s}})\phi_p \qquad (3.42) \\ &= -\left\langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \Psi_{\mathfrak{s}}\boldsymbol{e}_1\right\rangle_{-1,1} + \sum_{p\in N_{\mathfrak{m}}} \int_{\omega_{\mathfrak{s}}} r_1(\boldsymbol{u}_{\mathfrak{m}})\Psi_{\mathfrak{s}}\phi_p \\ &\quad - \sum_{p\in N_{\mathfrak{m}}^{\mathrm{sC}}} s_p c_p(\Psi_{\mathfrak{s}}) \int_{\gamma_{p,C}} \phi_p - \sum_{p\in N_{\mathfrak{m}}^{\mathrm{fC}}} \left\langle \mathcal{R}_{\mathfrak{m},1}, \phi_p\right\rangle c_p(\Psi_{\mathfrak{s}}). \end{aligned}$$

If the side $\mathfrak{s}$ is not contained in any patch $\gamma_{p,C}$ of semi- or full-contact nodes $\boldsymbol{p}$, the two last terms are zero and we can proceed similar to (3.41). Otherwise, in order to get rid of the last two terms we replace $\Psi_{\mathfrak{s}}$ by a suitable function $\theta_{\mathfrak{s}}$ such that $c_p(\theta_{\mathfrak{s}}) = 0$ for all semi- and full-contact nodes. The value $c_p(\cdot)$ for a semi-contact node $\boldsymbol{p}$ depends on $\tilde{\gamma}_{p,C}$ which is a strict subset of $\gamma_{p,C}$, compare Section 3.1.1. If $\gamma_{p,C}$ consists of two intervals we choose the inner third of $\gamma_{p,C}$ containing $\boldsymbol{p}$ as $\tilde{\gamma}_{p,C}$. If $\gamma_{p,C}$ is a union of triangles we refine each triangle once uniformly and generate $\tilde{\gamma}_{p,C}$ as the $2D$ patch enclosing $\boldsymbol{p}$ with respect to this subgrid. For example in Figure 3.1 the dark blue region is $\tilde{\gamma}_{p,C}$ for $\boldsymbol{p} = \boldsymbol{p}_1$. For a side $\mathfrak{s}$ with the nodes $\{\boldsymbol{p}_i\}_{i=1,..,n}$ we denote the sides of the subgrid containing $\boldsymbol{p}_i$ by $\mathfrak{s}_i$ and the middle part by $\mathfrak{s}_M$, see Figure 3.1. For the function $\theta_{\mathfrak{s}}$ we make the following ansatz



(a) Subgrid of an interval

(b) Subgrid of a triangle

Figure 3.1: Subgrid of boundary patches $\gamma_{p,C}$

$$\theta_{\mathfrak{s}} = \left(\sum_{i=1}^{n} a_i \Psi_i + a_M \Psi_M \right) \tag{3.43}$$

where n is the number of nodes of the side $\mathfrak{s}$ and Ψ_i and Ψ_M are side bubble functions to $\mathfrak{s}_i$ and $\mathfrak{s}_M$. The coefficients are determined so that

1. $\int_{\mathfrak{s}} 1 = \sum_{p \in N_{\mathfrak{m}} \setminus N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\mathfrak{s}} \theta_{\mathfrak{s}} \phi_p$
2. $\int_{\mathfrak{s}_i} \theta_{\mathfrak{s}} \phi_{p_i} = 0$ for all $\boldsymbol{p}_i$ semi-contact nodes
3. $\int_{\mathfrak{s}} \theta_{\mathfrak{s}} \phi_{p_i} = 0$ for all $\boldsymbol{p}_i$ full-contact nodes with $\mathfrak{s} = \bar{\mathfrak{s}}$.

As $\bar{\boldsymbol{p}}$ is not a full-contact node there is at least one contribution in the right hand side of the first condition. Inserting the ansatz (3.43) in the aforementioned conditions we get a solvable system of equations with $n+1$ coefficients (degrees of freedom) a_M, a_i, $i = \{1,..,n\}$ and $n+1$ conditions. At this point the special choice of $c_p(\varphi) := \frac{\int_{\tilde{\gamma}_{p,C}} \varphi \phi_p}{\int_{\tilde{\gamma}_{p,C}} \phi_p}$ as mean value on $\tilde{\gamma}_{p,C}$ for semi-contact nodes becomes clear because the choice $c_p(\varphi) = \frac{\int_{\gamma_{p,C}} \varphi\phi_p}{\int_{\gamma_{p,C}} \phi_p}$ as mean value over the whole patch $\gamma_{p,C}$ would lead to a contradiction of the conditions. In the second condition $\mathfrak{s}_i$ would be replaced by $\mathfrak{s}$ and the condition $\int_{\mathfrak{s}} \theta_{\mathfrak{s}}\phi_{p_i} = 0$ for all $\boldsymbol{p}_i$ of the side $\mathfrak{s}$ would imply $\sum\limits_{p\in N_{\mathfrak{m}}\setminus N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\mathfrak{s}} \theta_{\mathfrak{s}}\phi_p = 0$ so that the first condition could not be fulfilled.

As we assumed a mesh of simplices, $\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})$ is constant on $\mathfrak{s}$ so that $c_p(\theta_{\mathfrak{s}}) = 0$ implies $c_p(\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}) = 0$ and it follows from the first condition

$$\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 = \sum_{p\in N_{\mathfrak{m}}\setminus N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\phi_p. \tag{3.44}$$

Putting together (3.44), (3.42) with test function $\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}$ instead of $\Psi_{\mathfrak{s}}$ and exploiting the conditions $c_p(\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}) = 0$ for all actual contact nodes we end up with

$$\begin{aligned}
\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 &= \sum_{p\in N_{\mathfrak{m}}\setminus N_{\mathfrak{m}}^{\mathrm{fC}}} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\sigma_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\phi_p \\
&= -\left\langle \boldsymbol{\mathcal{G}}_{\mathfrak{m}}, \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\boldsymbol{e}_1\right\rangle_{-1,1,\omega_{\bar{p}}} + \int_{\omega_{\mathfrak{s}}} r_1(\boldsymbol{u}_{\mathfrak{m}})\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}} \\
&\leq \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_{\bar{p}}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\|_{1,\omega_{\mathfrak{s}}} + \|r_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\|_{\omega_{\mathfrak{s}}} \\
&\lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_{\bar{p}}} h_{\mathfrak{s}}^{-\frac{1}{2}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} + h_{\mathfrak{s}}^{\frac{1}{2}}\|r_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}.
\end{aligned} \tag{3.45}$$

In the last line of (3.45) we used the properties of the bubble functions (see Section 2.2.1) on the subgrid and the fact that $\tilde{\gamma}_{p,C}$ is a fixed portion of $\gamma_{p,C}$ so that $h_{\mathfrak{s}} = ch_{\mathfrak{s}_i}$ for a mesh-independent constant c. We divide by $h_{\mathfrak{s}}^{-\frac{1}{2}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}$ leading to

$$h_{\mathfrak{s}}^{\frac{1}{2}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \lesssim \|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_{\bar{p}}} + h_{\mathfrak{s}}\|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\mathfrak{s}}}.$$

By means of the triangle inequality and the shape-regularity, the upper bounds (3.39) and (3.40) of $\|\boldsymbol{\mathcal{G}}_{\mathfrak{m}}\|_{-1,\omega_{\bar{p}}}$ and $\|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_{\bar{p}}}$ we get

$$h_{\bar{p}}^{\frac{1}{2}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{\bar{p},C}} \lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_{\bar{p}}} + \|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_{\bar{p}}} + h_{\bar{p}}\|\bar{\boldsymbol{f}}-\boldsymbol{f}\|_{\omega_{\bar{p}}}.$$

Thus, we have proven Theorem 3.1.2 for $\eta_{5,\bar{p}}$ with $\bar{\boldsymbol{p}} \in N_{\mathfrak{m}}^{C}\setminus N_{\mathfrak{m}}^{\mathrm{fC}}$.

3.3.3 Lower bound in terms of the contribution $\eta_{6,p}$

We derive a lower bound of the local error measure in terms of the local contributions of $\eta_{6,p} = (s_p d_p)^{\frac{1}{2}}$. If $s_p = 0$ or $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{q}) = 0$ for all neighboring nodes of $\boldsymbol{p}$ we have $\eta_{6,p} = 0$. Therefore, we assume $s_p > 0$ and $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{q}) > 0$ for at least one node on $\gamma_{p,C}$. Let $\hat{\boldsymbol{q}}$ be a node which fulfills $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\hat{\boldsymbol{q}}) \geq (g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{q})$ for all neighboring nodes $\boldsymbol{q}$ of $\boldsymbol{p}$. Due to $s_p > 0$ we have $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{p}) = 0$. As we consider boundary

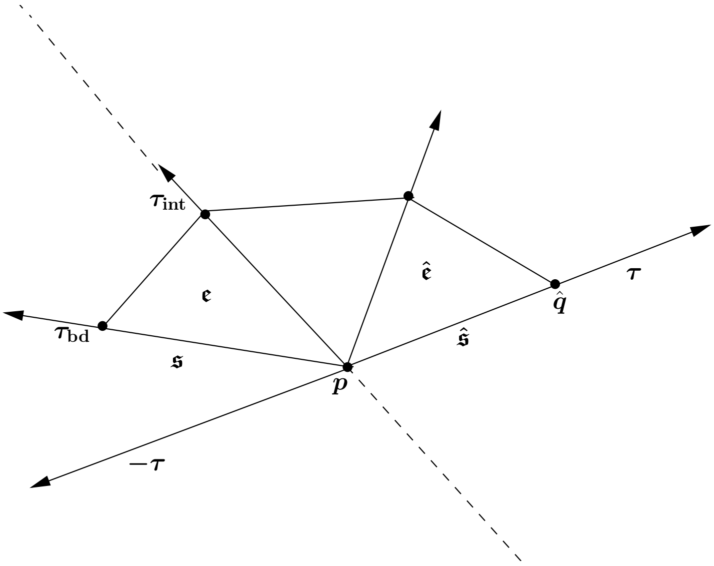


Figure 3.2: Construction of the linear combination of $-\boldsymbol{\tau}$ in $2D$

meshes of triangles or intervals the discrete functions are piecewise linear. Using Taylor expansion around $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{p}) = 0$ leads to

$$\begin{aligned}&(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\hat{\boldsymbol{q}})\\ &= \boldsymbol{\nabla}|_{\hat{\mathfrak{s}}}(g_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot (\hat{\boldsymbol{q}} - \boldsymbol{p}) \lesssim h_p \boldsymbol{\nabla}|_{\hat{\mathfrak{s}}}(g_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot \boldsymbol{\tau}\end{aligned} \tag{3.46}$$

where $\hat{\mathfrak{s}} \subset \gamma_{p,C}$ is a side containing the nodes $\hat{\boldsymbol{q}}$ and $\boldsymbol{p}$ and $\boldsymbol{\tau}$ is the unit tangential vector pointing from $\boldsymbol{p}$ to $\hat{\boldsymbol{q}}$. The corresponding element is denoted with $\hat{\mathfrak{e}}$.

First, we deal with a node $\boldsymbol{p}$ at a convex edge of the boundary. We assumed in Theorem 3.1.2 that the node $\boldsymbol{p}$ has at least one neighboring node in the interior. (This condition means that $\boldsymbol{p}$ belongs to more than one triangle for $d = 2$ and more than two tetrahedra for $d = 3$.) Thus, there exists another element $\mathfrak{e} \subset \omega_p$ with a boundary side $\mathfrak{s} \subset \partial\omega_p \cap \Gamma$. We define an extension from $g_{\mathfrak{m}}$ to a function in $\mathcal{H}_{\mathfrak{m}}$ by $\bar{g}_{\mathfrak{m}}(\boldsymbol{q}) = u_{\mathfrak{m},1}(\boldsymbol{q})$ for all $\boldsymbol{q} \in N_{\mathfrak{m}} \backslash N_{\mathfrak{m}}^C$. Due to the definition of $\bar{g}_{\mathfrak{m}}$, we have

$$\boldsymbol{\nabla}|_{\mathfrak{e}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot \boldsymbol{\tau}_{\mathrm{int}} = 0. \tag{3.47}$$

For the gradients in direction $\boldsymbol{\tau}_{\mathrm{bd}}$ pointing from $\boldsymbol{p}$ to a neighboring boundary node $\boldsymbol{q} \neq \hat{\boldsymbol{q}}$ we have

$$\boldsymbol{\nabla}|_{\mathfrak{e}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot \boldsymbol{\tau}_{\mathrm{bd}} \geq 0. \tag{3.48}$$

In the case $d = 2$ the line given by $\boldsymbol{p}$ and the vector $\boldsymbol{\tau}_{\mathrm{int}}$ divides the plane into half-planes with $\boldsymbol{\tau}$ on one side and $\boldsymbol{\tau}_{\mathrm{bd}}$ and $-\boldsymbol{\tau}$ on the other side, see Figure 3.2. Therefore,

$$-\boldsymbol{\tau} = \alpha\boldsymbol{\tau}_{\mathrm{int}} + \beta\boldsymbol{\tau}_{\mathrm{bd}} \tag{3.49}$$

with $\beta > 0$ and α arbitrary. If $d = 3$ imagine a plane $P_{\tau,\tau_{\mathrm{int}}}$ spanned by $\boldsymbol{\tau}$ and $\boldsymbol{\tau}_{\mathrm{int}}$. The cut of the plane with an opposite side $\mathfrak{s}$ defines a unit vector $\boldsymbol{\tau}_{\mathrm{bd}} := \beta_1\boldsymbol{\tau}_{\mathrm{bd},1} + \beta_2\boldsymbol{\tau}_{\mathrm{bd},2}$

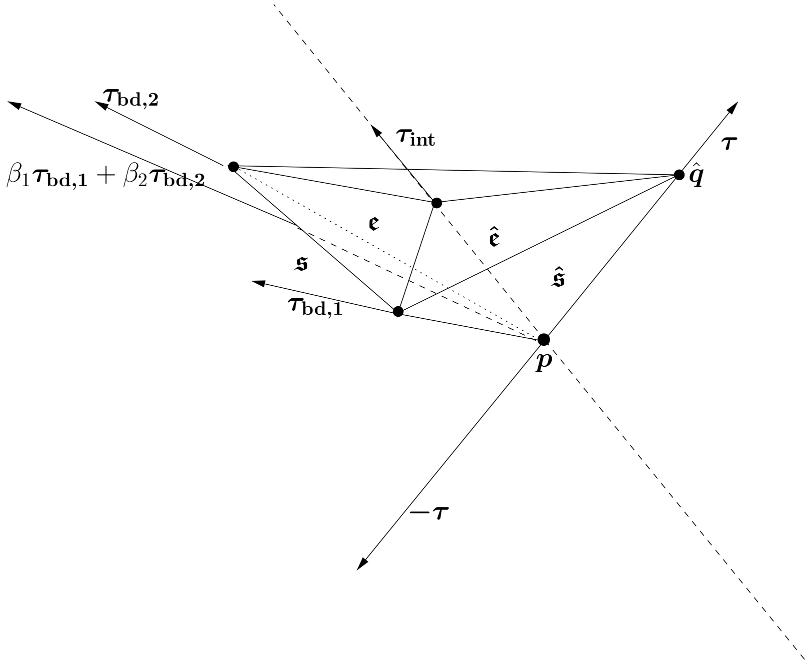


Figure 3.3: Construction of the linear combination of $-\boldsymbol{\tau}$ in $3D$

with $\beta_1, \beta_2 \geq 0$ and $\boldsymbol{\tau}_{\text{bd},1}, \boldsymbol{\tau}_{\text{bd},2}$ are the vectors along the edges of the side $\mathfrak{s}$ starting in $\boldsymbol{p}$, compare Figure 3.3. The line $\boldsymbol{\tau}_{\text{int}}$ divides this plane $P_{\tau,\tau_{\text{int}}}$ such that $-\boldsymbol{\tau}$ and $\boldsymbol{\tau}_{\text{bd}}$ lie in one half-plane and (3.49) holds with $\beta > 0$ and α arbitrary. Putting together (3.47), (3.48) and (3.49), we get

$$-\boldsymbol{\nabla}|_{\mathfrak{e}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot \boldsymbol{\tau} \geq 0 \tag{3.50}$$

and therefore adding (3.50) to (3.46) gives rise to

$$\begin{aligned}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})(\hat{\boldsymbol{q}}) &= \boldsymbol{\nabla}|_{\hat{\mathfrak{s}}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot (\hat{\boldsymbol{q}} - \boldsymbol{p}) \\ &\lesssim h_p(\boldsymbol{\nabla}|_{\hat{\mathfrak{e}}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}) - \boldsymbol{\nabla}|_{\mathfrak{e}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})) \cdot \boldsymbol{\tau}.\end{aligned} \tag{3.51}$$

For a contact node $\boldsymbol{p}$ at a concave edge the vector pointing from $\boldsymbol{p}$ in direction $-\boldsymbol{\tau}$ goes through another element $\mathfrak{e}$ belonging to ω_p. Due to the construction of $\bar{g}_{\mathfrak{m}}$ such that $(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{q}) \geq 0$ for all $\boldsymbol{q} \in N_{\mathfrak{m}}$, we have

$$-\boldsymbol{\nabla}|_{\mathfrak{e}}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}) \cdot \boldsymbol{\tau} \geq 0$$

and therefore, we can add this gradient in (3.46) getting the same result (3.51) as in the convex case.

Let $\mathfrak{e}_0 = \hat{\mathfrak{e}}$, $\mathfrak{e}_n = \mathfrak{e}$ and $\mathfrak{e}_i$, $i = 1, .., n-1$ be the elements between $\mathfrak{e}_0$ and $\mathfrak{e}_n$. The common sides of two neighboring elements are denoted with $\mathfrak{s}_i = \mathfrak{e}_{i-1} \cap \mathfrak{e}_i$. For $v_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m}}$ the difference in the gradients only occur in normal direction due to the continuity of finite element functions. Thus, the Euclidean norm $|\boldsymbol{\nabla}|_{\hat{\mathfrak{e}}} v_{\mathfrak{m}} - \boldsymbol{\nabla}|_{\mathfrak{e}} v_{\mathfrak{m}}|$ is bounded in terms of the interelement jumps $[[\boldsymbol{\nabla} v_{\mathfrak{m}}]]^I$ defined in (2.10).

$$\begin{aligned} |\boldsymbol{\nabla}|_{\hat{\mathfrak{e}}} v_{\mathfrak{m}} - \boldsymbol{\nabla}|_{\mathfrak{e}} v_{\mathfrak{m}}| &\leq |\boldsymbol{\nabla}|_{\mathfrak{e}_0} v_{\mathfrak{m}} - \boldsymbol{\nabla}|_{\mathfrak{e}_1} v_{\mathfrak{m}}| + |\boldsymbol{\nabla}|_{\mathfrak{e}_1} v_{\mathfrak{m}} - \boldsymbol{\nabla}|_{\mathfrak{e}_2} v_{\mathfrak{m}}| + \ldots + |\boldsymbol{\nabla}|_{\mathfrak{e}_{n-1}} v_{\mathfrak{m}} - \boldsymbol{\nabla}|_{\mathfrak{e}_n} v_{\mathfrak{m}}| \\ &\lesssim \sum_{i=1}^{n} h_p^{\frac{-(d-1)}{2}} \|[[\boldsymbol{\nabla} v_{\mathfrak{m}}]]^I\|_{\mathfrak{s}_i} \\ &\lesssim h_p^{\frac{-(d-1)}{2}} \|[[\boldsymbol{\nabla} v_{\mathfrak{m}}]]^I\|_{\gamma_{p,I}} \end{aligned} \tag{3.52}$$

Combining (3.51) and (3.52) with $v_{\mathfrak{m}} = \bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1}$ gives rise to

$$(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\hat{\boldsymbol{q}}) \lesssim h_p^{\frac{-d+2}{2}} \left(h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})]]^I\|_{\gamma_{p,I}} \right).$$

As $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\boldsymbol{q}) \leq (g_{\mathfrak{m}} - u_{\mathfrak{m},1})(\hat{\boldsymbol{q}})$ for all $\boldsymbol{q} \in \gamma_{p,C}$ we conclude

$$\begin{aligned} d_p &= \int_{\gamma_{p,C}} (g_{\mathfrak{m}} - u_{\mathfrak{m},1}) \phi_p \lesssim h_p^{d-1} h_p^{\frac{-d+2}{2}} \left(h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})]]^I\|_{\gamma_{p,I}} \right) \\ &\lesssim h_p^{\frac{d}{2}} \left(h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})]]^I\|_{\gamma_{p,I}} \right). \end{aligned}$$

For the upper bound of $s_p d_p$ we use Hölder's inequality and scaling arguments such that

$$\begin{aligned} & s_p d_p \\ &= \frac{\int_{\gamma_{p,I}} J_1^I(\boldsymbol{u}_{\mathfrak{m}})\phi_p + \int_{\omega_p} r_1(\boldsymbol{u}_{\mathfrak{m}})\phi_p + \int_{\gamma_{p,N}} J_1^N(\boldsymbol{u}_{\mathfrak{m}})\phi_p - \int_{\gamma_{p,C}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\phi_p}{\int_{\gamma_{p,C}} \phi_p} \int_{\gamma_{p,C}} (g_{\mathfrak{m}} - u_{\mathfrak{m},1})\phi_p \\ &\lesssim h_p^{-d+1} \left(\|J_1^I(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,I}} \underbrace{\|\phi_p\|_{\gamma_{p,I}}}_{\approx h_p^{\frac{d-1}{2}}} + \|r_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\omega_p} \underbrace{\|\phi_p\|_{\omega_p}}_{\approx h_p^{\frac{d}{2}}} \right. \\ &\quad \left. + \left\|J_1^N(\boldsymbol{u}_{\mathfrak{m}})\right\|_{\gamma_{p,N}} \underbrace{\|\phi_p\|_{\gamma_{p,N}}}_{\approx h_p^{\frac{d-1}{2}}} + \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\gamma_{p,C}} \underbrace{\|\phi_p\|_{\gamma_{p,C}}}_{\approx h_p^{\frac{d-1}{2}}} \right) \cdot \int_{\gamma_{p,C}} (g_{\mathfrak{m}} - u_{\mathfrak{m},1})\phi_p \\ &\lesssim (\eta_{2,p} + \eta_{1,p} + \eta_{3,p} + \eta_{5,p}) \cdot h_p^{-\frac{d}{2}} \cdot \int_{\gamma_{p,C}} (g_{\mathfrak{m}} - u_{\mathfrak{m},1})\phi_p \\ &\lesssim \left(\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_p} \right) \left(h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})]]^I\|_{\gamma_{p,I}} \right) \\ &\lesssim \left(\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{1,\omega_p} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_p} + h_p\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\omega_p} + h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla}(\bar{g}_{\mathfrak{m}} - u_{\mathfrak{m},1})]]^I\|_{\gamma_{p,I}} \right)^2 \end{aligned} \tag{3.53}$$

Thus, we have proven **Theorem 3.1.2**. We note that we would get the additional contribution $h_p^{\frac{1}{2}} \|\bar{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\gamma_{p,N}}$ if the actual contact boundary was not a strict subset of the potential contact boundary.

Remark 3.3.1. *If we consider the simplified Signorini problem (Problem 1.2.2),* $[[\boldsymbol{\nabla} u_{\mathfrak{m}}]]^I$ *is the jump term. Hence, by means of the triangle inequality and the local lower bound of the error measure in terms of* $\|[[\boldsymbol{\nabla} u_{\mathfrak{m}}]]^I\|$ *we can estimate (3.53) further*

$$s_p d_p \lesssim \left(\|u_{\mathfrak{m}} - u\|_{1,\omega_p} + \|\tilde{\lambda}_{\mathfrak{m}} - \lambda\|_{-1,\omega_p} + h_p \|\bar{f} - f\|_{\omega_p} + h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla} \bar{g}_{\mathfrak{m}}]]^I\|_{\gamma_{p,I}} \right)^2 . \quad (3.54)$$

For the proof of (3.54) we exploited the node values of the extension $\bar{g}_{\mathfrak{m}}(\boldsymbol{q})$ *at nodes which are no potential contact boundary nodes. As we assume that the actual contact boundary is a strict subset of the potential contact boundary only the node value of the extension* $\bar{g}_{\mathfrak{m}}(\boldsymbol{q})$ *at one neighboring interior node is relevant for the proof of (3.54). Thus, if* $g_{\mathfrak{m}}$ *is an affine function over a planar boundary it can be extended to an affine function* $\bar{g}_{\mathfrak{m}}$ *over the whole domain by defining* $\bar{g}_{\mathfrak{m}}(q)$ *at one interior node neighboring to* $\boldsymbol{p}$*, so that* $h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla} \bar{g}_{\mathfrak{m}}]]^I\|_{\gamma_{p,I}} = 0$*. In this sense we may regard* $h_p^{\frac{1}{2}} \|[[\boldsymbol{\nabla} \bar{g}_{\mathfrak{m}}]]^I\|_{\gamma_{p,I}}$ *as measure of the non-linear part of* $g_{\mathfrak{m}}$*.*

4 Numerical studies

The aim of this chapter is to show the numerical performance of our new residual-type a posteriori error estimator for contact problems. We have derived the error estimator for meshes of triangles and tetrahedra in Chapter 3. However, we see in Section 4.2 that the error estimator performs very well for different types of elements as hexahedra, pyramids or prisms.
In Section 4.1 we deal with some implementation aspects. It is noteworthy that the implementation of our new residual-type a posteriori error estimator fits in the framework of a standard residual error estimator because only a few contributions have to be added. In Section 4.2 we illustrate the performance of the new residual-type a posteriori error estimator. We investigate the adaptively refined grids and we compare the experimental order of convergence on adaptively and uniformly refined grids. In Sections 4.2.1 and 4.2.2 we consider the problems given in Section 1.2.2 for which we know the analytic contact stresses and the radius of the actual contact zone. Thus, we give the relative error of both quantities computed on sequences of adaptively refined grids to show the quick error reduction. In Section 4.2.3 the relevance and the distribution of the different local error estimator contributions are investigated for several examples in $3D$ with different kinds of obstacles.

4.1 Implementation aspects

The implementation of the new residual-type a posteriori error estimator has been carried out in the framework of the finite element toolbox `UG` [BBL$^+$97] and the obstacle toolbox `OBSLIB++`, see [Kra00].

4.1.1 Implementation of the error estimator

In the following, we denote by $\mathfrak{m}_k$ the mesh after k-times of adaptive mesh-refinement. An adaptive mesh-refinement process consists of the following steps

⓪ Set $k = 0$. Choose an initial coarse mesh $\mathfrak{m}_0$ and set a tolerance.

① **Solve** the discrete problem on $\mathfrak{m}_k$.

② **Estimate** the error by means of the local error estimator contributions $\eta_\mathfrak{e}$ for each element $\mathfrak{e}$.

③ **Mark** the elements depending on the magnitude of their local error estimator contributions $\eta_\mathfrak{e}$ if the estimated global error is above the tolerance. Otherwise stop the process.

④ **Refine** the mesh $\mathfrak{m}_k \rightsquigarrow \mathfrak{m}_{k+1}$, set $k = k + 1$ and go to step 2.

Although the reliability and efficiency of the error estimator is proven for meshes of triangles or tetrahedra, respectively, we test the performance of the error estimator for different kinds of elements. Hence, the starting grids (Step: ⓪) may consist of quadrilaterals, hexahedra, prisms and pyramids as well. We solve the discrete variational inequalities (Step: ①) with a non-smooth multigrid method, see [Kor97, Kra00].
As the computation of the error estimator contributions η_1, η_2, η_3 are like in the standard residual error estimator it is appropriate to use a hierarchy of classes, see Figure 4.1. In the base class of all kinds of error estimators the error estimator contributions for all elements are summed up in order to check if the given tolerance is achieved or not. If the given tolerance is not reached yet the elements are marked for refinement depending on the marking strategy (Step: ③). In our numerical experiments we use a mean value strategy. An element $\tilde{\mathfrak{e}}$ is marked for refinement if

$$\eta_{\tilde{\mathfrak{e}}}^2 \geq \theta^2 \frac{\sum\limits_{\mathfrak{e}\in\mathfrak{m}} \eta_{\mathfrak{e}}^2}{\sum\limits_{\mathfrak{e}\in\mathfrak{m}} 1}$$

where $\eta_{\mathfrak{e}}$ are the local estimator contributions of the elements and θ is a given parameter which we set to $\theta = 1.1$ in our numerical experiments. We recall that our error estimator is based on patches instead of elements. Further below we explain how the contributions to the elements can be computed.

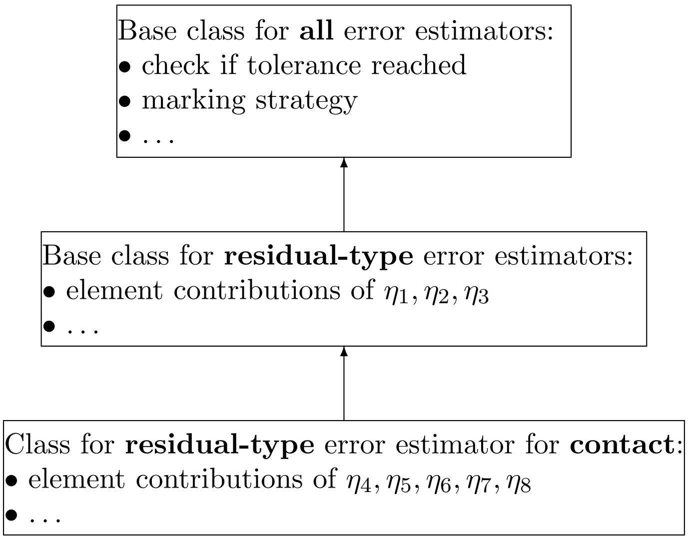


Figure 4.1: Hierarchy of classes for residual-type error estimator for contact

In the base class for residual-type error estimators, the element contributions of η_1, η_2, η_3 are computed. These are the contributions of the standard residual error estimator for linear elliptic equations, see Section 2.2.1. The class of our new residual-type a posteriori error estimator is derived from the base class for residual-type error estimators. It remains to compute the additional error estimator contributions $\eta_4, \eta_5, \eta_6, \eta_7, \eta_8$. In other words, the implementation of the new residual-type a posteriori error estimator can be added to an existing implementation of the standard residual error estimator.

For the computation of the error estimator contributions $\eta_2, \eta_3, \eta_4, \eta_5$, the stresses $\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}})$ have to be computed from the gradients of $\boldsymbol{u}_{\mathfrak{m}}$ and from Hooke's tensor, compare Section 1.1.2. As the discrete solution is given by

$$\boldsymbol{u}_{\mathfrak{m}}(\boldsymbol{x}) = \sum_{p \in N_{\mathfrak{m}}} u_{\mathfrak{m},i}(\boldsymbol{p}) \phi_p(\boldsymbol{x}) \boldsymbol{e}_i,$$

the first derivatives of the nodal basis functions ϕ_p have to be computed. We recall that we use the Einstein summation convention for repeated indices. Usually, the nodal basis functions $\hat{\phi}_p$ on the reference element $\hat{\mathfrak{e}}$ and the transformation $\boldsymbol{T} : \hat{\mathfrak{e}} \to \mathfrak{e}$ are given. The inverse of $\boldsymbol{T}$ denoted by $\boldsymbol{T}^{-1}$ maps points $\boldsymbol{x}$ in $\mathfrak{e}$ on points $\hat{\boldsymbol{x}} = \boldsymbol{T}^{-1}(\boldsymbol{x})$ in $\hat{\mathfrak{e}}$. The nodal basis function ϕ_p is given by the chain $\phi_p = \hat{\phi}_p \circ \boldsymbol{T}^{-1}$. Hence, the partial derivative of ϕ_p is

$$\frac{\partial \phi_p}{\partial x_i}(\boldsymbol{x}) = \left(\frac{\partial \hat{\phi}_p}{\partial \hat{x}_k} \circ \boldsymbol{T}^{-1}(\boldsymbol{x}) \right) \cdot \frac{\partial (\boldsymbol{T}^{-1}(\boldsymbol{x}))_k}{\partial x_i}.$$

The Jacobian matrix $\boldsymbol{\nabla T}^{-1}$ is computed from the Jacobian matrix $\boldsymbol{\nabla T}$ and the corresponding determinant $\det(\boldsymbol{\nabla T})$ by means of $\boldsymbol{\nabla T}^{-1}(\boldsymbol{x}) = (\boldsymbol{\nabla T})^{-1}(\hat{\boldsymbol{x}}) = (\boldsymbol{\nabla T})^{-1} \circ \boldsymbol{T}^{-1}(\boldsymbol{x})$ and $(\boldsymbol{\nabla T})^{-1} = \frac{1}{\det(\boldsymbol{\nabla T})} \operatorname{adj}(\boldsymbol{\nabla T})$.

The computation of $\operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}})$ occurring in the error estimator contribution η_1 if quadrilaterals, hexahedra, prisms or pyramids are used needs the second derivatives of the discrete solution. For example, in the two-dimensional case the first component of the divergence of the stresses is

$$\begin{aligned}
(\operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}}))_1 &= \frac{\partial \sigma_{11}(\boldsymbol{u}_{\mathfrak{m}})}{\partial x_1} + \frac{\partial \sigma_{12}(\boldsymbol{u}_{\mathfrak{m}})}{\partial x_2} \\
&= \frac{\partial}{\partial x_1} \left(H_{11} \epsilon_{11}(\boldsymbol{u}_{\mathfrak{m}}) + H_{12} \epsilon_{22}(\boldsymbol{u}_{\mathfrak{m}}) + 2H_{13} \epsilon_{12}(\boldsymbol{u}_{\mathfrak{m}}) \right) \\
&\quad + \frac{\partial}{\partial x_2} \left(H_{31} \epsilon_{11}(\boldsymbol{u}_{\mathfrak{m}}) + H_{32} \epsilon_{22}(\boldsymbol{u}_{\mathfrak{m}}) + 2H_{33} \epsilon_{12}(\boldsymbol{u}_{\mathfrak{m}}) \right) \\
&= H_{11} \frac{\partial}{\partial x_1} \left(\frac{\partial u_{\mathfrak{m},1}}{\partial x_1} \right) + H_{12} \frac{\partial}{\partial x_1} \left(\frac{\partial u_{\mathfrak{m},2}}{\partial x_2} \right) + H_{13} \frac{\partial}{\partial x_1} \left(\frac{\partial u_{\mathfrak{m},1}}{\partial x_2} + \frac{\partial u_{\mathfrak{m},2}}{\partial x_1} \right) \\
&\quad + H_{31} \frac{\partial}{\partial x_2} \left(\frac{\partial u_{\mathfrak{m},1}}{\partial x_1} \right) + H_{32} \frac{\partial}{\partial x_2} \left(\frac{\partial u_{\mathfrak{m},2}}{\partial x_2} \right) + H_{33} \frac{\partial}{\partial x_2} \left(\frac{\partial u_{\mathfrak{m},1}}{\partial x_2} + \frac{\partial u_{\mathfrak{m},2}}{\partial x_1} \right)
\end{aligned}$$

where ϵ_{ij} are the components of the strain tensor (1.10) and H_{ij} are the components of the matrix representation of Hooke's tensor defined in (1.16). The second partial derivatives of the nodal basis functions in terms of $\hat{\phi}_p$ and the transformation $\boldsymbol{T}^{-1}$ are given by

$$
\begin{aligned}
\frac{\partial^2 \phi_p}{\partial x_i \partial x_j}(\boldsymbol{x}) &= \frac{\partial^2 (\hat{\phi}_p \circ \boldsymbol{T}^{-1})}{\partial x_i \partial x_j}(\boldsymbol{x}) \\
&= \frac{\partial}{\partial x_i}\left(\left(\frac{\partial \hat{\phi}_p}{\partial \hat{x}_l} \circ \boldsymbol{T}^{-1}(\boldsymbol{x})\right) \cdot \frac{\partial (\boldsymbol{T}^{-1}(\boldsymbol{x}))_l}{\partial x_j}\right) \\
&= \left(\frac{\partial^2 \hat{\phi}_p}{\partial \hat{x}_l \hat{x}_k} \circ \boldsymbol{T}^{-1}(\boldsymbol{x})\right) \cdot \frac{\partial (\boldsymbol{T}^{-1}(\boldsymbol{x}))_k}{\partial x_i} \cdot \frac{\partial (\boldsymbol{T}^{-1}(\boldsymbol{x}))_l}{\partial x_j} \\
&\quad + \left(\frac{\partial \hat{\phi}_p}{\partial \hat{x}_l} \circ \boldsymbol{T}^{-1}(\boldsymbol{x})\right) \cdot \frac{\partial^2 (\boldsymbol{T}^{-1}(\boldsymbol{x}))_l}{\partial x_j \partial x_i}.
\end{aligned}
$$

Hence, the first and second derivatives of the nodal basis functions on the reference element and of the inverse transformation are needed. As we do not have the inverse transformation, the computation of the second derivative of $\boldsymbol{T}^{-1}$ is not straightforward. Exemplarily, we consider the two-dimensional case. As explained above $\boldsymbol{\nabla T}^{-1}(\boldsymbol{x}) = (\boldsymbol{\nabla T})^{-1}(\hat{\boldsymbol{x}})$ and

$$
(\boldsymbol{\nabla T})^{-1} = \frac{1}{\det(\boldsymbol{\nabla T})}\begin{pmatrix} (\boldsymbol{\nabla T})_{22} & -(\boldsymbol{\nabla T})_{12} \\ -(\boldsymbol{\nabla T})_{21} & (\boldsymbol{\nabla T})_{11} \end{pmatrix}.
$$

Thus, in order to compute the second derivative of $\boldsymbol{\nabla T}^{-1}$, the derivatives of $(\boldsymbol{\nabla T})_{ij}$ and of $\det(\boldsymbol{\nabla T})$ are required. For example, the second partial derivative $\frac{\partial \left(\boldsymbol{T}^{-1}(\boldsymbol{x})\right)_1}{\partial x_1 \partial x_1}$ is computed from

$$
\begin{aligned}
&\frac{\partial}{\partial x_1}(\boldsymbol{\nabla T}^{-1})_{11}(\boldsymbol{x}) \\
&= \frac{\partial}{\partial x_1}\left((\boldsymbol{\nabla T})_{22} \circ \boldsymbol{T}^{-1}(\boldsymbol{x}) \cdot \frac{1}{\det(\boldsymbol{\nabla T}) \circ \boldsymbol{T}^{-1}(\boldsymbol{x})}\right) \\
&= \frac{1}{(\det(\boldsymbol{\nabla T}) \circ \boldsymbol{T}^{-1}(\boldsymbol{x}))^2}\left\{\left(\frac{\partial}{\partial \hat{x}_i}(\boldsymbol{\nabla T})_{22} \circ \boldsymbol{T}^{-1}(\boldsymbol{x})\right) \cdot \frac{\partial (\boldsymbol{T}^{-1}(\boldsymbol{x}))_i}{\partial x_1} \cdot \det(\boldsymbol{\nabla T}) \circ \boldsymbol{T}^{-1}(\boldsymbol{x})\right. \\
&\quad \left. -(\boldsymbol{\nabla T})_{22} \circ \boldsymbol{T}^{-1}(\boldsymbol{x}) \cdot \frac{\partial}{\partial \hat{x}_i}\det(\boldsymbol{\nabla T}) \circ \boldsymbol{T}^{-1}(\boldsymbol{x}) \cdot \frac{\partial (\boldsymbol{T}^{-1}(\boldsymbol{x}))_i}{\partial x_1}\right\}.
\end{aligned}
$$

In our implementation of the error estimator we loop over all elements. For each element $\mathfrak{e}$ we compute $\int_{\mathfrak{e}} (\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}))^2 \, dx$ by means of the first and second derivatives of the transformation and the basis functions on the reference element as explained above. As the error estimator is based on patches and an element occurs in each patch of its nodes, we weight the element contribution $\int_{\mathfrak{e}} (\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}))^2 \, dx$ by the number of nodes belonging to $\mathfrak{e}$.

For each element we loop over the sides to compute the remaining error estimator contributions. Depending on the kind of side, interior side, Neumann boundary side or potential contact boundary side, we compute the integral of the square of the stresses or jump terms over the side and weight it by the number of corresponding nodes. Exemplarily, we consider η_5. The side contribution $\int_{\mathfrak{s}} (\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}))^2 \, da$ has to be weighted by the number of nodes $\boldsymbol{p} \in N_{\mathfrak{m}} \backslash N_{\mathfrak{m}}^{\mathrm{fC}}$ belonging to this side $\mathfrak{s}$. Therefore, we loop over the nodes to check if they are semi- or full-contact nodes or not in actual contact. Nodes are full-contact nodes if all neighboring nodes belonging to the potential contact boundary are in actual contact. Therefore, a list of patches which assigns all neighboring nodes to a node is convenient.

The classification of contact nodes is also required for the computation of the error estimator contribution η_6. For a semi-contact node $\boldsymbol{p}$ on a side $\mathfrak{s}$, we compute s_p and the part of d_p belonging to the side $\mathfrak{s}$. Instead of generating a subgrid as used in the proof of the lower bound, see Figure 3.1, we may compute the integral of $(g_{\mathfrak{m}} - u_{\mathfrak{m},1})\phi_p$ over a strict subset as, e.g., a fourth of the triangle containing the node $\boldsymbol{p}$. We proceed in the same way for the computation of the side contributions to η_7.

4.1.2 Evaluation and visualization

We shortly comment on the evaluation and visualization of the quantities which we investigate in the following sections for different numerical experiments.

In order to analyze our new residual-type error estimator in detail, we analyze the structure of the adaptively refined grid and the distribution of the local error estimator contributions. Therefore, we implemented a routine to write the grid and the data in a `vtk` file format. Thus, it can be visualized in `paraview`.

We do not have the exact solution $\boldsymbol{u}$ on Ω of our examples. But for some examples we know the exact contact stresses, compare Section 1.2.2. As already explained in Section 3.1.1 there is no obvious discrete counterpart of the contact stresses. We make use of the node values s_p of the discrete contact force density obtained by the lumped L^2-scalar product to define a numerical approximation of the exact contact stresses

$$s_{\mathfrak{m}} := \sum_{p \in N_{\mathfrak{m}}^C} s_p \phi_p. \tag{4.1}$$

This quantity defines a functional on $H^{\frac{1}{2}}(\Gamma_C)$

$$\langle s_{\mathfrak{m}}, \varphi_{\mathfrak{m},1} \rangle_{-\frac{1}{2},\frac{1}{2},\Gamma_C} = \int_{\Gamma_C} \left(\sum_{p \in N_{\mathfrak{m}}^C} s_p \phi_p \right) \varphi_{\mathfrak{m},1},$$

see also (3.9). We compare the exact contact stresses and the approximation $s_{\mathfrak{m}}$ for the examples presented in Sections 4.2.1 and 4.2.2.

As we do not have the exact solution we do not know the error of our numerical solution. Therefore, in order to investigate further the performance of our new residual type error estimator, we analyze the rate of convergence of the error estimator instead of the error on adaptively and uniformly refined grids. We note that $h \simeq (\# N_{\mathfrak{m}})^{-\frac{1}{d}}$ for a uniform grid. Let $\mathfrak{m}_k$ and $\mathfrak{m}_{k+1}$ be the grids of two successive steps of adaptive refinement. The corresponding sets of nodes are $N_{\mathfrak{m}_k}$, $N_{\mathfrak{m}_{k+1}}$ and the error estimators are η^k and

η^{k+1}. Then the experimental order of convergence eoc, fulfilling $\eta^k = (\#N_{\mathfrak{m},k})^{-\mathrm{eoc}}$ and $\eta^{k+1} = (\#N_{\mathfrak{m},k+1})^{-\mathrm{eoc}}$ is given by

$$\mathrm{eoc} := \frac{\ln(\eta^k) - \ln(\eta^{k+1})}{\ln(\#N_{\mathrm{m},k+1}) - \ln(\#N_{\mathrm{m},k})}. \tag{4.2}$$

This formula can be derived as follows

$$\begin{aligned} & \tfrac{\eta^k}{\eta^{k+1}} &&= \left(\frac{\#N_{\mathrm{m},k}}{\#N_{\mathrm{m},k+1}}\right)^{-\mathrm{eoc}} \\ \Leftrightarrow \quad & \ln\left(\tfrac{\eta^k}{\eta^{k+1}}\right) &&= -\mathrm{eoc}\ln\left(\frac{\#N_{\mathrm{m},k}}{\#N_{\mathrm{m},k+1}}\right) \\ \Leftrightarrow \quad & \left(\ln(\eta^k) - \ln(\eta^{k+1})\right) &&= -\mathrm{eoc}\left(\ln(\#N_{\mathrm{m},k}) - \ln(\#N_{\mathrm{m},k+1})\right). \end{aligned}$$

4.2 Numerical results

In this section we present several numerical examples to show the performance of our new residual-type error estimator for contact. We analyze the structure of the adaptively refined grids and the experimental order of convergence (4.2) for the adaptive strategy compared to uniform refinement.
In the first two sections we compute the numerical solution of the examples given in Section 1.2.2, compare Figures 1.1 and 1.2, in an adaptive mesh refinement process based on our new estimator for contact. For these examples we compute the analytic contact stresses and the radius of the actual contact zone. Hence, we give the relative error of both quantities and we visualize the exact contact stresses and the numerical approximation (4.1). It is the aim of adaptive mesh refinement by means of a posteriori error estimators to make possible a good approximation of the solution with a small number of nodes compared to uniform refinement. Therefore, the relative errors are plotted against the number of actual contact nodes for meshes generated by means of our new residual-type error estimator (Section 3.1.2) and by means of the standard residual error estimator (Section 2.2.1).
In the last section we study the new estimator in more detail. For this purpose, we analyze the relevance and the distribution of the different local error estimator contributions.

4.2.1 $2D$ examples with analytically given distribution of contact stresses

In Section 1.2.2 we presented two examples of contact between a rigid and an elastic body for which the distribution of the contact stresses is given analytically. In order to model these examples by one-body contact problems we consider a unit quadrilateral with the edges $(0,0), (0,1), (1,0), (1,1)$ which is moved towards the obstacle by means of Dirichlet values. We apply Dirichlet boundary values -0.01 in x_2-direction and 0.0 in x_1-direction at the top of the quadrilateral. The force density and the Neumann values are zero. The material of the quadrilateral is assumed to be linear elastic with the Poisson ratio $\nu = 0.25$ and the Young's modulus $E = 3 \cdot 10^5$. We note that the unit of the stresses is the same as the unit chosen for the Young's modulus, e.g., $\frac{kN}{mm^2}$. The potential contact boundary is the lower edge of the quadrilateral and the direction of constraints is $\boldsymbol{\nu} = -\boldsymbol{e}_2$.

In the first example, where the elastic quadrilateral comes into contact with a rigid punch the contact constraints $u_\nu \leq g = 0$ are imposed in a strip of the length 0.125 between 0.4375 and 0.5625 and in the second example where the contact with a wedge ($\alpha \approx 79°$) is modeled, $g(x_1) = 0.2 \cdot (x_1 - 0.5)$ if $x_1 \geq 0.5$ and $g(x_1) = -0.2 \cdot (x_1 - 0.5)$ otherwise. We refer to these examples by Example 1 and Example 2, respectively.

We note that the load P applied to the rigid body in the examples of Section 1.2.2 coincides with the integral of the contact stresses over the contact boundary $\int_{\Gamma_C} \hat{\sigma}_\nu \, dx_1$. Thus, we solve the contact problem on a very fine grid in a preprocess to get a good approximation of the total contact forces. In this way we get the value 1549.10 for Example 1, enabling the computation of the exact contact stresses

$$\hat{\sigma}_\nu^{\text{exact}}(\boldsymbol{x}) = \frac{1549.10}{\pi((0.0525)^2 - x_T^2)^{\frac{1}{2}}}. \tag{4.3}$$

in the open interval $(0.4375, 0.5625)$ for x_T. At the points $\boldsymbol{x}$ with $x_T = 0.4375$ or $x_T = 0.5625$ where the edges of the punch indent the elastic quadrilateral the value of the exact contact stresses would be infinite.

The starting grids of our experiments are uniformly refined meshes, consisting of either triangles (512 elements) or quadrilaterals (256 elements). Subsequently, the adaptive refinement process has been carried out 15 times. The experiment is called Example 1A if the starting grid consists of quadrilaterals and Example 1B if the starting grid is made of triangles. The final system has $540,924$ degrees of freedom in Example 1A and $438,334$ degrees of freedom in Example 1B.

Figure 4.2 shows the final grid and zooms in the contact area in Example 1A. The refinement is symmetric and the smallest elements are located in the semi-contact zone and the neighboring area where discontinuities are expected. Figure 4.3 shows the grid after 15 steps of refinement in Example 1B. In Figure 4.3(a) we have zoomed in the contact area and in Figure 4.3(b) we have further zoomed in the area where the right edge of the punch indents the elastic body. Although the structure is similar to the one with a starting grid made of quadrilaterals the grid is less symmetric which is due to the alignment of the triangles on the starting grid. However, the solution (Figure 4.3(c)) is symmetric.

In Figure 4.4 the error estimator is plotted against the number of nodes for uniformly (blue line) and adaptively (red line) refined grids with a logarithmic scale on both axis. The experimental order of convergence (4.2) is given by the slopes of the graphs. For both experiments the final eoc is about 0.5 for the adaptive strategy compared to ≈ 0.25 for the uniform refinement.

In Figure 4.5 we compare the node values s_p with the exact values of the contact stresses $\hat{\sigma}_\nu^{\text{exact}}(\boldsymbol{p})$. As we can see the values s_p (red dots in Figure 4.5) coincide very well with the exact contact stresses (blue dots in Figure 4.5). Next, we consider the relative error

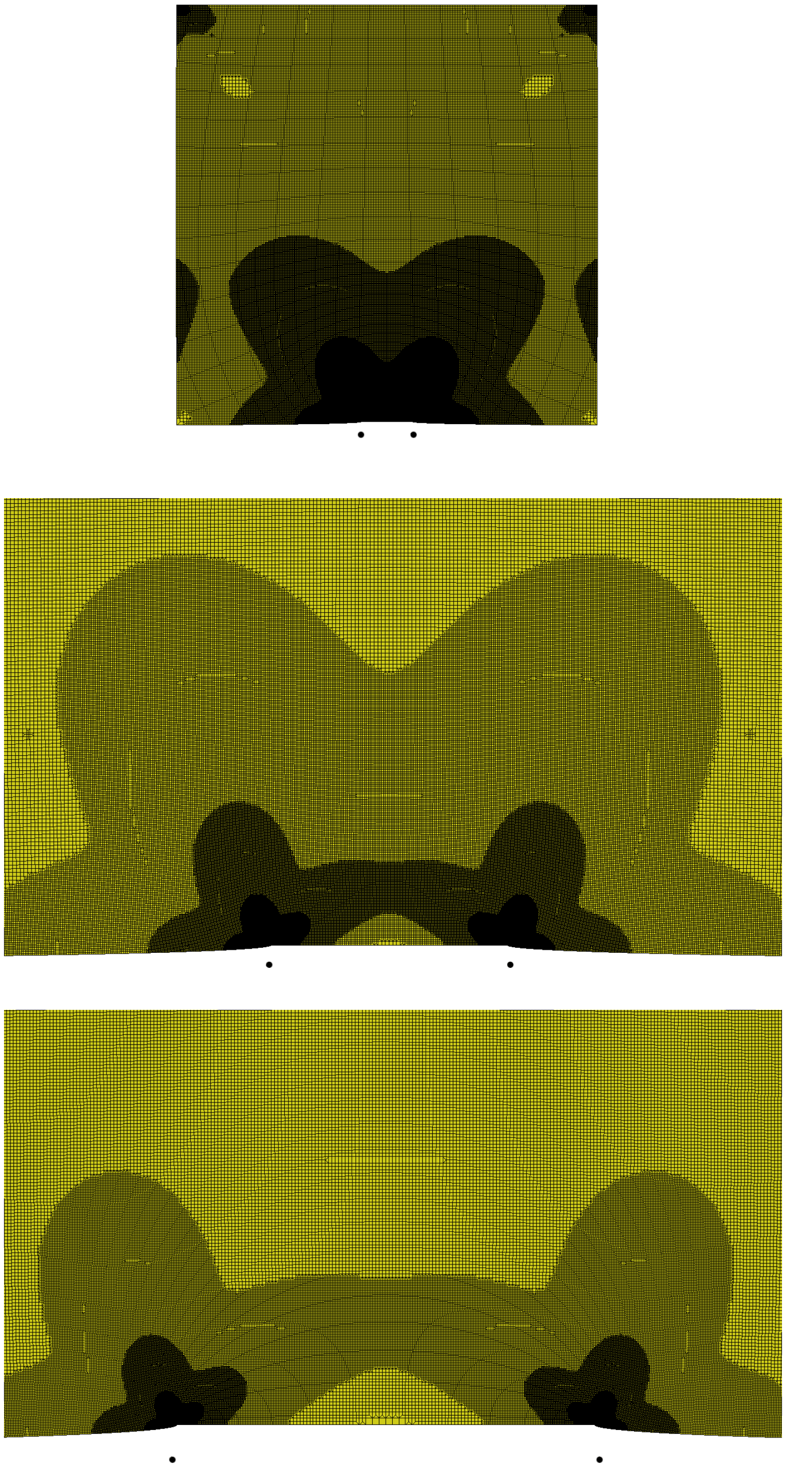
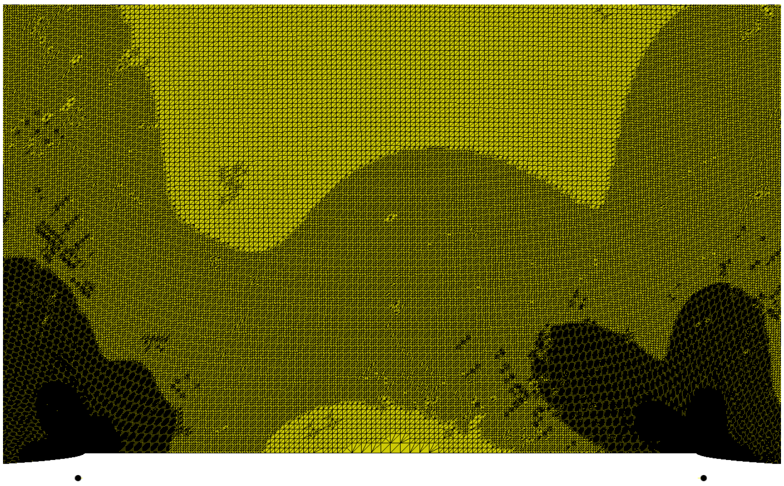
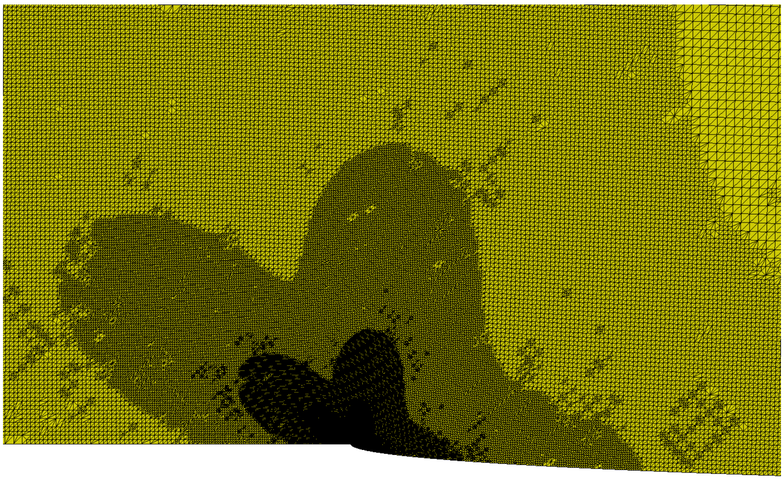


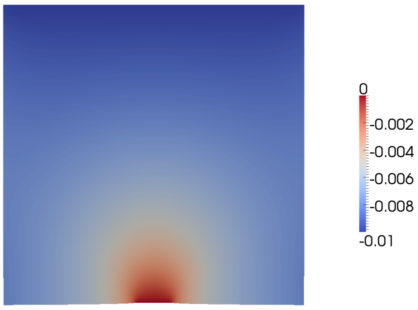
Figure 4.2: Example 1A: Final grid and zooms in the area around the actual contact boundary. The nodes mark the edges of the contact zone.



(a) Zoom in contact area of final grid

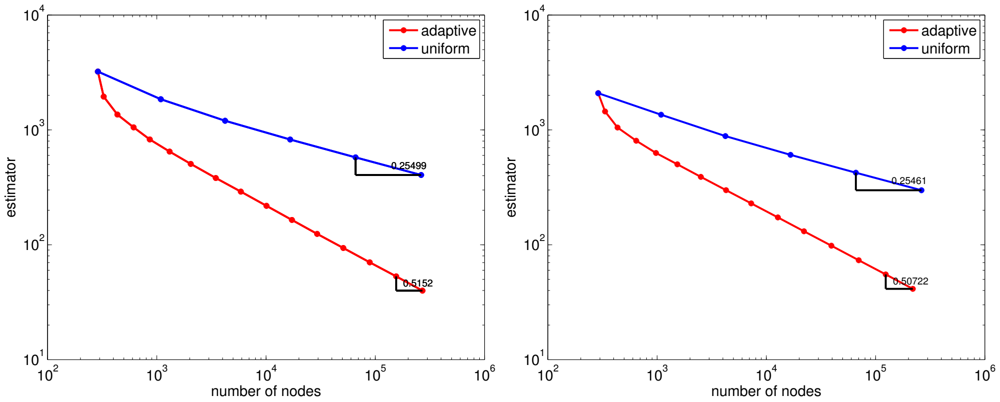


(b) Zoom in contact area of final grid where the edge of the punch indents the elastic body



(c) Component u_2 of the solution

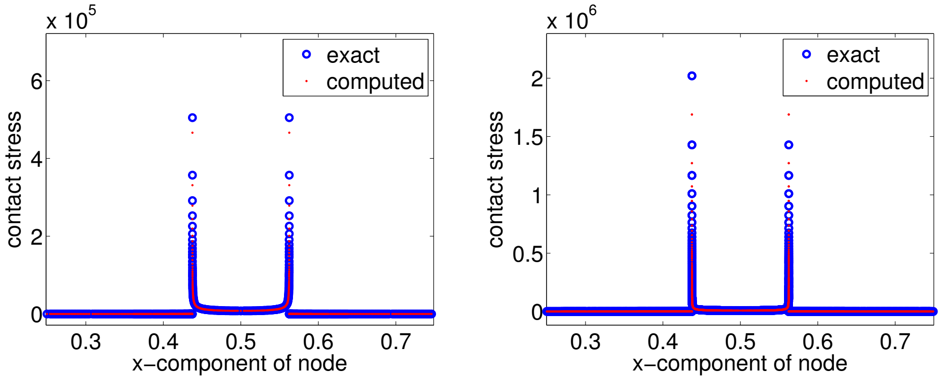
Figure 4.3: Example 1B



(a) Example 1A

(b) Example 1B

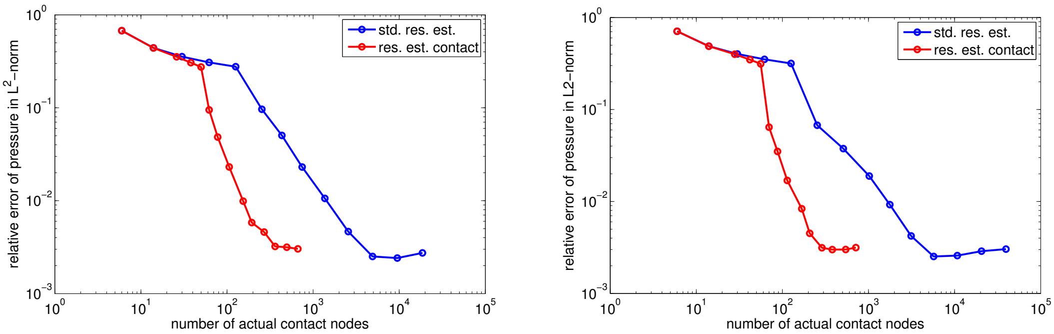
Figure 4.4: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)



(a) Example 1A

(b) Example 1B

Figure 4.5: Contact stresses



(a) Example 1A

(b) Example 1B

Figure 4.6: Relative error of pressure for $s_{\mathfrak{m}}$ computed on grids which are adaptively refined by means of our residual estimator for contact (res. est. contact) and the standard residual estimator (std. res. est.) plotted against $\#N^{\mathrm{aC}}_{\mathfrak{m}_k}$ (logarithmic scales)

in the L^2-norm

$$\frac{\|s_{\mathfrak{m}}(\boldsymbol{x}) - \hat{\sigma}_\nu^{\text{exact}}(\boldsymbol{x})\|_{\Gamma_C}}{\|\hat{\sigma}_\nu^{\text{exact}}(\boldsymbol{x})\|_{\Gamma_C}}, \tag{4.4}$$

plotted against the number of actual contact nodes in each adaptive refinement step. The red line in Figure 4.6 shows the relative error of $s_{\mathfrak{m}}$ computed on adaptively refined grids which are generated by means of our residual-type error estimator for contact. The blue line shows the relative error of $s_{\mathfrak{m}}$ computed on grids which are adaptively refined with the help of the standard residual error estimator, see Section 2.2.1. Using our new residual-type a posteriori estimator for contact we get a relative error of $\approx 3 \cdot 10^{-3}$ with solely 666 actual contact nodes in Example 1A and 380 in Example 1B, respectively. If the mesh is refined by means of the standard residual error estimator the same quality of approximation is obtained with $4,898$ actual contact nodes in Example 1A and $5,714$ in Example 1B, respectively.

In Example 2, where the rigid wedge indents the elastic quadrilateral, the cotangens of the angle α of the rigid wedge and the load applied to the rigid body, compare Figure 1.1(b) in Section 1.2.2, enter the formulas for the exact computation of the radius of the contact zone and the contact stresses. As in the foregoing example the load is given by the integral of the contact stresses over the contact area $\int_{\Gamma_C} \hat{\sigma}_\nu \, dx_1$ computed in a preprocess on a very fine grid. It amounts to 872.19. The cotangens of the angle of the wedge is 0.2. Putting these values and $\nu = 0.25$ and $E = 3 \cdot 10^5$ in the formula (1.28) in Section 1.2.2 we get the radius $b = 0.01362796875$ of the contact zone.

If the starting grid consists of quadrilaterals we call the experiment Example 2A and if the starting grid is made of triangles we refer to it by Example 2B. The adaptive refinement process has been carried out 14 times. The number of degrees of freedom is 510,588 in Example 2A and 512,850 in Example 2B.

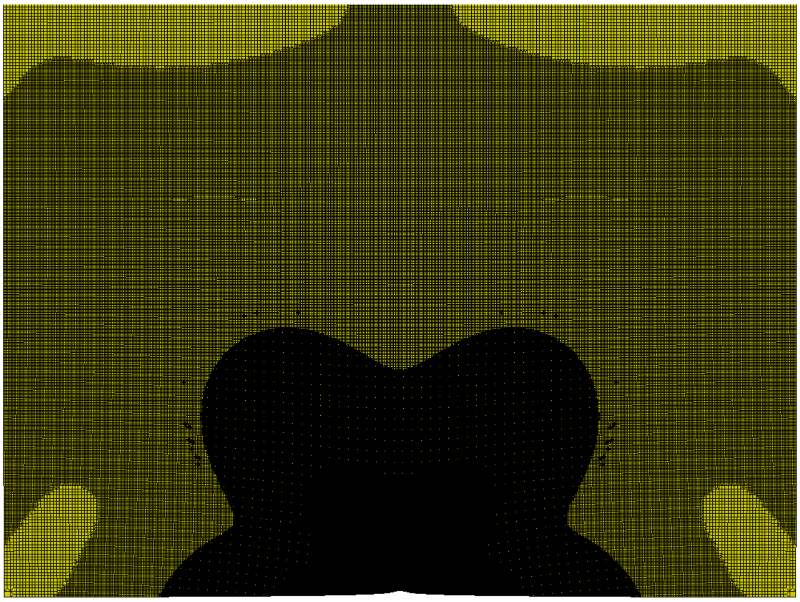
Figures 4.7 and 4.8 show zooms in the contact area for Example 2A. The refined grid

is symmetric and the smallest elements can be found mainly in the area where the tip of the wedge indents the elastic quadrilateral. At this point the exact normal stresses would be infinite. However, in small areas around the free boundary zone the elements have been refined very often, too, as can be seen in Figure 4.8(b). Just as in Example 1 the adaptively refined grid based on a starting grid of triangles is less symmetric, see Figure 4.9(a), although the solution is symmetric (Figure 4.9(b)).
The convergence of the error estimator on adaptively and uniformly refined grids is shown in Figure 4.10. The experimental order of convergence is about 0.502 for the adaptively refined grid and 0.436 and 0.399, respectively, for the uniformly refined grid.

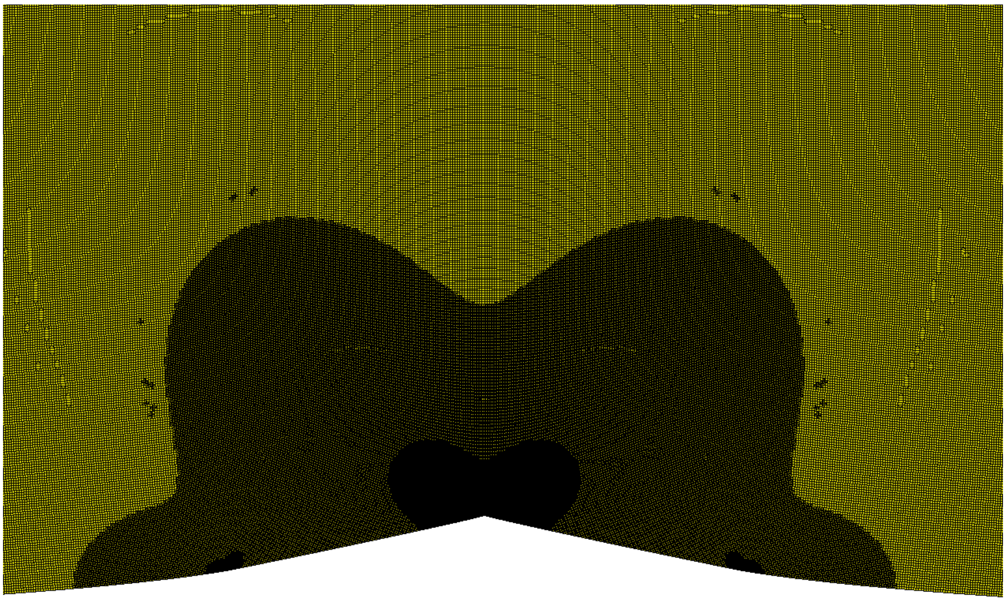
In Figure 4.11 the exact normal stresses $\hat{\sigma}_{\nu}^{\text{exact}}(\boldsymbol{p})$ (blue dots), computed by means of formula (1.29) in Section 1.2.2, are compared to the corresponding numerically computed quantities s_p (red dots). The values coincide very well. The L^2-norm of the relative error between $s_{\mathfrak{m}}(\boldsymbol{x})$ and $\hat{\sigma}_{\nu}^{\text{exact}}(\boldsymbol{x})$ as defined in (4.4) is given by the red lines in Figure 4.12 plotted against the number of actual contact nodes. The blue lines in Figure 4.12 refer to the same relative error but $s_{\mathfrak{m}}$ is computed on a mesh where the adaptive refinement is determined by the standard residual error estimator (Section 2.2.1). For both methods of adaptive refinement Figure 4.13 shows the relative error between b and the numerically computed radius of the contact zone plotted against the number of actual contact nodes. In order to visualize that the adaptive refinement process by means of the standard residual error estimator requires much more nodes in order to obtain the same accuracy as by means of our residual-type estimator for contact we have chosen logarithmic scales in Figures 4.12 and 4.13.

4.2.2 $3D$ example with analytically given distribution of contact stresses

The final example presented in Section 1.2.2 of a contact problem for which exact formulas of the contact stresses (1.31) and the radius of the contact zone (1.30) are given is a special case of the Hertzian contact problem, see Figure 1.2. We compare the exact and numerically computed values of contact stresses and radius of contact zone and we show the grids which are adaptively refined by means of our new residual-type error estimator for contact.
The Hertzian contact problem is modeled by a linear elastic cube which is pushed against the obstacle by means of Dirichlet values. We apply Dirichlet values -0.005 in x_3-direction and 0.0 in x_1- and x_2-direction at the top of the cube. As in the $2D$ examples the Neumann values and the force density is set to zero and we use the same material constants $E = 3 \cdot 10^5$ and $\nu = 0.25$. The potential contact boundary is the bottom of the cube and the direction of constraints is $\boldsymbol{\nu} = -\boldsymbol{e}_3$. The gap function describing the distance between the elastic cube and the rigid ball with radius R^M and midpoint $\boldsymbol{m}$ is given by $g = \sqrt{(R^M)^2 - (x_1 - m_1)^2 - (x_2 - m_2)^2} + m_3$. For the computation of the distribution of the contact stresses the value of the load P is needed. It is computed in a preprocess on a very fine grid and amounts to 88.28. The radius of the rigid ball representing the obstacle is $R^M = 0.75$. Thus, the exact radius of the contact zone is 0.615142, computed by (1.30).

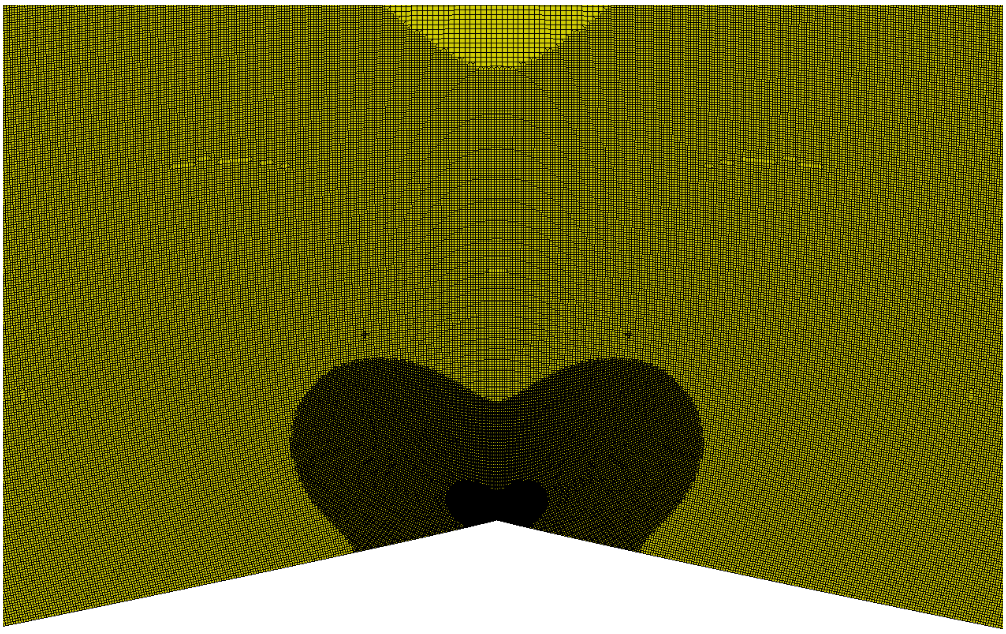


(a) Zoom of final grid

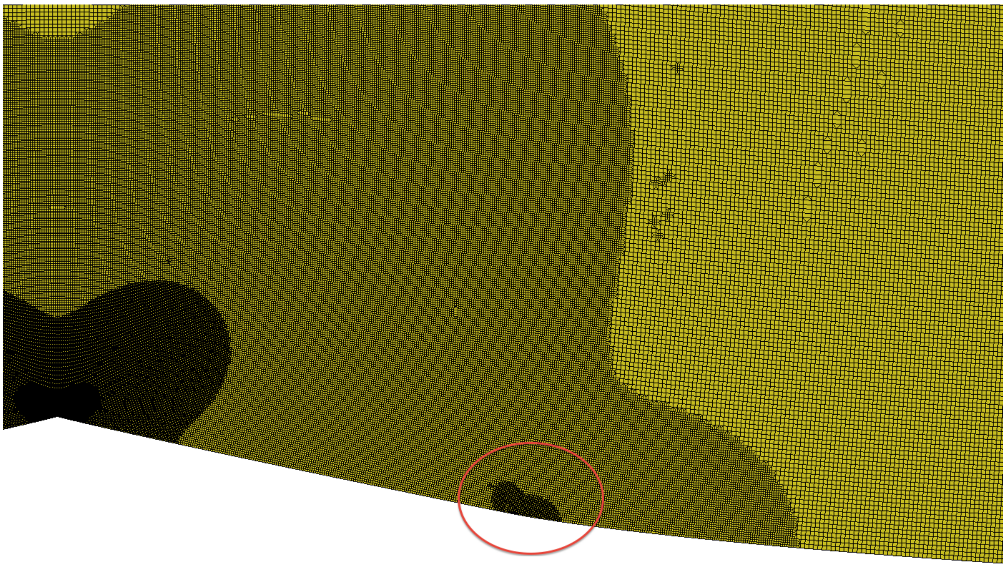


(b) Zoom in contact area of final grid

Figure 4.7: Example 2A

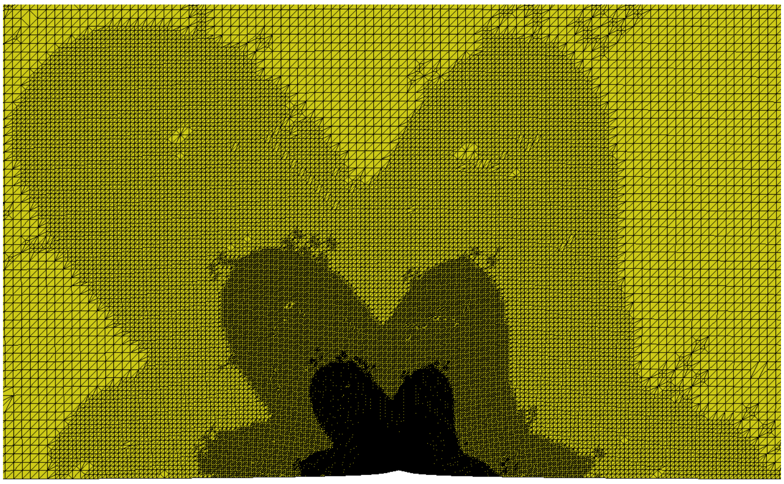


(a) Zoom in contact area of final grid

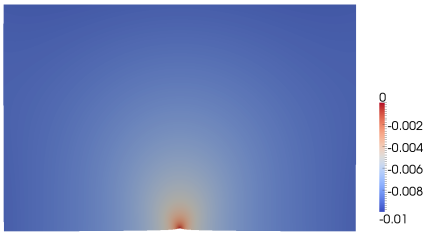


(b) Zoom in right side of contact area of final grid

Figure 4.8: Example 2A

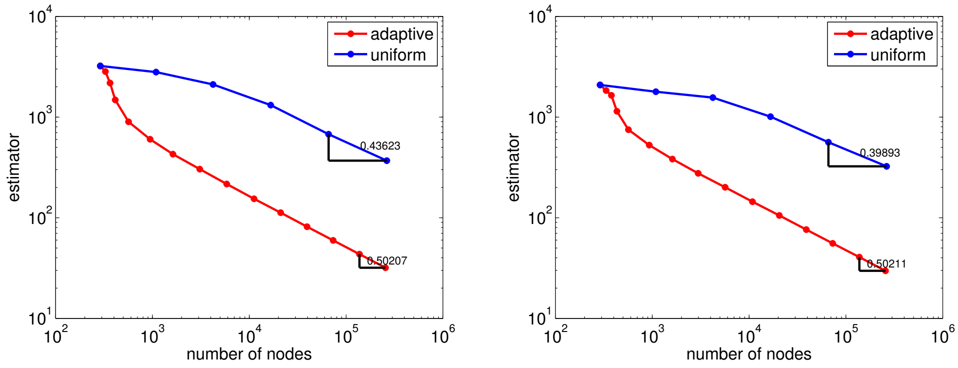


(a) Zoom in contact area of final grid



(b) Component u_2 of the solution

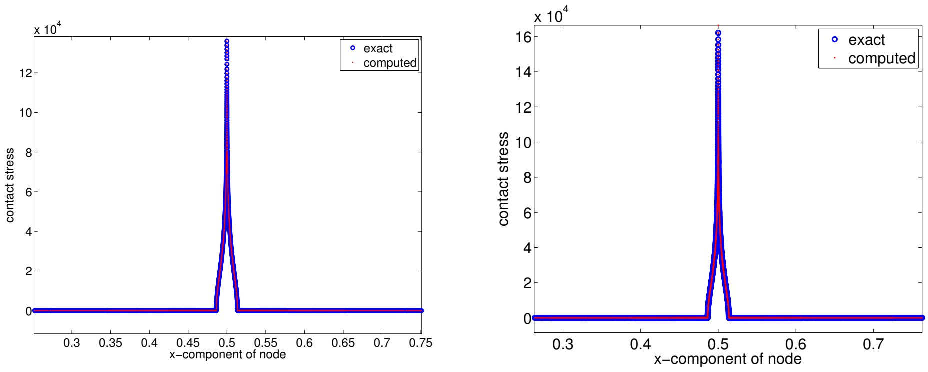
Figure 4.9: Example 2B



(a) Example 2A

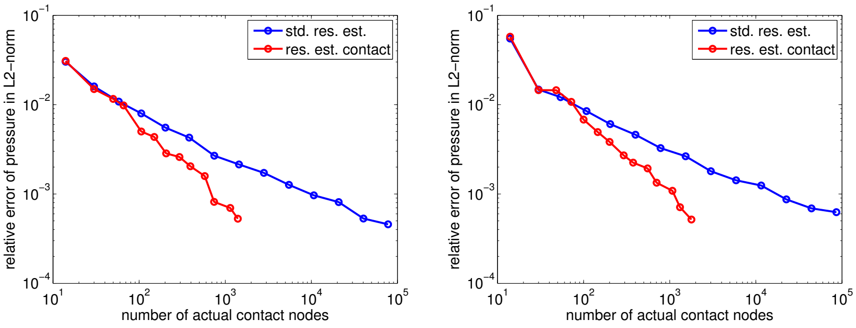
(b) Example 2B

Figure 4.10: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)



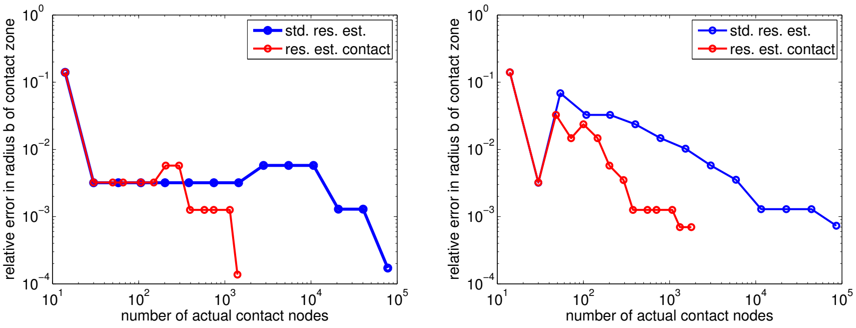
(a) Example 2A (b) Example 2B

Figure 4.11: Contact stresses



(a) Example 2A (b) Example 2B

Figure 4.12: Relative error of pressure for $s_{\mathfrak{m}}$ computed on grids which are adaptively refined by means of our residual estimator for contact (res. est. contact) and the standard residual estimator (std. res. est.) plotted against $\#N^{\mathrm{aC}}_{\mathfrak{m}_k}$ (logarithmic scales)



(a) Example 2A

(b) Example 2B

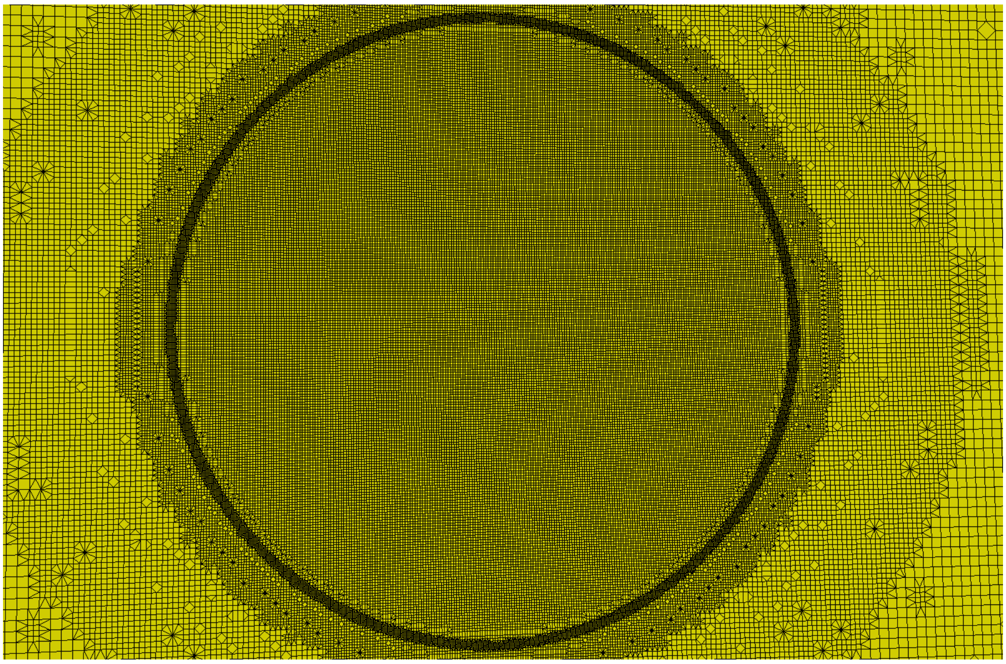
Figure 4.13: Relative error of radius of contact zone computed on grids which are adaptively refined by means of our residual estimator for contact (res. est. contact) and the standard residual estimator (std. res. est.) plotted against $\#N^{\mathrm{aC}}_{\mathfrak{m}_k}$ (logarithmic scales)

We consider two experiments Example 3A where the starting grid is a uniform mesh of hexahedra (64 elements) and Example 3B where the starting grid is a uniform mesh of tetrahedra (384 elements). The final grids have been 10 times adaptively refined. The number of degrees of freedom is 4,871,271 in Example 3A and 6,044,778 in Example 3B, respectively.

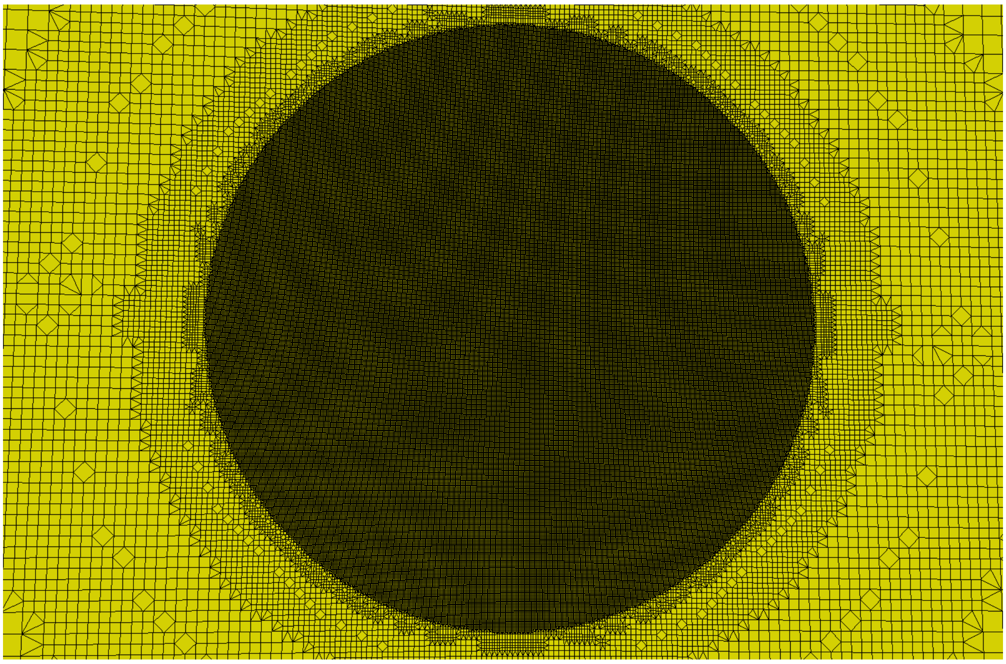
In Figure 4.14(a) we show a zoom of the actual contact boundary of the adaptively refined grid of Example 3A. The free boundary is refined more often as the area around full-contact nodes. In contrast, the adaptive refinement by means of the standard residual error estimator, see Figure 4.14(b), does not distinguish between the free boundary and the full-contact area. In this case the actual contact boundary is overrefined.

Figures 4.15(b) and (c) show zooms of the actual contact boundary of the adaptively refined grids by means of our residual-type estimator for contact and by means of the standard residual estimator where the starting grids consist of tetrahedra. The grid is less symmetric as in Example 3A where the starting grid is made of hexahedra. Just as in the $2D$ examples of Section 4.2.1 this is due to the alignment of the tetrahedra on the starting grid. However, the solution in Figure 4.15(a) is symmetric.

For both experiments, Example 3A and 3B, the rates of convergence of our new residual estimator for contact have been plotted in Figure 4.16 for the adaptively refined grids as well as for the uniformly refined grids. Although the experimental order of convergence for the adaptive refinement strategy is better as for the uniform refinement, the advantage is less as in the 2D examples of Section 4.2.1. The reason might be found in the regularity of the solutions. In both examples of Section 4.2.1 the contact stresses are expected to grow up to infinity where the edges of the punch or the tip of the wedge indent the elastic body. In contrast, the contact stresses $\hat{\sigma}_\nu(\boldsymbol{x})$ of the Hertzian contact problem are always finite, compare (1.31). The contact stresses in Example 3A and 3B can be seen in Figure 4.17.

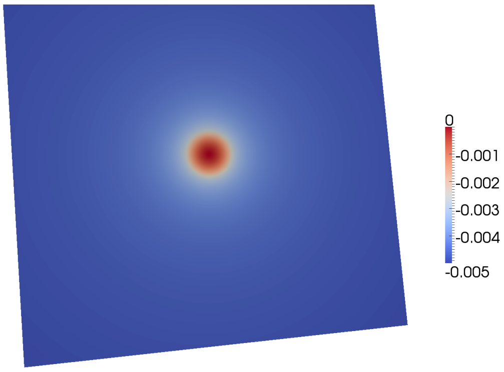


(a) New residual-type estimator for contact

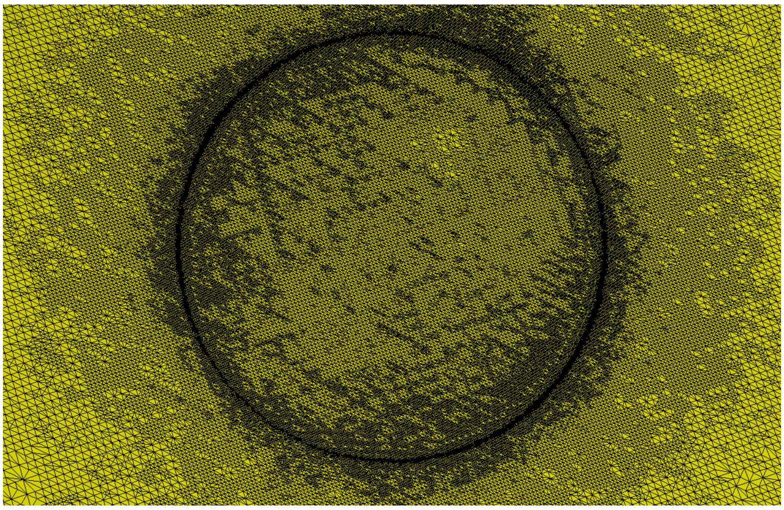


(b) Standard residual estimator

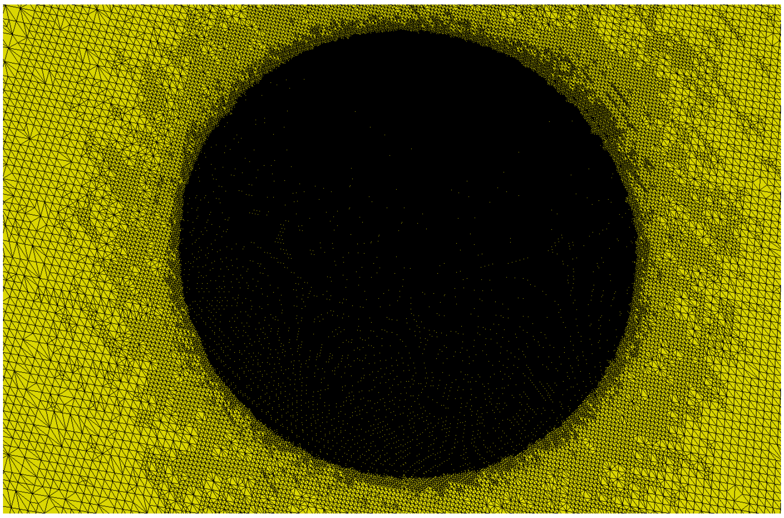
Figure 4.14: Example 3A: Zoom in contact area of final grid



(a) Component u_3 of the solution

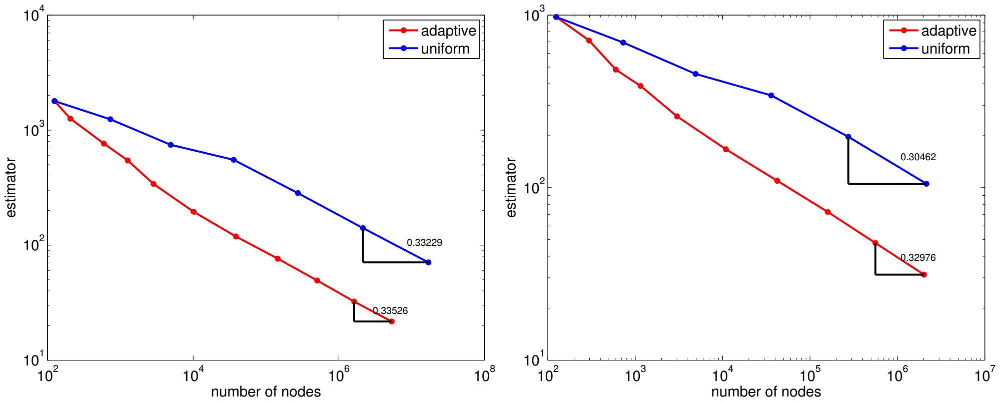


(b) New residual-type estimator for contact



(c) Standard residual estimator

Figure 4.15: Example 3B: Solution and zoom in contact area of final grid



(a) Example 3A

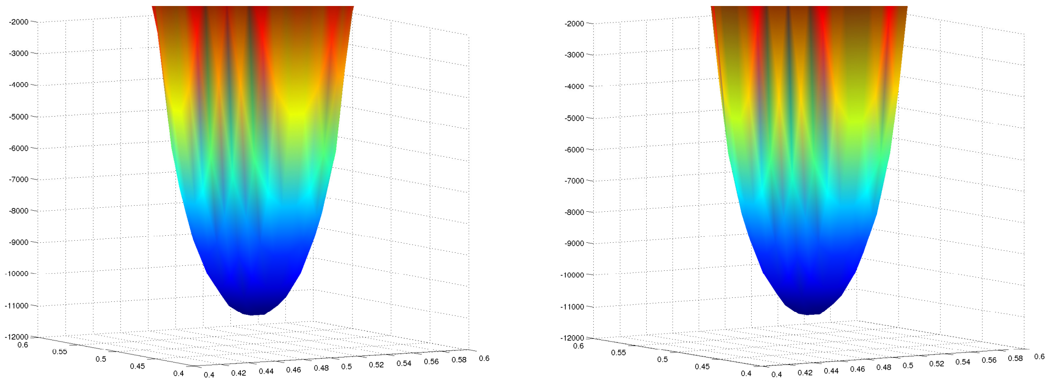
(b) Example 3B

Figure 4.16: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)

The L^2-norm of the relative error (4.4) between $s_{\mathfrak{m}}(\boldsymbol{x})$ and $\hat{\sigma}_\nu^{\text{exact}}(\boldsymbol{x})$ is plotted against the number of actual contact nodes in Figure 4.18 for both kinds of starting grids. As in the foregoing examples the red line shows the relative error of $s_{\mathfrak{m}}$ computed on grids which have been adaptively refined by means of our residual-type estimator for contact. The blue lines in Figure 4.18 represent the same quantity for $s_{\mathfrak{m}}$ computed on grids which have been adaptively refined by means of the standard residual error estimators. In order to illustrate that by means of our residual estimator for contact less nodes are required to obtain the same accuracy in the contact stresses as by means of the standard residual error estimator we have chosen logarithmic scales in Figure 4.18.

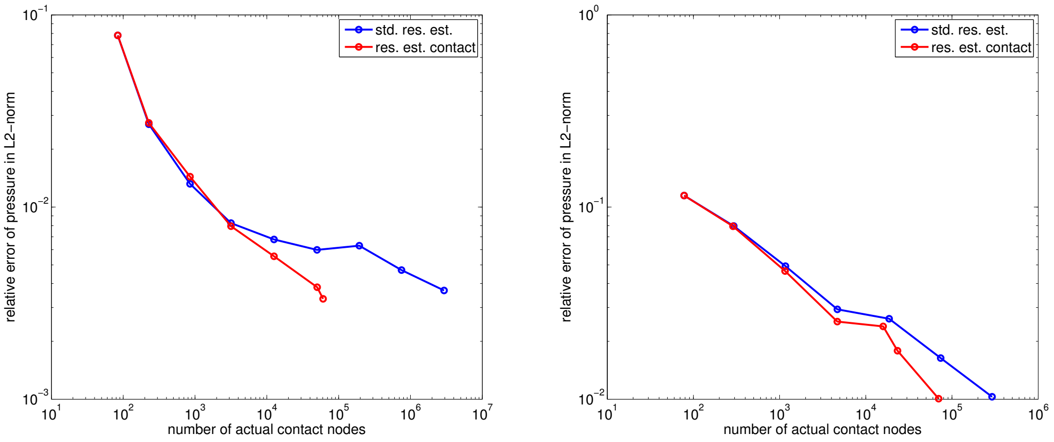
In Figure 4.19 we show the relative error between s_p and the exact contact stress $\hat{\sigma}_\nu(\boldsymbol{p})$, compare formula (1.32), at the node $\boldsymbol{p} = (0.5, 0.5, 0.0)$ where the north pole of the rigid ball indents the elastic cube. The quantity is plotted against the number of contact nodes for both methods of adaptive refinement by means of our new residual-type error estimator (red line) and the standard residual estimator (blue line). In Example 3A the decay of the error is very similar for both methods. However, in order to obtain the same accuracy the standard error estimator requires more actual contact nodes. In Example 3B the relative error does not decay continuously.

A similar behavior that the relative error increases in between can be observed for the relative error between b and the numerically computed radius of the actual contact zone, see Figure 4.20. For both methods of adaptive refinement by means of our residual-type estimator for contact and the standard residual estimator the relative error is the same if the starting grid consists of hexahedra. In contrast, the relative error decays more rapidly in Example 3B when using our residual-type estimator for contact. Just as in Figure 4.18 we have chosen logarithmic scales in Figure 4.19 and Figure 4.20 for the illustration of the decay of the relative errors as a function of the number of actual contact nodes.



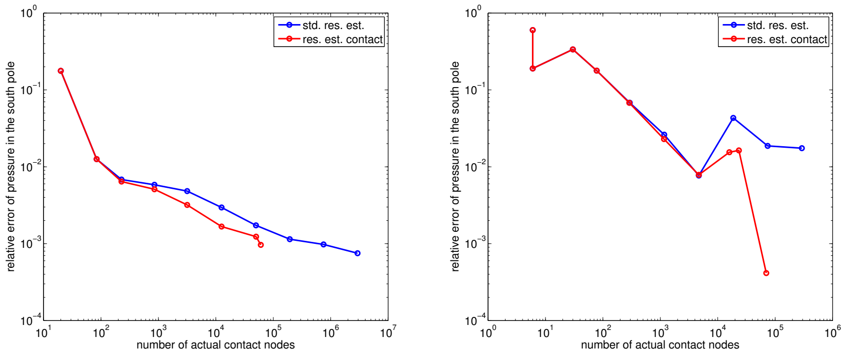
(a) Example 3A (b) Example 3B

Figure 4.17: Contact stresses



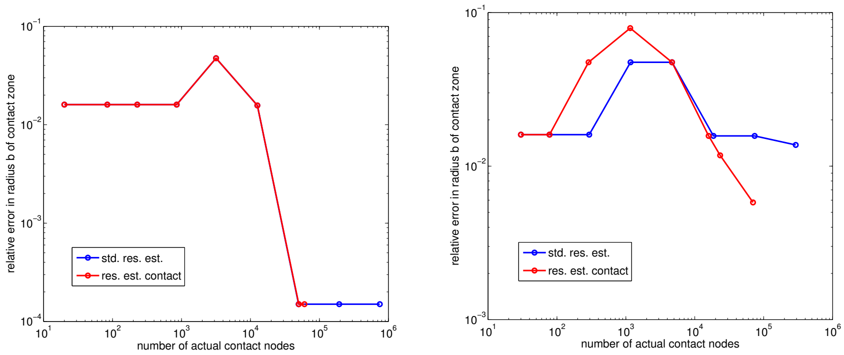
(a) Example 3A (b) Example 3B

Figure 4.18: Relative error of pressure for $s_{\mathfrak{m}}$ computed on grids which are adaptively refined by means of our residual estimator for contact (res. est. contact) and the standard residual estimator (std. res. est.) plotted against $\#N^{\mathrm{aC}}_{\mathfrak{m}_k}$ (logarithmic scales)



(a) Example 3A (b) Example 3B

Figure 4.19: Relative error of pressure at the node $\boldsymbol{p} = (0.5, 0.5, 1.0)$ computed on grids which are adaptively refined by means of our residual estimator for contact (res. est. contact) and the standard residual estimator (std. res. est.) plotted against $\#N^{\mathrm{aC}}_{\mathfrak{m}_k}$ (logarithmic scales)



(a) Example 3A (b) Example 3B

Figure 4.20: Relative error of the radius of the contact zone computed on grids which are adaptively refined by means of our residual estimator for contact (res. est. contact) and the standard residual estimator (std. res. est.) plotted against $\#N^{\mathrm{aC}}_{\mathfrak{m}_k}$ (logarithmic scales)

4.2.3 Relevance of the different error estimator contributions

In this section we analyze our residual-type estimator for contact in more detail. Therefore, we generate sequences of adaptively refined grids by means of our residual-type error estimator for different contact problems in $3D$. We choose different kinds of starting grids consisting of tetrahedra, hexahedra, or primarily of prisms or pyramids. We analyze the relevance of the different error estimator contributions η_i and the distribution of their local contributions $\eta_{i,p}$. We examine the structure of the refined grids and we have a look at the development of the number of actual, semi- and full-contact nodes in the steps of adaptive refinement. Further, we compare the rate of convergence of the estimator on adaptively and uniformly refined grids.

We consider five different experiments. In each of them the domain of computation is a unit cube of linear elastic material with the Young's modulus $E = 5 \cdot 10^2$ and the Poisson ratio $\nu = 0.3$. The given force density $\boldsymbol{f}$ is set to zero and the Neumann values are also zero. One side of the cube is the Dirichlet boundary and the opposite side is the potential contact boundary so that the body can be pressed against the obstacle by means of non-zero Dirichlet values in direction of the obstacle. The remaining sides are Neumann boundary sides. The direction of constraints $\boldsymbol{\nu}$ is perpendicular to the potential contact boundary. The Dirichlet values in direction $\boldsymbol{\nu}$ are set to 0.05 or 0.08 depending on the obstacle. For the other directions we choose zero Dirichlet values.

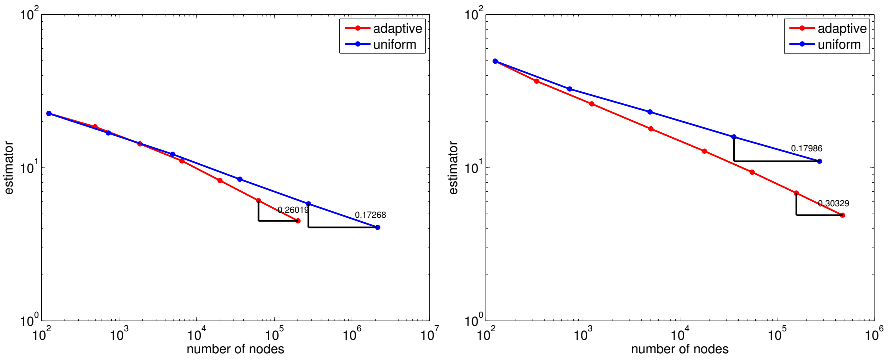
Example 4: $g = 0$ in a strip

The first two examples are $3D$ counterparts of the $2D$ examples in Section 4.2.1. In the first example the elastic body comes into contact with the obstacle in a strip of the width 0.6. Therefore, the gap function is $g = 0$ within this strip and set to $g = 0.5$ on the remaining potential contact boundary. We refer to this experiment as Example 4.

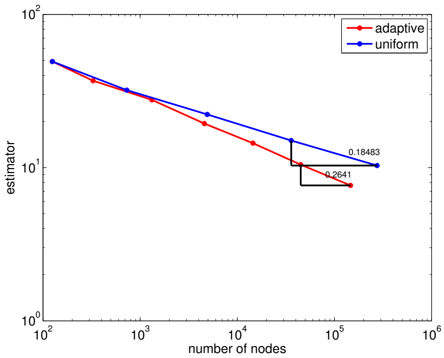
The problem is solved on three different kinds of starting grids, consisting of either tetrahedra, hexahedra or primarily of prisms. For this example Figure 4.21 shows the experimental order of convergence for the adaptive strategy and for the uniform refinement depending on the kind of starting grid. The best rate of convergence is obtained for a starting grid of hexahedra.

The magnitude of the different error estimator contributions varies strongly. While in the first step of refinement the error estimator contributions $\eta_1, \eta_2, \eta_4, \eta_5$ are larger than 10^2, the error estimator contributions η_6, η_7, η_8 are at most in a range of 10^{-2}. We note that the edges of the strip where $g = 0$ do not coincide with edges on the starting grid such that both cases $g_{\mathfrak{m}} > g$ and $g > g_{\mathfrak{m}}$ occur.

In Figure 4.22 the decay of the square of the dominant error estimator contributions η_1^2 (yellow line, referred to as "element residual"), η_2^2 (red line, referred to as "inner jump"), $\eta_4^2 + \eta_5^2$ (blue line, referred to as "contact stress") and the sum $\sum_{k=1}^{8} \eta_k^2$ (black line) is visualized. In the case the grid consists of tetrahedra, the error estimator contribution η_1 vanishes and η_2 mainly determines the adaptive refinement process. If the starting grid consists primarily of prisms, η_1 is dominant. Compared to η_1 and η_2, the contribution η_5 is small. Its influence is restricted to the free boundary zone as can be seen in Figure



(a) Starting grid of tetrahedra (b) Starting grid of hexahedra



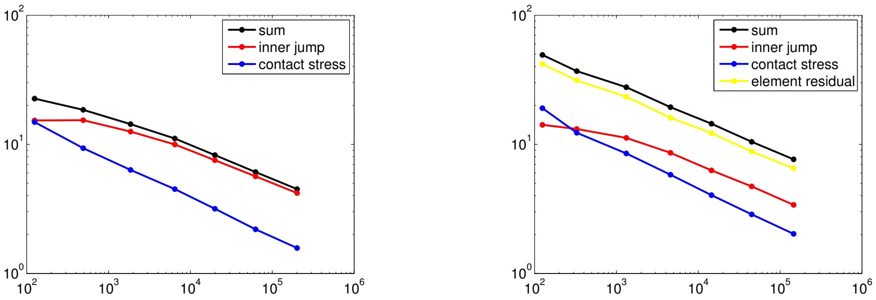
(c) Starting grid of prisms

Figure 4.21: Example 4: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)

4.23 in the steps $k = 2, 4, 6$ of adaptive refinement. However, in this example, η_5 is not the only responsible for the refinement of the elements in this region. Near the free boundary zone, where the edges of the punch indent the elastic body, the local error estimator contributions $\eta^2_{1,p}$ and $\eta^2_{2,p}$ are of the same magnitude as $\eta^2_{5,p}$ in the fourth refinement step which can be seen in Figure 4.24 in a cross-section through the body.
If the starting grid consists primarily of prisms the development of the number of actual contact nodes, denoted by $N^{\mathrm{aC}}_{\mathfrak{m}_k}$ and semi- and full-contact nodes is visualized in Figure 4.25. On the final grid the number of actual contact nodes is about half of the number of the potential contact nodes although the area of actual contact is more than fifty percent of Γ_C. This confirms once again that the actual contact boundary is not overrefined.

Example 5: a pyramid as obstacle

In the second example of this section, called Example 5, the obstacle is a pyramid so that the tip of the pyramid indents the elastic cube, see Figure 4.27(a). The angle between a triangular side and the base of the pyramid is 45°.
We choose starting grids made of hexahedra or primarily of pyramids. The experimental



(a) Starting grid of tetrahedra (b) Starting grid of prisms

Figure 4.22: Example 4: Dominant estimator contributions plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)

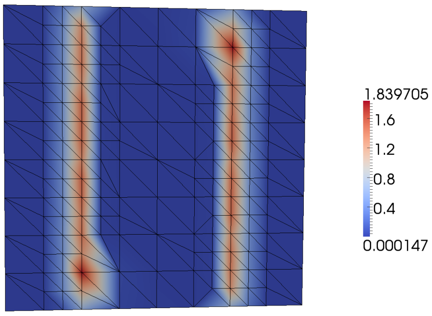
order of convergence for the adaptive strategy in contrast to the uniform refinement, see Figure 4.26, demonstrates clearly the merit of the error estimator. For both variants of starting grids Figures 4.27(c) and 4.28(a) show the solution $-u_\nu$ on the adaptively refined grids around the actual contact boundary. In Figure 4.28(b) we see the solution $-\boldsymbol{u} \cdot \boldsymbol{\nu}$ on the adaptively refined grid in a diagonal cut through the contact area.

The elements have been refined most frequently along the cross where the edges of the pyramid indent the elastic body and around the free boundary zone, which is given by the curved lines surrounding the cross. In Figure 4.27(b) the status of the contact nodes, full-contact nodes (red), semi-contact nodes (white) and the remaining potential contact nodes (blue) on the final grid after 9 steps of adaptive refinement, is visualized.

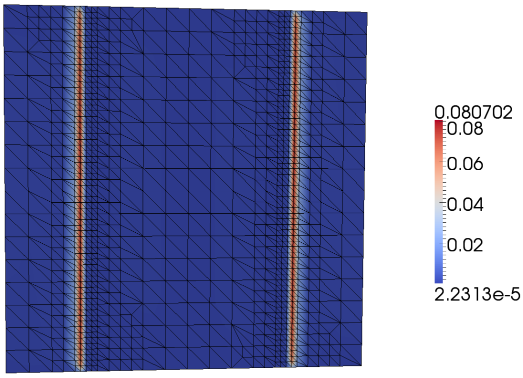
Just as in the foregoing example the largest error estimator contributions for the adaptive refinement are η_1, η_2 and η_5. In the first refinement step the value of η_1 is in the range of 10^3 and goes down to 10^{-1} in the last refinement step. In step 6 the local contributions $\eta_{1,p}$ vary between 0.07 and 0.0. The distribution of the local contributions $\eta_{1,p}^2$ in refinement step 6 can be seen in Figure 4.29(a). The value of η_2 is in the range of 10^2 in the first refinement step and about 10^{-1} in the last refinement step. The local contributions $\eta_{2,p}^2$ in refinement step 6 differ between 0.02 and 10^{-9}, see 4.29(b). From Figure 4.29 we can deduce that the refinement of the cross where the edges of the pyramid indent the elastic body is due to η_1 and η_2. The contribution η_5 is mainly responsible for the adaptive refinement around the free boundary zone. The distribution of the local contributions $\eta_{5,p}^2$ in refinement step 6 is given in Figure 4.29(c). The local contributions have a maximal value of $7 \cdot 10^{-3}$. We note that in this refinement step η_6 is already in a very small range of 10^{-11}.

Example 6: a ball as obstacle

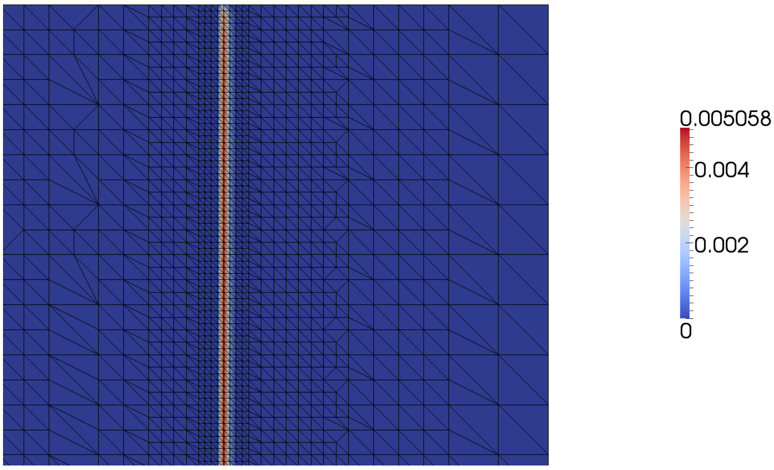
In the next experiment, called Example 6, the obstacle is a ball like in Section 4.2.2. This time we choose a grid of hexahedra. Figure 4.30(a) and (b) shows the final grid of the contact boundary as well as projected on the obstacle and Figure 4.30(c) shows a



(a) Adaptive refinement step $k = 2$

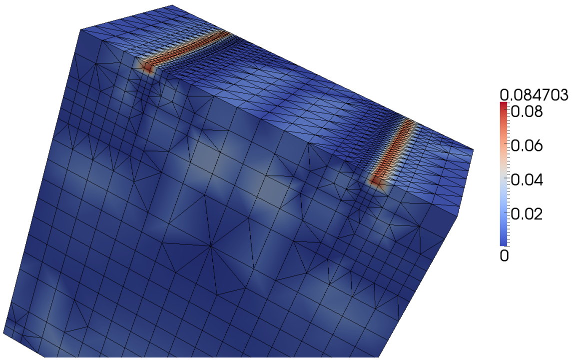


(b) Adaptive refinement step $k = 4$

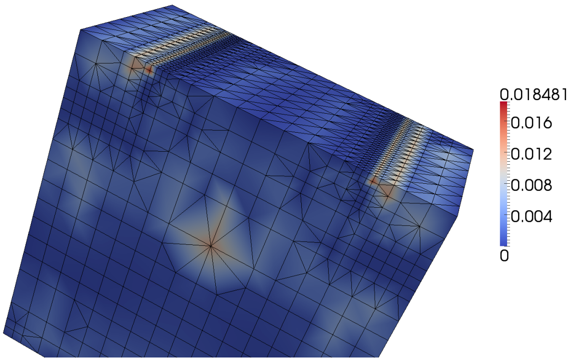


(c) Adaptive refinement step $k = 6$, zoom in one edge

Figure 4.23: Distribution of the local contributions $\eta_{5,p}^2$ in Example 4 with a starting grid of prisms



(a) $\eta_{1,p}^2$



(b) $\eta_{2,p}^2$

Figure 4.24: Refinement step $k = 4$ in Example 4 with a starting grid of prisms

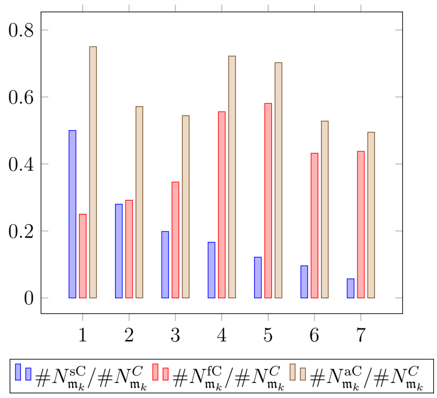
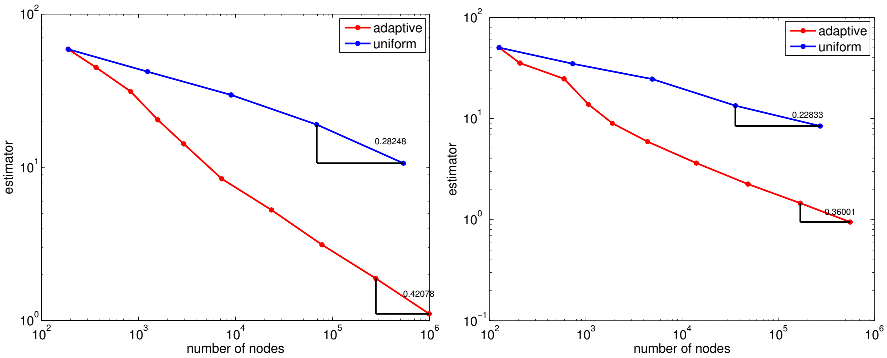


Figure 4.25: Ratio of the number of actual, semi- and full-contact nodes to the number of potential contact nodes in the adaptive refinement steps $k = 1, ..., 7$ in Example 4 with a starting grid of prisms



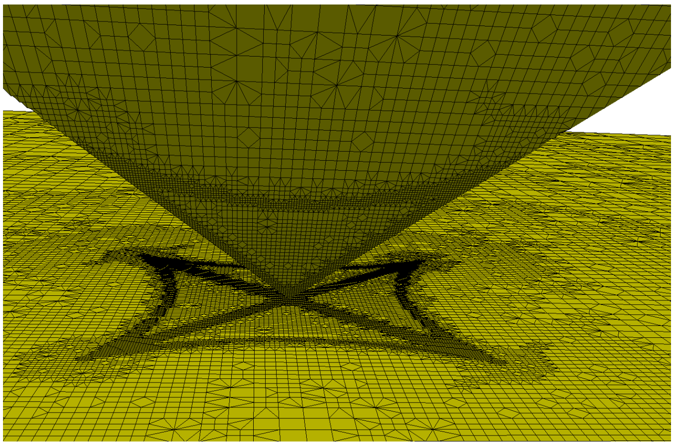
(a) Starting grid of pyramids (b) Starting grid of hexahedra

Figure 4.26: Example 5: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)

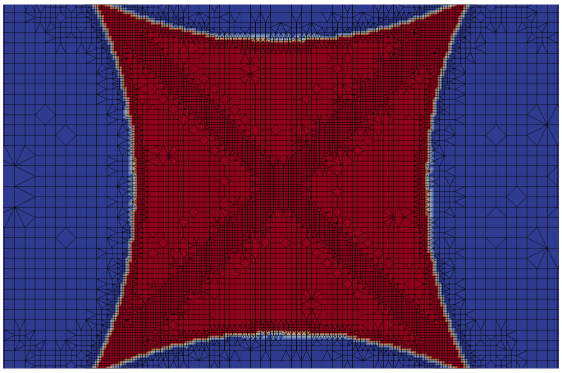
diagonal cut through the cube after seven steps of adaptive refinement. In Figure 4.31 we see the rate of convergence of the error estimator for adaptive and uniform refinement. On the basis of this example we have a closer look at the error estimator contributions η_5, η_6, η_8 which are related to the contact constraints. Exemplarily, we pick refinement step 4, see Figure 4.32(a) for the grid around the actual contact boundary. The largest of these error estimator contributions is η_5 with local values of $\eta_{5,p}^2$ up to 0.02 in the free boundary zone, see Figure 4.32(b). From the definition of η_6 we know that the local contributions are restricted to the free boundary zone which can be seen in Figure 4.32(c). As the ball which is the obstacle is convex, the finite element approximation $g_{\mathfrak{m}}$ is greater than or equal to g so that $\eta_{8,p}$ is non-zero in the whole actual contact zone, see Figure 4.32(d). In this refinement step the values of $\eta_{6,p}$ and $\eta_{8,p}$ are already very small.

Example 7: a hat as obstacle

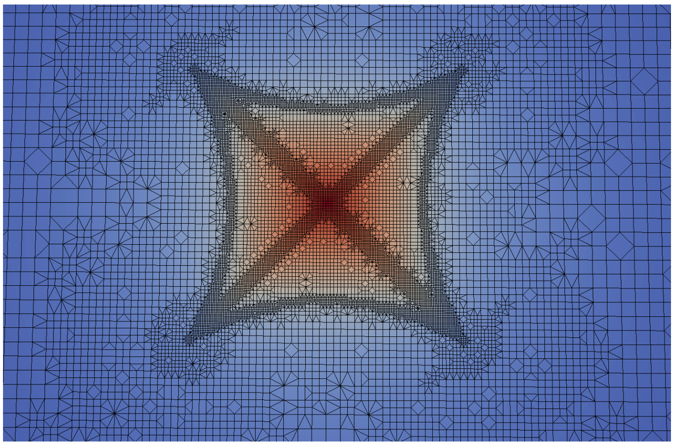
Next, we consider an example where both additional error estimator contributions η_7 and η_8 are non-zero. The obstacle looks like a hat; see, e.g., Figure 4.33(b). We call this experiment Example 7 and choose starting grids of pyramids or tetrahedra. For both kinds of starting grids the rate of convergence of the estimator is shown in Figure 4.34. We consider the choice of a starting grid consisting primarily of pyramids in more detail. The final grid and the solution can be seen in Figure 4.33. The distribution of the local error estimator contributions $\eta_{5,p}^2$, $\eta_{6,p}^2$, $\eta_{7,p}^2$ and $\eta_{8,p}^2$ which are related to the contact constraints is visualized in Figure 4.35 in the fourth adaptive refinement step. As in all foregoing examples the most important of these error estimator contributions is η_5 with a maximal local value of 0.03 for $\eta_{5,p}^2$ in refinement step 4. From the definition of η_6 it is clear that the local contributions are restricted to the free boundary zone, see Figure 4.35(c). Due to the shape of the hat and the structure of the grid the error estimator contributions $\eta_{7,p}$ and $\eta_{8,p}$ are also located near the free boundary zone, compare Figure 4.35(d) and (e).



(a) Final grid of contact boundary projected on obstacle, zoomed in actual contact area

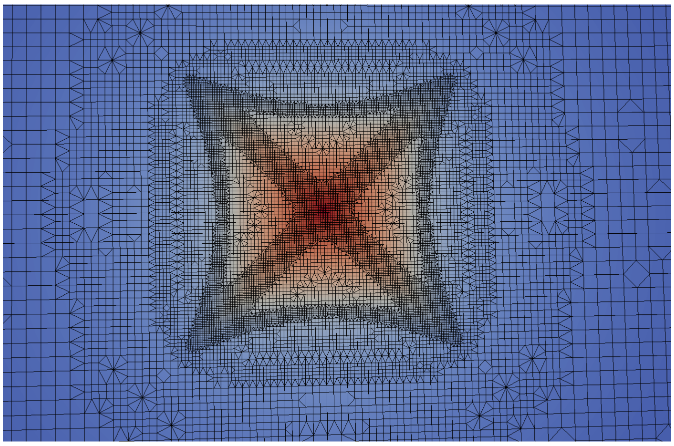


(b) Contact status of nodes on final grid: full-contact (red), semi-contact (white)

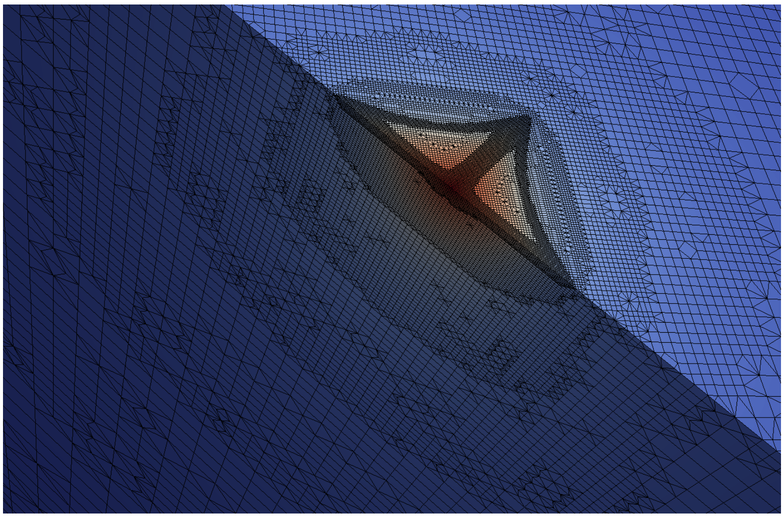


(c) Solution $(-u_\nu)$ on final grid around the actual contact boundary; ranges between -0.05 (blue) and 0.0 (red)

Figure 4.27: Example 5 with starting grid of pyramids

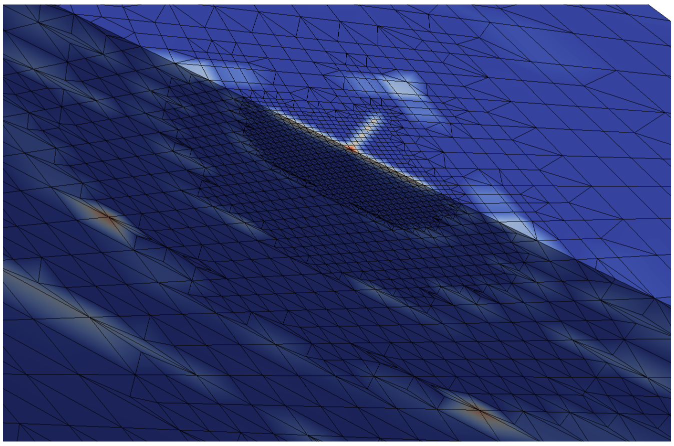


(a) Solution $(-u_\nu)$ on final grid around the actual contact boundary; ranges between -0.05 (blue) and 0.0 (red)

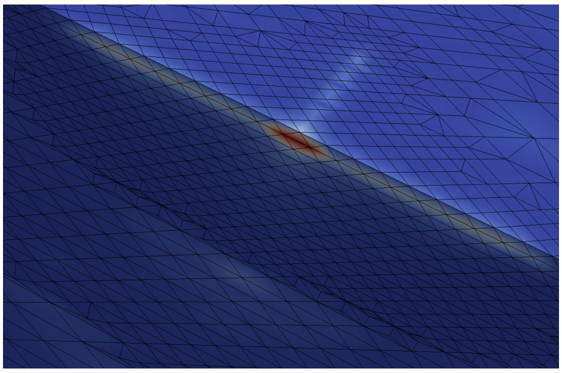


(b) Solution $(-\boldsymbol{u} \cdot \boldsymbol{\nu})$ on final grid in diagonal cut through contact area; ranges between -0.05 (blue) and 0.0 (red)

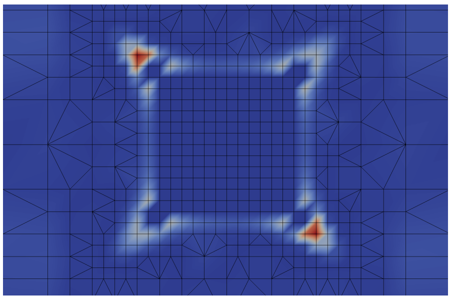
Figure 4.28: Example 5 with starting grid of hexahedra



(a) $\eta_{1,p}^2$ in refinement step 6; values of $\eta_{1,p}^2$ range between 0.0 (blue) and 0.07 (red)

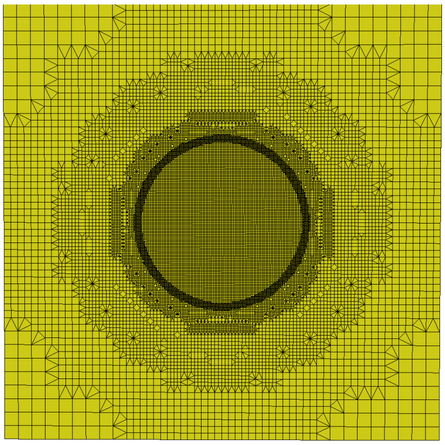


(b) $\eta_{2,p}^2$ in refinement step 6; values of $\eta_{2,p}^2$ range between 10^{-9} (blue) and 0.02 (red)

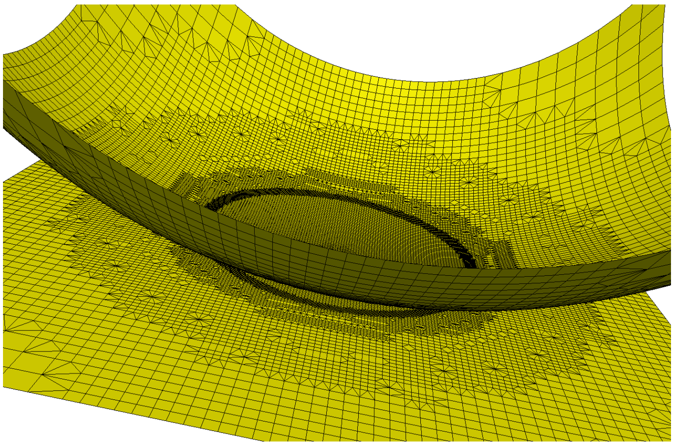


(c) $\eta_{5,p}^2$ on the contact boundary in refinement step 6; values of $\eta_{5,p}^2$ range between 10^{-9} (blue) and 0.007 (red)

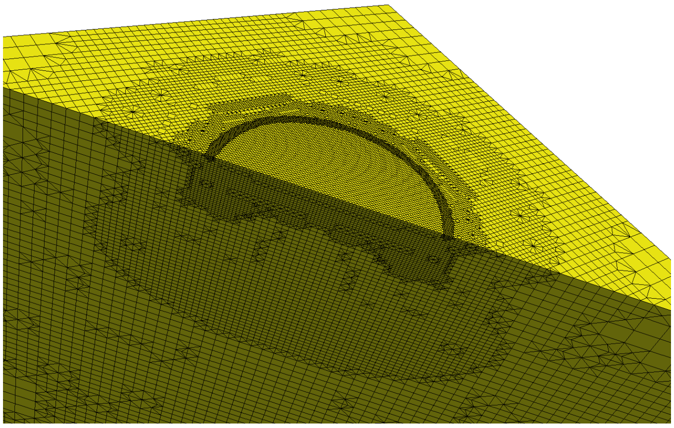
Figure 4.29: Example 5 with starting grid of pyramids



(a) Final grid of contact boundary



(b) Final grid of contact boundary projected on obstacle



(c) Final grid in a diagonal cut

Figure 4.30: Example 6 with starting grid of hexahedra

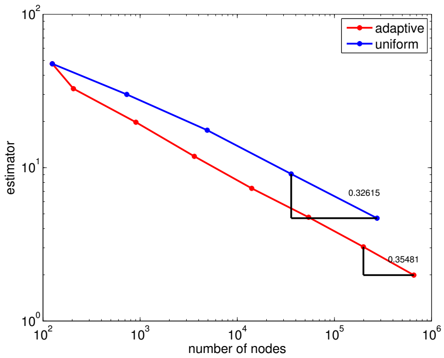
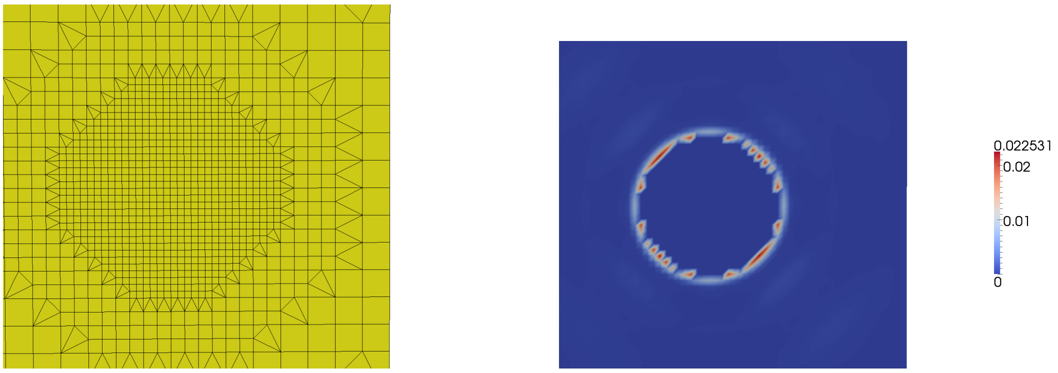
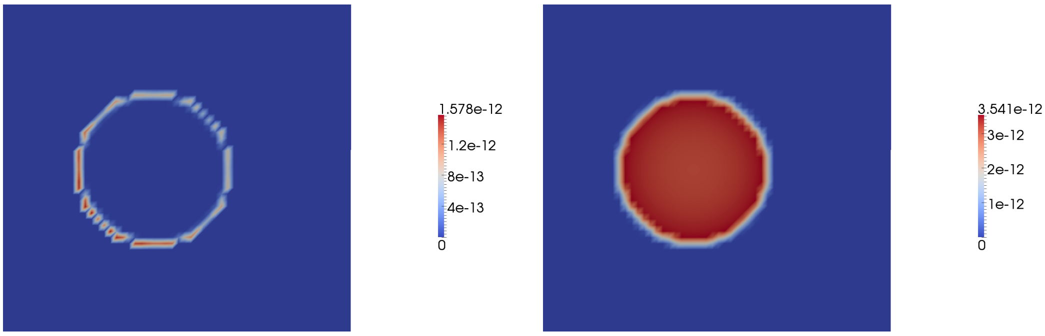


Figure 4.31: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales) in Example 6 with starting grid of hexahedra



(a) Grid in refinement step 4

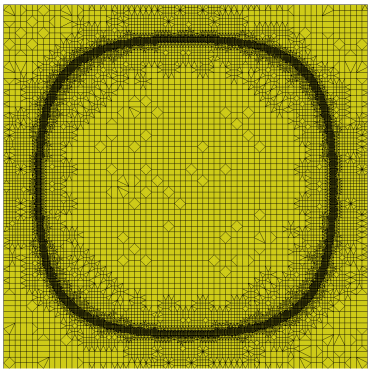
(b) $\eta^2_{5,p}$ in refinement step 4



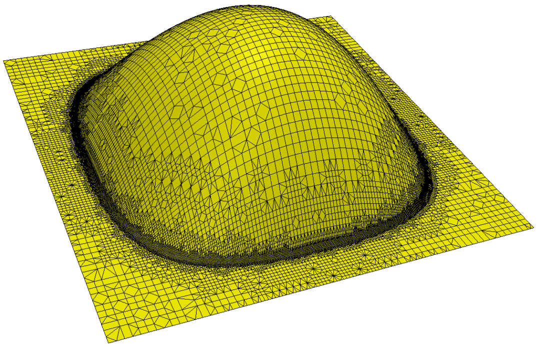
(c) $\eta^2_{6,p}$ in refinement step 4

(d) $\eta^2_{8,p}$ in refinement step 4

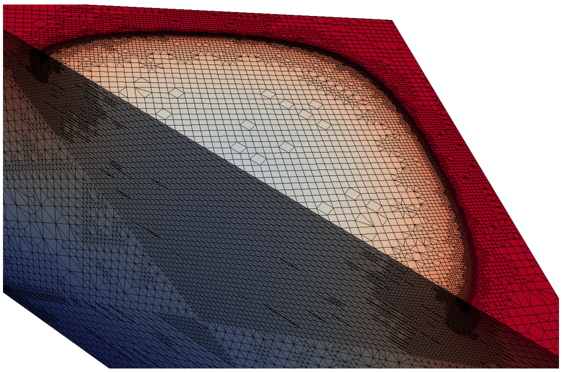
Figure 4.32: Example 6 with starting grid of hexahedra; view on the contact boundary



(a) Final grid of contact boundary

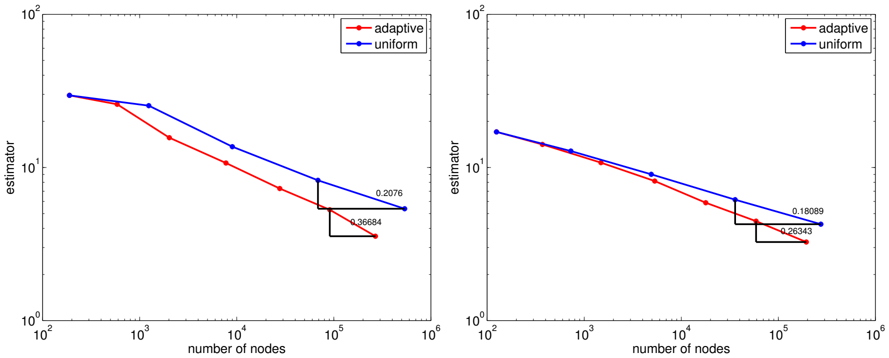


(b) Final grid of contact boundary projected on obstacle



(c) Solution $(-\boldsymbol{u} \cdot \boldsymbol{\nu})$

Figure 4.33: Example 7 with starting grid of pyramids



(a) Starting grid of pyramids (b) Starting grid of tetrahedra

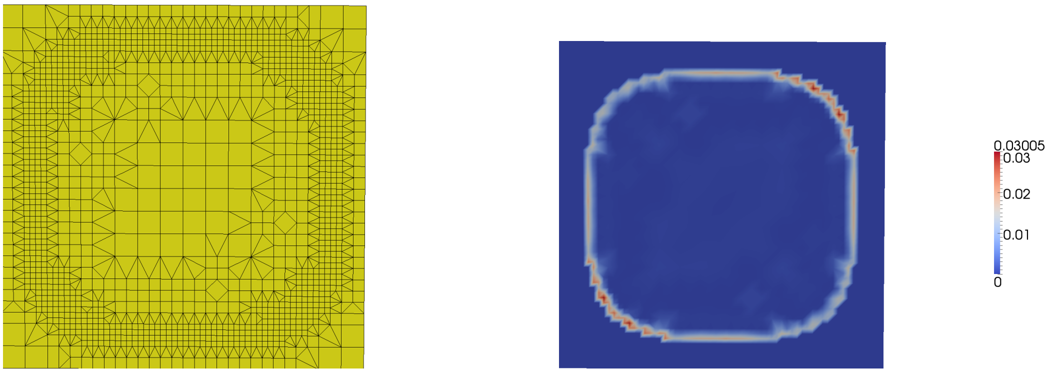
Figure 4.34: Example 7: Estimator plotted against $\#N_{\mathfrak{m}_k}$ (logarithmic scales)

Example 8: a wave as obstacle

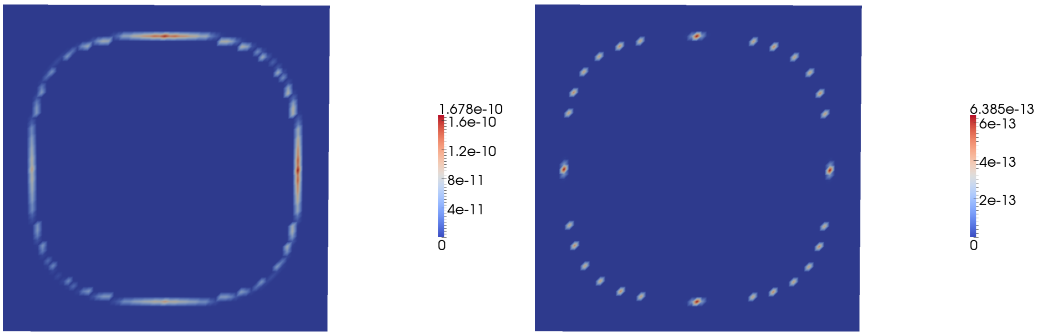
If we choose a sinus wave as obstacle (Example 8) the importance of the additional error estimator contribution η_8 gets evident. The starting grid consists of hexahedra. The grid projected from the contact boundary on the obstacle is shown for the starting grid and after three refinement steps in Figure 4.36. Due to the length of the sinus waves and due to the meshsize of the hexahedra on the starting grid the grid around three of the four contact regions is well resolved but not of the fourth one on the left side. There the local error estimator contributions $\eta^2_{8,p}$ are non-zero, see Figure 4.36(c), and induce further refinement and hence the detection of contact.

Example 9: "smiley"

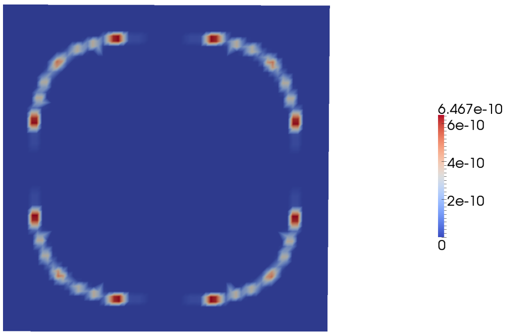
In our final example (Example 9) we choose an obstacle such that the refined mesh represents a smiley on the contact surface. Depending on the starting grids the smileys look different. Figure 4.37 shows the adaptively refined grids of the contact boundary after two, four and six steps of refinement for starting grids consisting primarily of pyramids or hexahedra.



(a) Grid in refinement step 4 (b) $\eta^2_{5,p}$ in refinement step 4

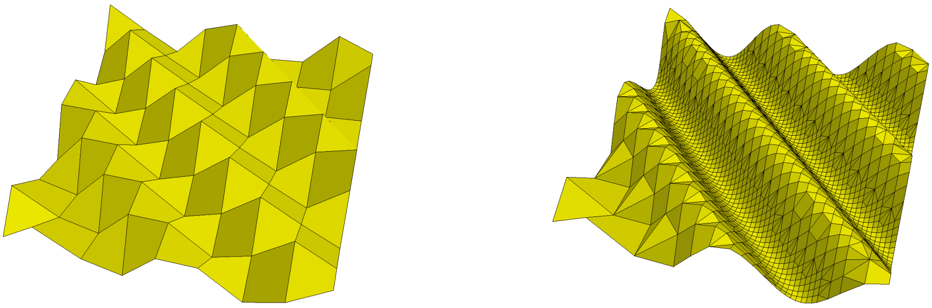


(c) $\eta^2_{6,p}$ in refinement step 4 (d) $\eta^2_{7,p}$ in refinement step 4



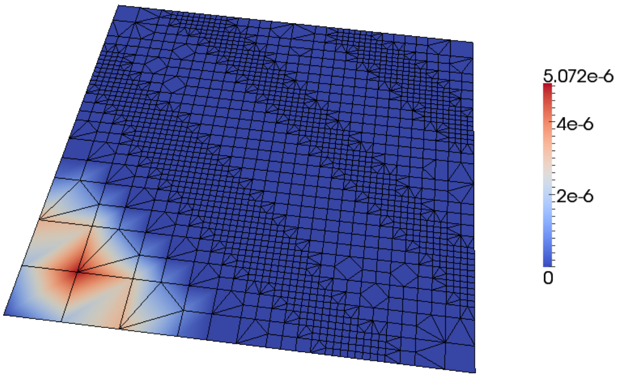
(e) $\eta^2_{8,p}$ in refinement step 4

Figure 4.35: Example 7 with starting grid of pyramids; view on the contact boundary



(a) Starting grid projected on the obstacle

(b) Adaptively refined grid projected on the obstacle in refinement step 3



(c) $\eta^2_{8,p}$ on the contact boundary in refinement step 3

Figure 4.36: Example 8 with starting grid of hexahedra

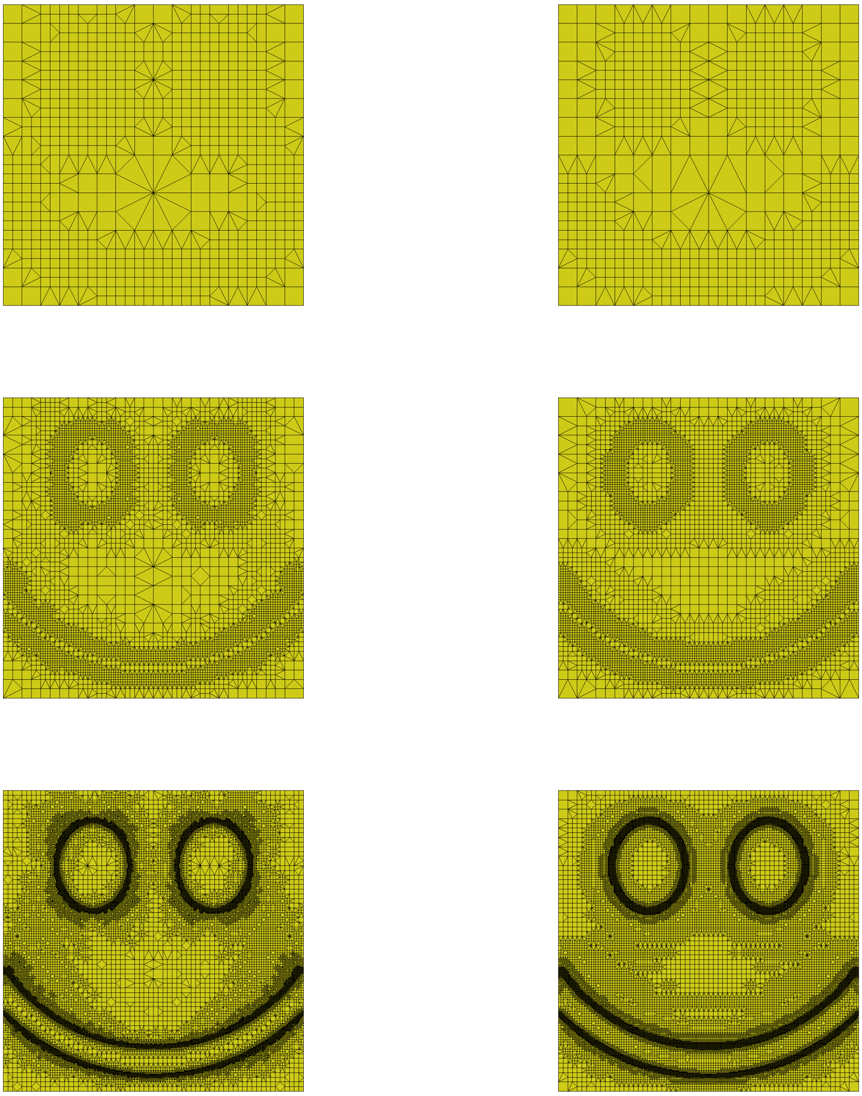


Figure 4.37: Grid of the contact boundary in Example 9 in refinement steps $2, 4, 6$; left: starting grid of pyramids; right: starting grid of hexahedra

5 Discretization in time of dynamic contact problems

The inherently non-smooth character of dynamic contact problems in linear elasticity [EJK05] turns the construction of suitable time integration schemes into a highly challenging task, as classical time discretization schemes cannot be applied in a straightforward way [AB08, Mor99, HTS+76, TP93, LC97, KROM99, LL02, DKE08, KLR08, HHW08, KW09a, DEP11]. In fact, many aspects which are all related to the non-smoothness originating from the contact interface have to be considered when constructing a time integration scheme. Besides solution related quantities as displacements, boundary stresses and velocities, which are obviously influenced by the non-penetration constraint, structural properties as energy and momentum conservation, symplecticity, and the hidden constraint, the persistency condition may be relevant. It is especially the interaction of all these aforementioned quantities, properties and constraints which complicates the construction of an appropriate integration scheme.

During the last decades numerous different time discretization schemes for elastodynamic contact problems have been developed; see, e.g., [HTS+76, TP93, LC97, KROM99, LL02, DKE08, KLR08, HHW08] each of them showing different advantages and disadvantages with respect to the above introduced criteria. Especially in the earlier works classical time-discretization schemes are improved by adding corrections to the velocities, accelerations and contact stresses, cf. [HTS+76, TP93, LL02] whereas recently the interplay of space and time discretization with respect to the energy conservation and the stability of the contact stresses has been considered. In [LC97] the focus lies on the energy conservation which is achieved by enforcing a discrete persistency condition instead of the non-penetration condition. A reason for the instability occurring in the contact stresses can be found in the double role of the contact nodes as degrees of freedom for the boundary and the interior. Therefore, modifications to the discretization in space are applied in the algorithms presented in [KLR08, HHW08, DEP11] whereas in [DKE08] a predictor step removes the artificial influence of the discretization in space to the contact stresses. Most of the above mentioned time discretization schemes are modifications of the classical Newmark scheme [New59]. Unfortunately, in the case of contact constraints, the classical Newmark scheme leads to instabilities at the contact boundary and energy blow-ups. In Section 5.2 we give a deeper insight into the causes of these instabilities and we present selected algorithms [KROM99, DKE08, LC97, KLR08, HHW08]. Therefore, we follow the lines of our article [KW09a]. We note that recently in [DEP11] a further comparison of some time integration schemes for contact problems has been published. In Section 5.3, we present our new space-time connecting discretization scheme which has been published in [KW11].

We note that another important approach for time-discretization schemes in mechanics are variational integrators [HLW06,KMOW00] which focuses on the symplectic structure. A recent approach, applied to contact problems, can be found in [GSO10].

5.1 Dynamic contact problems in linear elasticity

In this section, we give the strong and the weak formulations of the dynamic contact problem in linear elasticity. An important aspect in the time discretization of dynamic contact problems is the conservation of the physical properties in the discrete system. Therefore, we cannot solely consider the idealized one body contact problem as in the foregoing chapters. We discuss the conservation properties and comment on existence and uniqueness results.

5.1.1 Strong and weak formulation of dynamic contact problems

Dynamic contact problems describe the time-dependent contact, e.g., a collision, between two solid bodies. In the case of two-body contact problems both bodies Ω^S and Ω^M are deformable and the sought displacement field is $\boldsymbol{u} = (\boldsymbol{u}^S, \boldsymbol{u}^M)$. The displacements depend on the spatial variable $\boldsymbol{x} \in \Omega$ and on the temporal variable $t \in [0,T]$ where $[0,T]$ is the observed time interval. The velocity is denoted by $\dot{\boldsymbol{u}}$ and the acceleration is given by $\ddot{\boldsymbol{u}}$. The linearized non-penetration condition for two-body contact problems is given by $[u_\nu] \leq g$ on Γ_C^S, see Section 1.2.1. From Newton's axiom of action and reaction follows the balance of forces $\hat{\boldsymbol{\sigma}}(\boldsymbol{u}^M \circ \boldsymbol{\Phi}) = -\hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S)$ on Γ_C^S where $\boldsymbol{\Phi}$ is the bijective mapping between the potential contact boundaries, see Section 1.2.1. Hence, we get the general formulation of the dynamic frictionless two-body contact problem in linear elasticity

Problem 5.1.1. *Strong formulation of the dynamic two-body contact problem*

$$\begin{aligned}
\rho\ddot{\boldsymbol{u}} - \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}) &= \boldsymbol{f} && \text{in } \Omega \times [0,T] && (5.1)\\
\hat{\boldsymbol{\sigma}}(\boldsymbol{u}) &= \boldsymbol{\pi} && \text{on } \Gamma_N \times [0,T] && (5.2)\\
\boldsymbol{u} &= \boldsymbol{u}_D && \text{on } \Gamma_D \times [0,T] && (5.3)\\
\hat{\boldsymbol{\sigma}}(\boldsymbol{u}^M \circ \boldsymbol{\Phi}) &= -\hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S) && \text{on } \Gamma_C^S \times [0,T] && (5.4)\\
[u_\nu] &\leq g && \text{on } \Gamma_C^S \times [0,T] && (5.5)\\
\hat{\sigma}_\nu(\boldsymbol{u}) &\leq 0 && \text{on } \Gamma_C^S \times [0,T] && (5.6)\\
([u_\nu] - g) \cdot \hat{\sigma}_\nu(\boldsymbol{u}) &= 0 && \text{on } \Gamma_C^S \times [0,T] && (5.7)\\
\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}) &= \boldsymbol{0} && \text{on } \Gamma_C^S && (5.8)\\
\boldsymbol{u}(\boldsymbol{x},0) &= \boldsymbol{u}_0(\boldsymbol{x}) && \text{in } \Omega && (5.9)\\
\dot{\boldsymbol{u}}(\boldsymbol{x},0) &= \dot{\boldsymbol{u}}_0(\boldsymbol{x}) && \text{in } \Omega && (5.10)
\end{aligned}$$

where ρ is the density. We focus on frictionless contact. However, we comment shortly on the extensibility of the presented time discretization schemes to frictional contact later on. In the case of dynamic two-body contact problems the Coulomb friction law (1.19) changes to

$$\begin{aligned} &[\dot{\boldsymbol{u}}_T] = 0 \Rightarrow |\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u})| \leq \mathcal{F}|\hat{\sigma}_\nu(\boldsymbol{u})| \quad \text{on } \Gamma_C^S \times [0,T] \\ &[\dot{\boldsymbol{u}}_T] \neq 0 \Rightarrow \hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}) = -\mathcal{F}|\hat{\sigma}_\nu(\boldsymbol{u})| \frac{[\dot{\boldsymbol{u}}_T]}{|[\dot{\boldsymbol{u}}_T]|} \quad \text{on } \Gamma_C^S \times [0,T] \end{aligned} \tag{5.11}$$

where $[\dot{\boldsymbol{u}}_T] := \dot{\boldsymbol{u}}_T^S - \dot{\boldsymbol{u}}_T^M \circ \boldsymbol{\Phi}$ is the jump of the velocities in tangential direction.
A collision of the solid bodies can be caused by the application of a force density $\boldsymbol{f}$, Dirichlet values $\boldsymbol{u}_D$ or initial values for displacements $\boldsymbol{u}_0(\boldsymbol{x})$ and velocities $\dot{\boldsymbol{u}}_0(\boldsymbol{x})$. For the ease of presentation, we restrict our considerations to the case of prescribed initial velocities in normal direction and set $\boldsymbol{f} = \boldsymbol{0}$, $\Gamma_D = \emptyset$ and $\boldsymbol{u}_0(\boldsymbol{x}) = \boldsymbol{0}$. Therefore, the whole boundary is divided into the potential contact boundary where the non-penetration condition is enforced and the Neumann boundary where we assume $\boldsymbol{\pi} = \boldsymbol{0}$.
The weak formulation of the dynamic contact problem is a hyperbolic variational inequality which is derived from the strong formulation via integrating by parts

$$\langle \rho \ddot{\boldsymbol{u}}, \boldsymbol{\varphi} \rangle + a(\boldsymbol{u}, \boldsymbol{\varphi}) = \langle \hat{\boldsymbol{\sigma}}(\boldsymbol{u}), \boldsymbol{\varphi} \rangle_{\Gamma_C} \qquad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0 \tag{5.12}$$

taking $\boldsymbol{\varphi} = \boldsymbol{v} - \boldsymbol{u}$ as test function, exploiting the non-penetration condition, the sign of $\hat{\sigma}_\nu(\boldsymbol{u})$ and the fact that the tangential stresses are assumed to be zero

Problem 5.1.2. *Weak formulation*
For every $t \in (0,T]$ find $\boldsymbol{u}(\cdot,t) \in \boldsymbol{\mathcal{H}}$ with $\ddot{\boldsymbol{u}}(\cdot,t) \in \boldsymbol{L}^2(\Omega)$, so that

$$\langle \rho \ddot{\boldsymbol{u}}, \boldsymbol{v} - \boldsymbol{u} \rangle + a(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) \geq 0 \qquad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{K}} \tag{5.13}$$

with $\boldsymbol{\mathcal{K}} := \left\{ \boldsymbol{v} \in \boldsymbol{\mathcal{H}} \mid [\boldsymbol{v}] \cdot \boldsymbol{\nu} \leq g \text{ on } \Gamma_C^S \right\}$. In contrast to the static case, for the dynamic contact problems in linear elasticity many questions regarding the solution like existence, uniqueness and regularity are still open, cf. [EJK05]. Only for the special case that the underlying equation is the wave equation an existence and uniqueness result is given in [LS84] and a result on the weak solvability can be found in [Kim89]. The mathematical difficulty results mainly from the non-smoothness in time of the velocities and displacements at the contact boundary. However, if the material is assumed to be linear viscoelastic existence results can be derived, see [EJK05] and the references cited therein. For dynamic viscoelastic contact problems satisfying the Kelvin-Voigt constitutive law the stability under perturbation of the initial data has been studied in [KDS09].
We assume that a solution $\boldsymbol{u}$ of Problem 5.1.2 exists. For a clear presentation we define $\langle \boldsymbol{F}(\boldsymbol{u}), \cdot \rangle := a(\boldsymbol{u}, \cdot)$ and we set $\rho = 1$. By means of the subdifferential, see Definition 1.3.1, of the indicator functional $\mathcal{I}_{\mathcal{K}}(\cdot)$ defined in (1.40) we can reformulate (5.13) as variational inclusion

$$0 \in \langle \ddot{\boldsymbol{u}}, \boldsymbol{\varphi} \rangle + \langle \boldsymbol{F}(\boldsymbol{u}), \boldsymbol{\varphi} \rangle + \partial^{\text{sub}} \mathcal{I}_{\mathcal{K}}(\boldsymbol{u})(\boldsymbol{\varphi}) \qquad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0, \tag{5.14}$$

like in (1.43) for the case of static contact problems. In Section 1.3 we already noted that the element of the subdifferential $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{H}}^*$ fulfilling the variational inclusion, see

(1.44), has the meaning of a constraining force density, called contact force density. This interpretation is of crucial importance for understanding the difficulties connected to the discretization in time and space. Here and in the following, we use a different notation. We define the contact force density corresponding to Problem 5.1.2 as follows

$$\langle \boldsymbol{F}_{\mathrm{con}}(\boldsymbol{u}), \boldsymbol{\varphi} \rangle := \langle \ddot{\boldsymbol{u}}, \boldsymbol{\varphi} \rangle + \langle \boldsymbol{F}(\boldsymbol{u}), \boldsymbol{\varphi} \rangle \qquad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0. \tag{5.15}$$

Hence, Newton's equation of motion for the constrained case is given by

$$\ddot{\boldsymbol{u}} = -\boldsymbol{F}(\boldsymbol{u}) + \boldsymbol{F}_{\mathrm{con}}(\boldsymbol{u}). \tag{5.16}$$

We recall that the contact force density and the contact stresses, respectively, are a priori unknown and have to be determined as part of the solution process. On the basis of equation (5.16) discretization schemes are derived. We note that Newton's equation of motion (5.16) is the Euler-Lagrange equation to the Lagrangian function

$$\boldsymbol{L}(\boldsymbol{u}, \dot{\boldsymbol{u}}) := \frac{1}{2} \langle \dot{\boldsymbol{u}}, \dot{\boldsymbol{u}} \rangle - \frac{1}{2} a(\boldsymbol{u}, \boldsymbol{u}) - \mathcal{I}_{\mathcal{K}}(\boldsymbol{u}).$$

The whole energy $\mathcal{E}(\boldsymbol{u}) = \frac{1}{2} \langle \dot{\boldsymbol{u}}, \dot{\boldsymbol{u}} \rangle + \frac{1}{2} a(\boldsymbol{u}, \boldsymbol{u})$ of the system is given by the sum of the kinetic energy $\mathcal{E}_{\mathrm{kin}}(\dot{\boldsymbol{u}}) := \frac{1}{2} \langle \dot{\boldsymbol{u}}, \dot{\boldsymbol{u}} \rangle$ and the elastic potential energy.

5.1.2 Conservation properties

We discuss the influence of the contact constraints on the conservation properties. As the one-body contact problem is an unphysical idealization we have to consider the case of two-body contact problems. Further, we assume the solution $\boldsymbol{u}$ to be sufficiently smooth.

The proof of the conservation of the linear momentum $\boldsymbol{\mathcal{L}}(\boldsymbol{u}) := \int_\Omega \dot{\boldsymbol{u}}\, dx$ follows from (5.12) by means of a constant trial function $\boldsymbol{w}$

$$\begin{aligned} \left(\frac{d\boldsymbol{\mathcal{L}}(\boldsymbol{u})}{dt} \right) \cdot \boldsymbol{w} &= \int_\Omega \ddot{\boldsymbol{u}} \cdot \boldsymbol{w}\, dx \\ &= -\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \underbrace{\boldsymbol{\epsilon}(\boldsymbol{w})}_{0}\, dx + \int_{\Gamma_C^S} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S) \boldsymbol{w}\, da + \int_{\Gamma_C^M} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}^M) \boldsymbol{w}\, da \\ &= \int_{\Gamma_C^S} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S)(\boldsymbol{w} - \boldsymbol{w})\, da = 0. \end{aligned} \tag{5.17}$$

For the angular momentum $\boldsymbol{\mathcal{D}}(\boldsymbol{u}) := \int_\Omega \boldsymbol{u} \times \dot{\boldsymbol{u}}\, dx$ we take the trial function $\boldsymbol{v} = \boldsymbol{w} \times \boldsymbol{u}$ where $\boldsymbol{w}$ is constant. This leads to

$$\begin{aligned} \left(\frac{d\boldsymbol{\mathcal{D}}(\boldsymbol{u})}{dt} \right) \cdot \boldsymbol{w} &= \int_\Omega \ddot{\boldsymbol{u}} \cdot (\boldsymbol{w} \times \boldsymbol{u})\, dx \\ &= \underbrace{-\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\boldsymbol{w} \times \boldsymbol{u})\, dx}_{0} + \int_{\Gamma_C^S} \boldsymbol{w} \cdot [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S) \times (\boldsymbol{u}^S - \boldsymbol{u}^M \circ \boldsymbol{\Phi})]\, da. \end{aligned} \tag{5.18}$$

The last term disappears if the contact stresses $\hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S)$ are parallel to $(\boldsymbol{u}^S - \boldsymbol{u}^M \circ \boldsymbol{\Phi})$. Due to the approximation of the geometric non-penetration condition by means of the linearized non-penetration condition, see Section 1.2, usually, this condition is not fulfilled exactly.

Finally, we consider the whole energy $\mathcal{E}(\boldsymbol{u}) = \frac{1}{2}\langle \dot{\boldsymbol{u}}, \dot{\boldsymbol{u}} \rangle + \frac{1}{2} a(\boldsymbol{u}, \boldsymbol{u})$ of the system.

$$\begin{aligned} \frac{d\mathcal{E}(\boldsymbol{u})}{dt} &= \int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}) : \boldsymbol{\epsilon}(\dot{\boldsymbol{u}})\, dx + \int_\Omega \ddot{\boldsymbol{u}} \cdot \dot{\boldsymbol{u}}\, dx \\ &= \int_{\Gamma_C^S} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}^S)(\dot{\boldsymbol{u}}^S - \dot{\boldsymbol{u}}^M \circ \boldsymbol{\Phi})\, da \end{aligned}$$

This relation follows by means of (5.12) with $\dot{\boldsymbol{u}}$ as trial function. Hence, it is evident that the energy is conserved if the persistency condition

$$\hat{\sigma}_\nu(\boldsymbol{u}^S) \frac{d}{dt} \left([\boldsymbol{u}_\nu] - g \right) = 0 \tag{5.19}$$

is fulfilled. This condition states that the relative velocities in normal direction have to be zero at the actual contact boundary. It is easy to imagine that these relative velocities are zero if the contact boundaries stick together. But as we will see later on it is difficult to realize this condition in the fully discrete setting. In the case of friction, when $\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}^S) \neq 0$ holds, we can deduce that friction evokes dissipation, due to the fact that either $[\dot{\boldsymbol{u}}_T] = \boldsymbol{0}$ or $\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}^S) = -\mathcal{F} |\hat{\sigma}_\nu(\boldsymbol{u}^S)| \frac{[\dot{\boldsymbol{u}}_T]}{\| [\dot{\boldsymbol{u}}_T] \|}$, see [KW09b]. For the proof of dissipativity in the case of viscoelastic materials we refer to [Kla11].

5.2 Selected algorithms based on the Newmark scheme

One of the most common time discretization schemes in structural dynamics is the classical Newmark scheme. Unfortunately, in the case of contact constraints, the classical Newmark scheme evokes oscillations in the contact stresses, the displacements and the velocities at the contact boundary and even energy blow-ups may occur which spoil the accuracy of the solution. Reasons can be found in the interaction of the time and space discretization, in the violation of the persistency condition and in the discrete change of the velocity at the moment of impact. We give a deeper insight into the causes of these instabilities in Section 5.2.1. In order to avoid this unstable behavior of the discrete evolution, many time discretization schemes, especially modifications of the Newmark scheme have been developed. We follow the lines of our article [KW09a] to present selected approaches and to discuss the ideas and the pros and cons of the different algorithms in a unifying framework. By means of numerical results for one- and two-body contact problems we demonstrate the merits and drawbacks of the different algorithms.

5.2.1 Classical Newmark scheme for contact problems

Let $\tau > 0$ denote a discrete time step size and the solution at the discrete time $t^n = t^0 + n \cdot \tau$ is $\boldsymbol{u}^n$. The classical Newmark scheme, which is based on Taylor expansion of the displacements and the velocities, is given by

$$\begin{aligned}\boldsymbol{u}^{n+1} &= \boldsymbol{u}^n + \tau\dot{\boldsymbol{u}}^n + \frac{\tau^2}{2}\left((1-2\beta)\ddot{\boldsymbol{u}}^n + 2\beta\ddot{\boldsymbol{u}}^{n+1}\right)\\ \dot{\boldsymbol{u}}^{n+1} &= \dot{\boldsymbol{u}}^n + \tau\left((1-\gamma)\ddot{\boldsymbol{u}}^n + \gamma\ddot{\boldsymbol{u}}^{n+1}\right).\end{aligned} \tag{5.20}$$

The parameters γ and 2β can be chosen from the interval $[0,1]$. The special choice $2\beta = \gamma = 1/2$ leads to an algorithm conserving energy, linear momentum and angular momentum [HLW06, ST92] in the unconstrained case. For this choice of parameters the method is of second order consistency.

In the following, we use the relation between the accelerations and the forces, which is given by Newton's equation of motion (5.16). We define $\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) := \frac{1}{2}\boldsymbol{F}(\boldsymbol{u}^n) + \frac{1}{2}\boldsymbol{F}(\boldsymbol{u}^{n+1})$ and reformulate (5.20) for $2\beta = \gamma = 1/2$ to

$$\boldsymbol{u}^{n+1} = \boldsymbol{u}^n + \tau\dot{\boldsymbol{u}}^n - \frac{\tau^2}{2}\left(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\right) \tag{5.21}$$

$$\dot{\boldsymbol{u}}^{n+1} = \dot{\boldsymbol{u}}^n - \tau\left(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\right). \tag{5.22}$$

As $\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})$ is part of the solution, (5.21) has to be understood as the variational inclusion

$$\mathbf{0} \in \boldsymbol{u}^{n+1} - \boldsymbol{u}^n - \tau\dot{\boldsymbol{u}}^n + \frac{\tau^2}{2}\left(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n) + \frac{1}{2}\partial^{\text{sub}}\mathcal{I}_{\mathcal{K}}(\boldsymbol{u}^{n+1})\right).$$

This is a slight abuse of notation, that we will employ in this chapter.

The conservation of linear momentum can be verified easily. Therefore, we take a constant trial function $\boldsymbol{w}$ and we assume $\boldsymbol{u}^n, \boldsymbol{u}^{n+1}$ to be sufficiently smooth

$$\begin{aligned}\left(\boldsymbol{\mathcal{L}}(\boldsymbol{u}^{n+1}) - \boldsymbol{\mathcal{L}}(\boldsymbol{u}^n)\right)\cdot\boldsymbol{w} = -\tau &\left(\frac{1}{2}\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}^n) : \underbrace{\boldsymbol{\epsilon}(\boldsymbol{w})}_{=0}\,dx + \frac{1}{2}\int_\Omega \boldsymbol{\sigma}(\boldsymbol{u}^{n+1}) : \underbrace{\boldsymbol{\epsilon}(\boldsymbol{w})}_{=0}\,dx\right)\\ &+ \tau\left(\frac{1}{2}\int_{\Gamma_C^S} \hat{\sigma}_\nu(\boldsymbol{u}^n)\cdot\underbrace{[\boldsymbol{w}]\cdot\boldsymbol{\nu}}_{=0}\,da + \frac{1}{2}\int_{\Gamma_C^S}\hat{\sigma}_\nu(\boldsymbol{u}^{n+1})\cdot\underbrace{[\boldsymbol{w}]\cdot\boldsymbol{\nu}}_{=0}\,da\right).\end{aligned} \tag{5.23}$$

But unfortunately, due to the additional contact force density, the desirable energy conserving property of the classical Newmark scheme does not hold for the case of contact. This can be seen by taking the difference of the energy in two successive time steps. Combining equations (5.21) and (5.22), we derive

$$\begin{aligned}\mathcal{E}(\boldsymbol{u}^{n+1}) - \mathcal{E}(\boldsymbol{u}^n) &= \int_\Omega \left(\frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n) + \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\right)(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n)\,dx\\ &= \int_{\Gamma_C^S}\left(\frac{1}{2}\hat{\sigma}_\nu(\boldsymbol{u}^n) + \frac{1}{2}\hat{\sigma}_\nu(\boldsymbol{u}^{n+1})\right)[\boldsymbol{u}^{n+1} - \boldsymbol{u}^n]\cdot\boldsymbol{\nu}\,da.\end{aligned} \tag{5.24}$$

We do not have energy conservation in the case of frictionless contact unless $[\boldsymbol{u}^{n+1} - \boldsymbol{u}^n] \cdot \boldsymbol{\nu} = 0$ where $(\hat{\sigma}_\nu(\boldsymbol{u}^n) + \hat{\sigma}_\nu(\boldsymbol{u}^{n+1})) \neq 0$ at the contact boundary. This condition can be understood as a discrete variant of the persistency condition. But, as far as we know, no algorithm exists which is able to fulfill both the non-penetration condition and a kind of persistency condition, ensuring the energy conservation simultaneously.
Numerical experiments show that the classical Newmark scheme leads to energy blow-ups, spoiling the overall accuracy. In [KROM99] the contact forces are treated fully implicitly, leading to the following algorithm which we call contact-implicit Newmark scheme.

Algorithmus 5.2.1. ***Contact-implicit Newmark scheme***

$$\begin{aligned}
\boldsymbol{u}^{n+1}_{\text{pred}} &= \boldsymbol{u}^n + \tau\dot{\boldsymbol{u}}^n && (5.25)\\
\boldsymbol{u}^{n+1} &= \boldsymbol{u}^{n+1}_{\text{pred}} - \frac{\tau^2}{2}\Big(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\Big) && (5.26)\\
\dot{\boldsymbol{u}}^{n+1} &= \dot{\boldsymbol{u}}^n - \tau\Big(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\Big) && (5.27)
\end{aligned}$$

It can be proven that this algorithm is dissipative, thus avoiding energy blow-ups. The change of the total energy between two successive time steps is

$$\mathcal{E}(\boldsymbol{u}^{n+1}) - \mathcal{E}(\boldsymbol{u}^n) = \int_\Omega \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n)\,dx \leq 0$$

which follows from $\boldsymbol{u}^n \in \boldsymbol{\mathcal{K}}$ and the variational inequality $(\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1}), \boldsymbol{u}^{n+1} - \boldsymbol{v}) \leq 0$ for all $\boldsymbol{v} \in \boldsymbol{\mathcal{K}}$.
Until now, we only analyzed the time discrete system and figured out that the energy is not conserved. As we already mentioned in the introduction, another disadvantage of the classical Newmark scheme is the occurrence of oscillations at the contact boundary. The causes for these undesirable oscillations can be found in the interaction of the discretization in time and space. Therefore, we have a look at the space-time discrete system. We denote the mass matrix by $\boldsymbol{M}$. The space-discrete first equation of the classical Newmark scheme in the case of contact (5.21) is given by

$$\boldsymbol{M}\boldsymbol{u}^{n+1}_{\mathrm{m}} = \boldsymbol{M}\left(\boldsymbol{u}^n_{\mathrm{m}} + \tau\dot{\boldsymbol{u}}^n_{\mathrm{m}}\right) - \frac{\tau^2}{4}\boldsymbol{M}\left(\underbrace{(\boldsymbol{F} - \boldsymbol{F}_{\text{con}})(\boldsymbol{u}^n_{\mathrm{m}}) + (\boldsymbol{F} - \boldsymbol{F}_{\text{con}})(\boldsymbol{u}^{n+1}_{\mathrm{m}})}_{\text{expected equilibration}}\right). \qquad (5.28)$$

In the continuous case, to ensure the non-penetration condition, the contact force density $\boldsymbol{F}_{\text{con}}$ equilibrates $\boldsymbol{F}$ at the contact boundary in normal direction. Therefore, we would expect in the discrete version $(\boldsymbol{F} - \boldsymbol{F}_{\text{con}})(\boldsymbol{u}^n_{\mathrm{m}}) \cdot \boldsymbol{\nu} + (\boldsymbol{F} - \boldsymbol{F}_{\text{con}})(\boldsymbol{u}^{n+1}_{\mathrm{m}}) \cdot \boldsymbol{\nu} = 0$ at the contact boundary. But in the space-discrete case we have entries in the mass matrix at the contact boundary although in the continuous case the boundary has measure zero. Thus, the constraining force $\boldsymbol{F}_{\text{con}}$ to the non-penetration condition also reacts to $\boldsymbol{M}([\boldsymbol{u}^n_{\mathrm{m}} + \tau\dot{\boldsymbol{u}}^n_{\mathrm{m}}]) \cdot \boldsymbol{\nu}$ at the contact boundary if this part evokes penetration of the two bodies. The problem is that this contribution to $\boldsymbol{F}_{\text{con}}$ has no realistic physical meaning and it evokes the oscillations in the contact forces as can be seen from a simple combination of the two lines of the Newmark scheme under the assumption of $[\boldsymbol{u}_{\mathrm{m}}^{n+1}]\cdot\boldsymbol{\nu} = [\boldsymbol{u}_{\mathrm{m}}^{n}]\cdot\boldsymbol{\nu}$ which holds in the case of continuous contact

$$\begin{aligned}
[\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}]\cdot\boldsymbol{\nu} &= [\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}]\cdot\boldsymbol{\nu} - \tau\left(\boldsymbol{F}_{\text{rel}}^{1/2}(\boldsymbol{u}_{\mathrm{m}}^{n}, \boldsymbol{u}_{\mathrm{m}}^{n+1}) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}_{\mathrm{m}}^{n}) - \frac{1}{2}\boldsymbol{F}_{\text{con}}(\boldsymbol{u}_{\mathrm{m}}^{n+1})\right)\cdot\boldsymbol{\nu} \\
&= [\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}]\cdot\boldsymbol{\nu} + \tau\left(\frac{2}{\tau^2}\left([\boldsymbol{u}_{\mathrm{m}}^{n+1} - (\boldsymbol{u}_{\mathrm{m}}^{n} + \tau\dot{\boldsymbol{u}}_{\mathrm{m}}^{n})]\right)\right)\cdot\boldsymbol{\nu} \\
&= -[\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}]\cdot\boldsymbol{\nu}. \qquad (5.29)
\end{aligned}$$

This zigzagging in the velocities $[\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}]\cdot\boldsymbol{\nu}$ is translated to the contact force density by means of $[\boldsymbol{u}_{\mathrm{m}}^{n} + \tau\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}]\cdot\boldsymbol{\nu}$.
Unfortunately, these oscillations also occur for the contact-implicit Newmark scheme (Algorithm 5.2.1), see [DKE08, Section 2.2].

5.2.2 Prediction of the contact boundary

In [DKE08] Deuflhard et al. present a contact-stabilized Newmark scheme whose particular feature is the stable behavior of the contact forces. A special predictor step which is added to the contact-implicit Newmark scheme (Algorithm 5.2.1) prevents the unphysical part of the contact forces.
As already explained, the oscillations at the contact boundary originate from the unwanted interaction of space and time discretization. In order to overcome this deficiency, step (5.25) is replaced by the variational inclusion

$$\mathbf{0} \;\in\; \boldsymbol{u}_{\text{pred}}^{n+1} - (\boldsymbol{u}^{n} + \tau\dot{\boldsymbol{u}}^{n}) + \partial^{\text{sub}}\mathcal{I}_{\boldsymbol{\mathcal{K}}}(\boldsymbol{u}_{\text{pred}}^{n+1}) \qquad (5.30)$$

which we call predictor step. This variational inclusion requires the evaluation of the normal trace of an $\boldsymbol{L}^2$-function which is not possible in general. But corresponding to the fact that the discrete boundary gets assigned to a mass, finite element functions have boundary values. The solution $\boldsymbol{u}_{\text{pred}}^{n+1}$ of (5.30) is the same as the $\boldsymbol{L}^2$-projection of (5.25) onto the admissible set $\boldsymbol{\mathcal{K}}$. As $\boldsymbol{u}_{\text{pred}}^{n+1}$ of (5.30) is already contained in the admissible set $\boldsymbol{\mathcal{K}}$, $[\boldsymbol{u}_{\text{pred}}^{n+1}]\cdot\boldsymbol{\nu}$ is a predictor of $[\boldsymbol{u}^{n+1}]\cdot\boldsymbol{\nu}$. If $[\boldsymbol{u}_{\text{pred}}^{n+1}]\cdot\boldsymbol{\nu} = [\boldsymbol{u}^{n+1}]\cdot\boldsymbol{\nu}$ at the actual contact boundary, the artificial part of $\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\cdot\boldsymbol{\nu}$ is removed, compare the discussion to equation (5.28), and the contact force density $\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})\cdot\boldsymbol{\nu}$ equilibrates $\boldsymbol{F}^{1/2}(\boldsymbol{u}^{n}, \boldsymbol{u}^{n+1})\cdot\boldsymbol{\nu}$ in (5.26) which corresponds to Newton's Axiom of the equilibrium of forces. Further, the force equilibrium directly implies $[\dot{\boldsymbol{u}}^{n+1}]\cdot\boldsymbol{\nu} = [\dot{\boldsymbol{u}}^{n}]\cdot\boldsymbol{\nu}$ which means that there is no zigzagging as in the classical Newmark scheme.
We define $-\frac{\tau^2}{2}\boldsymbol{P}_{\text{con}}(\boldsymbol{u}_{\text{pred}}^{n+1}) \in \partial^{\text{sub}}\mathcal{I}_{\boldsymbol{\mathcal{K}}}(\boldsymbol{u}_{\text{pred}}^{n+1})$ as the element of $\partial^{\text{sub}}\mathcal{I}_{\boldsymbol{\mathcal{K}}}(\boldsymbol{u}_{\text{pred}}^{n+1})$ fulfilling

$$\left(\frac{\tau^2}{2}\boldsymbol{P}_{\text{con}}(\boldsymbol{u}_{\text{pred}}^{n+1}), \boldsymbol{\varphi}\right) = \left(\boldsymbol{u}_{\text{pred}}^{n+1} - \boldsymbol{u}^{n} - \tau\dot{\boldsymbol{u}}^{n}, \boldsymbol{\varphi}\right). \qquad (5.31)$$

This notation directly leads us to the following formulation of the contact-stabilized Newmark scheme.

Algorithmus 5.2.2. ***Contact-stabilized Newmark scheme***

$$
\begin{aligned}
\boldsymbol{u}^{n+1}_{\text{pred}} &= \boldsymbol{u}^n + \tau \dot{\boldsymbol{u}}^n + \frac{\tau^2}{2} \boldsymbol{P}_{\text{con}}(\boldsymbol{u}^{n+1}_{\text{pred}}) && (5.32)\\
\boldsymbol{u}^{n+1} &= \boldsymbol{u}^{n+1}_{\text{pred}} - \frac{1}{2}\tau^2 \left(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1}) \right) && (5.33)\\
\dot{\boldsymbol{u}}^{n+1} &= \dot{\boldsymbol{u}}^n - \tau \left(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1}) \right) && (5.34)
\end{aligned}
$$

Here, the change in the total energy between two successive time steps is

$$
\begin{aligned}
&\mathcal{E}(\boldsymbol{u}^{n+1}) - \mathcal{E}(\boldsymbol{u}^n) = \\
&\int_\Omega \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^{n+1})(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n)\, dx + \int_\Omega \boldsymbol{P}_{\text{con}}(\boldsymbol{u}^{n+1}_{\text{pred}})(\boldsymbol{u}^{n+1}_{\text{pred}} - \boldsymbol{u}^{n+1})\, dx \le 0
\end{aligned} \tag{5.35}
$$

which follows from the variational inequalities to (5.32) and (5.33), see [DKE08, Theorem 2.2]. Therewith the algorithm is dissipative. Moreover, the algorithm conserves the linear momentum.

In our article [KW09b] the contact-stabilized Newmark method is extended to the case of friction. Therein we use the time-discretized friction law

$$
\begin{aligned}
&[\boldsymbol{u}_T^{n+1} - \boldsymbol{u}_T^n] = \boldsymbol{0} \Rightarrow \|\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}^{n+1})\| \le \mathcal{F}|\hat{\sigma}_\nu(\boldsymbol{u}^{n+1})| \quad \text{on } \Gamma_C^S \times [0,T]\\
&[\boldsymbol{u}_T^{n+1} - \boldsymbol{u}_T^n] \ne \boldsymbol{0} \Rightarrow \hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}^{n+1}) = -\mathcal{F}|\hat{\sigma}_\nu(\boldsymbol{u}^{n+1})| \frac{[\boldsymbol{u}_T^{n+1} - \boldsymbol{u}_T^n]}{\|\,[\boldsymbol{u}_T^{n+1} - \boldsymbol{u}_T^n]\,\|} \quad \text{on } \Gamma_C^S \times [0,T].
\end{aligned} \tag{5.36}
$$

In a same manner as in (5.35) the dissipation of the algorithm in the frictional case can be proven, due to the implicit treatment of the contact and frictional stresses [KW09b]. As the normal and tangential stresses influence each other, the stability of the normal stresses due to the predictor step is very important for the frictional effects. The arising discrete (quasi-)variational problems are solved with a non-smooth multiscale method for Coulomb friction [Kra09, KW09b]. Therein the non-linearities are resolved directly without any dependency on penalty parameters.

In the case of viscoelastic materials described by the Kelvin-Voigt model consistency results for the contact-stabilized Newmark scheme are given in [KSD10] and an adaptive time step control is proposed in [KSD11].

An application of the contact-stabilized Newmark scheme to contact problems in non-linear elasticity can be found in [GKW10].

5.2.3 Enforcing a discrete persistency condition

In [LC97] Laursen and Chawla present an algorithm which conserves energy, linear momentum and angular momentum. They use the time discretization scheme

$$
\begin{aligned}
\boldsymbol{u}^{n+1} &= \boldsymbol{u}^n + \tau \dot{\boldsymbol{u}}^n + \alpha\tau^2 \left(\alpha \ddot{\boldsymbol{u}}^{n+1} + (1-\alpha)\ddot{\boldsymbol{u}}^n \right)\\
\dot{\boldsymbol{u}}^{n+1} &= \dot{\boldsymbol{u}}^n + \tau \left(\alpha \ddot{\boldsymbol{u}}^{n+1} + (1-\alpha)\ddot{\boldsymbol{u}}^n \right)
\end{aligned}
$$

proposed in [ST92] with $\alpha = 1/2$. This corresponds to the classical Newmark scheme with the parameters $2\beta = \gamma = 1/2$.
With the same arguments as in (5.17) or (5.23), respectively, the conservation of the linear momentum follows. Laursen and Chawla use their algorithm for the computation of large deformation and so, they do not approximate the non-penetration condition by a linearized non-penetration condition, see Section 1.2. Therefore, the conservation of the angular momentum (5.18) is given, too. As we know from (5.24), a discrete kind of the persistency condition (5.19) has to be fulfilled in order to conserve the energy in the case of contact. Laursen and Chawla enforce

$$\frac{1}{\tau} \cdot [\boldsymbol{u}^{n+1} - \boldsymbol{u}^n] \cdot \boldsymbol{\nu} \leq 0 \tag{5.37}$$

whenever contact or penetration occurred in the foregoing time step. Let μ_ν be the corresponding Lagrange multiplier. Then $\mu_\nu \frac{1}{\tau} \cdot [\boldsymbol{u}^{n+1} - \boldsymbol{u}^n] \cdot \boldsymbol{\nu}$ is a so-called algorithmic persistency condition.
To reformulate the algorithm proposed by Laursen and Chawla in our setting, let $\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1})$ correspond to the Lagrange multiplier μ_ν.

Algorithmus 5.2.3.

$$\begin{aligned}
\boldsymbol{u}^{n+1} &= \boldsymbol{u}^n + \tau \dot{\boldsymbol{u}}^n - \frac{\tau^2}{2}\Big(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1})\Big) \\
\dot{\boldsymbol{u}}^{n+1} &= \dot{\boldsymbol{u}}^n - \tau\Big(\boldsymbol{F}^{1/2}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1}) - \boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n, \boldsymbol{u}^{n+1})\Big)
\end{aligned}$$

The authors of [LC97] use penalty and augmented Lagrangian methods to solve the arising variational inequalities. The extension of this algorithm to frictional contact can be found in [CL98].

5.2.4 Removing the discrete mass from the contact boundary

In the previously presented algorithms the boundary gets assigned to a mass due to the discretization in space although in the continuous case the boundary has measure zero. The interplay of the discrete mass at the contact boundary and the time discretization evokes artificial contact forces as already explained in Section 5.2.1, see for example equation (5.28). Khenous et al. [KLR06,KLR08] introduce a mass redistribution method, in which mass from the contact boundary is moved to the interior. Thereby, the entire mass, the moment of inertia and the center of gravity remain the same. To keep the complexity as low as possible, the zero entries of the standard mass matrix are kept. The redistributed mass matrix $\tilde{\boldsymbol{M}}$ is computed as minimization problem of the distance to the standard mass matrix $\boldsymbol{M}$ under the constraints mentioned above. The method is proposed for one-body contact problems but is easily extendable to two-body contact problems.
In [KLR06, KLR08] the discretization takes place first in space, second in time. The linearized non-penetration condition is enforced. We indicate quantities belonging to

the part with mass by $(\cdot)_M$ and without mass by $(\cdot)_{\not M}$. We denote the stiffness matrix corresponding to the bilinear form $a(\cdot,\cdot)$ by $\boldsymbol{A}$. Further, the node values of the discrete Lagrange multiplier weighted by the mass are denoted by $\mu_{\nu,p}$, so that $\boldsymbol{u}_{\mathtt{m}} = \begin{pmatrix}\boldsymbol{u}_{\mathtt{m},M}\\ \boldsymbol{u}_{\mathtt{m},\not M}\end{pmatrix}$ is the solution of the following space-discrete system

$$\begin{pmatrix}\tilde{\boldsymbol{M}} & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\ddot{\boldsymbol{u}}_{\mathtt{m},M}\\ \ddot{\boldsymbol{u}}_{\mathtt{m},\not M}\end{pmatrix}+\begin{pmatrix}\boldsymbol{A}_{M,M} & \boldsymbol{A}^T_{\not M,M}\\ \boldsymbol{A}_{\not M,M} & \boldsymbol{A}_{\not M,\not M}\end{pmatrix}\begin{pmatrix}\boldsymbol{u}_{\mathtt{m},M}\\ \boldsymbol{u}_{\mathtt{m},\not M}\end{pmatrix}=\sum_{p\in\Gamma_C}\mu_{\nu,p}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{\nu}(p)\end{pmatrix}\tag{5.38}$$

with the initial values $\boldsymbol{u}_{\mathtt{m}}(\boldsymbol{p},0)=\boldsymbol{u}^0_{\mathtt{m}}(\boldsymbol{p})$ and $\dot{\boldsymbol{u}}_{\mathtt{m}}(\boldsymbol{p},0)=\dot{\boldsymbol{u}}^0_{\mathtt{m}}(\boldsymbol{p})$. It is a coupled system in the variables with and without mass

$$\begin{aligned}\tilde{\boldsymbol{M}}\ddot{\boldsymbol{u}}_{\mathtt{m},M}+\boldsymbol{A}_{M,M}\boldsymbol{u}_{\mathtt{m},M} &= -\,\boldsymbol{A}^T_{\not M,M}\boldsymbol{u}_{\mathtt{m},\not M}\\ \boldsymbol{A}_{\not M,M}\boldsymbol{u}_{\mathtt{m},M}+\boldsymbol{A}_{\not M,\not M}\boldsymbol{u}_{\mathtt{m},\not M} &= \sum_{p\in\Gamma_C}\mu_{\nu,p}\cdot\boldsymbol{\nu}(\boldsymbol{p}).\end{aligned}\tag{5.39}$$

The second equation is a variational inequality

$$(\boldsymbol{v}_{\mathtt{m},\not M}-\boldsymbol{u}_{\mathtt{m},\not M})^T\boldsymbol{A}_{\not M,\not M}\boldsymbol{u}_{\mathtt{m},\not M}\geq-(\boldsymbol{v}_{\mathtt{m},\not M}-\boldsymbol{u}_{\mathtt{m},\not M})^T\boldsymbol{A}_{\not M,M}\boldsymbol{u}_{\mathtt{m},M}.$$

The system (5.39) is well-posed and has a unique solution, see [KLR06, Theorem 3.1].

For the discretization in time the authors of [KLR06, KLR08] test the Crank-Nicholson method as well as the Newmark method with the parameters $\beta=\gamma=1/2$. Independently of the method, the redistributed mass matrix reduces the oscillations in the contact force density as well as the changes in the total energy. The Newmark scheme with $\gamma=\beta=1/2$

$$\begin{aligned}\boldsymbol{u}^{n+1} &= \boldsymbol{u}^n+\tau\dot{\boldsymbol{u}}+\tfrac{\tau^2}{2}\ddot{\boldsymbol{u}}^{n+1}\\ \dot{\boldsymbol{u}}^{n+1} &= \dot{\boldsymbol{u}}^n+\tfrac{\tau}{2}(\ddot{\boldsymbol{u}}^n+\ddot{\boldsymbol{u}}^{n+1})\end{aligned}\tag{5.40}$$

takes the accelerations purely implicitly in the first step. Combining (5.40) and (5.38) leads to the following equation for the computation of the displacements.

Algorithmus 5.2.4.

$$\begin{pmatrix}\tilde{\boldsymbol{M}}+\frac{\tau^2}{2}\boldsymbol{A}_{M,M} & \frac{\tau^2}{2}\boldsymbol{A}^T_{\not M,M}\\ \frac{\tau^2}{2}\boldsymbol{A}_{\not M,M} & \frac{\tau^2}{2}\boldsymbol{A}_{\not M,\not M}\end{pmatrix}\begin{pmatrix}\boldsymbol{u}^{n+1}_{\mathtt{m},M}\\ \boldsymbol{u}^{n+1}_{\mathtt{m},\not M}\end{pmatrix}=\begin{pmatrix}\tilde{\boldsymbol{M}}(\boldsymbol{u}^n_{\mathtt{m},M}+\tau\dot{\boldsymbol{u}}^n_{\mathtt{m},M})\\ \boldsymbol{0}\end{pmatrix}+\frac{\tau^2}{2}\boldsymbol{M}\boldsymbol{F}_{\text{con}}\left(\boldsymbol{u}^{n+1}_{\mathtt{m}}\right)$$

with

$$\boldsymbol{M}\boldsymbol{F}_{\text{con}}\left(\boldsymbol{u}^{n+1}_{\mathtt{m}}\right)=\sum_{p\in\Gamma_C}\lambda_{p,\nu}\begin{pmatrix}\boldsymbol{0}\\ \boldsymbol{\nu}(\boldsymbol{p})\end{pmatrix}.$$

Unfortunately, this algorithm requires extra computational costs, as the redistribution of the mass matrix necessitates a minimization problem under constraints in the dimension of the mass matrix. Hager et al. [HHW08] circumvent this problem by using special quadrature formulas on so-called macro-elements to compute the mass redistribution.

To explain the macro-elements, we assume a mesh $\mathcal{T}_h$ of the domain Ω to be given. The small strip of all elements $T_h \in \mathcal{T}_h$ which have an edge or a face on $\bar{\Gamma}_C$ is denoted by Ω_C and the remaining elements of $\mathcal{T}_h$ constitute Ω_I. Each macro-element $T_H \in \mathcal{T}_H$ is either an element $T_h \in \Omega_I$ or a union of an element $T_h \in \Omega_I$ with elements $T_h \in \Omega_C$. The quadrature formula used on the macro-elements has no quadrature points placed in $\Omega_C \cup \bar{\Gamma}_C$, which directly implies that the new mass matrix $\tilde{\boldsymbol{M}}$ has no entries for basis functions at the contact boundary. Furthermore, the quadrature formula is chosen such that it is exact on each macro-element $T_H \in \mathcal{T}_H$ for all quadratic functions on T_H. Due to these properties of the quadrature formula, the redistributed mass matrix $\tilde{\boldsymbol{M}}$ leads to the same total mass, center of gravity and moment of inertia as given by the standard mass matrix $\boldsymbol{M}$.

As we already know from the algorithm of Khenous et al., the modified mass matrix reduces the oscillations in the contact stresses. Further, Hager et al. use the method of modified mass matrix in the algorithm of Laursen and Chawla. In particular, the discrete persistency condition is enforced instead of the non-penetration condition. Therefore, the whole energy is conserved if the kinetic energy is computed with the redistributed mass matrix $\tilde{\mathcal{E}}_{\text{kin}} := \frac{1}{2}\dot{\boldsymbol{u}}\tilde{\boldsymbol{M}}\dot{\boldsymbol{u}}$. The algorithm is given by

Algorithmus 5.2.5.

$$\begin{aligned}
&\begin{pmatrix} \tilde{\boldsymbol{M}} + \frac{\tau^2}{4}\boldsymbol{A}_{M,M} & \frac{\tau^2}{4}\boldsymbol{A}^T_{\not{M},M} \\ \frac{\tau^2}{4}\boldsymbol{A}_{\not{M},M} & \frac{\tau^2}{4}\boldsymbol{A}_{\not{M},\not{M}} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^{n+1}_{\mathfrak{m},M} \\ \boldsymbol{u}^{n+1}_{\mathfrak{m},\not{M}} \end{pmatrix} \\
&= \begin{pmatrix} \tilde{\boldsymbol{M}}(\boldsymbol{u}^n_{\mathfrak{m},M} + \tau\dot{\boldsymbol{u}}^n_{\mathfrak{m},M}) \\ \boldsymbol{0} \end{pmatrix} - \begin{pmatrix} \frac{\tau^2}{4}\boldsymbol{A}_{M,M} & \frac{\tau^2}{4}\boldsymbol{A}^T_{\not{M},M} \\ \frac{\tau^2}{4}\boldsymbol{A}_{\not{M},M} & \frac{\tau^2}{4}\boldsymbol{A}_{\not{M},\not{M}} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}^n_{\mathfrak{m},M} \\ \boldsymbol{u}^n_{\mathfrak{m},\not{M}} \end{pmatrix} \\
&\quad + \frac{\tau^2}{2}\boldsymbol{M}\boldsymbol{F}_{\text{con}}\left(\boldsymbol{u}^n_{\mathfrak{m}}, \boldsymbol{u}^{n+1}_{\mathfrak{m}}\right)
\end{aligned}$$

$$\dot{\boldsymbol{u}}^{n+1}_{\mathfrak{m}} = -\dot{\boldsymbol{u}}^n_{\mathfrak{m}} + \frac{2}{\tau}(\boldsymbol{u}^{n+1}_{\mathfrak{m}} - \boldsymbol{u}^n_{\mathfrak{m}})$$

where $\boldsymbol{F}_{\text{con}}(\boldsymbol{u}^n_{\mathfrak{m}}, \boldsymbol{u}^{n+1}_{\mathfrak{m}})$ is the constraining force density to the constraint (5.37) which is enforced if contact occurred in the previous time step.

In [HHW08] this algorithm is designed for frictional contact problems with the time-discretized friction law (5.36). For the numerical resolution of the contact and frictional constraints, a primal-dual active set strategy is used.

We note that the modified mass lumping method has been further analyzed in, e.g., [HW09, DE09, DE11] and recently, the method has been investigated in a central difference scheme [DEP11].

5.2.5 Comparison and interpretation of the basic ideas

The classical Newmark scheme with $2\beta = \gamma = 1/2$ leads to oscillations in the energy for contact problems. In [KROM99] Kane et al. use the classical Newmark scheme with purely implicit treatment of the contact force density, see also [PKMO02]. It can be proven that this method is dissipative. But unfortunately, there are still oscillations in the contact forces. Deuflhard et al. [DKE08] improve this contact-implicit Newmark method with regard to the oscillations in the contact forces. One explanation for these oscillations is the interplay of the discretization in space and time. In the continuous case the force density $\boldsymbol{F}$ has to be equilibrated by means of the contact force density $\boldsymbol{F}_{\text{con}}$ such that penetration is avoided. After discretization in space and time, if $[\boldsymbol{u}^n + \tau\dot{\boldsymbol{u}}^n] \cdot \boldsymbol{\nu}$ provokes penetration, see (5.28), we have an artificial force with no physical meaning which has to be equilibrated by $\boldsymbol{F}_{\text{con}}$. In [DKE08] an additional predictor step which avoids the appearance of these unphysical contact forces is introduced. Numerical examples show a stable behavior of the contact forces.

From the point of view of Khenous et al. [KLR08] it is not simply the interaction of the time and space discretization which evokes the oscillations but the special energy conserving property of the time discretization in the unconstrained case together with the discretization in space. If for example a node is stopped at the contact boundary, its kinetic energy would be definitely lost. Thus, the energy conserving scheme makes the velocities oscillate (5.29) to keep the corresponding part of the kinetic energy. Consequently, not only the velocities oscillate but also the energy and especially the contact forces. Khenous et al. remove the mass from the contact boundary. This method reduces the oscillations in the contact forces significantly. But setting up the discrete system requires additional numerical effort as the redistributed mass matrix is computed via a minimization problem under constraints. In contrast, the predictor step in [DKE08] only requires a discrete $\boldsymbol{L}^2$-projection onto the discrete admissible set. Moreover, the discrete $\boldsymbol{L}^2$-projection can be applied easily also in the case of changing spatial discretizations, which means that the discretization can take place first in time, second in space with adaptive refinement in each time step. This is not the case in [KLR08]; they use the method of lines. In both methods numerical examples show that the energy tends to be conserved for smaller mesh and time steps sizes. But in the algorithm of Deuflhard et al. the dissipativity can be proven.

Laursen and Chawla allow the penetration of the bodies in order to get energy conservation [LC97]. At points, where the two bodies penetrated each other in the foregoing time step further penetration is stopped by enforcing the discrete persistency condition. The discrete persistency condition implies that the difference quotient is zero at contacting points and therefore, the energy is conserved. The simultaneous treatment of the non-penetration condition and the persistency condition seems to be a too great demand in the discrete setting. A discrete persistency condition can be fulfilled if the two bodies are allowed to penetrate each other, whereby smaller time-step sizes cause less penetration.

As this method conserves the energy exactly and the mass redistribution method of Khenous et al. reduces the oscillations in the contact forces, Hager et al. [HHW08] combine both ideas and introduce a new less expensive method to compute a redistributed mass matrix.

5.2.6 Numerical comparison

We consider numerical simulations to demonstrate the behavior of the different presented algorithms. Therefore, we implemented the algorithms for linear elastic one- and two-body contact problems in $3D$, although mostly $2D$ examples are found in the original literature. First, we show numerical examples of the classical Newmark scheme, the contact-implicit Newmark scheme (Algorithm 5.2.1), the contact-stabilized Newmark scheme (Algorithm 5.2.2), Algorithm 5.2.3 where the discrete persistency condition is enforced, and of different algorithms using a redistribution of the mass matrix (as Algorithm 5.2.4 and Algorithm 5.2.5). Second, we comment on the similarities and differences, the advantages and disadvantages of the methods.

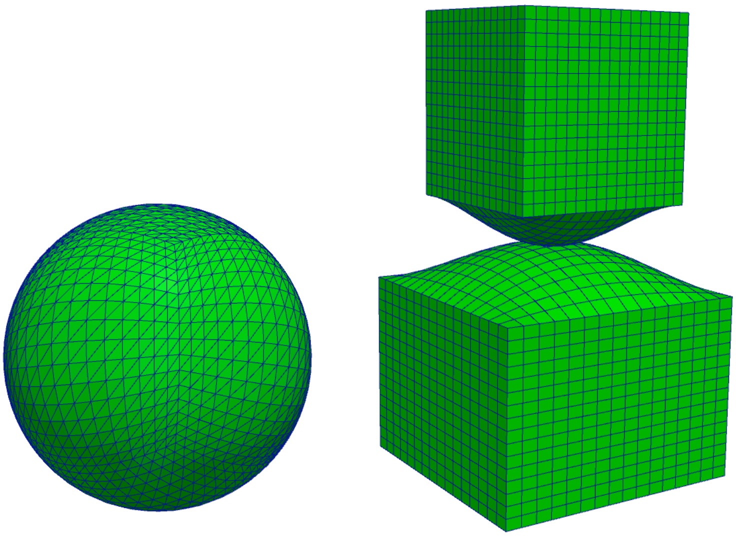
The implementation has been carried out within the framework of the finite element toolbox `UG` [BBL+97] and the obstacle toolbox `OBSLIB++` [Kra00]. Therein, the discrete variational inequalities are solved with a monotone multigrid method which treats the non-linearities efficiently [Kor97, Kra00].

As there is no analytical solution for a three-dimensional dynamic contact problem in linear elasticity, we cannot compare numerical and analytical solutions to validate the algorithms. Therefore, we are interested in the physical behavior of the system, which is reflected by displacements, velocities, energy and contact forces. In the foregoing sections we analyzed and compared the different algorithms and figured out that the temporal evolutions of energy, contact forces, contact stresses, and velocities at the contact boundary are crucial points to examine.

For the numerical examples we use two different geometries, see Figure 5.8. The unit ball, Geometry **A**, consists of a four times uniformly refined mesh of pyramids and is used for one-body contact problems, meaning that the ball comes in contact with a rigid foundation. The initial distance to the foundation is zero at the south pole. Geometry **B** is used for two-body contact problems. The side lengths of the cubes are 1.4 and 1.0, respectively. The mesh of hexahedra is also four times uniformly refined and the initial distance between the two bodies at the south pole is zero. The Poisson ratio is $\nu = 0.3$ and the Young's modulus is $E = 2 \cdot 10^5$ in Geometry **B** and $E = 5 \cdot 10^2$ in Geometry **A**. As in Section 5.1.1 assumed, the density is set to $\rho = 1$ and contact is enforced by means of an initial velocity which amounts to -2 for the ball in the one-body contact problem and -2 for the upper body and 2 for the lower body in the two-body contact problem. The time step size is chosen to $\tau = 5 \cdot 10^{-3}$ for the one-body contact problem and $\tau = 5 \cdot 10^{-4}$ for the two-body contact problem. We note that the parameters are dimensionless in our simulations.

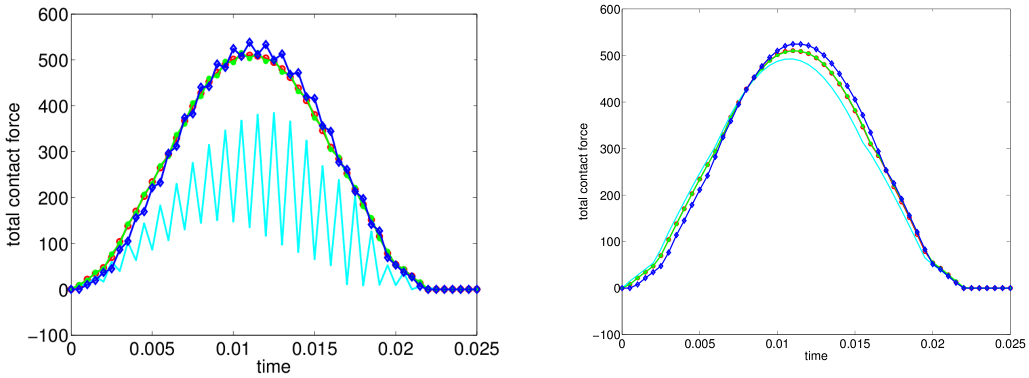
Classical Newmark scheme

We start with the classical Newmark scheme for the two-body contact problem. In Figure 5.3(a) the cyan colored line indicates the corresponding course of energy. The



(a) Geometry **A** (b) Geometry **B**

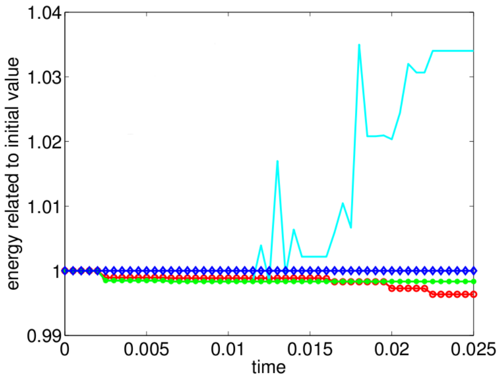
Figure 5.1: Meshes



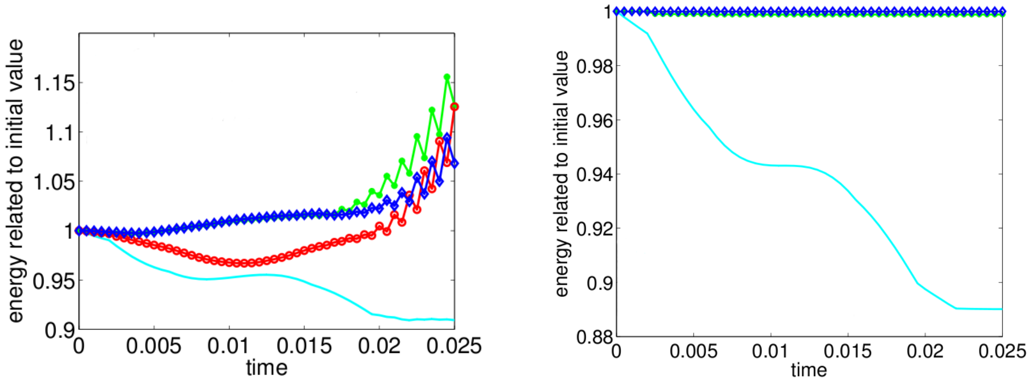
(a) cyan: classical Newmark; green (dots): contact-implicit Newmark; red (circles): contact-stabilized Newmark; blue (diamonds): Algorithm 5.2.3

(b) cyan: Algorithm 5.2.4; green (dots): contact-implicit with redistributed mass; red (circles): contact-stabilized Newmark; blue (diamonds): Algorithm 5.2.5

Figure 5.2: Total contact force



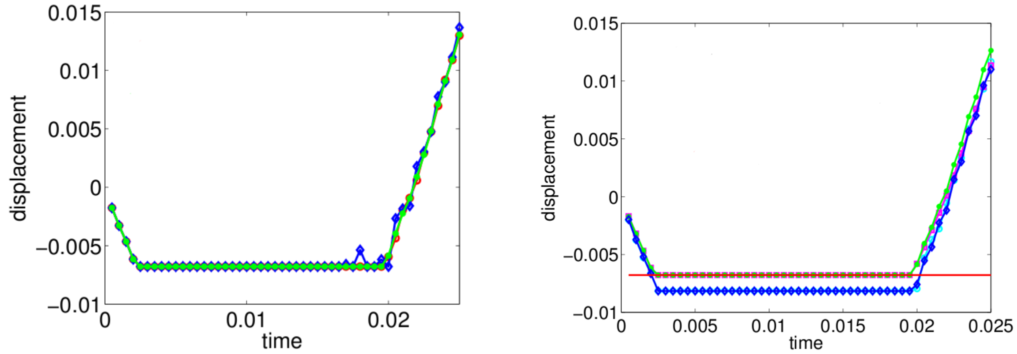
(a) $\mathcal{E} = \frac{1}{2}\dot{\boldsymbol{u}}\boldsymbol{M}\dot{\boldsymbol{u}} + \frac{1}{2}\boldsymbol{u}\boldsymbol{A}\boldsymbol{u}$; cyan: classical Newmark; green (dots): contact-implicit Newmark; red (circles): contact-stabilized Newmark; blue (diamonds): Algorithm 5.2.3



(b) $\mathcal{E} = \frac{1}{2}\dot{\boldsymbol{u}}\boldsymbol{M}\dot{\boldsymbol{u}} + \frac{1}{2}\boldsymbol{u}\boldsymbol{A}\boldsymbol{u}$; cyan: Algorithm 5.2.4; green (dots): contact-implicit Newmark with redistributed mass; red (circles): contact-implicit Newmark with zero entries in the mass matrix at the contact boundary but not redistributed; blue (diamonds): Algorithm 5.2.5

(c) $\tilde{\mathcal{E}} = \frac{1}{2}\dot{\boldsymbol{u}}\tilde{\boldsymbol{M}}\dot{\boldsymbol{u}} + \frac{1}{2}\boldsymbol{u}\boldsymbol{A}\boldsymbol{u}$; cyan: Algorithm 5.2.4; green (dots): contact-implicit Newmark with redistributed mass; blue (diamonds): Algorithm 5.2.5

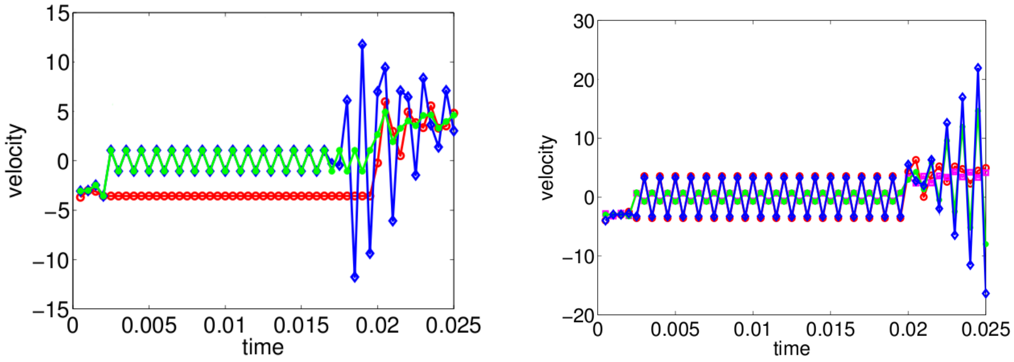
Figure 5.3: Energy related to initial value



(a) blue (diamonds): classical Newmark; green (dots): contact-implicit Newmark ;red (circles): contact-stabilized Newmark

(b) straight red line: distance; cyan (circles): Algorithm 5.2.3; magenta (squares): Algorithm 5.2.4; green (dots): contact-implicit with redistributed mass; blue (diamonds): Algorithm 5.2.5

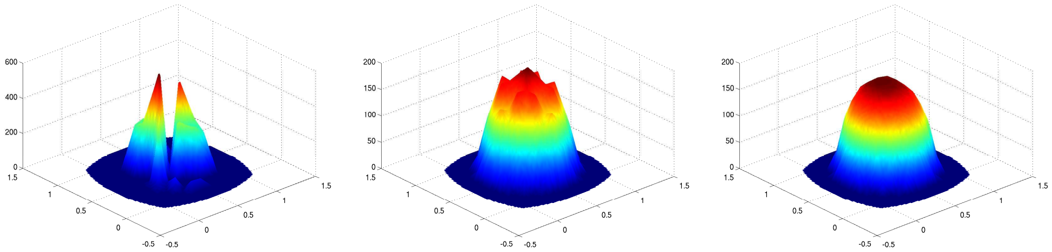
Figure 5.4: Relative displacement



(a) blue (diamonds): classical Newmark; green (dots): contact-implicit Newmark ;red (circles): contact-stabilized Newmark

(b) red (circles): Algorithm 5.2.3; magenta (squares): Algorithm 5.2.4; green (dots): contact-implicit with redistributed mass; blue (diamonds): Algorithm 5.2.5

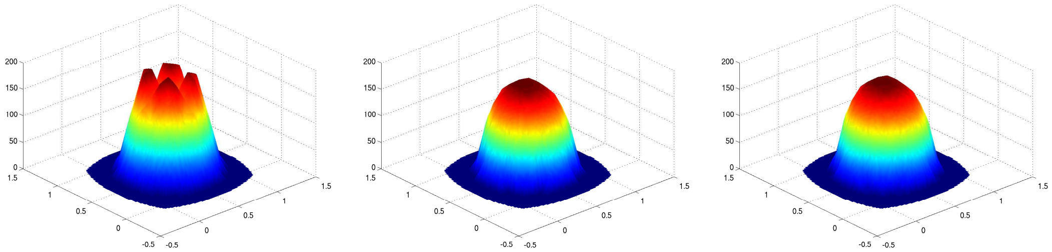
Figure 5.5: Relative velocity



(a) classical Newmark

(b) contact-implicit Newmark

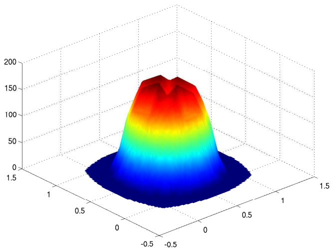
(c) contact-stabilized Newmark



(d) Algorithm 5.2.3

(e) Algorithm 5.2.4

(f) contact-implicit with redistributed mass



(g) Algorithm 5.2.5

Figure 5.6: Contact stresses at time step $n = 37$

increase is up to 3.5% of the initial amount. The blue line with diamonds in Figure 5.4(a) shows the trajectory of the relative displacement $[\boldsymbol{u}^{n+1}] \cdot \boldsymbol{\nu}$ and the blue line with diamonds in Figure 5.5(a) shows the trajectory of the relative velocity $[\dot{\boldsymbol{u}}^{n+1}] \cdot \boldsymbol{\nu}$ of an arbitrarily chosen point which gets into contact. In the beginning the relative velocity shows the uniform zigzagging as expected from (5.29). In the end, when the points bounce and finally detach, the zigzagging is irregular. Here, the increase of the energy may play a role. Even the displacements show some irregularities during and after the process of detaching. The development of the total contact force is shown in Figure 5.2(a) indicated by the cyan colored line. To have a closer look at the behavior of the contact stresses, we show the contact stresses in an arbitrarily chosen time step in Figure 5.6(a). As we want to exclude undesirable effects stemming from the mortar method, the contact stresses are taken from the one-body contact problem.

Contact-implicit Newmark scheme (Algorithm 5.2.1)

In Figure 5.3(a) the green dotted line shows the course of energy for the two-body contact problem using the contact-implicit Newmark scheme of Kane et al. As stated in Subsection 5.2.2 this method is dissipative. The relative loss of energy is around 0.16%. Most of the loss of energy is due to the moment of impact. Similar as in the classical Newmark scheme we still have the oscillations in the relative velocities (see green line with dots in Figure 5.5(a)). This can be calculated in the same way as in (5.29). But in contrast to the classical Newmark scheme during the process of detaching there is no bouncing in the relative displacements (see Figure 5.4(a), green line with dots) and the amplitudes of the oscillations are more moderate. In the classical Newmark scheme the relative velocity oscillates between -12 and 12, whereas in the contact-implicit Newmark scheme the largest oscillation in the relative velocities after the detaching is between 2 and 5, see Figure 5.5(a). The development of the total contact force is shown in Figure 5.2(a) indicated by the green line with dots. There are still oscillations but the amplitude is very small compared to the classical Newmark scheme. Figure 5.6(b) shows the contact stresses at an arbitrarily chosen time step for the one-body contact problem. Oscillations can be observed, where the contact stresses are large.

Contact-stabilized Newmark scheme (Algorithm 5.2.2)

In Figure 5.3(a) (red line with circles) the dissipative behavior (5.35) of the contact-stabilized Newmark scheme of Deuflhard et al. can be seen. The system looses more energy during the process of detaching than during impact. The loss of energy is around 0.4%. In Section 5.2.2, we explained that due to the predictor step the zigzagging of the relative velocities during contact is removed. The relative velocity remains the same during contact. But after detaching there are some oscillations. This can be seen in Figure 5.5(a) (red line with circles). As expected from the name of the method the contact forces do not show any oscillations (Figure 5.2(a), red line with circles).
The smooth behavior of the contact stresses is shown in Figure 5.6(c) for the numerical example of the one-body contact problem.

Algorithm 5.2.3, enforcing a discrete persistency condition

Algorithm 5.2.3 of Laursen and Chawla is the contact-implicit Newmark scheme with the discrete persistency condition as constraint instead of the non-penetration condition. Therefore, the energy is conserved (blue line with diamonds in Figure 5.3(a)). The trajectory of the relative displacement in Figure 5.4(b) (cyan colored line with circles) reminds of the contact-implicit and contact-stabilized Newmark scheme, but the relative displacements evoke penetration of the two bodies as the non-penetration condition is not enforced. The straight red line indicates the initial gap at the chosen node. The behavior of the relative velocity, Figure 5.5(b) (red line with circles) also reminds of the contact-implicit Newmark scheme but the amplitudes are larger. In Algorithm 5.2.3 the constraints $[\boldsymbol{u}^{n+1}] \cdot \boldsymbol{\nu} \leq [\boldsymbol{u}^n] \cdot \boldsymbol{\nu}$ have to be enforced when contact or penetration has occurred in the previous time step. Therefore, $[\boldsymbol{u}^{n+1}] \cdot \boldsymbol{\nu} = [\boldsymbol{u}^n] \cdot \boldsymbol{\nu}$ whenever the constraint in time step $n+1$ is active. Substituting this in (5.29) leads to $[\dot{\boldsymbol{u}}^{n+1}] \cdot \boldsymbol{\nu} = -[\dot{\boldsymbol{u}}^n] \cdot \boldsymbol{\nu}$ which means that the amplitude of the oscillation corresponds to the amount of the relative velocity before contact occurred. The temporal evolution of the contact forces (Figure 5.2(a), blue line with diamonds) and contact stresses (Figure 5.6(d)) are worse in the sense of more oscillations and larger amplitudes than for the contact-implicit Newmark scheme but still better than for the classical Newmark scheme.

Algorithms using a redistribution of the mass matrix

In this subsection, we deal with three different algorithms using the idea of the redistributed mass, as firstly proposed in [KLR06]. Therein Khenous et al. use a Newmark scheme with the parameters $\beta = \gamma = 1/2$, see Algorithm 5.2.4. However, Algorithm 5.2.5 of Hager et al. is the contact-implicit Newmark scheme with a redistributed mass and combined with Algorithm 5.2.3, which means that in the unconstrained case the algorithm corresponds to the classical Newmark scheme with the parameters $2\beta = \gamma = 1/2$ as in all the other presented methods. First, we investigate Algorithm 5.2.4, second, the contact-implicit Newmark scheme with redistribution of the mass, which is Algorithm 5.2.5 with the non-penetration condition instead of the persistency condition and finally, Algorithm 5.2.5 with the persistency condition.

For the computation of the redistribution of the mass matrix such that the whole mass of the system and the center of gravity are conserved we use an Uzawa method.

In all the three subsequent algorithms we consider two different ways to compute the kinetic energy which is part of the whole energy. Usually, the kinetic energy is computed as $\mathcal{E}_{\text{kin}} = \frac{1}{2}\dot{\boldsymbol{u}}\boldsymbol{M}\dot{\boldsymbol{u}}$, where $\boldsymbol{M}$ is the standard mass matrix. However, in [HHW08] the kinetic energy was computed with the help of the redistributed mass matrix $\tilde{\boldsymbol{M}}$, i.e., $\tilde{\mathcal{E}}_{\text{kin}} := \frac{1}{2}\dot{\boldsymbol{u}}\tilde{\boldsymbol{M}}\dot{\boldsymbol{u}}$, which means that the velocities at the potential contact boundary are simply ignored, whereas the strains are taken into account for the elastic energy, because the stiffness matrix corresponds to the standard mass matrix $\boldsymbol{M}$. As long as a node remains in contact, the displacements do not change and therefore, the difference quotient $\frac{1}{\tau}\left([\boldsymbol{u}^{n+1} - \boldsymbol{u}^n] \cdot \boldsymbol{\nu}\right)$ is zero, but the algorithms provide non-zero velocities, see Figure 5.4(b). Even at nodes which are not actually in contact the velocities are ignored.

The whole energy corresponding to $\tilde{\mathcal{E}}_{\text{kin}}$ is denoted by $\tilde{\mathcal{E}}$.

Algorithm 5.2.4:
The whole energy $\tilde{\mathcal{E}}$ relative to the initial amount is shown in Figure 5.3(c) by the cyan colored line. Although there is no proof of dissipativity for this algorithm, the energy always decreases in this numerical example. The relative loss of energy amounts to 11.5%, which is about 30 times of the relative loss of energy caused by the contact-stabilized Newmark scheme. The course of the energy $\mathcal{E}$ relative to the initial amount can be seen in Figure 5.3(b) by the cyan colored line. The loss of energy amounts to nearly 9% of the initial amount. It is not always dissipative. The behavior of the relative velocity (Figure 5.5(b), magenta colored line with squares) is very similar to the contact-implicit Newmark scheme. The evolution of the contact forces and contact stresses, respectively, is smooth, which can be seen in Figure 5.2(b) (cyan colored line) and Figure 5.6(e).

Contact-implicit Newmark scheme with redistributed mass:
The energy $\tilde{\mathcal{E}}$ related to the initial value can be seen in Figure 5.3(c) (green line with dots). The loss of energy is up to 0.12%. In contrast, the energy $\mathcal{E}$ increases heavily up to 15%, see Figure 5.3(b) (green line with dots). To check if the redistribution of the mass to the interior is responsible or not for this energy blow-up we make another experiment replacing $\tilde{\boldsymbol{M}}$ in the algorithm by the standard mass matrix with zero entries at the contact boundary but without redistribution of the mass in the interior. However, the behavior in the course of energy is similar, see Figure 5.3(b) (red line with circles). The green lines with dots in Figures 5.4(b) and 5.5(b) show the trajectories of the relative displacements and velocities at a fixed node. After detaching the amplitude of the oscillations in the relative velocity at the contact boundary is very large. Similar to the foregoing algorithm the contact forces (green line with dots in Figure 5.2(b)), and contact stresses, Figure 5.6(f), are smooth.

Algorithm 5.2.5:
As the persistency condition is enforced in Algorithm 5.2.5, the energy is conserved if the kinetic energy is computed with the redistributed mass matrix, see blue line with diamonds in Figure 5.3(c). Energy conservation does not hold if we use the standard mass matrix for the computation of the kinetic energy, see blue line with diamonds in Figure 5.3(b). The energy blow-up amounts up to 10%. Just as in the algorithm of Laursen and Chawla the displacements evoke penetration (blue line with diamonds in Figure 5.4(b)). The red straight line is the initial distance between the two bodies at the observed node. During contact the amplitude of the oscillations of the relative velocity (blue line with diamonds 5.5(b)) at the contact boundary is the same as in Algorithm 5.2.3 but after detaching the amplitude tops any of the other algorithm as the value of the relative velocity in normal direction changes between -18 and 22. The total contact forces, blue line with diamonds in Figure 5.2(b), are smooth, but if we have a closer look at the contact stresses for the one-body contact problem the stresses show slight oscillations, Figure 5.6(g).

Conclusion

Apart from the classical Newmark scheme all the algorithms show a similar behavior of the relative displacements at contact nodes. But in Algorithms 5.2.3 and 5.2.5 the non-penetration condition is violated. Using the classical Newmark scheme we observe some kinks in the process of detaching.

In all the algorithms the relative velocities at the contact boundary show oscillations at some time. The contact-stabilized Newmark scheme is the only one which does not evoke oscillations in the relative velocities during the actual contact phase. Once a node is in contact the relative velocity remains the same. After detaching there are a few oscillations as in the contact-implicit Newmark scheme, Algorithm 5.2.3 and Algorithm 5.2.4. The other algorithms show more oscillations with larger amplitudes after detaching, especially Algorithm 5.2.5. During contact the two algorithms enforcing the discrete persistency condition instead of the non-penetration condition evoke oscillations in the relative velocities, which have an amplitude of the amount of the relative velocities before contact occurred. This numerical observation can be easily verified by reformulating the velocity update $\dot{\boldsymbol{u}}^{n+1} = -\dot{\boldsymbol{u}}^n + \frac{2}{\tau}(\boldsymbol{u}^{n+1} - \boldsymbol{u}^n)$ and exploiting $[\boldsymbol{u}^{n+1}] \cdot \boldsymbol{\nu} = [\boldsymbol{u}^n] \cdot \boldsymbol{\nu}$ whenever constraints are active. For all other algorithms the amplitude of the relative velocity is less or equal to the amount of the relative velocity before contact occurred. Unfortunately, none of the investigated algorithms shows the desirable behavior of the relative velocities, which should be zero in the case of contact to fulfill the persistency condition.

The behavior of the velocities influences the energy. The energy of the classical Newmark scheme oscillates and increases up to 3.5% in our example, whereas the contact-implicit and contact-stabilized Newmark schemes are provably dissipative. The relative loss of energy amounts to 0.16% and 0.4%. In Algorithm 5.2.3 the energy is conserved because the discrete persistency condition is enforced. The same holds for Algorithm 5.2.5 if the kinetic energy is computed with the redistributed mass matrix. Depending on whether the kinetic energy is computed with the standard or the redistributed mass matrix the energy may increase or decrease in the different algorithms employing a redistribution of the mass matrix.

The total contact forces show a smooth behavior for the contact-stabilized Newmark scheme and all variants of algorithms using a redistribution of the mass matrix (see Figure 5.2(b)). The contact stresses reflect this stable behavior, except for Algorithm 5.2.5, where the persistency condition is enforced. There are slight oscillations in the contact stresses. Using the contact-implicit Newmark scheme we observe some oscillations in the contact stresses and forces. In Algorithm 5.2.3 there are even more oscillations and of course the classical Newmark scheme shows so much oscillations with such large amplitudes that the real behavior can be recognized hardly.

5.3 A family of space-time connecting discretization schemes with local impact detection

As explained in the foregoing section the performance of classical time discretization schemes based on Taylor expansion like the classical Newmark scheme is disturbed seriously by the non-smoothness of the displacements and velocities in time at the actual contact boundary.

On impact, further displacement is abruptly avoided and the velocity should be zero, if the point is not detaching immediately. In the time-continuous setting the zero value of the velocity is a consequence of the constant value in time of the displacements during contact, i.e., $u_\nu = g$. Unfortunately, a classical time discretization scheme is not able to reproduce this property. Therefore, both constraints, the non-penetration condition and the persistency condition (5.19), meaning that either the contact stresses or the velocities at the contact boundary are zero, have to be enforced separately.

For an adaptation of classical time discretization schemes to dynamic contact problems it would be desirable to localize the moment of impact where regularity cannot be expected. For a one-dimensional object it would be possible to detect the moment of impact and in consequence to choose a suitable time step. But for an arbitrary two or three-dimensional object we have an infinite number of points coming into contact at individual times. Although in the discrete setting, the number of contact nodes is finite, still each contact node has its own impact time and the actual contact zone is part of the solution process and not known a priori.

To illustrate this difficulty, let us consider one time step $[t, t+\tau]$. Within this time step, different nodes $\boldsymbol{p}_1, \boldsymbol{p}_2, \dots$ on the contact boundary might become active (i.e. in contact) or inactive at different sub-times $t \leq t+\tau_1 \leq \cdots \leq \ldots \leq t+\tau$. On the one hand, a subdivision of the time intervall $[t, t+\tau]$ into $[t, t+\tau_1], [t+\tau_1, t+\tau_2], \dots$ is out of question since the number of nodes on the contact boundary might become arbitrarily large due to refinement in space. On the other hand, each single impact of a boundary node gives rise to a non-smooth event, i.e., an abrupt change of the velocity, which has to be dealt with carefully.

A choice of very small time step sizes may reduce the error caused by the time discretization but is numerically very expensive as for each time step the whole non-linear system has to be solved. This disadvantage might be significantly reduced by an adaptive time step control as, e.g., in [KSD10], but close to an impact still many time steps have to be used if a standard time discretization is employed.

Therefore, we follow the lines of our article [KW11] to present our new space-time connecting discretization scheme which due to a predictor step allows for the implicit computation of individual impact times for each node on the contact interface. These information from the space discrete setting incorporated in the time discretization provide reasonable approximations to the displacements, the contact stresses and the velocities at the contact boundary. The displacements fulfill the non-penetration condition, the contact stresses are stable as in [DKE08] and the computation of the velocities is motivated by the persistency condition.

This space-time connecting discretization scheme as well as other methods proposed for contact problems are elements of a family of modified Newmark schemes depending on matrix-valued algorithmic parameters $\boldsymbol{\alpha}, \boldsymbol{\beta}$. In our analysis we show for which choice of parameters the resulting algorithms are dissipative and we discuss the influence of the choice of parameters on the stability of the contact stresses as well as on the behavior of the velocities. Consequently, we provide a unifying theory for some contact-related and dissipative modifications of the classical Newmark scheme.

5.3.1 Time discretization depending on local impact times

For the derivation of the algorithm we consider a one-body contact problem discretized in space. In the following, the discrete quantities for displacement, velocity, acceleration and force have to be understood as vectors of the dimension $\#N_{\mathfrak{m}} \cdot d$. The matrices $\boldsymbol{M}$ and $\boldsymbol{A}$ are the mass matrix and the stiffness matrix representing $(\cdot,\cdot)$ and $a(\cdot,\cdot)$ in the basis of linear finite elements and the chosen Cartesian coordinate systems. We assume the mass matrix to be lumped.

In Section 5.2.2 we presented the contact-stabilized Newmark scheme [DKE08] which enables a prediction of the actual contact zone. We may use this idea to compute individual impact time step sizes $\tau^*(\boldsymbol{p}) \leq \tau$ which gives us useful information for corrections of the contact-velocities. In the fully discrete setting the predictor step (5.32) in Algorithm 5.2.2 is

$$\boldsymbol{M}\boldsymbol{u}_{\mathfrak{m},\mathrm{pred}}^{n+1} = \boldsymbol{M}\boldsymbol{u}_{\mathfrak{m}}^{n} + \tau \boldsymbol{M}\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n} + \frac{\tau^2}{2}\boldsymbol{M}\boldsymbol{P}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathfrak{m},\mathrm{pred}}^{n+1}\right). \tag{5.41}$$

Here and in the following, we set $\boldsymbol{e}_1(\boldsymbol{p}) := \boldsymbol{\nu}(\boldsymbol{p})$. As $\boldsymbol{M}$ is a diagonal matrix the effects of the constraints are restricted to the contact boundary in the predictor step. For each contact node $\boldsymbol{p} \in N_{\mathfrak{m}}^{C}$ and the corresponding direction of constraint $\boldsymbol{e}_1(\boldsymbol{p})$ exists an impact time step size $\tau^*(\boldsymbol{p})$, fulfilling

$$\boldsymbol{u}_{\mathfrak{m},\mathrm{pred}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \boldsymbol{u}_{\mathfrak{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) + \tau^*(\boldsymbol{p})\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}).$$

The relation between the impact time step size $\tau^*(\boldsymbol{p})$ and the linear residual corresponding to the discrete variational inequality of the predictor step is given by

$$\frac{\tau^2}{2}\left(\boldsymbol{M}\boldsymbol{P}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathfrak{m},\mathrm{pred}}^{n+1}\right)\right)(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = (\tau^*(\boldsymbol{p}) - \tau)\left(\left(\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n}\right)(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})\right) \leq 0 \tag{5.42}$$

for each node $\boldsymbol{p} \in N_{\mathfrak{m}}^{C}$. We define a $(\#N_{\mathfrak{m}} \cdot d) \times (\#N_{\mathfrak{m}} \cdot d)$ diagonal matrix $\boldsymbol{\tau}^*$ with $d \times d$ block diagonal matrices such that

$$\boldsymbol{e}_i(\boldsymbol{p}) \cdot \boldsymbol{\tau}^* \cdot \boldsymbol{e}_i(\boldsymbol{p}) = \begin{cases} \tau^*(\boldsymbol{p}), & \text{if} \quad \boldsymbol{p} \in N_{\mathfrak{m}}^{C} \quad \text{and} \quad i = 1 \\ \tau, & \text{otherwise} \end{cases}$$

and

$$\boldsymbol{\tau} = \tau\, \boldsymbol{id}.$$

With this notation we reformulate equation (5.41) to

$$\boldsymbol{M}\boldsymbol{u}_{\mathbb{m},\mathrm{pred}}^{n+1} = \boldsymbol{M}\boldsymbol{u}_{\mathbb{m}}^{n} + \tau\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} + \boldsymbol{M}\left(\boldsymbol{\tau^*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} = \boldsymbol{M}\boldsymbol{u}_{\mathbb{m}}^{n} + \boldsymbol{\tau}^*\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}.$$

We define a family of modified Newmark schemes which makes use of the local impact time step sizes $\tau^*(\boldsymbol{p})$. The family depends on additional $(\#N_{\mathbb{m}} \cdot d) \times (\#N_{\mathbb{m}} \cdot d)$ diagonal matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ where the entries of the $d \times d$ diagonal matrices $\boldsymbol{\alpha}(\boldsymbol{p})$ are given by

$$\boldsymbol{e}_i(\boldsymbol{p}) \cdot \boldsymbol{\alpha}(\boldsymbol{p}) \cdot \boldsymbol{e}_i(\boldsymbol{p}) = \begin{cases} \alpha(\boldsymbol{p}), & \text{if} \quad \boldsymbol{p} \in N_{\mathbb{m}}^C \quad \text{and} \quad i = 1 \\ 0, & \text{otherwise.} \end{cases}$$

The choices for $\boldsymbol{\alpha}(\boldsymbol{p})$ are discussed in the following and it will turn out that good choices for $\boldsymbol{\beta}$ depend on $\boldsymbol{\alpha}$.

Algorithmus 5.3.1. *Family of modified Newmark schemes*

$$\boldsymbol{M}\boldsymbol{u}_{\mathbb{m},\mathrm{pred}}^{n+1} = \boldsymbol{M}\boldsymbol{u}_{\mathbb{m}}^{n} + \tau\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} + \boldsymbol{M}\left(\boldsymbol{\tau^*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} \tag{5.43}$$

$$\begin{aligned} \boldsymbol{M}\boldsymbol{u}_{\mathbb{m}}^{n+1} =& \boldsymbol{M}\boldsymbol{u}_{\mathbb{m}}^{n} + \tau\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} + \boldsymbol{\beta}\boldsymbol{M}\left(\boldsymbol{\tau}^* - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} \\ &- \frac{\tau^2}{2}\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathbb{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}\right) - \boldsymbol{\alpha}\frac{2}{\tau^2}\boldsymbol{M}\left(\boldsymbol{\tau}^* - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}\right) \end{aligned} \tag{5.44}$$

$$\begin{aligned} \boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1} =& \boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} \\ &- \tau\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathbb{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}\right) - \boldsymbol{\alpha}\frac{2}{\tau^2}\boldsymbol{M}\left(\boldsymbol{\tau}^* - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}\right) \end{aligned} \tag{5.45}$$

We remind the reader that the term $-\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}\right)$ is a priori unknown and computed as part of the solution process of the variational inequality corresponding to (5.44). Analogously, the term $-\frac{\tau^2}{2}\boldsymbol{M}\boldsymbol{P}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathbb{m},\mathrm{pred}}^{n+1}\right) = -\boldsymbol{M}\left(\boldsymbol{\tau^*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}$ is a priori unknown and computed as part of the solution process of the variational inequality corresponding to (5.43). We note that despite the lumping of the mass matrix the non-penetration condition influences the displacements in the whole body due to the elastic reaction in (5.44).

We derive suitable choices for the parameters $\alpha(\boldsymbol{p})$ and $\beta(\boldsymbol{p})$. We first investigate the dependence of the course of energy on the parameters $\alpha(\boldsymbol{p})$ and $\beta(\boldsymbol{p})$ as well as the stability of the contact stresses, which lead to some conditions on the parameters. Finally, the precise determination of the parameters is motivated by the persistency condition (5.19) which says that the velocities at the contact boundary should be zero during a contact phase.

It is well-known that up to now, no time-discretization for contact-problems exists which conserves the energy while enforcing the non-penetration condition without numerical instabilities showing up in the contact stresses. Therefore, we analyze the family of modified Newmark schemes under the aspect of dissipativity. We compute the difference of energy between two successive time steps in the following proposition.

Proposition 5.3.1. *The difference in energy $\mathcal{E}(\boldsymbol{u})$ between two successive time steps is*

$$\begin{aligned}
\mathcal{E}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)-\mathcal{E}\left(\boldsymbol{u}_{\mathrm{m}}^{n}\right) = {} & \left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)+\boldsymbol{\alpha}\frac{2}{\tau^{2}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\\
&+\boldsymbol{\beta}\frac{2}{\tau^{2}}\left(\boldsymbol{u}_{\mathrm{m,pred}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)\cdot\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\\
&+\boldsymbol{\beta}\left(\boldsymbol{\beta}-\boldsymbol{id}\right)\frac{2}{\tau^{2}}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\cdot\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}.
\end{aligned}$$

Proof. We need the following reformulations derived from Algorithm 5.3.1

$$\begin{aligned}
\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}-\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n} &= -\tau\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n}+\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)-\boldsymbol{\alpha}\frac{2}{\tau^{2}}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\\
&= \frac{2}{\tau}\left(\boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n}-\tau\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}-\boldsymbol{\beta}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\\
\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}+\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n} &= \frac{2}{\tau}\left(\boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n}-\boldsymbol{\beta}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)
\end{aligned}$$

to compute the difference in energy between two successive time steps

$$\begin{aligned}
&\mathcal{E}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)-\mathcal{E}\left(\boldsymbol{u}_{\mathrm{m}}^{n}\right)\\
&=\frac{1}{2}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}\cdot\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}-\frac{1}{2}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\cdot\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}+\frac{1}{2}\boldsymbol{u}_{\mathrm{m}}^{n+1}\cdot\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1}-\frac{1}{2}\boldsymbol{u}_{\mathrm{m}}^{n}\cdot\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n}\\
&=\frac{1}{2}\left(\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}+\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{M}\left(\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}-\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)+\frac{1}{2}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}+\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{A}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\\
&=-\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}-\boldsymbol{\beta}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\cdot\\
&\quad\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n}+\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)-\boldsymbol{\alpha}\frac{2}{\tau^{2}}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\\
&\quad+\ \frac{1}{2}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{A}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}+\boldsymbol{u}_{\mathrm{m}}^{n}\right)\\
&=\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)+\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{\alpha}\frac{2}{\tau^{2}}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\\
&\quad-\boldsymbol{\beta}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\cdot\frac{2}{\tau^{2}}\boldsymbol{M}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}-\tau\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}-\boldsymbol{\beta}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\\
&=\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)+\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}\right)\cdot\boldsymbol{\alpha}\frac{2}{\tau^{2}}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\\
&\quad-\boldsymbol{\beta}\frac{2}{\tau^{2}}\boldsymbol{M}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\cdot\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}-\boldsymbol{u}_{\mathrm{m}}^{n}-\tau\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}-\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)\\
&\quad+\boldsymbol{\beta}\frac{2}{\tau^{2}}\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\cdot\boldsymbol{M}\left(\boldsymbol{\beta}-\boldsymbol{id}\right)\left(\boldsymbol{\tau}^{*}-\boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{\left(\boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{u}_{\mathbb{m}}^{n}\right) \cdot \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}\right)}_{\langle 1 \rangle} + \underbrace{\boldsymbol{\alpha}\frac{2}{\tau^2}\left(\boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{u}_{\mathbb{m}}^{n}\right) \cdot \boldsymbol{M}\left(\boldsymbol{\tau}^{*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}}_{\langle 2 \rangle} \\
&\quad + \underbrace{\boldsymbol{\beta}\frac{2}{\tau^2}\left(\boldsymbol{u}_{\mathbb{m},\mathrm{pred}}^{n+1} - \boldsymbol{u}_{\mathbb{m}}^{n+1}\right) \cdot \boldsymbol{M}\left(\boldsymbol{\tau}^{*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}}_{\langle 3 \rangle} \\
&\quad + \underbrace{\boldsymbol{\beta}\left(\boldsymbol{\beta} - \boldsymbol{id}\right)\frac{2}{\tau^2}\left(\boldsymbol{\tau}^{*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} \cdot \boldsymbol{M}\left(\boldsymbol{\tau}^{*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}}_{\langle 4 \rangle}\,.
\end{aligned}
$$

qed.

The variational inequality corresponding to (5.44) implies the sign condition $\left(\boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{v}_{\mathbb{m}}\right) \cdot \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}\right) \leq 0$ and therefore, $\langle 1 \rangle \leq 0$. In the same way the sign condition $\left(\boldsymbol{u}_{\mathbb{m},\mathrm{pred}}^{n+1} - \boldsymbol{v}_{\mathbb{m}}\right) \cdot \boldsymbol{M}\left(\boldsymbol{\tau}^{*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n} \leq 0$ follows from (5.43), so that $\langle 3 \rangle \leq 0$ for $\boldsymbol{\beta} \geq 0$. As long as $0 \leq \beta(\boldsymbol{p}) \leq 1$ for all $\boldsymbol{p} \in N_{\mathbb{m}}^{C}$, $\langle 4 \rangle \leq 0$ because $\boldsymbol{M}$ is a positive definite matrix. Only $\langle 2 \rangle$ cannot be treated in such a straightforward way and requires more consideration. It follows from (5.42) that $\left(\boldsymbol{M}\left(\boldsymbol{\tau}^{*} - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathbb{m}}^{n}\right)(\boldsymbol{p}) \leq 0$ for all $\boldsymbol{p} \in N_{\mathbb{m}}^{C}$. Thus, $\alpha(\boldsymbol{p})$ has to be chosen such that $\alpha(\boldsymbol{p})\left(\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) - \boldsymbol{u}_{\mathbb{m}}^{n}(\boldsymbol{p})\right) \cdot \boldsymbol{e}_1(\boldsymbol{p})\right) \geq 0$. We distinguish between actual contact nodes, i.e., $\boldsymbol{u}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = g_{\mathbb{m}}(\boldsymbol{p})$ which are either already in contact or where impact occurs, i.e., $\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) - \boldsymbol{u}_{\mathbb{m}}^{n}(\boldsymbol{p})\right) \cdot \boldsymbol{e}_1(\boldsymbol{p}) \geq 0$ and those nodes which are not in contact, i.e., $\boldsymbol{u}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) \neq g_{\mathbb{m}}(\boldsymbol{p})$. We recall that we like to choose $\boldsymbol{\alpha}$ such that the velocities at actual contact boundary nodes are zero, i.e., $\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$ as motivated by the persistency condition. Therefore, if we have no actual contact $\boldsymbol{u}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) \neq g_{\mathbb{m}}(\boldsymbol{p})$ we choose $\alpha(\boldsymbol{p}) = 0$ so that the velocity is not changed. If we have an actual contact node, i.e., $\left(\boldsymbol{u}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) - \boldsymbol{u}_{\mathbb{m}}^{n}(\boldsymbol{p})\right) \cdot \boldsymbol{e}_1(\boldsymbol{p}) \geq 0$ we choose $\alpha(\boldsymbol{p}) \geq 0$. We conclude that the aim to avoid energy blow-ups restricts the choice of the parameters $\alpha(\boldsymbol{p})$ and $\beta(\boldsymbol{p})$. We have to choose $0 \leq \beta(\boldsymbol{p}) \leq 1$ and $0 \leq \alpha(\boldsymbol{p})$ for an actual contact node and $\alpha(\boldsymbol{p}) = 0$ for a node which is not in contact.

Apart from the energy, the linear momentum and the angular momentum should be constants of the physical system. Due to the fact that we use a linearized non-penetration condition we cannot expect conservation of angular momentum, compare Section 5.1.2. But the linear momentum $\boldsymbol{\mathcal{L}}\left(\boldsymbol{u}_{\mathbb{m}}\right) \cdot \boldsymbol{e}_i := \boldsymbol{M}\dot{\boldsymbol{u}}_{\mathbb{m}} \cdot \boldsymbol{e}_i$ for all $i = 1, \ldots, d$ is conserved even for the classical Newmark scheme in the case of contact. We verify that this result also holds for the family of modified Newmark schemes independent of the choice of parameters $\alpha(\boldsymbol{p})$ and $\beta(\boldsymbol{p})$. Therefore, we have to assume a two-body contact problem because the one-body contact problem where a body comes into contact with a rigid foundation is an idealized situation

$$
\begin{aligned}
&\boldsymbol{e}_i \cdot \left(\boldsymbol{\mathcal{L}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right) - \boldsymbol{\mathcal{L}}\left(\boldsymbol{u}_{\mathrm{m}}^{n}\right)\right) \\
&= -\tau \left(\underbrace{\frac{1}{2}\boldsymbol{e}_i \cdot \boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n}}_{=0} + \underbrace{\boldsymbol{e}_i \cdot \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1}}_{=0} \right. \\
&\left. - \underbrace{\boldsymbol{e}_i \cdot \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)}_{=0} - \underbrace{\boldsymbol{\alpha}\frac{2}{\tau^2}\boldsymbol{e}_i \cdot \boldsymbol{M}\left(\boldsymbol{\tau}^* - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}}_{=0} \right) = 0 \quad \forall i = 1, \dots, d.
\end{aligned}
\tag{5.46}
$$

The expressions $\boldsymbol{e}_i \cdot \boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}$ are zero because $\boldsymbol{e}_i$ is constant. In the case of two-body contact the contact force density on both contacting zones have the same value but opposite sign such that they sum up to zero. As $\frac{\tau^2}{2}\boldsymbol{P}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m,pred}}^{n+1}\right)(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \left(\tau^*(\boldsymbol{p}) - \tau\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ is also a constraining force density which is equal on both contacting sides the last expression in (5.46) sums up to zero, too.

In the contact-stabilized Newmark scheme the predictor is used in the computation of the solution

$$
\boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n+1} = \boldsymbol{M}\boldsymbol{u}_{\mathrm{m,pred}}^{n+1} - \frac{\tau^2}{2}\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1} - \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)\right). \tag{5.47}
$$

Hence, the contact forces $\boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)$ solely equilibrate the elastic reaction $\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1}$ which is evoked by the non-penetration condition, compare the discussion in Section 5.2.2. To get stable contact stresses like in the contact-stabilized Newmark scheme we choose $\alpha(\boldsymbol{p}) = 1 - \beta(\boldsymbol{p})$ such that equation (5.47) and (5.44) are the same.

Therefore, it suffices to choose a set of parameters $\alpha(\boldsymbol{p})$ with $0 \leq \alpha(\boldsymbol{p}) \leq 1$ which is zero if $\boldsymbol{p}$ is not an actual contact node. We derive a choice of $\alpha(\boldsymbol{p})$ for an actual contact node $\boldsymbol{p}$, motivated by the idea to improve the velocities such that the velocity at an actual contact node is zero as expected from the persistency condition. We consider only nodes which are active in the predictor step because otherwise $\alpha(\boldsymbol{p})\left(\boldsymbol{\tau}^* - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$ independent of the choice of $\alpha(\boldsymbol{p})$. Therefore, $\boldsymbol{u}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \boldsymbol{u}_{\mathrm{m,pred}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = g_{\mathrm{m}}(\boldsymbol{p})$ which implies that force equilibrium $\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1} - \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)\right)(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$ holds and the velocity update (5.45) reduces to

$$
\begin{aligned}
\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) &= \dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) + \frac{2}{\tau}\alpha(\boldsymbol{p})\left(\tau^*(\boldsymbol{p}) - \tau\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) \\
&= \left(1 + \alpha(\boldsymbol{p})\frac{2}{\tau}\left(\tau^*(\boldsymbol{p}) - \tau\right)\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}).
\end{aligned}
\tag{5.48}
$$

Therefrom we get $\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$ if we choose $\alpha(\boldsymbol{p}) = \frac{\tau}{2(\tau - \tau^*(\boldsymbol{p}))}$. For $\tau^*(\boldsymbol{p}) \leq \frac{\tau}{2}$ we have $\alpha(\boldsymbol{p}) \leq 1$ which is necessary for the dissipativity of the algorithm. If $\tau^*(\boldsymbol{p}) \geq \frac{\tau}{2}$ we keep $\alpha(\boldsymbol{p}) = 1$ which we get for $\tau^*(\boldsymbol{p}) = \frac{\tau}{2}$ from $\alpha(\boldsymbol{p}) = \frac{\tau}{2(\tau - \tau^*(\boldsymbol{p}))}$. This choice of $\alpha(\boldsymbol{p})$ for $\tau^*(\boldsymbol{p}) > \frac{\tau}{2}$ does not enforce the velocity to be zero, i.e., $\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$, but

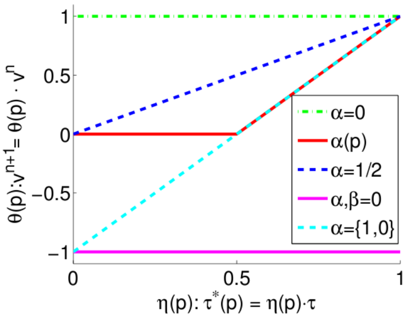


Figure 5.7: $\theta(\boldsymbol{p}) : \dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \theta(\boldsymbol{p})\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ is plotted against $\eta(\boldsymbol{p}) : \tau^*(\boldsymbol{p}) = \eta(\boldsymbol{p})\tau$

we get the maximum reduction of the absolute value of the velocity. We note that if a node $\boldsymbol{p}$ is already in contact, i.e., $\boldsymbol{u}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = g_{\mathrm{m}}(\boldsymbol{p})$, it follows $\tau^*(\boldsymbol{p}) = 0 \leq \frac{\tau}{2}$ such that $\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$ which means that the velocities are zero at all nodes which are in contact in the actual and the foregoing time step. In Figure 5.7 (red line) the factor $\theta(\boldsymbol{p})$ fulfilling $\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \theta(\boldsymbol{p})\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ is plotted against $\eta(\boldsymbol{p})$ fulfilling $\tau^*(\boldsymbol{p}) = \eta(\boldsymbol{p})\tau$.

We call Algorithm 5.3.1 with the above described choice of $\alpha(\boldsymbol{p})$ and $\beta(\boldsymbol{p})$ improved contact-stabilized Newmark scheme.

Algorithmus 5.3.2. *Improved contact-stabilized Newmark scheme*

$$\begin{aligned}
\boldsymbol{M}\boldsymbol{u}_{\mathrm{m,pred}}^{n+1} &= \boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n} + \boldsymbol{\tau}^*\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\\
\boldsymbol{M}\boldsymbol{u}_{\mathrm{m}}^{n+1} &= \boldsymbol{M}\boldsymbol{u}_{\mathrm{m,pred}}^{n+1} - \frac{\tau^2}{2}\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1} - \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right)\right)\\
\boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n+1} &= \boldsymbol{M}\dot{\boldsymbol{u}}_{\mathrm{m}}^{n} - \tau\left(\frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n} + \frac{1}{2}\boldsymbol{A}\boldsymbol{u}_{\mathrm{m}}^{n+1} - \boldsymbol{M}\boldsymbol{F}_{\mathrm{con}}\left(\boldsymbol{u}_{\mathrm{m}}^{n+1}\right) - \boldsymbol{\alpha}\frac{2}{\tau^2}\boldsymbol{M}\left(\boldsymbol{\tau}^* - \boldsymbol{\tau}\right)\dot{\boldsymbol{u}}_{\mathrm{m}}^{n}\right)
\end{aligned}$$

In the following, we present other choices of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ which arise from the foregoing discussions. If possible we relate the resulting algorithms to existing methods proposed for dynamic contact problems in the literature. For the simplest choice $\boldsymbol{\alpha} = \boldsymbol{0}$ and $\boldsymbol{\beta} = \boldsymbol{0}$, where no parameter dependence is given, the ansatz reduces to the contact-implicit Newmark scheme as used in [KROM99]. This method is dissipative but gives rise to artificial contact stresses because $\alpha(\boldsymbol{p}) \neq (1-\beta(\boldsymbol{p}))$. The contact stresses oscillate during a contact phase because the sign of the velocity changes in each time step if $\boldsymbol{u}_{\mathrm{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \boldsymbol{u}_{\mathrm{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ as explained in Section 5.2.1. The magenta colored line in Figure 5.7 illustrates this effect.

The contact-stabilized Newmark scheme is recovered by the choice $\boldsymbol{\alpha} = \boldsymbol{0}$ and $\boldsymbol{\beta} = \boldsymbol{id}$. This method is energy dissipative and the velocity remains the same during a contact phase $\dot{\boldsymbol{u}}^{n}(\boldsymbol{p})\cdot\boldsymbol{e}_1(\boldsymbol{p}) = \dot{\boldsymbol{u}}^{n+1}(\boldsymbol{p})\cdot\boldsymbol{e}_1(\boldsymbol{p})$, see Figure 5.7 (green line). This is closely connected to the stability of the contact stresses and therefore very advantageous but it is not the behavior we would expect from the persistency condition.

In the improved contact-stabilized Newmark scheme we choose $\alpha(\boldsymbol{p}) = 1$ and $\beta(\boldsymbol{p}) = (1 - \alpha(\boldsymbol{p})) = 0$ if $\tau^*(\boldsymbol{p}) \geq \frac{\tau}{2}$. Therefore, the question arises what happens if we take the same choice $\alpha(\boldsymbol{p}) = 1$ for each actual contact node with $\boldsymbol{u}_{\mathfrak{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = g_{\mathfrak{m}}(\boldsymbol{p})$ without regarding $\tau^*(\boldsymbol{p})$. The relation between the old and new velocity (5.48) changes to

$$\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = \left(1 + \frac{2}{\tau}\left(\tau^*(\boldsymbol{p}) - \tau\right)\right) \dot{\boldsymbol{u}}_{\mathfrak{m}}^{n}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}). \tag{5.49}$$

In the case $\tau^*(\boldsymbol{p}) \leq \frac{\tau}{2}$ the velocity changes the sign pointing into the direction of detachment (compare Figure 5.7, cyan colored line) which leads to oscillations in the contact stresses as in the contact-implicit Newmark scheme. In Figure 5.7 as well as in the numerical examples we refer to this algorithm by "$\alpha = \{1, 0\}$", because $\alpha(\boldsymbol{p})$ is either 1 or 0 if the node is in contact or not.

Another obvious choice of $\alpha(\boldsymbol{p})$ is $\frac{1}{2}$ and $\beta(\boldsymbol{p}) = (1 - \alpha(\boldsymbol{p})) = \frac{1}{2}$ for each potential contact node. From the improved contact-stabilized Newmark scheme we know that $\alpha(\boldsymbol{p}) = \frac{1}{2}$ implies $\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) = 0$ if the body is already in contact ($\tau^*(\boldsymbol{p}) = 0$) and $\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}(\boldsymbol{p}) = g_{\mathfrak{m}}(\boldsymbol{p})$. Choosing $\alpha(\boldsymbol{p}) = \frac{1}{2}$ for impact nodes ($\tau^*(\boldsymbol{p}) \neq 0$) reduces the velocity as long as $\boldsymbol{u}_{\mathfrak{m}}^{n+1}(\boldsymbol{p}) = \boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}(\boldsymbol{p}) = g_{\mathfrak{m}}(\boldsymbol{p})$, see Figure 5.7 (dark blue line) but not as much as in the improved contact-stabilized Newmark scheme. No oscillations are provoked as for the choice $\alpha(\boldsymbol{p}) = 1$. Due to $\boldsymbol{\alpha} = \boldsymbol{\beta} = \frac{1}{2}\boldsymbol{id}$ the mass matrix can be employed without lumping to get a dissipative scheme and the expressions $\langle 2 \rangle$ and $\langle 3 \rangle$ in the difference of energy between two successive time steps (see proof of Proposition 5.3.1) sum up to

$$\langle 2 \rangle + \langle 3 \rangle = \boldsymbol{\alpha}\left(\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1} - \boldsymbol{u}_{\mathfrak{m}}^{n}\right) \cdot \boldsymbol{M}\boldsymbol{P}_{\text{con}}\left(\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}\right) \leq 0.$$

The fact, that no difference is made in the choice of $\alpha(\boldsymbol{p})$ for non-contact and contact nodes leads to a dependence of the new velocity $\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n+1}(\boldsymbol{p})$ on the predictor $\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}(\boldsymbol{p})$

$$\dot{\boldsymbol{u}}_{\mathfrak{m}}^{n+1} = \frac{\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1} - \boldsymbol{u}_{\mathfrak{m}}^{n}}{\tau} - \tau\left(\frac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{u}_{\mathfrak{m}}^{n} + \frac{1}{2}\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{u}_{\mathfrak{m}}^{n+1} - \boldsymbol{F}_{\text{con}}\left(\boldsymbol{u}_{\mathfrak{m}}^{n+1}\right)\right) \tag{5.50}$$

even if the predictor is in contact $\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}(\boldsymbol{p}) = g_{\mathfrak{m}}(\boldsymbol{p})$ but not the solution $\boldsymbol{u}_{\mathfrak{m}}^{n+1}(\boldsymbol{p}) \neq g_{\mathfrak{m}}(\boldsymbol{p})$. This dependence seems not to be reasonable. In the numerical results in Section 5.3.3 we see that potential contact nodes $\boldsymbol{p} \in N_{\mathfrak{m}}^{C}$ exist which are active for the predictor $\boldsymbol{u}_{\mathfrak{m},\text{pred}}^{n+1}$ but not for the solution $\boldsymbol{u}_{\mathfrak{m}}^{n+1}$. We like to remark that the velocity update (5.50) has also been used in [Kla11].

5.3.2 Implementation aspects

Before we give numerical results for the time discretization schemes studied in the previous section we like to shortly comment on implementation aspects. If not otherwise stated all numerical experiments have been carried out in the obstacle toolbox

`OBSLIB++` [Kra00] in the framework of the finite element toolbox `UG` [BBL+97]. The variational inequalities (5.43) and (5.44) have been solved with a monotone multigrid method [Kor97, KK01]. We note that the predictor step (5.43) needs only marginal computing time due to the lumped mass matrix.

For the modifications concerning the velocity in the algorithms with $\boldsymbol{\alpha} \neq \mathbf{0}$ we first compute the velocity

$$\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1} = \dot{\boldsymbol{u}}_{\mathbb{m}}^{n} - \tau \left(\frac{1}{2} \boldsymbol{M}^{-1} \boldsymbol{A} \boldsymbol{u}_{\mathbb{m}}^{n} + \frac{1}{2} \boldsymbol{M}^{-1} \boldsymbol{A} \boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{F}_{\text{con}} \left(\boldsymbol{u}_{\mathbb{m}}^{n+1} \right) \right) \tag{5.51}$$

as in the methods with $\boldsymbol{\alpha} = \mathbf{0}$ and in a second step we correct this velocity (5.51) depending on the choice of $\boldsymbol{\alpha}$ as follows:

1. $\alpha(\boldsymbol{p})$ - improved contact-stabilized Newmark scheme

   for all $\boldsymbol{p} \in N_{\mathbb{m}}^{C}$ which are active for the predictor and the solution

   - if $\boldsymbol{u}_{\mathbb{m},\text{pred}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) \leq \left(\boldsymbol{u}_{\mathbb{m}}^{n}(\boldsymbol{p}) + \frac{\tau}{2} \dot{\boldsymbol{u}}_{\mathbb{m}}^{n}(\boldsymbol{p}) \right) \cdot \boldsymbol{e}_1(\boldsymbol{p})$, set the velocity to $\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) := 0$
   - otherwise set $\boldsymbol{r}_{\mathbb{m}}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) := -\frac{\tau^2}{2} \boldsymbol{P}_{\text{con}} \left(\boldsymbol{u}_{\mathbb{m},\text{pred}}^{n+1} \right) (\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$, which is the linear residual corresponding to the variational inequality of the predictor step, and add $-\frac{2}{\tau^2} \boldsymbol{r}_{\mathbb{m}}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ to $\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ computed in (5.51)

2. $\alpha = 1/2$

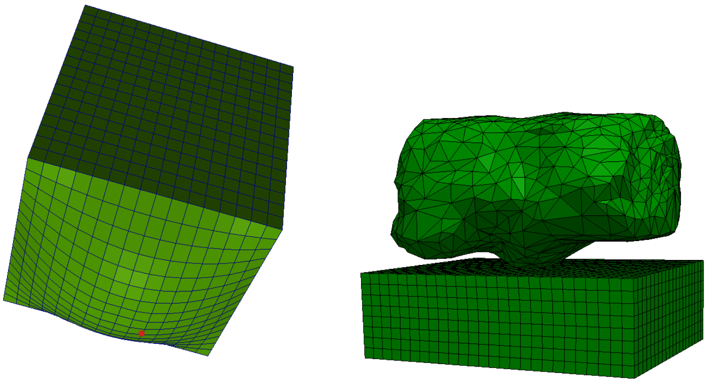
$$\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1} = \frac{\boldsymbol{u}_{\mathbb{m},\text{pred}}^{n+1} - \boldsymbol{u}_{\mathbb{m}}^{n}}{\tau} - \tau \left(\frac{1}{2} \boldsymbol{M}^{-1} \boldsymbol{A} \boldsymbol{u}_{\mathbb{m}}^{n} + \frac{1}{2} \boldsymbol{M}^{-1} \boldsymbol{A} \boldsymbol{u}_{\mathbb{m}}^{n+1} - \boldsymbol{F}_{\text{con}} \left(\boldsymbol{u}_{\mathbb{m}}^{n+1} \right) \right)$$

3. $\alpha = \{1, 0\}$

   for all $\boldsymbol{p} \in N_{\mathbb{m}}^{C}$ which are active for the predictor and the solution

   - set $\boldsymbol{r}_{\mathbb{m}}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p}) := -\frac{\tau^2}{2} \boldsymbol{P}_{\text{con}} \left(\boldsymbol{u}_{\mathbb{m},\text{pred}}^{n+1} \right) (\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$, which is the linear residual corresponding to the variational inequality of the predictor step, and add $-\frac{2}{\tau^2} \boldsymbol{r}_{\mathbb{m}}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ to $\dot{\boldsymbol{u}}_{\mathbb{m}}^{n+1}(\boldsymbol{p}) \cdot \boldsymbol{e}_1(\boldsymbol{p})$ computed in (5.51)

Remark 5.3.1. Rothe's method and mesh refinement
Up to now, for the derivation and the analysis of the algorithms we assumed a fixed mesh in space over all time steps because it enables a straightforward proof of the dissipativity of the algorithm. However, the presented time-discretizations can be used in the context of Rothe's method which means that we may discretize first in time, second in space, allowing for remeshing in each time step. Fortunately, neither the contact-stabilization nor the velocity corrections are adversely affected if the mesh is changed between time steps n *and* $n+1$ *but the proof of dissipativity is not guaranteed. Let* $\tilde{\boldsymbol{u}}_{\mathbb{m}}^{n}$ *be the representation of the solution* $\boldsymbol{u}_{\mathbb{m}}^{n}$ *on another mesh which is generated by coarsening and refinement. Then* $\tilde{\boldsymbol{u}}_{\mathbb{m}}^{n}$ *is not necessarily contained in the admissible set* $\boldsymbol{\mathcal{K}}_{\mathbb{m}}$ *corresponding to the new mesh. This problem can be solved by replacing* $\tilde{\boldsymbol{u}}_{\mathbb{m}}^{n}$ *with an admissible solution* $\tilde{\boldsymbol{u}}_{\mathbb{m},\text{admissible}}^{n}$ *in*



(a) Geometry **1** - The marked node is used for evaluating displacements and velocities, see Figure 5.12.

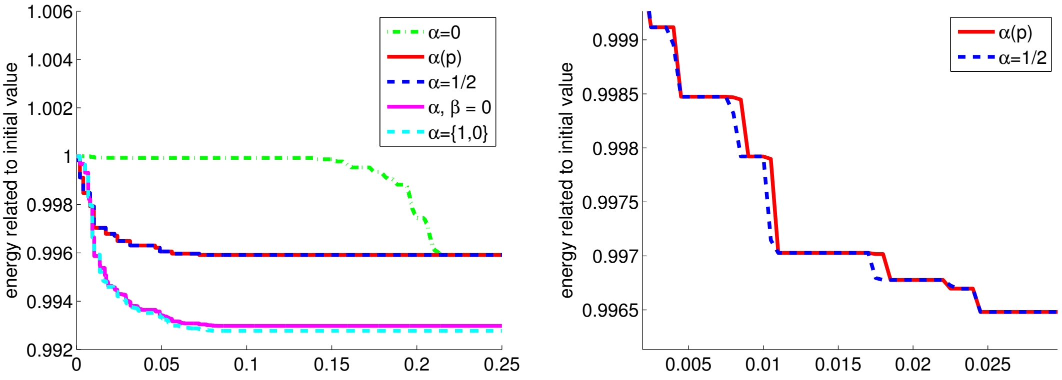
(b) Geometry **2**

Figure 5.8: Geometries and meshes

the algorithm to compute $\boldsymbol{u}_{\mathtt{m}}^{n+1}$*. Such an admissible solution can be easily generated by a projection of* $\tilde{\boldsymbol{u}}_{\mathtt{m}}^{n}$ *onto the admissible set. Then* $\mathcal{E}(\boldsymbol{u}_{\mathtt{m}}^{n+1}) - \mathcal{E}(\tilde{\boldsymbol{u}}_{\mathtt{m},\text{admissible}}^{n}) \leq 0$*. In this sense the algorithm is still dissipative, but* $\mathcal{E}(\tilde{\boldsymbol{u}}_{\mathtt{m},\text{admissible}}^{n}) \neq \mathcal{E}(\boldsymbol{u}_{\mathtt{m}}^{n})$ *in general.*

5.3.3 Numerical results

In this section we demonstrate the numerical behavior of the different elements of the family of modified Newmark schemes (Algorithm 5.3.1) as explained in Section 5.3.1. We denote the different methods by the chosen parameters. This means that $\alpha(\boldsymbol{p})$ refers to the so-called improved contact-stabilized Newmark scheme (Algorithm 5.3.2) emphasizing that the parameters are individually chosen for each node $\boldsymbol{p}$. The contact-stabilized Newmark scheme is denoted by $\alpha = 0$, the contact-implicit Newmark scheme by $\alpha, \beta = 0$ because it is the only analyzed scheme which does not fulfill the property $\alpha = (1 - \beta)$. The two other methods presented in Section 5.3.1 are called $\alpha = 1/2$ and $\alpha = \{1, 0\}$ because $\alpha = 1$ holds only for active contact nodes.
Since up to now no existence results for the hyperbolic system of dynamic contact problems are available, we measure the quality of temporal discretization schemes as usual along structural indicators as conservation of energy, displacements and velocities at the contact interface, stability of the contact stresses and closely related to it the number of contact nodes. In our numerical examples we consider one-body contact problems with different time-step sizes. The material constants are chosen to be $E = 5 \cdot 10^2$, $\nu = 0.3$ and $\rho = 1$. The meshes of the geometries shown in Figure 5.8 consist of hexahedra. In Geometry **1** the rounded side comes into contact with a plane rigid foundation. In Figure 5.8(b) the domain of computation is the block and the obstacle is a real tibial head. If not indicated otherwise we have chosen Geometry **1**, four times uniformly refined with 12,552 degrees of freedom, for the subsequent numerical studies.



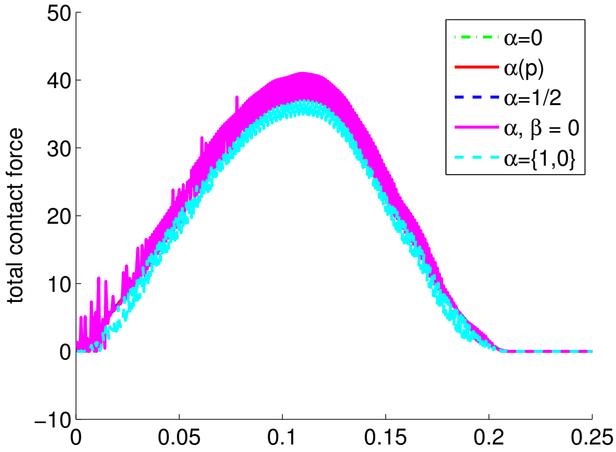
(a) $\tau = 5 \cdot 10^{-6}$

(b) $\tau = 5 \cdot 10^{-4}$

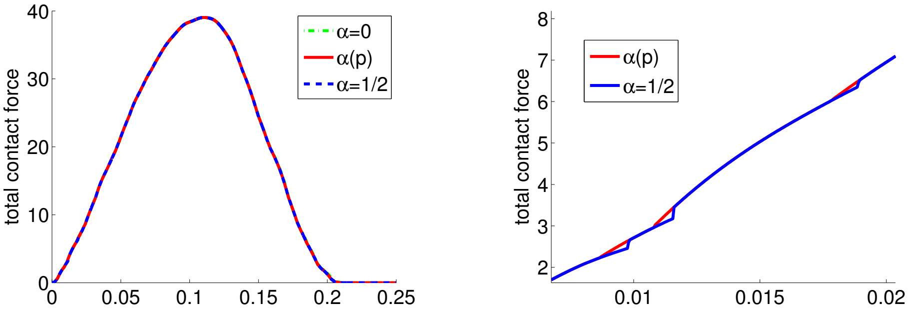
Figure 5.9: Energy related to the initial value

As we know from Section 5.3.1 all five time discretization schemes are dissipative which can be seen in Figure 5.9(a) for a time step size of $\tau = 5 \cdot 10^{-6}$. The major amount of energy loss is due to the moment of impact except for the contact-stabilized Newmark scheme ($\alpha = 0$) where it is the moment of detaching. For this choice of time step and mesh size the methods with $\alpha, \beta = 0$ and $\alpha = \{1, 0\}$ loose more energy than the other methods. It seems as if the curves of energy loss for the methods with $\alpha = 1/2$ and $\alpha(\boldsymbol{p})$ coincide but in Figure 5.9(b) a zoom shows that they are slightly different which is due to the different velocities. The dependance of the energy loss on the choice of time step and mesh size will be discussed later.

In contrast to the course of energy the behavior of the contact forces of all different methods obviously classifies the five methods into more and less stable algorithms. The contact forces appearing in the method with $\alpha, \beta = 0$ and $\alpha = \{1, 0\}$ oscillate very often, see Figure 5.10(a), as expected from the discussion in Section 5.3.1 about the changing of sign in the velocities. The three other methods ($\alpha = 0$, $\alpha = 1/2$ and $\alpha(\boldsymbol{p})$) show a stable behavior of the contact forces (Figure 5.10(b)) and the lines seem to coincide for all three methods. But as can be seen in the zoom (Figure 5.10(c)) the method $\alpha = 1/2$ gives rise to small kinks. Closely related to the oscillations and the different behavior of the contact forces is the number of contact nodes during the evolution and the difference in the number of nodes which are active for the predictor and for the solution. In Figure 5.11(a) we see that the number of contact nodes oscillates strongly for $\alpha, \beta = 0$ and $\alpha = \{1, 0\}$. Due to this unstable behavior of the contact forces and contact nodes for these two methods we restrict our further discussion to the remaining three methods. In all the remaining methods ($\alpha(\boldsymbol{p})$, $\alpha = 0$, $\alpha = 1/2$) we have a predictor step. We analyze the number of contact nodes which are active for the predictor but not for the solution (Figure 5.11(b)) and the other way round which are active for the solution and not for the predictor (Figure 5.11(c)). The number of contact nodes which are active for the solution but not for the predictor (Figure 5.11(c)) is zero for the contact-



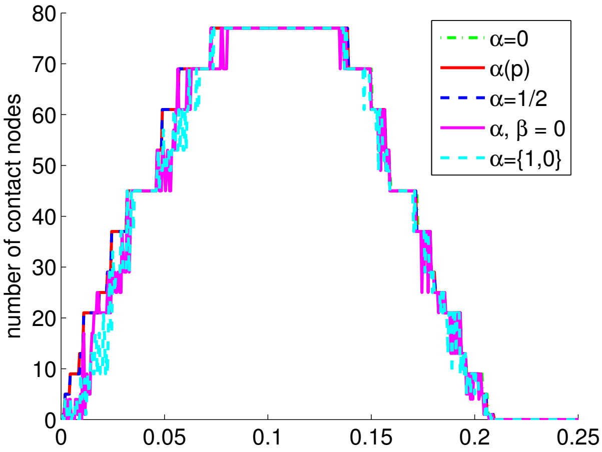
(a) $\tau = 5 \cdot 10^{-4}$



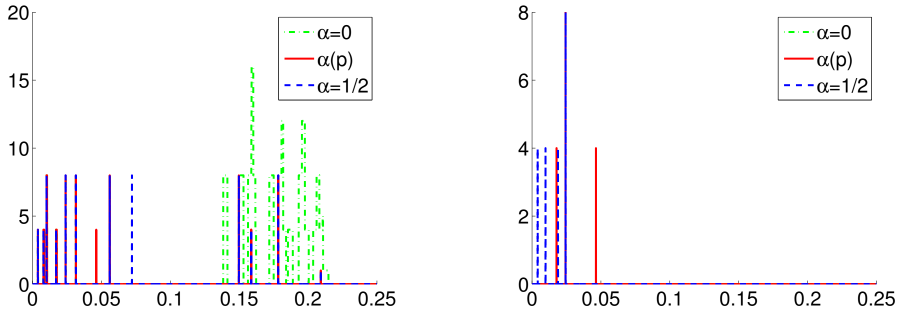
(b) $\tau = 5 \cdot 10^{-5}$

(c) zoom of Figure 5.10(b)

Figure 5.10: Contact forces



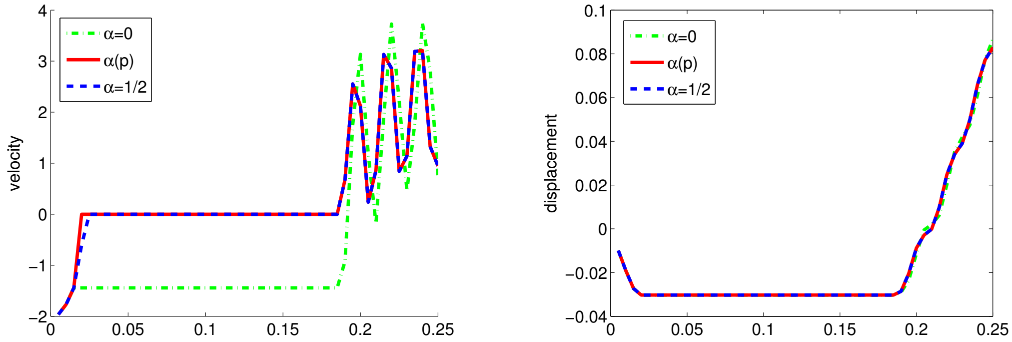
(a) $\tau = 5 \cdot 10^{-4}$



(b) number of nodes which are active for the predictor but not for the solution; $\tau = 5 \cdot 10^{-5}$

(c) number of nodes which are active for the solution but not for the predictor; $\tau = 5 \cdot 10^{-5}$

Figure 5.11: Number of contact nodes



(a) $\tau = 5 \cdot 10^{-3}$, velocity

(b) $\tau = 5 \cdot 10^{-3}$, displacement

Figure 5.12: Velocity and displacement at a single contact node

stabilized Newmark scheme ($\alpha = 0$), i.e., the active set of the solution is a subset of the active set of the predictor. In fact, this is important to avoid oscillations in the contact forces. The two other methods show a small number of nodes which are active for the solution but not for the predictor (Figure 5.11(c)), especially during impact. It seems as if this behavior has no adverse effect on the contact forces due to the modification of the velocities. In Figure 5.11(b) for $\alpha = 0$ it is conspicuous that during the process of detaching the number of active nodes for the predictor is much larger as the number of active nodes for the solution. This is due to the velocity which remained the velocity of impact during the contact phase. A discrepancy between the predictor and the solution for active nodes in the predictor step enlarges the loss of energy as follows from $\langle 3 \rangle$ in the proof of Proposition 5.3.1 which in turn explains why the large part of energy loss is due to the moment of detaching in the method with $\alpha = 0$ in contrast to the others.

In Figure 5.12 we show the trajectories of velocities (Figure 5.12(a)) and displacements (Figure 5.12(b)) at a single contact node which is depicted in Figure 5.8(a). The velocity is set to zero immediately for the method with $\alpha(\boldsymbol{p})$. Nodes where $\tau^*(\boldsymbol{p}) > \frac{\tau}{2}$ also occur. At such nodes the velocities would be zero one time step later as always for the method with $\alpha = 1/2$ (blue line).

Due to the good behavior of the improved contact-stabilized Newmark scheme ($\alpha(\boldsymbol{p})$) concerning contact forces and velocities it is our favorite method and we restrict our further more detailed studies to this special method and a comparison to the method with $\alpha = 0$.

A closer look at the process of detaching for the number of contact nodes, for the contact forces and the displacements themselves (Figure 5.12(b)) foreshadows that the body detaches earlier using the new method with $\alpha(\boldsymbol{p})$ instead of the method with $\alpha = 0$. This effect would be welcome because in [DEP11] it is shown for a $1D$ problem that the contact-stabilized Newmark scheme detaches later than the exact solution. Therefore, we compute the example of [DEP11, Section 3.1] where an elastic $1D$ bar of length $L = 10$ is dropped from a height $h_0 = 5$ with a starting velocity $-v_0$. The amount of the

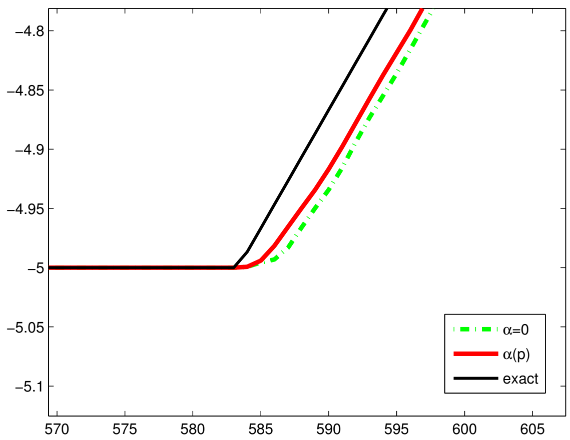


Figure 5.13: 1D example - displacements at the moment of detaching

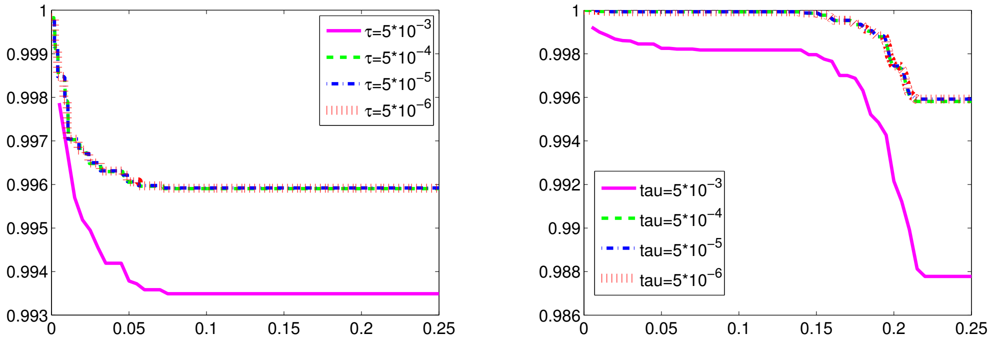
starting velocity is chosen to $v_0 = 10$, the elastic modulus is $E = 900$ and the density is set to $\rho = 1$. The exact solution at the single contact node is given by

$$u(x,t) := \begin{cases} h_0 - v_0 \cdot t, & \text{if} \quad t \leq \frac{h_0}{v_0} \\ 0 & \text{if} \quad \frac{h_0}{v_0} \leq t \leq \left(\frac{h_0}{v_0} + 2 \cdot \frac{L}{c_0}\right) \\ v_0 \cdot \left(t - \left(\frac{h_0}{v_0} + 2 \cdot \frac{L}{c_0}\right)\right) & \text{if} \quad \left(\frac{h_0}{v_0} + 2 \cdot \frac{L}{c_0}\right) \leq t \end{cases}$$

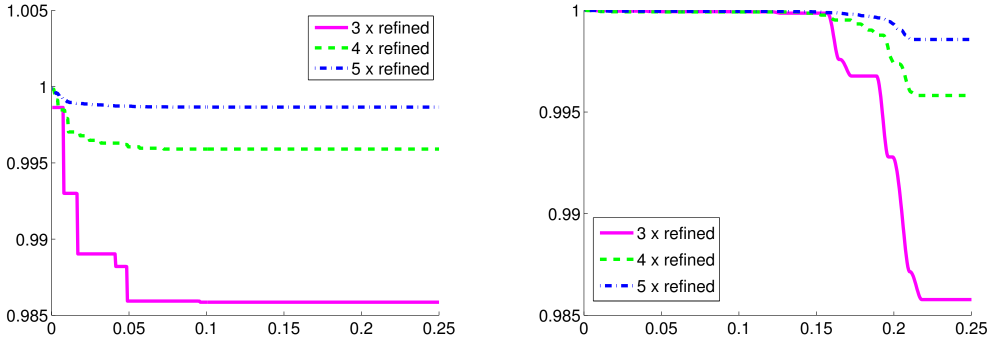
where $c_0 = \sqrt{\frac{E}{\rho}}$ is the wave speed. We choose the time step size $\tau = 5 \cdot 10^{-3}$ and a mesh size of $h = 0.1$ to compute the numerical solutions in a `Matlab` implementation. The black line refers to the exact solution, the green dotted line to the solution with $\alpha = 0$ and the red line to the solution with $\alpha(\boldsymbol{p})$ in Figure 5.13 which shows a zoom around the detaching time plotted against the number of time steps. Although the method with $\alpha(\boldsymbol{p})$ does not coincide with the exact solution it is closer to it than the solution computed with the help of the contact-stabilized Newmark scheme ($\alpha = 0$).

As indicated in the beginning of this section the course of energy depends on the choice of time step sizes and mesh sizes. Therefore, we present a small numerical study in Figure 5.14. For both methods ($\alpha(\boldsymbol{p})$ and $\alpha = 0$) smaller mesh sizes decrease the loss of energy. Smaller time step sizes do not affect the loss of energy so significantly. Having again a closer look at the proof of Proposition 5.3.1 we see that the entries of the mass matrix corresponding to the contact boundary heavily influence the loss of energy. Therefore, it is obvious that a smaller mesh size decrease the loss of energy.

In the first column of Figure 5.15 we see the contact stresses for the improved contact-stabilized Newmark scheme at discrete times $n\tau$. The displacements at the contact boundary can be seen in the second column and the velocities in the third column. For



(a) $\alpha(\boldsymbol{p})$ - varying τ (b) $\alpha = 0$ - varying τ

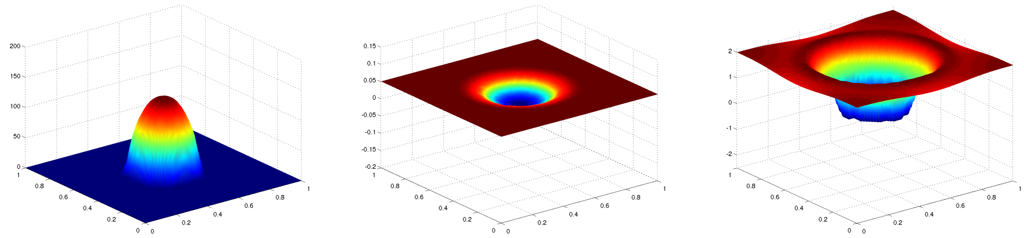


(c) $\alpha(\boldsymbol{p})$ - varying h (d) $\alpha = 0$ - varying h

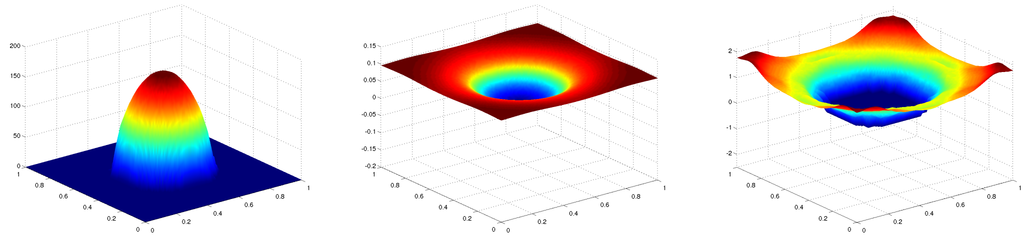
Figure 5.14: Energy related to initial value dependent on different time step and mesh sizes

the computation we have chosen Geometry **1** five times uniformly refined and the time step size $\tau = 5 \cdot 10^{-4}$. The smoothness of the contact stresses is obvious. The velocities are zero at the actual contact boundary due to the construction of our algorithm and except in the process of detaching the behavior is also very smooth. However, especially during the process of detaching, our improved contact-stabilized Newmark scheme ($\alpha(\boldsymbol{p})$) shows a significantly better behavior of the velocities with respect to their stability compared to the contact-stabilized Newmark scheme ($\alpha = 0$). This can be seen in Figure 5.16 where the velocities for both methods with a time steps size of $\tau = 5 \cdot 10^{-4}$ are plotted at different time steps n. Here, the color of the plots is adapted to the maximum and minimum value of all the different plots which are compared. The oscillations occurring in the method with $\alpha = 0$ are due to the artificial behavior during contact where the velocities remain the velocities of impact which have the wrong sign for detaching.

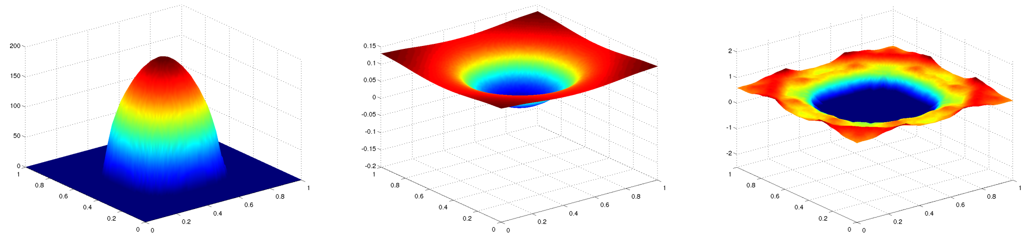
As the main advantage of the improved contact-stabilized Newmark scheme ($\alpha(\boldsymbol{p})$) is the correction of the velocity at the contact boundary, we show the velocities for a problem with a very complex gap function (Figure 5.17). Here a cube gets into contact with a tibial head, see Figure 5.8(b). Due to the complexity of this obstacle we cannot use a linearized non-penetration condition, see Section 1.2. Therefore, a signed-distance function is used as explained in [KM11]. For the sake of completeness, the first two rows in Figure 5.17 show the contact stresses and displacements for the same time steps n as the velocities. The time step size is $\tau = 5 \cdot 10^{-3}$. Just as in the foregoing example the velocities are zero in the case of contact and the contact forces do not oscillate.



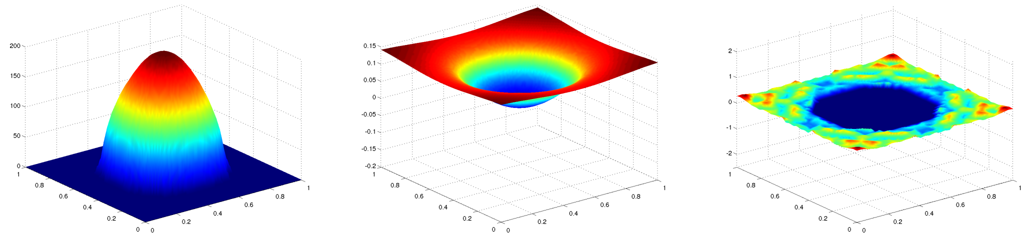
(a) $n = 50$ (b) $n = 50$ (c) $n = 50$



(d) $n = 100$ (e) $n = 100$ (f) $n = 100$

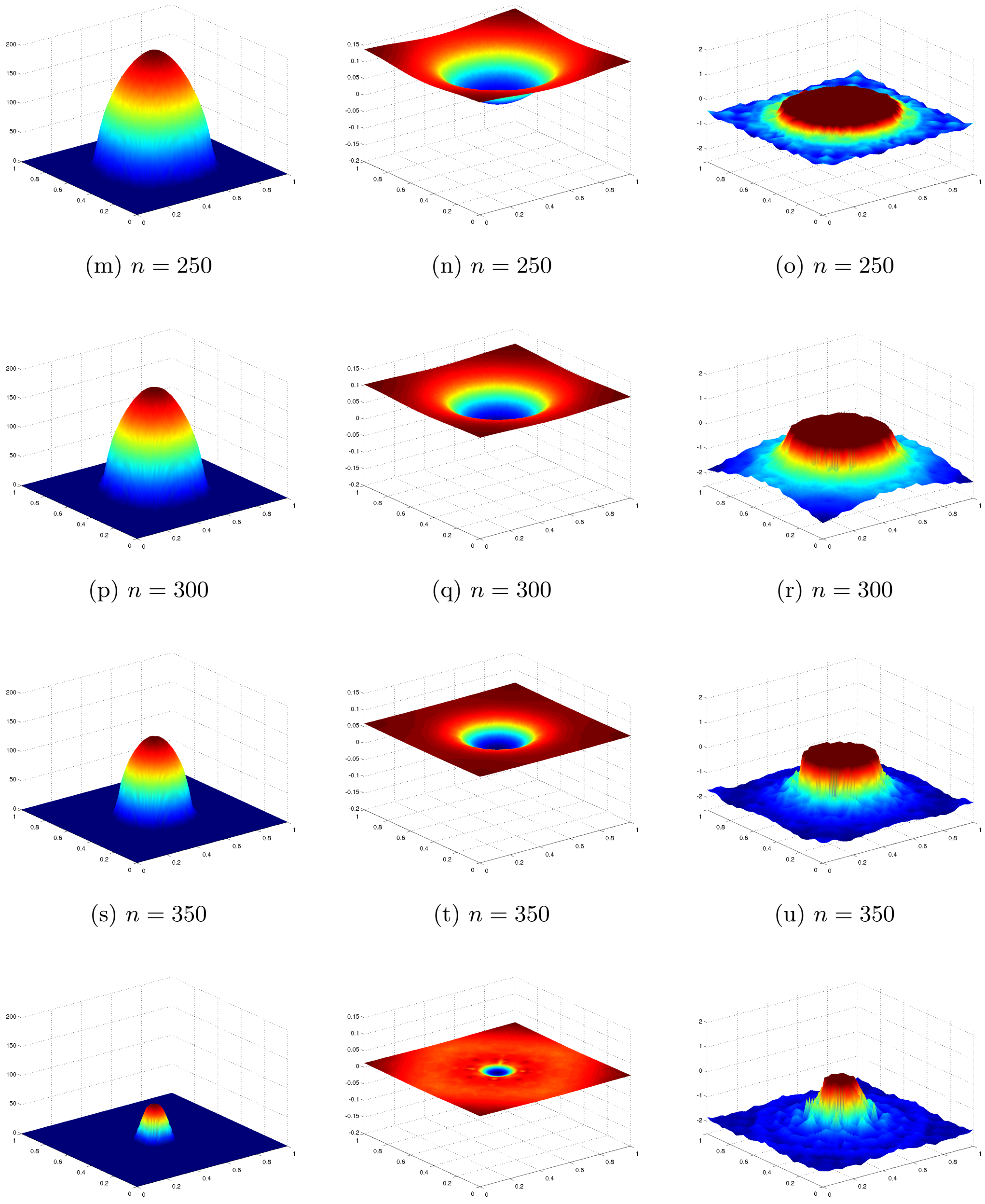


(g) $n = 150$ (h) $n = 150$ (i) $n = 150$



(j) $n = 200$ (k) $n = 200$ (l) $n = 200$

Figure 5.15: $\alpha(\boldsymbol{p})$: Contact stresses, displacements and velocities



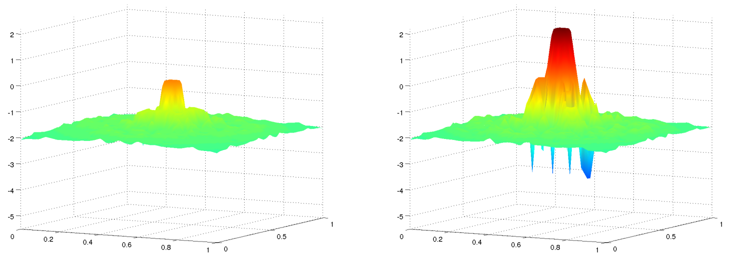
(m) $n = 250$ (n) $n = 250$ (o) $n = 250$

(p) $n = 300$ (q) $n = 300$ (r) $n = 300$

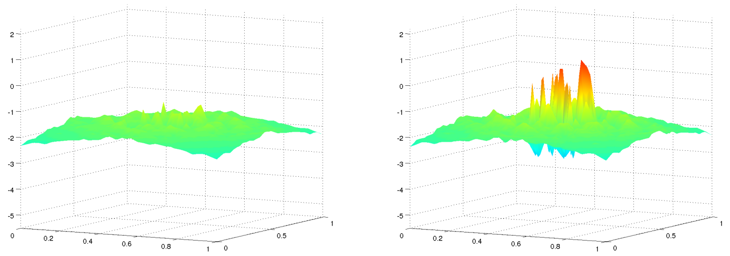
(s) $n = 350$ (t) $n = 350$ (u) $n = 350$

(v) $n = 400$ (w) $n = 400$ (x) $n = 400$

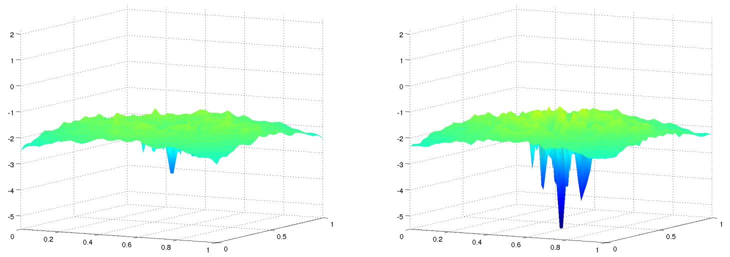
Figure 5.15: $\alpha(\boldsymbol{p})$: Contact stresses, displacements and velocities



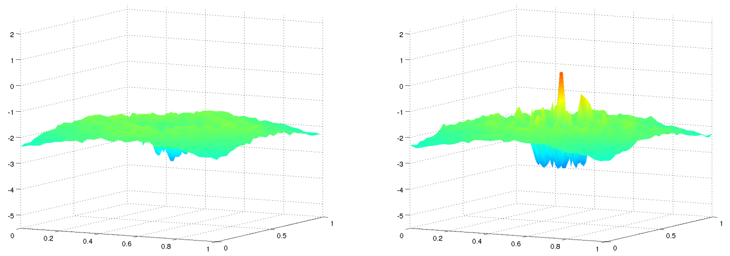
(a) $\alpha(\boldsymbol{p})$, $n = 414$ (b) $\alpha = 0$, $n = 414$



(c) $\alpha(\boldsymbol{p})$, $n = 424$ (d) $\alpha = 0,n = 424$

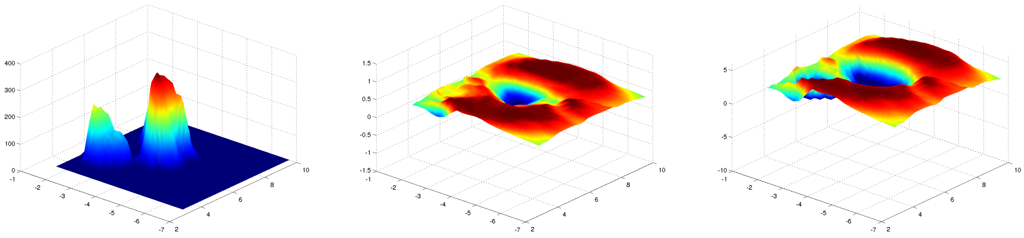


(e) $\alpha(\boldsymbol{p})$, $n = 430$ (f) $\alpha = 0,n = 433$

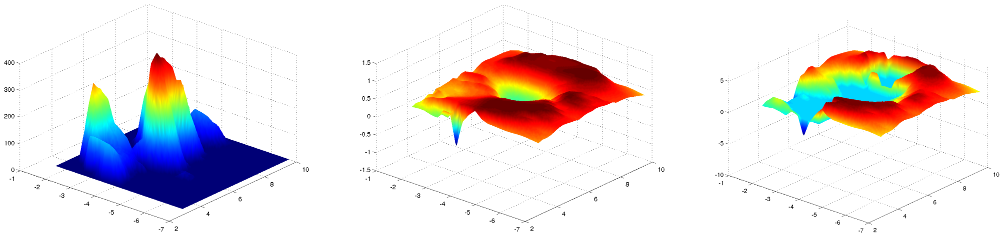


(g) $\alpha(\boldsymbol{p})$, $n = 433$ (h) $\alpha = 0,n = 442$

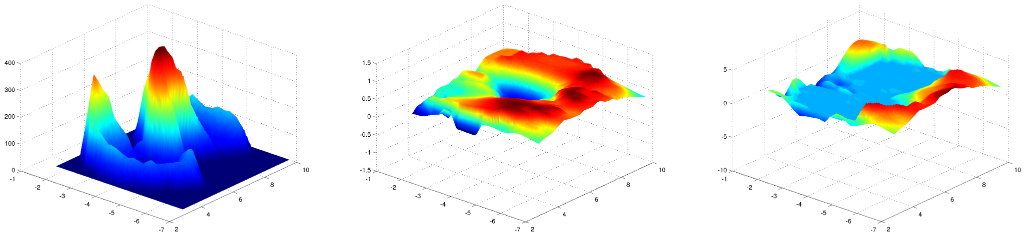
Figure 5.16: Velocities during process of detaching; $\tau = 5 \cdot 10^{-4}$



(a) Contact stresses, $n = 26$ (b) Displacements, $n = 26$ (c) Velocities, $n = 26$

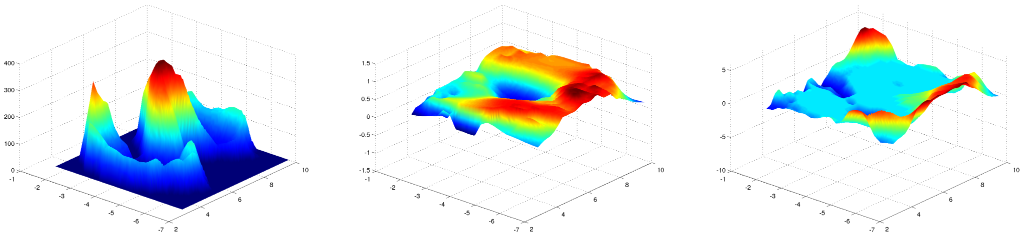


(d) Contact stresses,$n = 36$ (e) Displacements, $n = 36$ (f) Velocities, $n = 36$

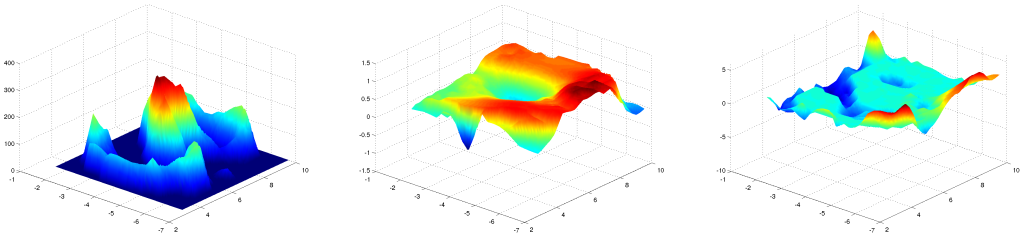


(g) Contact stresses,$n = 46$ (h) Displacements, $n = 46$ (i) Velocities, $n = 46$

Figure 5.17: $\alpha(\boldsymbol{p})$; complex obstacle; $\tau = 5 \cdot 10^{-3}$



(j) Contact stresses,$n = 56$ (k) Displacements, $n = 56$ (l) Velocities, $n = 56$



(m) Contact stresses,$n = 70$ (n) Displacements, $n = 70$ (o) Velocities, $n = 70$

Figure 5.17: $\alpha(\boldsymbol{p})$; complex obstacle; $\tau = 5 \cdot 10^{-3}$

Bibliography

[AB08] V. Acary and B. Brogliato. *Numerical methods for non smooth systems.* Lecture Notes in Applied and Computational Mechanics 35. Springer, 2008.

[Ada75] R. Adams. *Sobolev spaces.* Academic Press, New York, 1975.

[Ant05] S. Antman. *Nonlinear problems of elasticity.* Springer, New York, 2005.

[AO00] M. Ainsworth and J. Oden. *A posteriori error estimation in finite element analysis.* Wiley, New York, 2000.

[Bat86] K. Bathe. *Finite-Elemente-Methoden.* Springer, Berlin, 1986.

[BBHL99] F. Ben Belgacem, P. Hild, and P. Laborde. Extension of the mortar finite element method to a variational inequality modeling unilateral contact. *Math. Models Methods Appl. Sci.*, 9(2):287–303, 1999.

[BBL$^+$97] P. Bastian, K. Birken, S. Lang, N. Neuß, H. Rentz-Reichert, and C. Wieners. UG – a flexible software toolbox for solving partial differential equations. *Comput. Vis. Sci.*, 1:27–40, 1997.

[BC04] S. Bartels and C. Carstensen. Averaging techniques yield reliable a posteriori finite element error control for obstacle problems. *Numer. Math.*, 99(2):225–249, 2004.

[BCH07] D. Braess, C. Carstensen, and R. Hoppe. Convergence analysis of a conforming adaptive finite element method for an obstacle problem. *Numer. Math.*, 107(3):455–471, 2007.

[BGK87] P. Boieri, F. Gastaldi, and D. Kinderlehrer. Existence, uniqueness, and regularity results for the two-body contact problem. *Appl. Math. Optim.*, 15(3):251–277, 1987.

[BHR77] F. Brezzi, W. Hager, and P. Raviart. Error estimates for the finite element solution of variational inequalities. *Numer. Math.*, 28(4):431–443, 1977.

[BHS08] D. Braess, R. Hoppe, and J. Schöberl. A posteriori estimators for obstacle problems by the hypercircle method. *Comput. Vis. Sci.*, 11(4-6):351–362, 2008.

[Bra05] D. Braess. A posteriori error estimators for obstacle problems—another look. *Numer. Math.*, 101(3):415–421, 2005.

[Bro95] M. Brokate. Mathematische Elastizitäts- und Plastizitätstheorie. Vorlesungsskriptskript, 1995.

[Car06] C. Carstensen. Clément interpolation and its role in adaptive finite element error control. In *Partial differential equations and functional analysis*, volume 168 of *Operator Theory Advances and Applications*, pages 27–43. Birkhäuser, Basel, 2006.

[CHLS02] P. Coorevits, P. Hild, K. Lhalouani, and T. Sassi. Mixed finite element methods for unilateral problems: convergence analysis and numerical studies. *Math. Comp.*, 71(237):1–25, 2002.

[Cia88] P. Ciarlet. *Mathematical Elasticity. Vol. I.* Studies in Mathematics and its Applications 20. North-Holland Publishing Co., Amsterdam, 1988.

[CL98] V. Chawla and T. Laursen. Energy consistent algorithms for frictional contact problems. *Internat. J. Numer. Methods Engrg.*, 42(5):799–827, 1998.

[Clé75] P. Clément. Approximation by finite element functions using local regularization. *RAIRO Anal. Numér.*, 9(R-2):77–84, 1975.

[CN00] Z. Chen and R. Nochetto. Residual type a posteriori error estimates for elliptic obstacle problems. *Numer. Math.*, 84(4):527–548, 2000.

[DE09] D. Doyen and A. Ern. Convergence of a space semi-discrete modified mass method for the dynamic Signorini problem. *Commun. Math. Sci.*, 7(4):1063–1072, 2009.

[DE11] D. Doyen and A. Ern. Analysis of the modified mass method for the dynamic signorini problem with coulomb friction. *SIAM J. Numer. Anal.*, 49(5):2039–2056, 2011.

[DEP11] D. Doyen, A. Ern, and S. Piperno. Time-integration schemes for the finite element dynamic Signorini problem. *SIAM J. Sci. Comput.*, 33(1):223–249, 2011.

[DK09] T. Dickopf and R. Krause. Efficient simulation of multi-body contact problems on complex geometries: a flexible decomposition approach using constrained minimization. *Internat. J. Numer. Methods Engrg.*, 77(13):1834–1862, 2009.

[DKE08] P. Deuflhard, R. Krause, and S. Ertel. A contact-stabilized Newmark method for dynamical contact problems. *Internat. J. Numer. Methods Engrg.*, 73(9):1274–1290, 2008.

[Eck96] C. Eck. *Existenz und Regularität der Lösungen für Kontaktprobleme mit Reibung.* PhD thesis, Mathematisches Institut der Universität Stuttgart, 1996.

[EGK08] C. Eck, H. Garcke, and P. Knabner. *Mathematische Modellierung.* Springer, Berlin, 2008.

[EJK05] C. Eck, J. Jarušek, and M. Krbec. *Unilateral contact problems.* Pure and Applied Mathematics 270. Chapman & Hall/CRC, Boca Raton, FL, 2005. Variational methods and existence theorems.

[ET76] I. Ekeland and R. Temam. *Convex analysis and variational problems.* North-Holland Publishing Co., Amsterdam, 1976.

[Eva10] L. Evans. *Partial differential equations.* Graduate Studies in Mathematics 19. AMS, Providence, RI, 2010.

[FdV79] B. Fraeijs de Vebeuke. *A course in elasticity.* Applied Mathematical Sciences 29. Springer, 1979.

[FV03] F. Fierro and A. Veeser. A posteriori error estimators for regularized total variation of characteristic functions. *SIAM J. Numer. Anal.*, 41(6):2032–2055, 2003.

[GKW10] C. Groß, R. Krause, and M. Walloth. Nonlinear decomposition methods in elastodynamics. In *Numerical Mathematics and Advanced Applications 2009*, pages 407–415. Springer, Berlin, 2010.

[Glo84] R. Glowinski. *Numerical methods for nonlinear variational problems.* Springer Series in Computational Physics. Springer, New York, 1984.

[GSO10] M. Gonzalez, B. Schmidt, and M. Ortiz. Energy-stepping integrators in Lagrangian mechanics. *Internat. J. Numer. Methods Engrg.*, 82(2):205–241, 2010.

[Gur81] M. Gurtin. *An introduction to continuum mechanics.* Mathematics in Science and Engineering 158. Academic Press, New York, 1981.

[Has83] J. Haslinger. Approximation of the Signorini problem with friction, obeying the Coulomb law. *Math. Methods Appl. Sci.*, 5(3):422–437, 1983.

[Her81] H. Hertz. Über die Berührung fester elastischer Körper. *Journal für die reine und angewandte Mathematik*, 92:156–171, 1881.

[HH81] J. Haslinger and I.n Hlaváček. Contact between elastic bodies. II. Finite element analysis. *Apl. Mat.*, 26(4):263–290, 1981.

[HHN96] J. Haslinger, I. Hlaváček, and J. Nečas. Numerical methods for unilateral problems in solid mechanics. In *Handbook of numerical analysis, Vol. IV*, pages 313–485. North-Holland, Amsterdam, 1996.

[HHW08] C. Hager, S. Hüeber, and B. Wohlmuth. A stable energy-conserving approach for frictional contact problems based on quadrature formulas. *Internat. J. Numer. Methods Engrg.*, 73(2):205–225, 2008.

[Hil00] P. Hild. Numerical implementation of two nonconforming finite element methods for unilateral contact. *Comput. Methods Appl. Mech. Engrg.*, 184(1):99–123, 2000.

[HL02] P. Hild and P. Laborde. Quadratic finite element methods for unilateral contact problems. *Appl. Numer. Math.*, 41(3):401–421, 2002.

[HLW06] E. Hairer, C. Lubich, and G. Wanner. *Geometric numerical integration.* Springer Series in Computational Mathematics 31. Springer, Berlin, 2006. Structure-preserving algorithms for ordinary differential equations.

[HMW05] S. Hüeber, M. Mair, and B. Wohlmuth. A priori error estimates and an inexact primal-dual active set strategy for linear and quadratic finite elements applied to multibody contact problems. *Appl. Numer. Math.*, 54(3-4):555–576, 2005.

[HN05] P. Hild and S. Nicaise. A posteriori error estimations of residual type for Signorini's problem. *Numer. Math.*, 101(3):523–549, 2005.

[HN07] P. Hild and S. Nicaise. Residual a posteriori error estimators for contact problems in elasticity. *M2AN Math. Model. Numer. Anal.*, 41(5):897–923, 2007.

[HR07] P. Hild and Y. Renard. An error estimate for the Signorini problem with Coulomb friction approximated by finite elements. *SIAM J. Numer. Anal.*, 45(5):2012–2031, 2007.

[HS45] J. Harding and I. Sneddon. The elastic stresses produced by the indentation of the plane surface of a semi-infinite elastic solid by a rigid punch. *Proc. Cambridge Philos. Soc.*, 41:16–26, 1945.

[HTS$^+$76] T. Hughes, R. Taylor, J. Sackmann, A. Curnier, and W. Kanoknukulchai. A finite element method for a class of contact-impact problems. *Comput. Methods Appl. Mech. Engrg.*, 8(3):249–276, 1976.

[Hub04] M. Huber. Zur Theorie der Berührung fester elastischer Körper. *Annalen der Physik*, 319(6):153–163, 1904.

[HW05] S. Hüeber and B. Wohlmuth. An optimal a priori error estimate for nonlinear multibody contact problems. *SIAM J. Numer. Anal.*, 43(1):156–173, 2005.

[HW09] C. Hager and B. Wohlmuth. Analysis of a space-time discretization for dynamic elasticity problems based on mass-free surface elements. *SIAM J. Numer. Anal.*, 47(3):1863–1885, 2009.

[Joh85] K. Johnson. *Contact mechanics.* Cambridge University Press, 1985.

[KDS09] C. Klapproth, P. Deuflhard, and A. Schiela. A perturbation result for dynamical contact problems. *Numer. Math. Theory Methods Appl.*, 2(3):237–257, 2009.

[Kim89] J. Kim. A boundary thin obstacle problem for a wave equation. *Comm. Partial Differential Equations*, 14(8-9):1011–1026, 1989.

[Kin81] D. Kinderlehrer. Remarks about Signorini's problem in linear elasticity. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 8(4):605–645, 1981.

[KK01] R. Kornhuber and R. Krause. Adaptive multigrid methods for Signorini's problem in linear elasticity. *Comput. Vis. Sci.*, 4(1):9–20, 2001.

[Kla11] C. Klapproth. *Adaptive Numerical Integration of Dynamical Contact Problems.* PhD thesis, Freie Universität Berlin, 2011.

[KLR06] H. Khenous, P. Laborde, and Y. Renard. Comparison of two approaches for the discretization of elastodynamic contact problems. *C. R. Math. Acad. Sci. Paris*, 342(10):791–796, 2006.

[KLR08] H. Khenous, P. Laborde, and Y. Renard. Mass redistribution method for finite element contact problems in elastodynamics. *Eur. J. Mech. A Solids*, 27(5):918–932, 2008.

[KM11] R. Krause and C. Mohr. Level set based multi-scale methods for large deformation contact problems. *Appl. Numer. Math.*, 61(4):428–442, 2011.

[KMOW00] C. Kane, J. E. Marsden, M. Ortiz, and M. West. Variational integrators and the Newmark algorithm for conservative and dissipative mechanical systems. *Internat. J. Numer. Methods Engrg.*, 49(10):1295–1325, 2000.

[KO88] N. Kikuchi and J. Oden. *Contact problems in elasticity: a study of variational inequalities and finite element methods.* SIAM, Philadelphia, PA, 1988.

[Kor96] R. Kornhuber. A posteriori error estimates for elliptic variational inequalities. *Comput. Math. Appl.*, 31(8):49–60, 1996.

[Kor97] R. Kornhuber. *Adaptive monotone multigrid methods for nonlinear variational problems.* Advances in Numerical Mathematics. Teubner, Stuttgart, 1997.

[Kra09] R. Krause. A nonsmooth multiscale method for solving frictional two-body contact problems in 2D and 3D with multigrid efficiency. *SIAM J. Sci. Comput.*, 31(2):1399–1423, 2008/09.

[Kra00] R. Krause. *Monotone Multigrid Methods for Signorini's Problem with Friction.* PhD thesis, Freie Universität Berlin, 2000.

[KROM99] C. Kane, E. A. Repetto, M. Ortiz, and J. E. Marsden. Finite element analysis of nonsmooth contact. *Comput. Methods Appl. Mech. Engrg.*, 180(1-2):1–26, 1999.

[KS80] D. Kinderlehrer and G. Stampacchia. *An introduction to variational inequalities and their applications.* Pure and Applied Mathematics 88. Academic Press, New York, 1980.

[KSD10] C. Klapproth, A. Schiela, and P. Deuflhard. Consistency results on Newmark methods for dynamical contact problems. *Numer. Math.*, 116(1):65–94, 2010.

[KSD11] C. Klapproth, A. Schiela, and P. Deuflhard. Adaptive timestep control for the contact-stabilized Newmark method. *Numer. Math.*, 119(1):49–81, 2011.

[KVW12] R. Krause, A. Veeser, and M. Walloth. Efficient and reliable residual-type a posteriori error estimator for the Signorini problem. Technical report, Institute of Computational Science, Università della Svizzera italiana, February 2012.

[KW09a] R. Krause and M. Walloth. Presentation and Comparison of Selected Algorithms for Dynamic Contact Based on the Newmark Scheme. Technical report, Institute of Computational Science, Università della Svizzera italiana, December 2009. Accepted for publication in *Applied Numerical Mathematics.*

[KW09b] R. Krause and M. Walloth. A time discretization scheme based on Rothe's method for dynamical contact problems with friction. *Comput. Methods Appl. Mech. Engrg.*, 199(1-4):1–19, 2009.

[KW11] R. Krause and M. Walloth. A family of space-time connecting discretization schemes with local impact detection for elastodynamic contact problems. *Comput. Methods Appl. Mech. Engrg.*, 200(47-48):3425–3438, 2011.

[KZ11] R. Kornhuber and Q. Zou. Efficient and reliable hierarchical error estimates for the discretization error of elliptic obstacle problems. *Math. Comp.*, 80(273):69–88, 2011.

[Lau02] T. Laursen. *Computational contact and impact mechanics.* Springer, Berlin, 2002.

[LC97] T. Laursen and V. Chawla. Design of energy conserving algorithms for frictionless dynamic contact problems. *Internat. J. Numer. Methods Engrg.*, 40(5):863–886, 1997.

[LL02] T. Laursen and G. Love. Improved implicit integrators for transient impact problems—geometric admissibility within the conserving framework. *Internat. J. Numer. Methods Engrg.*, 53(2):245–274, 2002.

[Lov39] A. Love. Boussinesq's problem for a rigid cone. *Quart. J. Math., Oxford Ser.*, 10:161–175, 1939.

[LPR91] C. Licht, E. Pratt, and M. Raous. Remarks on a numerical method for unilateral contact including friction. In *Unilateral problems in structural analysis IV*, Internat. Ser. Numer. Math. 101, pages 129–144. Birkhäuser, Basel, 1991.

[LS67] J. Lions and G. Stampacchia. Variational inequalities. *Comm. Pure Appl. Math.*, 20:493–519, 1967.

[LS84] G. Lebeau and M. Schatzman. A wave problem in a half-space with a unilateral constraint at the boundary. *J. Differential Equations*, 53(3):309–361, 1984.

[MNvPZ07] K. Moon, R. Nochetto, T. von Petersdorff, and C. Zhang. A posteriori error analysis for parabolic variational inequalities. *M2AN Math. Model. Numer. Anal.*, 41(3):485–511, 2007.

[Mor99] J. Moreau. Numerical aspects of the sweeping process. *Comput. Methods Appl. Mech. Engrg.*, 177(3-4):329–349, 1999.

[New59] N. Newmark. A method of computation for structural dynamics. *J. Eng. Mech. Div.*, 85:67–94, 1959.

[NJH80] J. Nečas, J. Jarušek, and J. Haslinger. On the solution of the variational inequality to the Signorini problem with small friction. *Boll. Un. Mat. Ital. B (5)*, 17(2):796–811, 1980.

[NSV03] R. Nochetto, K. Siebert, and A. Veeser. Pointwise a posteriori error control for elliptic obstacle problems. *Numer. Math.*, 95(1):163–195, 2003.

[NSV05] R. Nochetto, K. Siebert, and A. Veeser. Fully localized a posteriori error estimators and barrier sets for contact problems. *SIAM J. Numer. Anal.*, 42(5):2118–2135, 2005.

[NSV09] R. Nochetto, K. Siebert, and A. Veeser. Theory of Adaptive Finite Element Methods: An Introduction. In *Multiscale, Nonlinear and Adaptive Approximation*, pages 409–542. Springer, Berlin, 2009.

[NvPZ10] R. Nochetto, T. von Petersdorff, and C. Zhang. A posteriori error analysis for a class of integral equations and variational inequalities. *Numer. Math.*, 116(3):519–552, 2010.

[NW02] R. Nochetto and L. Wahlbin. Positivity preserving finite element approximation. *Math. Comp.*, 71(240):1405–1419, 2002.

[Ogd84] R. Ogden. *Nonlinear elastic deformations.* Ellis Horwood Ltd., Chichester, 1984.

[PKMO02] A. Pandolfi, C. Kane, J. Marsden, and M. Ortiz. Time-discretized variational formulation of non-smooth frictional contact. *Internat. J. Numer. Methods Engrg.*, 53(8):1801–1829, 2002.

[PP10] M. Page and D. Praetorius. Convergence of adaptive FEM for some elliptic obstacle problem. Technical report, Institute of Analysis and Scientific Computing, Vienna University of Technology, 2010.

[Rod87] J. Rodrigues. *Obstacle problems in mathematical physics.* North-Holland Publishing Co., Amsterdam, 1987.

[Sch89] R. Schumann. Regularity for Signorini's problem in linear elasticity. *Manuscripta Math.*, 63(3):255–291, 1989.

[Sch02] F. Schuricht. Variational approach to contact problems in nonlinear elasticity. *Calc. Var. Partial Differential Equations*, 15(4):433–449, 2002.

[SH83] A. Sackfield and D. Hills. Some useful results in the classical hertz contact problem. *Journal of strain analysis*, 18(2):101–105, 1983.

[Sne48] I. Sneddon. Boussinesq's problem for a rigid cone. *Proc. Cambridge Philos. Soc.*, 44:492–507, 1948.

[SS05] A. Schmidt and K. Siebert. *Design of adaptive finite element software.* Lecture Notes in Computational Science and Engineering 42. Springer, Berlin, 2005.

[ST92] J. Simo and N. Tarnow. The discrete energy-momentum method. Conserving algorithms for nonlinear elastodynamics. *Z. Angew. Math. Phys.*, 43(5):757–792, 1992.

[Ste08] O. Steinbach. *Numerical approximation methods for elliptic boundary value problems.* Springer, New York, 2008.

[SV07] K. Siebert and A. Veeser. A unilaterally constrained quadratic minimization with adaptive finite elements. *SIAM J. Optim.*, 18(1):260–289, 2007.

[Sän05] A. Sändig. Mathematische Methoden in der Kontinuumsmechanik, Vorlesungsskript, 2005.

[TP93] R. Taylor and P. Papadopoulos. On a finite element method for dynamic contact/impact problems. *Internat. J. Numer. Methods Engrg.*, 36(12):2123–2140, 1993.

[Vee01] A. Veeser. Efficient and reliable a posteriori error estimators for elliptic obstacle problems. *SIAM J. Numer. Anal.*, 39(1):146–167, 2001.

[Ver96] R. Verfürth. *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques.* Wiley - Teubner, 1996.

[Ver98] R. Verfürth. A review of a posteriori error estimation techniques for elasticity problems. In *Advances in adaptive computational methods in mechanics*, Stud. Appl. Mech. 47, pages 257–274. Elsevier, Amsterdam, 1998.

[Ver99] R. Verfürth. Error estimates for some quasi-interpolation operators. *M2AN Math. Model. Numer. Anal.*, 33(4):695–713, 1999.

[Ver11a] R. Verfürth. Adaptive Finite Element Methods, Vorlesungsskript, 2011.

[Ver11b] R. Verfürth. Numerik II Finite Elemente, Vorlesungsskript, 2011.

[WK03] B. Wohlmuth and R. Krause. Monotone multigrid methods on nonmatching grids for nonlinear multibody contact problems. *SIAM J. Sci. Comput.*, 25(1):324–347, 2003.

[Wlo82] J. Wloka. *Partielle Differentialgleichungen.* Teubner, Stuttgart, 1982.

[Woh00] B. Wohlmuth. A mortar finite element method using dual spaces for the Lagrange multiplier. *SIAM J. Numer. Anal.*, 38(3):989–1012, 2000.

[Woh07] B. Wohlmuth. An a posteriori error estimator for two-body contact problems on non-matching meshes. *J. Sci. Comput.*, 33(1):25–45, 2007.

[Wri02] P. Wriggers. *Computational Contact Mechanics.* J. Wiley & Sons Ltd., 2002.

[WW09] A. Weiss and B. Wohlmuth. A posteriori error estimator and error control for contact problems. *Math. Comp.*, 78(267):1237–1267, 2009.

[WW10] A. Weiss and B. Wohlmuth. A posteriori error estimator for obstacle problems. *SIAM J. Sci. Comput.*, 32(5):2627–2658, 2010.

[Yun91] T. Yun. The exact integral equation of Hertz's contact problem. *Appl. Math. Mech.*, 12(2):165–169, 1991.

[ZVKG11] Q. Zou, A. Veeser, R. Kornhuber, and C. Gräser. Hierarchical error estimates for the energy functional in obstacle problems. *Numer. Math.*, 117(4):653–677, 2011.