A reliable, efficient and localized error estimator for a discontinuous Galerkin method for the Signorini problem

Mirjam Walloth

Fachbereich Mathematik, TU Darmstadt, Dolivostraße 15, D-64293 Darmstadt

---

Abstract

We present a new residual-type a posteriori error estimator for the discontinuous finite element solution of contact problems. The theoretical results are derived for two and three-dimensional domains and arbitrary gap functions. The estimator yields upper and lower bounds to a suitable error norm which measures the error in the displacements and in a quantity related to the contact stresses and the actual contact zone. In the derivation of the error estimator the local properties of the discontinuous solution are exploited appropriately so that, on the one hand, the error estimator has no contributions related to the non-linearity in the interior of the actual contact zone and, on the other hand, the critical region between the actual and non-actual contact zone can be well refined.

Keywords: Signorini problem, residual-type a posteriori error estimator, discontinuous Galerkin method, Galerkin functional, full-contact zone.

---

1. Introduction

For the numerical simulation of processes in engineering and natural science adaptive mesh refinement is an indispensable tool to reach certain accuracy of the discrete solution for given computational resources. Adaptive mesh refinement is steered by a posteriori error estimators. It is desirable that a posteriori error estimators are equivalent to the error, such that they are reliable, i.e. giving an upper bound to the error and efficient, i.e. giving a lower bound to the error.

We consider the numerical simulation of Signorini's problem [1] which models the contact between a linear elastic body and a rigid body. The penetration of the bodies is avoided by inequality constraints at the potential contact boundary. In the weak formulation the problem is described by a variational inequality due to the constraints. The numerical solution of discrete variational

Email address: walloth@mathematik.tu-darmstadt.de (Mirjam Walloth)

inequalities is expensive. Usually, primal-dual-active set strategies or monotone multigrid methods are used. In this work we assume that a piecewise linear discontinuous Galerkin finite element discretization has been chosen. Discontinuous finite elements are more flexible, e.g. for unstructured grids, different polynomial degrees on each element and parallel computing. However, the number of degrees of freedom is much higher than for continuous finite elements, thus, increasing the computational costs. Therefore, adaptive mesh refinement is an important tool.

Residual based error estimators are explicit error estimators which can be computed directly from the finite element approximation and the given data, hence no further computation is required. In this article we present a new residual-type error estimator for the discontinuous finite element solution of contact problems. The theory is derived for two and three-dimensional domains and even non-discrete gap functions are considered. Exemplarily, we deal with the symmetric interior penalty method [2, 3]. It will turn out that in the case no actual contact occurs, the new error estimator coincides with the standard residual error estimator for linear elliptic problems. The new error estimator is also an extension of the latest error estimator for the continuous finite element solution of contact problems [4] to the non-conforming case.

Solving variational inequalities does not only provide the solution which in continuums mechanics is the displacement or deformation but also the constraining force which here in the case of contact problems corresponds to the contact stresses at the contact boundary. Even the actual contact zone where the contact stresses are non-zero ist a priori not known. The article [5] reveals that sharp a posteriori error estimators can be obtained by involving the error in the constraining forces in the error measure. Therein a Galerkin functional is introduced replacing the role of the residual for linear equations.

In the present work we propose a suitable Galerkin functional and an error measure for both unknowns in order to derive a sharp a posteriori error estimator. In contrast to continuous finite elements the nonconformity of the discretization has to be considered. Therefore, we exploit the definition of a so-called smoothing function which is a continuous finite element function constructed from the discontinuous finite element solution. An important point in this work is the definition of the discrete counterpart of the contact force density as a functional on $\boldsymbol{H}^1$, which we call quasi-discrete contact force density. In Section 3.1 we give a short excursion to the meaning of the linear residual and the discrete contact force as well as to different definitions of this discrete counterpart of the contact force density in the literature, e.g. in [5, 6, 7, 8, 9, 4, 10, 11] where it has been used for the derivation of a posteriori error estimators. We point out that applied to discontinuous finite elements the two existing approaches are very close and even in the interior of the actual contact zone, called full-contact, these two approaches coincide. Moreover, the works [12, 9, 4] reveal that a definition of the quasi-discrete contact force density which distinguishes between areas of the contact boundary where the bodies are fully or only partially in contact, gives rise to a localization of the error estimator contributions related to the non-linearity. In the present work this insight is translated to the discontinuous finite element solution, for the first time. In contrast to the case of continuous finite elements for contact problems the definition of full-contact is more localized and a nonlinear definition of the quasi-discrete contact force density [4] is avoidable due to the elementwise definition of the basis functions.

In the derivation of the new a posteriori error estimator we exploit local properties of the variational inequality where in some parts the non-penetration condition and the contact force density have to be considered, while in the rest the variational inequality reduces locally to a variational equation. We emphasize that an appropriate splitting of the arising terms after integration by parts of the Galerkin functional is important to maximize cancellation effects. Thereby, we avoid any error estimator contribution related to the non-linearity in the area of so-called full-contact in contrast to the work [11]. In consequence, the error estimator perceives that at the boundary which is in full-contact, where the solution equals the gap function, adaptive refinement cannot improve the solution. Here, we like to mention that in the case of arbitrary non-discrete gap functions error estimator contributions related to the obstacle approximation occur.

The proofs of efficiency are shown for all error estimator contributions occurring for discrete gap functions. They are mainly based on classical arguments known from residual error estimation except for one error estimator contribution which is associated to the complementarity condition. Such an error estimator contribution also occurs in the upper bound of the error estimator proposed in [11]. Therein it is argued that this error estimator contribution is of higher order. We like to emphasize that in the present work we explicitly derive a lower bound in terms of the error estimator contribution associated to the complementarity condition. The argument that terms are of higher order is only used for data oscillation.

Finally, numerical examples confirm our theoretical results. The convergence rate of the error estimator as well as refined meshes are shown for different examples. The mesh is refined more at the free boundary as in the area of full-contact, so that the transition zone between actual contact and non-contact is well resolved. It is obvious that the area of full-contact is not overrefined, especially compared to the standard residual estimator. Thus, the computational resources can be used for other areas where high errors are predicted.

2. The Signorini contact problem

The Signorini contact problem describes the contact of a linear elastic body with a rigid obstacle. The linear elastic body is represented by a domain $\Omega \subset \mathbb{R}^d$, $d = 2, 3$. The boundary $\Gamma = \partial\Omega$ 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. 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 $\boldsymbol{u} : \Omega \to \mathbb{R}^d$ which are vector-valued. Their components are printed in normal type and are indicated by subindices, e.g., u_i. The Cartesian basis vectors of $\mathbb{R}^d$ are denoted by $\boldsymbol{e}_i$, $i = 1, \ldots, d$.

As the body is assumed to consist of linear elastic material, the stress tensor $\boldsymbol{\sigma}(\boldsymbol{u}) : \Omega \to \mathbb{R}^{d\times d}$ obeys Hooke's law

$$\sigma_{ij}(\boldsymbol{u}) = E_{ijml}\epsilon_{ml}(\boldsymbol{u}) \tag{1}$$

where $\boldsymbol{\epsilon}$ is the linearized strain tensor given by

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

and E_{ijml} are the components of Hooke's tensor which is symmetric, elliptic and bounded.

In linear elasticity the non-penetration condition can be approximated by the so-called linearized non-penetration condition, compare [13] and [1]. The gap function describing the distance between the elastic body and the rigid body is given by $g : \Gamma_C \to \mathbb{R}$ and the direction of constraints are denoted by $\boldsymbol{\nu}$. Thus, the linearized non-penetration condition is $u_\nu \leq g$ where $u_\nu := \boldsymbol{u}\cdot\boldsymbol{\nu}$. The non-penetration condition evokes so-called contact stresses which are boundary stresses in direction of the constraints at the actual contact boundary. We use the notation $\hat{\boldsymbol{\sigma}}(\boldsymbol{u}) := \boldsymbol{\sigma}(\boldsymbol{u})\boldsymbol{n}$ for boundary stresses where $\boldsymbol{n}$ is the unit outward normal to the boundary. Hence, the contact stresses are given by $\hat{\sigma}_\nu(\boldsymbol{u}) := \hat{\boldsymbol{\sigma}}(\boldsymbol{u})\cdot\boldsymbol{\nu}$. As we neglect frictional effects the frictional stresses $\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}) := \hat{\boldsymbol{\sigma}}(\boldsymbol{u}) - \hat{\sigma}_\nu(\boldsymbol{u})\cdot\boldsymbol{\nu}$ are assumed to be zero.

The linear elastic body might be subjected to a volume force density $\boldsymbol{f}$, to surface forces $\boldsymbol{\pi}$ and to Dirichlet values $\boldsymbol{u}^D$. The complete problem formulation is given in Problem 1.

Problem 1. *Strong formulation of the Signorini contact problem*
Find $\boldsymbol{u} : \bar{\Omega} \to \mathbb{R}^d$ *such that*

$$\begin{aligned}
-\mathrm{div}\boldsymbol{\sigma}(\boldsymbol{u}) &= \boldsymbol{f} && \text{in } \Omega\\
\hat{\boldsymbol{\sigma}}(\boldsymbol{u}) &= \boldsymbol{\pi} && \text{on } \Gamma^N\\
\boldsymbol{u} &= \boldsymbol{u}^D && \text{on } \Gamma^D\\
u_\nu &\leq g && \text{on } \Gamma^C\\
\hat{\sigma}_\nu(\boldsymbol{u}) &\leq 0 && \text{on } \Gamma^C\\
(u_\nu - g)\cdot\hat{\sigma}_\nu(\boldsymbol{u}) &= 0 && \text{on } \Gamma^C\\
\hat{\boldsymbol{\sigma}}_T(\boldsymbol{u}) &= \boldsymbol{0} && \text{on } \Gamma^C.
\end{aligned}$$

In the following, we assume that the actual contact boundary, where $u_\nu = g$, is a strict subset of the potential contact boundary.

2.1. Weak formulation

The solution space of the weak formulation is the subspace

$$\boldsymbol{\mathcal{H}} := \{\boldsymbol{v} \in \boldsymbol{H}^1(\Omega) \mid \mathrm{tr}|_{\Gamma^D}(\boldsymbol{v}) = \boldsymbol{u}^D\}$$

of $\boldsymbol{H}^1(\Omega) := (H^1(\Omega))^d$ where tr is the trace operator. For convenience in the discrete approximation of the Dirichlet values we assume $\boldsymbol{u}^D$ to be continuous and piecewise linear on Γ^D. Whenever it is clear from the context that the restriction to the boundary requires the trace operator we omit the special notation. The space of test functions is given by $\boldsymbol{\mathcal{H}}_0 := \{\boldsymbol{\varphi} \in \boldsymbol{H}^1(\Omega) \mid \mathrm{tr}|_{\Gamma^D}(\boldsymbol{\varphi}) = \boldsymbol{0}\}$ and its dual is $\boldsymbol{\mathcal{H}}^*$. For a gap function $g \in H^{\frac{1}{2}}(\Gamma^C)$ we define the admissible set

$$\boldsymbol{\mathcal{K}} := \{\boldsymbol{v} \in \boldsymbol{\mathcal{H}} \mid v_\nu \leq g \text{ on } \Gamma^C\}. \tag{2}$$

We assume the force density $\boldsymbol{f}$ and the Neumann data $\boldsymbol{\pi}$ to be L^2-functions on Ω or Γ^N, respectively. Further, the directions of constraints $\boldsymbol{\nu}$ are given by a measurable vector field with absolute value $|\boldsymbol{\nu}(\boldsymbol{x})| = 1$. 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 with $\langle\cdot,\cdot\rangle_{-\frac{1}{2},\frac{1}{2}}$ and the corresponding norms are $\|\cdot\|_{\frac{1}{2}}$ and $\|\cdot\|_{-\frac{1}{2}}$. Later on, we need restrictions to subdomains which are indicated by a further subindex, e.g., $\|\cdot\|_{1,\omega}$ for $\omega \subset \Omega$. Finally, we define the symmetric bilinear form

$$a(\cdot,\cdot) := \int_\Omega \boldsymbol{\sigma}(\cdot) : \boldsymbol{\epsilon}(\cdot), \tag{3}$$

which is associated with the elastic energy.

The variational inequality in Problem 2 may be derived from the strong formulation (Problem 1) by integration by parts and exploiting $\hat{\sigma}_\nu(\boldsymbol{u})(v_\nu - u_\nu) \geq 0$ for $\boldsymbol{u}, \boldsymbol{v} \in \boldsymbol{\mathcal{K}}$.

Problem 2. *Variational inequality of the Signorini problem*
We seek a solution $\boldsymbol{u} \in \boldsymbol{\mathcal{K}}$ *such that*

$$a(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) \geq \langle \boldsymbol{f}, \boldsymbol{v} - \boldsymbol{u}\rangle + \langle \boldsymbol{\pi}, \boldsymbol{v} - \boldsymbol{u}\rangle_{\Gamma^N} \quad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{K}}. \tag{4}$$

The unique solvability of Problem 2 follows from the Theorem of Lions and Stampacchia, see e.g., [1, Theorem 2.1].

It exists a distribution $\boldsymbol{\lambda} \in \boldsymbol{\mathcal{H}}^*$ which turns the variational inequality (4) in an equation

$$\langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle + \langle \boldsymbol{\pi}, \boldsymbol{\varphi}\rangle_{\Gamma^N} - a(\boldsymbol{u}, \boldsymbol{\varphi}) = \langle \boldsymbol{\lambda}, \boldsymbol{\varphi}\rangle_{-1,1} \quad \forall \boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0. \tag{5}$$

From the optimization point of view $\boldsymbol{\lambda}$ is the Lagrange multiplier while from the physical point of view $\boldsymbol{\lambda}$ has the meaning of a constraining force density on Γ^C which we call contact force density. The contact force density is directly related to the contact stresses

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

which follows from the generalized Green's formula, see e.g., [1, Theorem 5.9]. Due to the variational inequality the contact force density fulfills the weak sign condition $\langle \boldsymbol{\lambda}, \boldsymbol{v} - \boldsymbol{u}\rangle_{-1,1} \leq 0$.

2.2. Discrete formulation

In the discrete setting the domain Ω is polygonal and the grid is a regular simplicial mesh $\mathfrak{M}$. It is taken from a shape-regular family, meaning that the ratio of the diameter of any element to the diameter of its inscribed circle is uniformly bounded. The polygonal 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. Further, we denote the set of all elements having a contact boundary side by $\mathfrak{M}^C$ and the set of all other elements by $\mathfrak{M}^I$. For the ease of presentation we assume that each element $\mathfrak{e} \in \mathfrak{M}^C$ has exactly one potential contact boundary side. The set of all sides $\mathfrak{s}$ (edges in $2D$ or faces in $3D$) is denoted by $\mathfrak{S}$ and we distinguish between the set $\mathfrak{S}^D$ of Dirichlet boundary sides, $\mathfrak{S}^N$ of Neumann boundary sides, $\mathfrak{S}^C$ of potential contact boundary sides and the set of interior sides $\mathfrak{S}^I$.

The set of all nodes $\boldsymbol{p}$ of the mesh is given by $\mathfrak{N}$ and we distinguish between the set $\mathfrak{N}^D$ of nodes on the Dirichlet boundary, the set $\mathfrak{N}^N$ of nodes at the Neumann boundary, the set $\mathfrak{N}^C$ of nodes at the potential contact boundary and the set of interior nodes $\mathfrak{N}^I$. If we refer to a set of nodes of a subset of the mesh we specify it by an additional subindex, as e.g. $\mathfrak{N}_\mathfrak{e}$ denotes the set of all nodes belonging to an element $\mathfrak{e}$. Further, we define a patch ω_p as the interior of the union of all elements sharing the node 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 γ_p^I. For Dirichlet and contact boundary nodes we denote the intersections between Γ and $\partial\omega_p$ by $\gamma_p^D := \Gamma^D \cap \partial\omega_p$ and $\gamma_p^C := \Gamma^C \cap \partial\omega_p$, respectively. Further, we will make use of $\omega_\mathfrak{s}$ which is the union of all elements sharing a side $\mathfrak{s}$.

The piecewise linear discontinuous finite element space corresponding to the mesh $\mathfrak{M}$ is denoted by

$$\boldsymbol{\mathcal{V}}_\mathfrak{m} := \mathbb{P}_d^1(\mathfrak{M}) = \{\boldsymbol{v}_\mathfrak{m} \in L^2(\Omega) \mid \forall \mathfrak{e} \in \mathfrak{M} \quad \boldsymbol{v}_\mathfrak{m}|_\mathfrak{e} \in \mathbb{P}_d^1(\mathfrak{e})\}.$$

Let ϕ_p be the linear finite element basis function. Then we define for all nodes p and all elements $\mathfrak{e}$ the piecewise linear basis functions by $\phi_{p,\mathfrak{e}} = \begin{cases} \phi_p & \text{on } \mathfrak{e} \\ 0 & \text{otherwise} \end{cases}$.

Thus $\boldsymbol{v}_\mathfrak{m}|_\mathfrak{e} \in \mathbb{P}_d^1(\mathfrak{e})$ can be represented by

$$\boldsymbol{v}_\mathfrak{m}|_\mathfrak{e} = \sum_{p \in \mathfrak{N}_\mathfrak{e}} \sum_{i=1}^{d} v_{\mathfrak{m},i}(\boldsymbol{p}) \phi_{p,\mathfrak{e}} \boldsymbol{e}_i.$$

We assume the direction of constraints $\boldsymbol{\nu}$ to be constant so that $v_{\mathfrak{m},\nu}|_\mathfrak{e} := \boldsymbol{v}_\mathfrak{m}|_\mathfrak{e} \cdot \boldsymbol{\nu}$ restricted to an element $\mathfrak{e}$ is a linear finite element function. Further, we set $\boldsymbol{\nu}$ to the first coordinate direction $\boldsymbol{e}_1$. The discrete approximation $g_\mathfrak{m}$ of the gap function g is assumed to be a continuous piecewise linear function. Then the discrete admissible set is given by

$$\boldsymbol{\mathcal{K}}_\mathfrak{m} := \{\boldsymbol{v}_\mathfrak{m} \in \boldsymbol{\mathcal{V}}_\mathfrak{m} \mid v_{\mathfrak{m},1}|_\mathfrak{e}(p) \leq g_\mathfrak{m}(p) \quad \forall p \in \mathfrak{N}_\mathfrak{e} \quad \forall \mathfrak{e} \in \mathfrak{M}^C\}.$$

For the formulation of the discontinuous Galerkin method we need the definitions of jumps and mean values. Let $\mathfrak{s}$ be an interior edge and $\mathfrak{e}_1$ and $\mathfrak{e}_2$ be two neighboring elements sharing the side $\mathfrak{s}$ and $\boldsymbol{n}^{\mathfrak{e}_1}$, $\boldsymbol{n}^{\mathfrak{e}_2}$ the two unit outward normals. The mean value on the common side is defined by $\{\boldsymbol{w}\} = \frac{1}{2}(\boldsymbol{w}|_{\mathfrak{e}_1} + \boldsymbol{w}|_{\mathfrak{e}_2})$ where $\boldsymbol{w}$ can be scalar, vector or tensor-valued. If w is scalar-valued the jump can be defined as $[w] = w|_{\mathfrak{e}_1} - w|_{\mathfrak{e}_2}$ which is also scalar-valued or it can be defined as $[w] = w|_{\mathfrak{e}_1}\boldsymbol{n}^{\mathfrak{e}_1} + w|_{\mathfrak{e}_2}\boldsymbol{n}^{\mathfrak{e}_2}$ which is vector-valued. If $\boldsymbol{w}$ is vector-valued the jumps can be either defined component-by-component like in the scalar case or alternatively defined as a matrix $[\![\boldsymbol{v}]\!] = (\boldsymbol{v}|_{\mathfrak{e}_1} \otimes \boldsymbol{n}^{\mathfrak{e}_1}) + (\boldsymbol{v}|_{\mathfrak{e}_2} \otimes \boldsymbol{n}^{\mathfrak{e}_2})$ where $\otimes$ is the dyadic product. For matrices $\boldsymbol{\tau}$ the jump term is defined as $[\boldsymbol{\tau}] = \boldsymbol{\tau}|_{\mathfrak{e}_1}\boldsymbol{n}^{\mathfrak{e}_1} + \boldsymbol{\tau}|_{\mathfrak{e}_2}\boldsymbol{n}^{\mathfrak{e}_2}$ which is a vector. For boundary edges $\mathfrak{s}$ we define $\{\boldsymbol{w}\} = [\boldsymbol{w}] = \boldsymbol{w}$.

For the discrete problem formulation we consider the symmetric interior penalty method, see e.g. [2]. The bilinear form is given by

$$\begin{aligned} a^{\mathsf{sip}}(\boldsymbol{v}_{\mathfrak{m}}, \boldsymbol{w}_{\mathfrak{m}}) = & \sum_{\mathfrak{e}\in\mathfrak{M}} \int_{\mathfrak{e}} \boldsymbol{\sigma}(\boldsymbol{v}_{\mathfrak{m}}) : \boldsymbol{\epsilon}(\boldsymbol{w}_{\mathfrak{m}}) \\ & - \sum_{\mathfrak{s}\in\mathfrak{S}^I\cup\mathfrak{S}^D} \int_{\mathfrak{s}} \{\hat{\boldsymbol{\sigma}}(\boldsymbol{v}_{\mathfrak{m}})\}[\boldsymbol{w}_{\mathfrak{m}}] + [\boldsymbol{v}_{\mathfrak{m}}]\{\hat{\boldsymbol{\sigma}}(\boldsymbol{w}_{\mathfrak{m}})\} \\ & + \sum_{\mathfrak{s}\in\mathfrak{S}^I\cup\mathfrak{S}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{v}_{\mathfrak{m}}][\boldsymbol{w}_{\mathfrak{m}}] \end{aligned}$$

where $\hat{\eta}$ is a positive parameter. We note that the scalar product $\{\hat{\boldsymbol{\sigma}}(\boldsymbol{v}_{\mathfrak{m}})\}[\boldsymbol{w}_{\mathfrak{m}}]$ can be rewritten as the Frobenius scalar product $\{\boldsymbol{\sigma}\} : [\![\boldsymbol{w}_{\mathfrak{m}}]\!]$. Further the scalar product $[\boldsymbol{v}_{\mathfrak{m}}][\boldsymbol{w}_{\mathfrak{m}}] = \sum_{i=1}^{d}[v_{\mathfrak{m},i}][w_{\mathfrak{m},i}]$ can be reformulated to the Frobenius scalar product $[\![\boldsymbol{v}_{\mathfrak{m}}]\!] : [\![\boldsymbol{w}_{\mathfrak{m}}]\!]$.

The right hand side is given by

$$\begin{aligned} F^{\mathsf{sip}}(\boldsymbol{w}_{\mathfrak{m}}) = & \sum_{\mathfrak{e}\in\mathfrak{M}} \int_{\mathfrak{e}} \boldsymbol{f}\boldsymbol{w}_{\mathfrak{m}} + \sum_{\mathfrak{s}\in\mathfrak{S}^N} \int_{\mathfrak{s}} \boldsymbol{\pi}\boldsymbol{w}_{\mathfrak{m}} \\ & - \sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} \boldsymbol{u}^D \hat{\boldsymbol{\sigma}}(\boldsymbol{w}_{\mathfrak{m}}) + \sum_{\mathfrak{s}\in\mathfrak{S}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} \boldsymbol{u}^D \boldsymbol{w}_{\mathfrak{m}}. \end{aligned}$$

Thus we can state the discrete problem formulation.

Problem 3. *Discrete variational inequality of the Signorini problem*
Find $\boldsymbol{u}_{\mathfrak{m}} \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}$ fulfilling the variational inequality

$$a^{\mathsf{sip}}(\boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}) \geq F^{\mathsf{sip}}(\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}) \quad \forall \boldsymbol{v}_{\mathfrak{m}} \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}. \tag{6}$$

The discrete formulation Problem 3 is consistent, i.e. the exact solution $\boldsymbol{u}$ fulfills the discrete variational inequality under the regularity assumption $\boldsymbol{u} \in \boldsymbol{\mathcal{H}} \cap \boldsymbol{H}^2(\Omega)$.

The norm corresponding to the bilinearform is

$$\|\boldsymbol{v}\|_{a,\mathsf{sip}} = \left(\sum_{\mathfrak{e}\in\mathfrak{M}} \|\boldsymbol{\epsilon}(\boldsymbol{v})\|^2_{L^2(\mathfrak{e})} + \sum_{\mathfrak{s}\in\mathfrak{S}^I\cup\mathfrak{S}^D} \frac{1}{h_{\mathfrak{s}}} \|[\boldsymbol{v}]\|^2_{L^2(\mathfrak{s})} \right)^{\frac{1}{2}}.$$

With respect to this norm the discrete coercivity of the bilinearform can be shown for a sufficiently large parameter $\hat{\eta}$. In [3] it is shown that $\hat{\eta} \gtrsim 6(\mu+\lambda)$ where μ, λ are the Lamé constants. Thus, as the bilinearform is bounded and elliptic on $\boldsymbol{\mathcal{V}}_{\mathfrak{m}}$ and $\boldsymbol{\mathcal{K}}_{\mathfrak{m}}$ is a closed and convex subset of $\boldsymbol{\mathcal{V}}_{\mathfrak{m}}$ the well-posedness in the discrete setting is guaranteed. For the two-dimensional case an a priori error estimate can be found in [3].

3. A posteriori error estimator

In this section we state the main results. The error measure, the Galerkin functional and the error estimator contributions will be defined and finally the Theorems of reliability and efficiency are formulated. The proofs of the Theorems are postponed to Sections 4 and 5. In order to define the error measure and the Galerkin functional (Section 3.2) we motivate and define the quasi-discrete contact force density in Section 3.1.

For the subsequent analysis we need a continuous approximation of the discontinuous solution. We call this function smoothing function and denote it by $\hat{\boldsymbol{u}}_{\mathfrak{m}}$. It is a function of the linear finite element space with incorporated Dirichlet values

$$\boldsymbol{\mathcal{H}}_{\mathfrak{m}} := \{\boldsymbol{v}_{\mathfrak{m}} \in \boldsymbol{\mathcal{C}}^0(\bar{\Omega}) \mid \forall \mathfrak{e} \in \mathfrak{m},\ \boldsymbol{v}_{\mathfrak{m}}|_{\mathfrak{e}} \in \mathbb{P}^1_d(\mathfrak{e}) \text{ and } \boldsymbol{v}_{\mathfrak{m}} = \boldsymbol{u}^D_{\mathfrak{m}} \text{ on } \Gamma_D\}.$$

Usually, for Dirichlet nodes $\hat{\boldsymbol{u}}_{\mathfrak{m}}(p) := \boldsymbol{u}^D_{\mathfrak{m}}(p)$ and for all remaining nodes the node values are defined as the average $\hat{\boldsymbol{u}}_{\mathfrak{m}}(p) := \frac{1}{\#(\mathfrak{e}\subset\omega_p)} \sum_{\mathfrak{e}\in\omega_p} \boldsymbol{u}_{\mathfrak{m}}|_{\mathfrak{e}}(p)$, see e.g. in [14].

Here, we have to adapt this definition at least for nodes $p \in \mathfrak{N}^C$ and the direction of constraints. We define $\hat{u}_{\mathfrak{m},1}(p) := \max\{u_{\mathfrak{m},1}|_{\mathfrak{s}}(p) : \mathfrak{s} \in \gamma^C_p\}\ \forall p \in \mathfrak{N}^C$. Thus, there exist $\mathfrak{s}^*$ such that $\hat{u}_{\mathfrak{m},1}(p) = u_{\mathfrak{m},1}|_{\mathfrak{s}^*}(p)$. For a contact boundary node and the tangential coordinate directions we can e.g. choose the node value $\hat{u}_{\mathfrak{m},i}(p) = u_{\mathfrak{m},i}(p)|_{\mathfrak{s}^*}$ where $\mathfrak{s}^*$ has been taken from the definition of $\hat{u}_{\mathfrak{m},1}(p)$ or choose the average as for the other nodes.

In the derivation of the error estimator we will make use of the following result. The proof follows similar to [10] and [14].

Lemma 1. *Let $\mathfrak{M}$ be a shape regular mesh and $v_{\mathfrak{m}} \in \boldsymbol{\mathcal{V}}_{\mathfrak{m}}$. It holds*

$$\sum_{\mathfrak{e}\in\mathfrak{M}} h_{\mathfrak{e}}^{-2} \|\hat{\boldsymbol{v}}_{\mathfrak{m}} - \boldsymbol{v}_{\mathfrak{m}}\|^2_{L^2(\mathfrak{e})} \lesssim \left(\sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} [\boldsymbol{v}_{\mathfrak{m}}]^2\right) + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} (\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}^D_{\mathfrak{m}})^2\right)$$

and

$$\sum_{\mathfrak{e}\in\mathfrak{M}} \|\nabla(\hat{\boldsymbol{v}}_{\mathfrak{m}} - \boldsymbol{v}_{\mathfrak{m}})\|^2_{L^2(\mathfrak{e})} \lesssim \left(\sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} [\boldsymbol{v}_{\mathfrak{m}}]^2\right) + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} (\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}^D_{\mathfrak{m}})^2\right).$$

Remark 1. *For the more general case that $\boldsymbol{u}^D \in \mathcal{C}^0(\Gamma^D)$ a variant of Lemma 1 can be found in [15].*

3.1. Quasi-discrete contact force density

As in [4] we aim to find an upper and lower bound of the error measure

$$\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1} \tag{7}$$

where $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}^*$ called quasi-discrete contact force density is a functional on $\boldsymbol{\mathcal{H}}_0$, approximating the discrete counterpart of the contact force density (5). This quantity will be motivated and defined in this Section.

In order to give a better understanding of the role of $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}^*$ and of the error measure we make a short excursion to conforming finite elements and existing residual-type error estimators for obstacle and contact problems.

We note that in the case of conforming finite elements and linear equations the quantity $\langle \boldsymbol{R}(\boldsymbol{u}_{\mathfrak{m}}), \cdot \rangle_{-1,1} := F(\cdot) - a(\boldsymbol{u}_{\mathfrak{m}}, \cdot)$, called linear residual, is used for the derivation of the error estimator. In this case if $\boldsymbol{R}(\boldsymbol{u}_{\mathfrak{m}})$ is tested against functions in $\boldsymbol{\mathcal{H}}_0$ it is a measure of the error. In the nonlinear case of obstacle and contact problems, if $\boldsymbol{R}(\boldsymbol{u}_{\mathfrak{m}})$ is tested against discrete functions it represents the difference between the discrete equation and the discrete inequality. Thus, we call it discrete contact force density and denote it by $\boldsymbol{\lambda}_{\mathfrak{m}}$. In consequence, in the case of obstacle and contact problems, the linear residual is no appropriate measure of the error. It has parts of the error as well as of the discrete contact force density.

Further, in the case of variational inequalities not only the displacements $\boldsymbol{u}$ but also the contact force density $\boldsymbol{\lambda}$ is an unknown of the system. The error measure (7) accounts for the errors in both quantities and the linear residual will be replaced by a Galerkin functional, see Section 3.2. Therefore, the discrete counterpart of the contact force density $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} \in \boldsymbol{\mathcal{H}}^*$ has to be defined.

In [5, 6, 7] for obstacle and contact problems discretized with continuous finite elements the discrete approximation of the contact force density is defined as a discrete function $\Lambda_{\mathfrak{m}} = \sum_{p \in \mathfrak{N}^C} s_p \phi_p$ with $s_p = \frac{\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \phi_p \boldsymbol{e}_1 \rangle_{-1,1}}{\int \phi_p}$. By definition it is a functional on $\boldsymbol{\mathcal{H}}_0$. We call $\Lambda_{\mathfrak{m}}$ lumped linear residual and s_p are the node values. Recently, this method has been used for discontinuous finite element discretizations in [10, 11]. We denote that $s_p \geq 0$ if $u_{\mathfrak{m},1}(p) = g_{\mathfrak{m}}(p)$ and $s_p = 0$ if $u_{\mathfrak{m},1}(p) < g_{\mathfrak{m}}(p)$, thus fulfilling nodewise a discrete counterpart of the sign and complementarity conditions of $\boldsymbol{\lambda}$ or $\hat{\sigma}_\nu(\boldsymbol{u})$, respectively. As $\Lambda_{\mathfrak{m}} = \sum_{p \in \mathfrak{N}^C} s_p \phi_p$ is a discrete function a complementarity condition cannot be fulfilled in the so-called semi-contact zone which consists of elements having nodes which are in contact and nodes which are not in contact. It is only valid in so-called full-contact areas where $u_{\mathfrak{m},1} = g_{\mathfrak{m}}$ and in non-actual-contact areas where $u_{\mathfrak{m},1} < g_{\mathfrak{m}}$.

For a posteriori error estimation it would be very advantageous, especially for the efficiency and the localization of estimator contributions, if the quasi-discrete contact force density is defined differently for the different areas of full- and semi-contact. Such an approach has been first used for the derivation of an a posteriori error estimator in [12] and applied to the finite element discretization of obstacle and contact problems in [8, 9, 4]. As far as we know, this article is the first one exploiting this approach for the derivation of a posteriori error estimators for a discontinuous Galerkin method for contact problems.

We call an element full-contact element, if $u_{1,\mathfrak{m}}|_{\mathfrak{e}}(p) = g_{\mathfrak{m}}|_{\mathfrak{e}}(p)$ for all contact boundary nodes $p \in \mathfrak{N}_{\mathfrak{e}}^{C}$. This implies that $\hat{u}_{1,\mathfrak{m}}|_{\mathfrak{e}}(p) = g_{\mathfrak{m}}|_{\mathfrak{e}}(p)$ $\forall p \in \mathfrak{N}_{\mathfrak{e}}^{C}$. We call an element semi-contact element if $u_{1,\mathfrak{m}}|_{\mathfrak{e}}(p) = g_{\mathfrak{m}}|_{\mathfrak{e}}(p)$ for at least one node $p \in \mathfrak{N}_{\mathfrak{e}}^{C}$ but not for all nodes. As in the case of discontinuous finite elements the support of a basis function is the element while in the case of continuous finite elements the support of a basis function is the patch ω_p, the definition of full-contact is more localized for the discontinuous than for the continuous finite element discretization, compare [4]. The set of all full-contact elements is denoted by $\mathfrak{M}^{fC}$. The set of all semi-contact elements is denoted by $\mathfrak{M}^{sC}$. The elements of $\mathfrak{M}^{C}$ which are neither in full- nor in semi-contact belong to the set of non-actual contact elements $\mathfrak{M}^{nC}$.

We define the quasi-discrete contact force density $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}$ as follows

$$\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1} := \sum_{\mathfrak{e}\in\mathfrak{M}^C} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_1) \int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \phi_{p,\mathfrak{e}} \tag{8}$$

where $c_{p,\mathfrak{e}}(\varphi_1)$ is defined differently depending on the contact status of the elements. For full-contact elements $\mathfrak{e} \in \mathfrak{M}^{fC}$ and the corresponding contact boundary side $\mathfrak{s} \subset \mathfrak{S}_{\mathfrak{e}}^{C}$ we define for all nodes $p \in \mathfrak{N}_{\mathfrak{e}}^{C} \cap \mathfrak{s}$

$$c_{p,\mathfrak{e}}(\varphi_1) = \frac{\int_{\mathfrak{s}} \varphi_1 \phi_{p,\mathfrak{e}}}{\int_{\mathfrak{s}} \phi_{p,\mathfrak{e}}}. \tag{9}$$

For semi-contact and non-actual contact elements $\mathfrak{e} \in \mathfrak{M}^{sC} \cup \mathfrak{M}^{nC}$ and the corresponding contact boundary side $\mathfrak{s} \subset \mathfrak{S}_{\mathfrak{e}}^{C}$ we define for all nodes $p \in \mathfrak{N}_{\mathfrak{e}}^{C} \cap \mathfrak{s}$

$$c_{p,\mathfrak{e}}(\varphi_1) = \frac{\int_{\mathfrak{b}} \varphi_1 \phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}} \tag{10}$$

where $\mathfrak{b} \subsetneq \mathfrak{s}$, e.g. the side belonging to p obtained by uniform refinement of $\mathfrak{s}$.

Further $s_{p,\mathfrak{e}} := \frac{\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \phi_{p,\mathfrak{e}} \boldsymbol{e}_1 \rangle_{-1,1,\mathfrak{e}}}{\int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \phi_{p,\mathfrak{e}}}$ is the node value of the lumped linear residual. Using integration by parts (6) can be reformulated to

$$\begin{aligned}
&\sum_{\mathfrak{e}\in\mathfrak{M}} \int_{\mathfrak{e}} (\operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}}) + \boldsymbol{f}) \cdot (\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}) - \sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})]\{\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}\} \\
&- \sum_{\mathfrak{s}\in\mathfrak{S}^N} \int_{\mathfrak{s}} (\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}) - \boldsymbol{\pi})\{\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}\} - \sum_{\mathfrak{s}\in\mathfrak{S}^C} \int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\{\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}\} \\
&+ \sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}]\{\hat{\boldsymbol{\sigma}}(\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}})\} + \sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)\{\hat{\boldsymbol{\sigma}}(\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}})\} \\
&- \sum_{\mathfrak{s}\in\mathfrak{S}^I} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}][\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}] - \sum_{\mathfrak{s}\in\mathfrak{S}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)(\boldsymbol{v}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}) \leq 0.
\end{aligned} \tag{11}$$

From now on, we use the abbreviation $\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}}) := \boldsymbol{f} + \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}})$ for the element residual.

For $\mathfrak{e} \in \mathfrak{M}^C$, $p \in \mathfrak{N}_{\mathfrak{e}}^C$ and $\boldsymbol{v}_{\mathfrak{m}} = \boldsymbol{u}_{\mathfrak{m}} - \phi_{p,\mathfrak{e}}\boldsymbol{e}_1 \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}$ in (11), we get the following representation of $\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \phi_{p,\mathfrak{e}}\boldsymbol{e}_1\rangle_{-1,1,\mathfrak{e}}$

$$
\begin{aligned}
&\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \phi_{p,\mathfrak{e}}\boldsymbol{e}_1\rangle_{-1,1,\mathfrak{e}} \\
&= \int_{\mathfrak{e}} r_1(\boldsymbol{u}_{\mathfrak{m}})\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \int_{\mathfrak{s}} [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})]\frac{1}{2}\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^N} \int_{\mathfrak{s}} (\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) - \pi_1)\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\phi_{p,\mathfrak{e}} \\
&\qquad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m},1}]\phi_{p,\mathfrak{e}}\boldsymbol{n}_{\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^D)\phi_{p,\mathfrak{e}} \\
&\qquad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}]\frac{1}{2}\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_1) + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_1) \geq 0.
\end{aligned}
\tag{12}
$$

Thus, $s_{p,\mathfrak{e}} \geq 0 \ \forall \mathfrak{e} \in \mathfrak{M}^C$. As for all $\mathfrak{e} \in \mathfrak{M}^{nC}$ one can choose test functions $\boldsymbol{v}_{\mathfrak{m}} = \boldsymbol{u}_{\mathfrak{m}} \pm \varepsilon\phi_{p,\mathfrak{e}}\boldsymbol{e}_1 \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}$ in the variational inequality (11), it follows that $s_{p,\mathfrak{e}} = 0 \ \forall \mathfrak{e} \in \mathfrak{M}^{nC}$.

Inserting this representation (12) of $s_{p,\mathfrak{e}}$ in (8) we get the formulation

$$
\begin{aligned}
&\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1} \\
&= \sum_{\mathfrak{e}\in\mathfrak{M}^C} \Bigg(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} \Bigg(\int_{\mathfrak{e}} (r_1(\boldsymbol{u}_{\mathfrak{m}}))c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \int_{\mathfrak{s}} [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})]\frac{1}{2}c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^N} \int_{\mathfrak{s}} (\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) - \pi_1)c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} \\
&\qquad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m},1}]c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}}\boldsymbol{n}_{\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^D)c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} \\
&\qquad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}]\frac{1}{2}c_{p,\mathfrak{e}}(\varphi_1)\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_1) + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)c_{p,\mathfrak{e}}(\varphi_1)\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_1) \Bigg) \Bigg).
\end{aligned}
\tag{13}
$$

As $\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})_1$ is constant on each element it follows from the definition of $c_{p,\mathfrak{e}}(\varphi_1)$ for full-contact elements that

$$
\sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} = \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\varphi_1\phi_{p,\mathfrak{e}} \quad \forall \mathfrak{e} \in \mathfrak{M}^{fC}.
$$

Exploiting this fact and $s_{p,\mathfrak{e}} = 0 \ \forall \mathfrak{e} \in \mathfrak{M}^{nC}$, we get the alternative formulation

$$
\begin{aligned}
&\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1} \\
&= \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_1) \int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \phi_{p,\mathfrak{e}} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^{fC}} \left(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} \left(\int_{\mathfrak{e}} r_1(\boldsymbol{u}_{\mathfrak{m}}) c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \int_{\mathfrak{s}} [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})] \frac{1}{2} c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} \right.\right. \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{N}} \int_{\mathfrak{s}} (\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) - \pi_1) c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\varphi_1\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m},1}] c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} \boldsymbol{n}_{\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{D}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^{D}) c_{p,\mathfrak{e}}(\varphi_1)\phi_{p,\mathfrak{e}} \\
&\quad \left.\left. + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}] \frac{1}{2} c_{p,\mathfrak{e}}(\varphi_1) \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_1) + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{D}} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^{D}) c_{p,\mathfrak{e}}(\varphi_1) \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_1) \right)\right).
\end{aligned} \tag{14}
$$

We note that the definition (14) of the quasi-discrete contact force density reminds us of the definitions in [12, 9, 4] while the definition (8) is similar to the lumped linear residuals used in e.g. [5, 11]. The difference can only be found in a factor $\rho(\varphi_1) = \left(\frac{\int_{\mathfrak{s}} \phi_{p,\mathfrak{e}}}{\int_{\mathfrak{s}} \varphi_1 \phi_{p,\mathfrak{e}}} \frac{\int_{\mathfrak{b}} \varphi_1 \phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}} \right)$ which is the ratio of the integral means $c_{p,\mathfrak{e}}(\varphi_1)$ over $\mathfrak{s}$ and the subset $\mathfrak{b}$. The contributions for semi-contact elements are weighted by this ratio.

$$
\begin{aligned}
\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1} &= \sum_{\mathfrak{e}\in\mathfrak{M}^{C}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_1) \int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \phi_{p,\mathfrak{e}} \\
&= \sum_{\mathfrak{e}\in\mathfrak{M}^{fC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} s_{p,\mathfrak{e}} \int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \varphi_1\phi_{p,\mathfrak{e}} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} s_{p,\mathfrak{e}} \rho(\varphi_1) \int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \varphi_1\phi_{p,\mathfrak{e}}
\end{aligned}
$$

In the following we will use the representation (8) as it is easier in the subsequent analysis.

3.2. Galerkin functional

Similar to the linear case where the error is bounded by the linear residual we want to exploit the dual norm of the Galerkin functional defined by

$$
\langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{\varphi} \rangle_{-1,1} := a\left(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}\right) + \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi} \right\rangle_{-1,1} \quad \forall \boldsymbol{\varphi} \in \boldsymbol{H}_0^1(\Omega) \tag{15}
$$

for the derivation of upper and lower bounds. From the definition of the Galerkin functional it follows directly

$$\|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1}^2 \lesssim \|\boldsymbol{G}_{\mathfrak{m}}\|_{-1}^2 + \|\boldsymbol{\epsilon}(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}})\|^2 \tag{16}$$

and

$$\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1,\mathfrak{e}} \lesssim \|\boldsymbol{\epsilon}(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\mathfrak{e}}. \tag{17}$$

The latter will be exploited in the proofs of the local lower bounds.

For the upper bound it remains to bound $\|\boldsymbol{\epsilon}(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}})\|^2$ which also occurs in the error norm $\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}}^2$.

From the triangle inequality, Young's inequality, Lemma 1 and the ellipticity of Hooke's tensor follows

$$\begin{aligned}
\|\boldsymbol{\epsilon}(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}})\|^2 &\lesssim \|\boldsymbol{\epsilon}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\|^2 + \|\boldsymbol{\epsilon}(\hat{\boldsymbol{u}}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}})\|^2 \\
&\lesssim \|\boldsymbol{\epsilon}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\|^2 + \underbrace{\left(\sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} [\boldsymbol{u}_{\mathfrak{m}}]^2\right) + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)^2\right)}_{(*)}.
\end{aligned} \tag{18}$$

As the contributions $(*)$ occur in the error norm as well as in the error estimator, it remains to bound further $\|\boldsymbol{\epsilon}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\|^2$ by means of Lemma 1 like in (18), of the boundedness of Hooke's tensor and of the weighted Young's inequality

$$\begin{aligned}
&\|\boldsymbol{\epsilon}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\|^2 \lesssim a(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}}) \\
&= a(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}}) + a(\boldsymbol{u}_{\mathfrak{m}} - \hat{\boldsymbol{u}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}}) \\
&= \langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \rangle_{-1,1} - \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right\rangle_{-1,1} + a(\boldsymbol{u}_{\mathfrak{m}} - \hat{\boldsymbol{u}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}}) \\
&\leq C_1 \left(\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1} + \|\boldsymbol{\epsilon}(\boldsymbol{u}_{\mathfrak{m}} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\|\right) \|\boldsymbol{\epsilon}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\| - \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right\rangle_{-1,1} \\
&\lesssim \frac{C_1}{2} \left(\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1}^2 + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{s}}} [\boldsymbol{u}_{\mathfrak{m}}]^2\right) + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{s}}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)^2\right)\right) \\
&\qquad + \frac{1}{2}\|\boldsymbol{\epsilon}(\boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}})\|^2 - \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right\rangle_{-1,1} \\
&\leq C_1 \left(\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1}^2 + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{s}}} [\boldsymbol{u}_{\mathfrak{m}}]^2\right) + \left(\sum_{\mathfrak{s}\in\mathfrak{S}^D} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{s}}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D)^2\right)\right) \\
&\qquad - 2\left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right\rangle_{-1,1}.
\end{aligned} \tag{19}$$

3.3. Error estimator and main results

The error estimator

$$\eta := \sum_{k=1}^{9} \eta_k, \tag{20}$$

for which we prove efficiency and reliability in the following sections consists of the following contributions:

$$\eta_1 := \left(\sum_{\mathfrak{e}\in\mathfrak{M}} \eta_{1,\mathfrak{e}}^2\right)^{\frac{1}{2}}, \qquad \eta_{1,\mathfrak{e}} := h_{\mathfrak{e}}\|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}}$$

$$\eta_2 := \left(\sum_{\mathfrak{s}\in\mathfrak{S}^I} \eta_{2,\mathfrak{s}}^2\right)^{\frac{1}{2}}, \qquad \eta_{2,\mathfrak{s}} := h_{\mathfrak{s}}^{\frac{1}{2}}\|[\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})]\|_{\mathfrak{s}}$$

$$\eta_3 := \left(\sum_{\mathfrak{s}\in\mathfrak{S}^N} \eta_{3,\mathfrak{s}}^2\right)^{\frac{1}{2}}, \qquad \eta_{3,\mathfrak{s}} := h_{\mathfrak{s}}^{\frac{1}{2}}\|\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}) - \boldsymbol{\pi}\|_{\mathfrak{s}}$$

$$\eta_4 := \left(\sum_{\mathfrak{s}\in(\mathfrak{S}^I\cup\mathfrak{S}^D)} \eta_{4,\mathfrak{s}}^2\right)^{\frac{1}{2}}, \qquad \eta_{4,\mathfrak{s}} := \begin{cases} h_{\mathfrak{s}}^{-\frac{1}{2}}\|[\boldsymbol{u}_{\mathfrak{m}}]\|_{\mathfrak{s}} & \text{if } \mathfrak{s}\in\mathfrak{S}^I \\ h_{\mathfrak{s}}^{-\frac{1}{2}}\|\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D\|_{\mathfrak{s}} & \text{if } \mathfrak{s}\in\mathfrak{S}^D \end{cases}$$

$$\eta_5 := \left(\sum_{\mathfrak{s}\in\mathfrak{S}^C} \sum_{i=2}^{d} \eta_{5,\mathfrak{s}}^2\right)^{\frac{1}{2}}, \qquad \eta_{5,\mathfrak{s}} := h_{\mathfrak{s}}^{\frac{1}{2}}\|\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}$$

$$\eta_6 := \left(\sum_{\mathfrak{e}\in\mathfrak{M}^{nC}\cup\mathfrak{M}^{sC}} \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \eta_{6,\mathfrak{s}}^2\right)^{\frac{1}{2}}, \ \eta_{6,\mathfrak{s}} := h_{\mathfrak{s}}^{\frac{1}{2}}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}$$

$$\eta_7 := \left(\sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \eta_{7,\mathfrak{e}}^2\right)^{\frac{1}{2}}, \qquad \eta_{7,\mathfrak{e}} := \left(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} \int_{\mathfrak{b}} (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}}\right)^{\frac{1}{2}}, \quad \mathfrak{b}\subset\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C$$

$$\eta_8 := \left(\sum_{\mathfrak{e}\in\mathfrak{M}^{sC}\cup\mathfrak{M}^{fC}} \eta_{8,\mathfrak{e}}^2\right)^{\frac{1}{2}}, \ \eta_{8,\mathfrak{e}} := \left(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}((g - g_{\mathfrak{m}})^+) \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}}\right)^{\frac{1}{2}}, \quad \mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C$$

$$\eta_9 := \|(\hat{u}_{\mathfrak{m},1} - g)^+\|_{\frac{1}{2},\Gamma_C}$$

In the following we will make use of the abbreviation

$$d_{p,\mathfrak{e}} := \int_{\mathfrak{b}} (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}} \tag{21}$$

in η_7. We denote the positive part of a function by $\varphi^+ := \max\{\varphi, 0\}$ and the negative part by $\varphi^- := \max\{-\varphi, 0\}$ such that $\varphi = \varphi^+ - \varphi^-$. Note that in η_7, η_9 the smoothing function $\hat{u}_{\mathfrak{m},1}$ occurs. The error estimator contributions are like the ones in [4], except η_4 which is special for discontinuous Galerkin methods. Further, in the absence of any contact, we have $\eta_7 = \eta_8 = \eta_9 = 0$ and η_6 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 $\boldsymbol{\pi} = \mathbf{0}$. If contact occurs this standard residual error estimator for linear equations would overestimate the error because the expected boundary stresses in the actual contact zone are non-zero. Except for the error estimator parts $\eta_{8,\mathfrak{e}}, \eta_9$ which disappear for discrete gap functions we provide lower bounds. For a discussion about the meaning of additional error estimator contributions like $\eta_{8,\mathfrak{e}}, \eta_9$ in the upper bound we refer to [5].

In [11] the error estimator contribution $h_{\mathfrak{e}}^{\frac{1}{2}} \|\sigma_1(\boldsymbol{u}_{\mathfrak{m}}) - \Lambda_{\mathfrak{m}}\|_{\mathfrak{s}}$ occurs instead of η_6. Thus, in [11] there is a non-zero contribution from the contact boundary in the area of full-contact. In the proofs in Section 4 we will see, how these contributions are avoided by splitting the different contributions occurring in the Galerkin functional in a way that cancellations are maximized.

In Sections 4 and 5 we will give the proofs of the following theorems about reliability and efficiency of the error estimator η defined in (20) and its local contributions.

Theorem 1 (Reliability). *The error estimator η provides an upper bound of the error measure (7):*

$$\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1} \lesssim \eta.$$

In a posteriori error estimation $\bar{\boldsymbol{f}}$ and $\bar{\boldsymbol{\pi}}$ denote piecewise constant approximations of $\boldsymbol{f}$ and $\boldsymbol{\pi}$. We recall that $h_{\mathfrak{e}}\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}}$ and $h_{\mathfrak{s}}^{\frac{1}{2}}\|\bar{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\mathfrak{s}}$ are formally of higher order.

Theorem 2 (Efficiency). *For the different local error estimator contributions $\eta_{k,\mathfrak{e}}$ with $k = 1, 5, 6$ the following local lower bounds hold*

$$\eta_{k,\mathfrak{e}} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip},\mathfrak{e}} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\mathfrak{e}} + h_{\mathfrak{e}}\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}}$$

and

$$\eta_{2,\mathfrak{s}} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip},\omega_{\mathfrak{s}}} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\omega_{\mathfrak{s}}} + \sum_{\mathfrak{e}\subset\omega_{\mathfrak{s}}} h_{\mathfrak{e}}\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}}$$

$$\eta_{3,\mathfrak{s}} \lesssim \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip},\mathfrak{e}} + \|\boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1,\mathfrak{e}} + h_{\mathfrak{e}}\|\bar{\boldsymbol{f}} - \boldsymbol{f}\|_{\mathfrak{e}} + h_{\mathfrak{s}}^{\frac{1}{2}}\|\bar{\boldsymbol{\pi}} - \boldsymbol{\pi}\|_{\mathfrak{s}}.$$

Under the assumption that for each $p \in \mathfrak{N}_{\mathfrak{e}}^C$ with $\mathfrak{e} \in \mathfrak{M}^{sC}$ there exists a neighboring interior node $\boldsymbol{p} \in \mathfrak{N}^I$ and for a suitable extension $\bar{g}_{\mathfrak{m}} \in \mathcal{H}_{\mathfrak{m}}$ of $g_{\mathfrak{m}}$ to a finite element function on Ω, the following estimate holds for η_7:

$$\eta_7 \lesssim \|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}}+\|\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}\|_{-1}+\sum_{\mathfrak{e}} h_{\mathfrak{e}}\|\bar{\boldsymbol{f}}-\boldsymbol{f}\|_{\mathfrak{e}}+\sum_{\mathfrak{e}\in\mathfrak{M}^{sC}}\sum_{p\in\mathfrak{N}_{\mathfrak{e}}} h_{\mathfrak{e}}^{\frac{1}{2}} \left\| \left[\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix} \right] \right\|_{\gamma_{p,I}}$$

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

The local estimator contributions

$$\eta_{4,\mathfrak{s}} = h_{\mathfrak{s}}^{-\frac{1}{2}} \|[\boldsymbol{u}_{\mathfrak{m}}]\|_{\mathfrak{s}} \ \forall \mathfrak{s} \in \mathfrak{S}^I \quad \text{and} \quad \eta_{4,\mathfrak{s}} = h_{\mathfrak{s}}^{-\frac{1}{2}} \|\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D\|_{\mathfrak{s}} \ \forall \mathfrak{s} \in \mathfrak{S}^D$$

are part of the norm

$$\|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}} = \left(\sum_{\mathfrak{e} \in \mathfrak{M}} \|\boldsymbol{\epsilon}(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}})\|_{L^2(\mathfrak{e})}^2 + \sum_{\mathfrak{s} \in \mathfrak{S}^I \cup \mathfrak{S}^D} \frac{1}{h_{\mathfrak{s}}} \|[\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}]\|_{L^2(\mathfrak{s})}^2 \right)^{\frac{1}{2}}$$

as $[\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}] = [\boldsymbol{u}_{\mathfrak{m}}]$ for all $\mathfrak{s} \in \mathfrak{S}^I$ and $[\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}] = [\boldsymbol{u}_{\mathfrak{m}}^D - \boldsymbol{u}_{\mathfrak{m}}]$ for all $\mathfrak{s} \in \mathfrak{S}^D$ as we assumed $\boldsymbol{u}^D$ to be continuous and piecewise linear. Thus, the global bound $\eta_4 \leq \|\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}}$ follows directly.

We note that Theorem 2 also deals with the error estimator contributions related to the non-linearity. While for η_6 standard arguments from residual error estimation for linear problems have been adapted, the argumentation for η_7 is different as can be seen in the proofs of Section 5. In [11] an error estimator contribution of the same type as η_7 occurs in the upper bound. We like to emphasize that in the present work a proof for the lower bound in terms of η_7 in both dimensions $d = 2, 3$ is given.

4. Reliability of the error estimator

4.1. Upper bound of the Galerkin functional

In this Section we derive an upper bound of the Galerkin functional tested against a function $\boldsymbol{\varphi} \in \boldsymbol{\mathcal{H}}_0$. First, we use integration by parts, the identity $\int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{v})\boldsymbol{w}] = \int_{\mathfrak{s}} \{\hat{\boldsymbol{\sigma}}(\boldsymbol{v})\}[\boldsymbol{w}] + \int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{v})]\{\boldsymbol{w}\}$ for all $\mathfrak{s} \in \mathfrak{S}^I$ and exploit $[\boldsymbol{\varphi}] = 0$.

$$
\begin{aligned}
&\langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{\varphi}\rangle_{-1,1} \\
&\quad = a(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}}, \boldsymbol{\varphi}) + \left\langle \boldsymbol{\lambda} - \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi}\right\rangle_{-1,1} \\
&\quad = \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle + \sum_{\mathfrak{e}\in\mathfrak{M}} \int_{\mathfrak{e}} \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi}] \\
&\qquad\quad + \sum_{\mathfrak{s}\in\mathfrak{S}^N} \int_{\mathfrak{s}} (\boldsymbol{\pi} - \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}))\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}^C} \int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} - \left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi}\right\rangle_{-1,1} \\
&\quad = \langle \boldsymbol{f}, \boldsymbol{\varphi}\rangle + \sum_{\mathfrak{e}\in\mathfrak{M}} \int_{\mathfrak{e}} \operatorname{div}\boldsymbol{\sigma}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} \{\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\} \underbrace{[\boldsymbol{\varphi}]}_{=0} - \sum_{\mathfrak{s}\in\mathfrak{S}^I} \int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})]\{\boldsymbol{\varphi}\} \\
&\qquad\quad + \sum_{\mathfrak{s}\in\mathfrak{S}^N} \int_{\mathfrak{s}} (\boldsymbol{\pi} - \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}))\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}^C} \int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} - \left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi}\right\rangle_{-1,1} \\
&\quad = \sum_{\mathfrak{e}\in\mathfrak{M}} \Bigg(\int_{\mathfrak{e}} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}^I_{\mathfrak{e}}} \int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})]\frac{1}{2}\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}^N_{\mathfrak{e}}} \int_{\mathfrak{s}} (\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}) - \boldsymbol{\pi})\boldsymbol{\varphi} \\
&\qquad\quad - \sum_{\mathfrak{s}\in\mathfrak{S}^C_{\mathfrak{e}}} \int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} \Bigg) - \left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi}\right\rangle_{-1,1} \\
&\qquad\quad - \sum_{\mathfrak{e}\in\mathfrak{M}} \sum_{\mathfrak{s}\in\mathfrak{S}^I_{\mathfrak{e}}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}] \underbrace{[\boldsymbol{\varphi}]}_{0} - \sum_{\mathfrak{e}\in\mathfrak{M}} \sum_{\mathfrak{s}\in\mathfrak{S}^D_{\mathfrak{e}}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}^D_{\mathfrak{m}}) \underbrace{\boldsymbol{\varphi}}_{0} \\
&\quad = \sum_{\mathfrak{e}\in\mathfrak{M}} \Bigg(\sum_{i=1}^{d} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} \Bigg(\int_{\mathfrak{e}} r_i(\boldsymbol{u}_{\mathfrak{m}})\varphi_i\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}^I_{\mathfrak{e}}} \int_{\mathfrak{s}} [\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})]\frac{1}{2}\varphi_i\phi_{p,\mathfrak{e}} \\
&\qquad\quad - \sum_{\mathfrak{s}\in\mathfrak{S}^N_{\mathfrak{e}}} \int_{\mathfrak{s}} (\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}}) - \pi_i)\varphi_i\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}^C_{\mathfrak{e}}} \int_{\mathfrak{s}} \hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})\varphi_i\phi_{p,\mathfrak{e}} \\
&\qquad\quad - \sum_{\mathfrak{s}\in\mathfrak{S}^I_{\mathfrak{e}}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m},i}]\varphi_i\boldsymbol{n}_{\mathfrak{e}}\phi_{p,\mathfrak{e}} - \sum_{\mathfrak{s}\in\mathfrak{S}^D_{\mathfrak{e}}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (u_{\mathfrak{m},i} - u^D_{\mathfrak{m},i})\varphi_i\phi_{p,\mathfrak{e}} \Bigg)\Bigg) \\
&\qquad\quad - \left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}, \boldsymbol{\varphi}\right\rangle_{-1,1}
\end{aligned} \tag{22}
$$

For nodes $p \in \mathfrak{N}\backslash\mathfrak{N}^C$ and all coordinate directions and for nodes $p \in \mathfrak{N}^C$ and the directions $i \neq 1$ we can choose test functions $\boldsymbol{v}_{\mathfrak{m}} = \boldsymbol{u}_{\mathfrak{m}} \pm \phi_{p,\mathfrak{e}}\boldsymbol{e}_i \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}}$ in the variational inequality (11) concluding for all elements $\langle \boldsymbol{\lambda}_{\mathfrak{m}}, \phi_{p,\mathfrak{e}}\boldsymbol{e}_i\rangle_{-1,1,\mathfrak{e}} = 0$ as in the linear case without constraints where $\boldsymbol{\lambda}_{\mathfrak{m}}$ is the linear residual. Thus, we can add for the aforementioned nodes and coordinate directions

$$
\langle \boldsymbol{\lambda}_{\mathfrak{m}}, c_{p,\mathfrak{e}}(\varphi_i)\phi_{p,\mathfrak{e}}\boldsymbol{e}_i\rangle_{-1,1} = 0
$$

where

$$c_{p,\mathfrak{e}}(\varphi_i) = \frac{\int_{\mathfrak{e}} \varphi_i \phi_{p,\mathfrak{e}}}{\int_{\mathfrak{e}} \phi_{p,\mathfrak{e}}}. \tag{23}$$

Further, we exploit the definition of $\tilde{\boldsymbol{\lambda}}_{\mathfrak{m}}$ (13) to get

$$\begin{aligned}
&\langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{\varphi}\rangle_{-1,1} \\
&= \sum_{\mathfrak{e}\in\mathfrak{M}} \Bigg(\sum_{i=1}^{d} \Bigg(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}} \Bigg(\int_{\mathfrak{e}} r_i(\boldsymbol{u}_{\mathfrak{m}})(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \int_{\mathfrak{s}} [\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})] \frac{1}{2} (\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{N}} \int_{\mathfrak{s}} (\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}}) - \pi_i)(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \int_{\mathfrak{s}} \hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m},i}](\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}} \boldsymbol{n}_{\mathfrak{e}} \\
&\quad - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{D}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \int_{\mathfrak{s}} (u_{\mathfrak{m},i} - u_{\mathfrak{m},i}^{D})(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}} \Bigg) \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \int_{\mathfrak{s}} [\boldsymbol{u}_{\mathfrak{m}}] \frac{1}{2} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i) \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}} \boldsymbol{e}_i) \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{D}} \int_{\mathfrak{s}} (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^{D}) \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i) \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}} \boldsymbol{e}_i) \Bigg) \Bigg).
\end{aligned}$$

We recall that for $\mathfrak{e} \in \mathfrak{M}^{fC}$ and the direction of constraints which has been chosen $i = 1$ the part $\sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})_i(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}$ vanishes. We like to emphasize that due to the right splitting of the contributions of the Galerkin functional we can exploit cancellation properties. Thus, there will be no contribution in the error estimator stemming from the contact stresses in the full-contact zone.

Applying Hölder's inequality to the Galerkin functional and rearranging the different parts we end up with the bound

$$
\begin{aligned}
&\langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{\varphi}\rangle_{-1,1} \\
&\leq \sum_{\mathfrak{e}\in\mathfrak{M}} \left(\sum_{i=1}^{d} \left(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}} (\|r_i(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} \|(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}\|_{\mathfrak{e}} \right.\right. \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \|[\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})]\|_{\mathfrak{s}} \frac{1}{2} \|(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}\|_{\mathfrak{s}} \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{N}} \|\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}}) - \pi_i\|_{\mathfrak{s}} \|(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}\|_{\mathfrak{s}} \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \|[u_{\mathfrak{m},i}]\|_{\mathfrak{s}} \|(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}\|_{\mathfrak{s}} \\
&\quad \left. + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{D}} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \|u_{\mathfrak{m},i} - u_{\mathfrak{m},i}^{D}\|_{\mathfrak{s}} \|(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}\|_{\mathfrak{s}} \right) \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \|[\boldsymbol{u}_{\mathfrak{m}}]\|_{\mathfrak{s}} \frac{1}{2} \| \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i) \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}} \boldsymbol{e}_i)\|_{\mathfrak{s}} \\
&\quad \left.\left. + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{D}} \|\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^{D}\|_{\mathfrak{s}} \| \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i) \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}} \boldsymbol{e}_i)\|_{\mathfrak{s}} \right)\right) \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^{C}} \left(\sum_{i=2}^{d} \left(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}} \left(\sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \|\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \|(\varphi_i - c_{p,\mathfrak{e}}(\varphi_i))\phi_{p,\mathfrak{e}}\|_{\mathfrak{s}} \right)\right)\right) \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^{nC}\cup\mathfrak{M}^{sC}} \left(\sum_{p\in\mathfrak{N}_{\mathfrak{e}}} \left(\sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \|(\varphi_1 - c_{p,\mathfrak{e}}(\varphi_1))\phi_{p,\mathfrak{e}}\|_{\mathfrak{s}} \right)\right) \qquad (24)
\end{aligned}
$$

The mean values (23) fulfill the standard L^2-approximation properties

$$
\begin{aligned}
\|\varphi_i - c_{p,\mathfrak{e}}(\varphi_i)\|_{\mathfrak{e}} &\lesssim h_{\mathfrak{e}} \|\boldsymbol{\nabla}\varphi_i\|_{\mathfrak{e}} \\
\|\varphi_i - c_{p,\mathfrak{e}}(\varphi_i)\|_{\mathfrak{s}} &\lesssim h_{\mathfrak{e}}^{\frac{1}{2}} \|\boldsymbol{\nabla}\varphi_i\|_{\mathfrak{e}},
\end{aligned} \qquad (25)
$$

e.g., [16]. The L^2-approximation properties hold also for the constants $c_{p,\mathfrak{e}}(\varphi_i)$ defined in (9) and (10) for semi- and full-contact elements, see [17].

It remains to bound $\| \sum_{p\in\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_i)\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{s}}$ in terms of $\|\boldsymbol{\nabla}\varphi_i\|_{\mathfrak{e}}$.

Lemma 2. *For an arbitrary function $\boldsymbol{\varphi}$, an element $\mathfrak{e}$, a node $p \in \mathfrak{N}_{\mathfrak{e}}$ and $c_{p,\mathfrak{e}}(\varphi_i)$ defined as in (23),(9) or (10) for all coordinate directions $\boldsymbol{e}_i$ with $\varphi_i = \boldsymbol{\varphi}\cdot\boldsymbol{e}_i$ the following stability estimate holds*

$$
\| \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i)\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{s}} \lesssim h_{\mathfrak{e}}^{-\frac{1}{2}} \|\nabla\varphi_i\|_{\mathfrak{e}}.
$$

Proof. First we use a scaling argument

$$\| \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i)\hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{s}} \lesssim h_{\mathfrak{e}}^{-\frac{1}{2}} \| \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{e}}. \tag{26}$$

To further bound $\| \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{e}}$ we exploit $0 = \boldsymbol{\sigma}(\boldsymbol{e}_i) = \boldsymbol{\sigma}(\sum_{p\in\mathfrak{e}} \phi_{p,\mathfrak{e}}\boldsymbol{e}_i) = \sum_{p\in\mathfrak{e}} \boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)$ and the local L^2-approximation properties.

$$\begin{aligned}
&\| \sum_{p\in\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{e}}^2 \\
&= \int_{\mathfrak{e}} \left(\sum_{p\in\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i) \right)^2 \\
&\lesssim \int_{\mathfrak{e}} \left| \sum_{p\in\mathfrak{e}} \left(c_{p,\mathfrak{e}}(\varphi_i) - \varphi_i\right) \boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i) \right| \left| \sum_{p\in\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i) \right| \\
&\lesssim \sum_{p\in\mathfrak{e}} \|\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{L^\infty(\mathfrak{e})} \|c_{p,\mathfrak{e}}(\varphi_i) - \varphi_i\|_{L^2(\mathfrak{e})} \left\| \sum_{p\in\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i) \right\|_{L^2(\mathfrak{e})} \\
&\lesssim \frac{1}{h_{\mathfrak{e}}} h_{\mathfrak{e}} \|\nabla\varphi_i\|_{\mathfrak{e}} \| \sum_{p\in\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_i)\boldsymbol{\sigma}(\phi_{p,\mathfrak{e}}\boldsymbol{e}_i)\|_{\mathfrak{e}}.
\end{aligned} \tag{27}$$

Thus, combining (26) and (27) we end up with the desired result. ☐

Lemma 3. *The Galerkin functional defined in (15) and the error estimator contributions defined in Section 3.3 satisfy*

$$\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1} \lesssim \left(\sum_{k=1}^{6} \eta_k^2 \right)^{\frac{1}{2}}. \tag{28}$$

Proof. Applying in (24) the result of Lemma 2, the L^2-approximation properties

Hölder's inequality and the shape regularity, we get

$$
\begin{aligned}
&\langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{\varphi} \rangle \\
&\lesssim \left(\left(\sum_{\mathfrak{e}\in\mathfrak{M}} (h_{\mathfrak{e}} \|\boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}})^2 \right)^{\frac{1}{2}} + \left(\sum_{\mathfrak{e}\in\mathfrak{M}} \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \|[\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})]\|_{\mathfrak{s}} \right)^2 \right)^{\frac{1}{2}} \right. \\
&\quad + \left(\sum_{\mathfrak{e}\in\mathfrak{M}} \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^N} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \|\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}) - \boldsymbol{\pi}\|_{\mathfrak{s}} \right)^2 \right)^{\frac{1}{2}} \\
&\quad + \left(\sum_{\mathfrak{e}\in\mathfrak{M}} \left(\sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \left(h_{\mathfrak{e}}^{-\frac{1}{2}} \|[\boldsymbol{u}_{\mathfrak{m}}]\|_{\mathfrak{s}} \right)^2 + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \left(h_{\mathfrak{e}}^{-\frac{1}{2}} \|\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D\|_{\mathfrak{s}} \right)^2 \right) \right)^{\frac{1}{2}} \\
&\quad + \left(\sum_{\mathfrak{e}\in\mathfrak{M}^C} \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \sum_{i=2}^{n} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \|\hat{\sigma}_i(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \right)^2 \right)^{\frac{1}{2}} \\
&\quad \left. + \left(\sum_{\mathfrak{e}\in\mathfrak{M}^{nC}\cup\mathfrak{M}^{sC}} \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^C} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} \right)^2 \right)^{\frac{1}{2}} \right) \|\nabla\boldsymbol{\varphi}\|_{L^2(\Omega)} \\
&\lesssim \left(\sum_{k=1}^{6} \eta_k^2 \right)^{\frac{1}{2}} \|\boldsymbol{\varphi}\|_{1,\Omega}. \qquad (29)
\end{aligned}
$$

□

4.2. Upper bound

In Section 3.2 we have shown that in order to give an upper bound of $\|\boldsymbol{\epsilon}(\boldsymbol{u} - \boldsymbol{u}_{\mathfrak{m}})\|^2$ we have to bound $\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1}$ and $\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} - \boldsymbol{\lambda}_{\mathfrak{m}}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right\rangle$. As we have derived the upper bound of $\|\boldsymbol{G}_{\mathfrak{m}}\|_{-1}$ in the foregoing section it remains to bound $\left\langle \tilde{\boldsymbol{\lambda}}_{\mathfrak{m}} - \boldsymbol{\lambda}, \boldsymbol{u} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right\rangle$. We recall that $\langle \lambda_i, \varphi_i \rangle = \left\langle \tilde{\lambda}_{\mathfrak{m},i}, \varphi_i \right\rangle = 0$ for $i \neq 1$.

First, we consider the case where the gap function is discrete $g = g_{\mathfrak{m}}$.

Lemma 4. *We assume $g = g_{\mathfrak{m}}$. Let the continuous approximation $\hat{\boldsymbol{u}}_{\mathfrak{m}}$ defined as in Section 3 and the quasi-discrete contact force density as in Section 3.1 then it holds*

$$\left\langle \tilde{\lambda}_{\mathfrak{m},1} - \lambda_1, u_1 - \hat{u}_{\mathfrak{m},1} \right\rangle_{-1,1} \lesssim \eta_7^2$$

Proof. If $g_{\mathfrak{m}} = g$ it holds $\hat{\boldsymbol{u}}_{\mathfrak{m}} \in \boldsymbol{\mathcal{K}}_{\mathfrak{m}} \subset \boldsymbol{\mathcal{K}}$ and thus $\langle \lambda_1, \hat{u}_{\mathfrak{m},1} - u_1 \rangle_{-1,1} \leq 0$. It remains to bound $\left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_1 - \hat{u}_{\mathfrak{m},1} \right\rangle_{-1,1}$.

Due to the definition of the continuous approximation $\hat{\boldsymbol{u}}_{\mathfrak{m}}$ for the contact boundary side $\mathfrak{s} \in \mathfrak{S}_{\mathfrak{e}}^C$ of a full-contact element $\mathfrak{e} \in \mathfrak{M}^{fC}$ it holds $\hat{u}_{\mathfrak{m},1}|_{\mathfrak{s}} = g_{\mathfrak{m}}|_{\mathfrak{s}}$.

$$
\begin{aligned}
\left\langle \tilde{\lambda}_{\mathfrak{m},1}, u_1 - \hat{u}_{\mathfrak{m},1} \right\rangle_{-1,1} &= \sum_{\mathfrak{e}\in\mathfrak{M}^C} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}(u_1 - \hat{u}_{\mathfrak{m},1}) \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}} \\
&= \sum_{\mathfrak{e}\in\mathfrak{M}^{fC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} \underbrace{s_{p,\mathfrak{e}}}_{\geq 0} \underbrace{\frac{\int_{\mathfrak{s}} (u_1 - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}}}{\int_{\mathfrak{s}} \phi_{p,\mathfrak{e}}}}_{\leq 0} \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} \frac{\int_{\mathfrak{b}} (u_1 - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}} \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^{nC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} \underbrace{s_{p,\mathfrak{e}}}_{=0} \frac{\int_{\mathfrak{b}} (u_1 - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}} \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}} \\
&\lesssim \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} \left(\underbrace{\frac{\int_{\mathfrak{b}} (u_1 - g_{\mathfrak{m}})\phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}}}_{\leq 0} + \frac{\int_{\mathfrak{b}} (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}} \right) \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}} \\
&\lesssim \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} \frac{\int_{\mathfrak{b}} (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}} \int_{\mathfrak{s}} \phi_{p,\mathfrak{e}} \\
&\lesssim \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} d_{p,\mathfrak{e}}
\end{aligned}
$$

□

Second, we consider the general case for an arbitrary gap function $g \in H^{\frac{1}{2}}(\Gamma_C)$. Thus, in contrast to the foregoing case $\hat{u}_{\mathfrak{m},1} \leq g$ may not hold. We start with the same idea as in [4, Section 4.3] applied to the continuous approximation $\hat{u}_{\mathfrak{m},1}$. We define

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

and a harmonic extension $\tilde{w}$ of $w := \hat{u}_{\mathfrak{m},1} - \hat{u}_{\mathfrak{m},1}^* \in H^{\frac{1}{2}}(\Gamma_C)$ so that the stability estimate (see e.g., [18], pp. 70–71)

$$\|\hat{u}_{\mathfrak{m},1} - \hat{u}_{\mathfrak{m},1}^*\|_1 \lesssim \|\hat{u}_{\mathfrak{m},1} - \hat{u}_{\mathfrak{m},1}^*\|_{\frac{1}{2},\Gamma_C} \tag{30}$$

holds. We set $\hat{u}_{\mathfrak{m},1}^* := \hat{u}_{\mathfrak{m},1} - \tilde{w} \in \mathcal{H}$.

Lemma 5. *Let $g \in H^{\frac{1}{2}}(\Gamma_C)$. The continuous approximation $\hat{\boldsymbol{u}}_{\mathfrak{m}}$ is defined as in Section 3 and the quasi-discrete contact force density as in Section 3.1. Then it holds*

$$\left\langle \tilde{\lambda}_{\mathfrak{m},1} - \lambda_1, u_1 - \hat{u}_{\mathfrak{m},1} \right\rangle_{-1,1} \lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|_{-1}^2 + \eta_7^2 + \eta_8^2 + \eta_9^2$$

Proof. The upper bound of $\langle\lambda_1, \hat{u}_{\mathfrak{m},1} - u_1\rangle$ follows as in [4, Section 4.3]

$$\begin{aligned}
\langle\lambda_1, \hat{u}_{\mathfrak{m},1} - u_1\rangle_{-1,1} &= \left\langle\lambda_1, \hat{u}_{\mathfrak{m},1} - \hat{u}^*_{\mathfrak{m},1}\right\rangle_{-1,1} + \underbrace{\left\langle\lambda_1, \hat{u}^*_{\mathfrak{m},1} - u_1\right\rangle_{-1,1}}_{\leq 0} \\
&= \left\langle\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}, \hat{u}_{\mathfrak{m},1} - \hat{u}^*_{\mathfrak{m},1}\right\rangle_{-1,1} + \left\langle\tilde{\lambda}_{\mathfrak{m},1}, \hat{u}_{\mathfrak{m},1} - \hat{u}^*_{\mathfrak{m},1}\right\rangle_{-1,1} \\
&\lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|\hat{u}_{\mathfrak{m},1} - \hat{u}^*_{\mathfrak{m},1}\|^2_{\frac{1}{2},\Gamma_C} + \left\langle\tilde{\lambda}_{\mathfrak{m},1}, \hat{u}_{\mathfrak{m},1} - \hat{u}^*_{\mathfrak{m},1}\right\rangle_{-1,1}.
\end{aligned}$$

Thus,

$$\left\langle\tilde{\lambda}_{\mathfrak{m},1} - \lambda_1, u_1 - \hat{u}_{\mathfrak{m},1}\right\rangle_{-1,1} \lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|\hat{u}_{\mathfrak{m},1} - \hat{u}^*_{\mathfrak{m},1}\|^2_{\frac{1}{2}} + \left\langle\hat{\lambda}_{\mathfrak{m},1}, u_1 - \hat{u}^*_{\mathfrak{m},1}\right\rangle_{-1,1}.$$

For further estimations we use the following identities on the contact boundary

$$u_1 - \hat{u}^*_{\mathfrak{m},1} = (u_1 - g) + (g - \hat{u}^*_{\mathfrak{m},1}) \leq (g - \hat{u}_{\mathfrak{m},1})^+$$

and

$$(g - \hat{u}_{\mathfrak{m},1})^+ \leq (g - g_{\mathfrak{m}})^+ + (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})^+ \leq (g - g_{\mathfrak{m}})^+ + (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1}).$$

Putting all together we end up with

$$\begin{aligned}
\left\langle\tilde{\lambda}_{\mathfrak{m},1} - \lambda_1, u_1 - \hat{u}_{\mathfrak{m},1}\right\rangle_{-1,1} &\lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|(\hat{u}_{\mathfrak{m},1} - g)^+\|^2_{\frac{1}{2},\Gamma_C} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^C}\sum_{p\in\mathfrak{N}^C_{\mathfrak{e}}} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}((g - \hat{u}_{\mathfrak{m},1})^+)\int_{\mathfrak{s}\in\mathfrak{S}^C_{\mathfrak{e}}} \phi_{p,\mathfrak{e}} \\
&\lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \frac{1}{2}\|(\hat{u}_{\mathfrak{m},1} - g)^+\|^2_{\frac{1}{2},\Gamma_C} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^C}\sum_{p\in\mathfrak{N}^C_{\mathfrak{e}}} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}((g - g_{\mathfrak{m}})^+)\int_{\mathfrak{s}\in\mathfrak{S}^C_{\mathfrak{e}}} \phi_{p,\mathfrak{e}} \\
&\quad + \sum_{\mathfrak{e}\in\mathfrak{M}^C}\sum_{p\in\mathfrak{N}^C_{\mathfrak{e}}} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}(g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})\int_{\mathfrak{s}\in\mathfrak{S}^C_{\mathfrak{e}}} \phi_{p,\mathfrak{e}} \\
&\lesssim \frac{1}{2}\|\lambda_1 - \tilde{\lambda}_{\mathfrak{m},1}\|^2_{-1} + \eta_9^2 + \eta_8^2 + \eta_7^2
\end{aligned}$$

□

Theorem 1 is established by inserting the results of Lemmata 3, 4, 5 in (19) and (16), respectively.

5. Efficiency of the error estimator

In this section we give the proof of Theorem 2. Therefore, it is useful to rewrite the definition of the Galerkin functional (15) as follows

$$\begin{aligned}\langle \boldsymbol{G}_{\mathfrak{m}}, \boldsymbol{\varphi}\rangle_{-1,1} = \sum_{\mathfrak{e}\in\mathfrak{M}} \Bigg(\int_{\mathfrak{e}} \boldsymbol{r}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{I}} \int_{\mathfrak{s}} [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})]\frac{1}{2}\boldsymbol{\varphi} \\ - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{N}} \int_{\mathfrak{s}} (\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}}) - \boldsymbol{\pi})\boldsymbol{\varphi} - \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \int_{\mathfrak{s}} \hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})\boldsymbol{\varphi} \Bigg) \\ - \sum_{\mathfrak{e}\in\mathfrak{M}^{C}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} s_{p,\mathfrak{e}} c_{p,\mathfrak{e}}(\varphi_1) \int_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^{C}} \phi_{p,\mathfrak{e}}.\end{aligned}$$

To prove that $\eta_{1,\mathfrak{e}}, \eta_{2,\mathfrak{s}}, \eta_{3,\mathfrak{s}}$ and $\eta_{5,\mathfrak{s}}$ are bounded from above by the error measure (7) (plus data oscillation) the properties of the element bubble functions $\Psi_{\mathfrak{e}}$ and side bubble functions $\Psi_{\mathfrak{s}}$, see, e.g., [19] are used. Due to the definition of the quasi-discrete contact force density, especially the mean values $c_{p,\mathfrak{e}}(\varphi_1)$ for all $p \in \mathfrak{N}_{\mathfrak{e}}^{C}$ and $\mathfrak{e} \in \mathfrak{M}^{C}$, it follows that $c_{p,\mathfrak{e}}(\Psi_{\mathfrak{e}}) = 0$ and $c_{p,\mathfrak{e}}(\Psi_{\mathfrak{s}}) = 0$ for all $\mathfrak{s} \notin \mathfrak{S}^{C}$. Thus, it is obvious that the proof follows as in the case of a linear elliptic problem where $\boldsymbol{G}_{\mathfrak{m}}$ replaces the linear residual.

Next, we consider $\eta_{6,\mathfrak{s}}$ for all contact boundary sides $\mathfrak{s}$ belonging to a semi- or non-actual contact boundary element. Imagine a triangle or intervall $\mathfrak{s}$ to be refined uniformly. The four new triangles or three new intervals are denoted by $\mathfrak{s}_M$ and $\mathfrak{s}_p$ where $\mathfrak{s}_p$ are the new sides belonging to the nodes p and we set $\mathfrak{b}$, occuring in the definition of the mean values $c_{p,\mathfrak{e}}(\varphi_i) = \frac{\int_{\mathfrak{b}} \varphi_i \phi_{p,\mathfrak{e}}}{\int_{\mathfrak{b}} \phi_{p,\mathfrak{e}}}$ to $\mathfrak{b} := \mathfrak{s}_p$. We define a bubble function $\theta_{\mathfrak{s}} = a_M \Psi_M$ belonging to $\mathfrak{s}_M$. Thus, $c_{p,\mathfrak{e}}(\theta_{\mathfrak{s}}) = 0$ as $\theta_{\mathfrak{s}} = 0$ on $\mathfrak{b}$. The coefficient a_M is chosen such that $\int_{\mathfrak{s}} 1 = \int_{\mathfrak{s}} a_M \Psi_M$. Thus,

$$\begin{aligned}\|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}^2 &= \int_{\mathfrak{s}} \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}} \\ &= -\langle \boldsymbol{G}_{\mathfrak{m}}, \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\boldsymbol{e}_1\rangle_{-1,1,\mathfrak{e}} + \int_{\mathfrak{e}} r_1(\boldsymbol{u}_{\mathfrak{m}})\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}} \\ &\quad + \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^{C}} s_{p,\mathfrak{e}}\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})|_{\mathfrak{s}} \underbrace{c_{p,\mathfrak{e}}(\theta_{\mathfrak{s}})}_{=0} \int_{\mathfrak{s}\in\mathfrak{S}_{e}^{C}} \phi_{p,\mathfrak{e}} \\ &\le \|\boldsymbol{G}_{\mathfrak{m}}\|_{-1,\mathfrak{e}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\|_{1,\mathfrak{e}} + \|r_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\theta_{\mathfrak{s}}\|_{\mathfrak{e}} \\ &\lesssim \|\boldsymbol{G}_{\mathfrak{m}}\|_{-1,\mathfrak{e}} h_{\mathfrak{s}}^{-\frac{1}{2}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}} + h_{\mathfrak{s}}^{\frac{1}{2}} \|r_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{e}} \|\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})\|_{\mathfrak{s}}.\end{aligned}$$

Thus, it follows from Lemma 3 and from the bound of $\eta_{1,\mathfrak{e}}$

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

It remains to bound $\eta_{7,\mathfrak{e}} = (s_{p,\mathfrak{e}} d_{p,\mathfrak{e}})^{\frac{1}{2}}$ in terms of the error measure. If $\mathfrak{e}$ is a full-contact element we have $d_{p,\mathfrak{e}} = 0$ and if we have a non-actual contact element we have $s_{p,\mathfrak{e}} = 0$. In consequence it suffices to consider semi-contact elements and a node p with $s_{p,\mathfrak{e}} > 0$ and $(g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})(p) = 0$. We choose $\hat{q}$ as the node maximizing $(g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})(\hat{q}) > 0$ for all nodes $q \in \mathfrak{s}$ and we define an extension of $g_{\mathfrak{m}}$ to a function in $H^1(\Omega)$ by means of $\bar{g}_{\mathfrak{m}}(q) = \hat{u}_{\mathfrak{m},1}(q)$ for all $q \in \mathfrak{N}_{\mathfrak{e}} \backslash \mathfrak{N}_{\mathfrak{e}}^C$. The function under consideration $(\bar{g}_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})$ is continuous such that we can use arguments from the case of continuous linear finite element functions. Thus, it holds as in [4]

$$(g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})(\hat{q}) \lesssim h_{\mathfrak{e}}^{\frac{-d+3}{2}} \left\| \left[\hat{\boldsymbol{\sigma}} \left(\begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right) \right] \right\|_{\gamma_{p,I}} .$$

In a next step we apply the result of Lemma 1 for discontinuous finite elements

$$\begin{aligned}
(g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})(\hat{q}) &\lesssim h_{\mathfrak{e}}^{\frac{-d+3}{2}} \left\| \left[\hat{\boldsymbol{\sigma}} \left(\begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix} - \hat{\boldsymbol{u}}_{\mathfrak{m}} \right) \right] \right\|_{\gamma_{p,I}} \\
&\lesssim h_{\mathfrak{e}}^{\frac{-d+3}{2}} \left(\left\| \left[\hat{\boldsymbol{\sigma}} \left(\begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix} - \boldsymbol{u}_{\mathfrak{m}} \right) \right] \right\|_{\gamma_{p,I}} + \| [\hat{\boldsymbol{\sigma}} (\boldsymbol{u}_{\mathfrak{m}} - \hat{\boldsymbol{u}}_{\mathfrak{m}})] \|_{\gamma_{p,I}} \right) \\
&\lesssim h_{\mathfrak{e}}^{\frac{-d+3}{2}} \left(\left\| \left[\hat{\boldsymbol{\sigma}} \left(\begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix} - \boldsymbol{u}_{\mathfrak{m}} \right) \right] \right\|_{\gamma_{p,I}} + \| [\nabla (\boldsymbol{u}_{\mathfrak{m}} - \hat{\boldsymbol{u}}_{\mathfrak{m}})] \|_{\gamma_{p,I}} \right) \\
&\lesssim h_{\mathfrak{e}}^{\frac{-d+2}{2}} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} + h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} (\boldsymbol{u}_{\mathfrak{m}})] \|_{\gamma_p^I} \right. \\
&\qquad \left. + h_{\mathfrak{e}}^{\frac{1}{2}} h_{\mathfrak{e}}^{-\frac{1}{2}} \| \nabla (\boldsymbol{u}_{\mathfrak{m}} - \hat{\boldsymbol{u}}_{\mathfrak{m}}) \|_{\omega_p} \right) \\
&\lesssim h_{\mathfrak{e}}^{\frac{-d+2}{2}} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} + h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} (\boldsymbol{u}_{\mathfrak{m}})] \|_{\gamma_p^I} \right. \\
&\qquad \left. + \left(\sum_{q \in \omega_p} \sum_{\mathfrak{s} \subset \gamma_q^I} \int_{\mathfrak{s}} \frac{1}{h_{\mathfrak{e}}} [\boldsymbol{u}_{\mathfrak{m}}]^2 \right)^{\frac{1}{2}} \right) .
\end{aligned}$$

For the ease of presentation we here exploited the fact that $\Gamma_C \cap \Gamma_D = \emptyset$. Otherwise contributions from the Dirichlet boundary would occur.

Exploiting the fact that $\hat{q}$ is the node maximizing $(g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})$ on $\mathfrak{s}$ we deduce

$$
\begin{aligned}
d_{p,\mathfrak{e}} &= \int_{\mathfrak{s}_p} (g_{\mathfrak{m}} - \hat{u}_{\mathfrak{m},1})\phi_{p,\mathfrak{e}} \\
&\lesssim h_{\mathfrak{e}}^{d-1} h_{\mathfrak{e}}^{\frac{-d+2}{2}} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} + h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})] \|_{\gamma_p^I} + \left(\sum_{q\in\omega_p} \sum_{\mathfrak{s}\subset\gamma_q^I} \int_{\mathfrak{s}} \frac{1}{h} [\boldsymbol{u}_{\mathfrak{m}}]^2 \right)^{\frac{1}{2}} \right) \\
&\lesssim h_{\mathfrak{e}}^{\frac{d}{2}} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} + h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}}(\boldsymbol{u}_{\mathfrak{m}})] \|_{\gamma_p^I} + \left(\sum_{q\in\omega_p} \sum_{\mathfrak{s}\subset\gamma_q^I} \int_{\mathfrak{s}} \frac{1}{h} [\boldsymbol{u}_{\mathfrak{m}}]^2 \right)^{\frac{1}{2}} \right) \\
&= h_{\mathfrak{e}}^{\frac{d}{2}} \left(h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} + \sum_{\mathfrak{s}\in\gamma_p^I} \eta_{2,\mathfrak{s}} + \sum_{q\in\omega_p} \sum_{\mathfrak{s}\subset\gamma_q^I} \eta_{4,\mathfrak{s}} \right). \qquad (31)
\end{aligned}
$$

For the upper bound of $s_{p,\mathfrak{e}}$ we use the representation (12), Hölder's inequality and scaling arguments

$$
\begin{aligned}
s_{p,\mathfrak{e}} &= \frac{\langle \lambda_{\mathfrak{m}}, \phi_{p,\mathfrak{e}} \rangle_{-1,1,\mathfrak{e}}}{\int_{\mathfrak{s}} \phi_{p,\mathfrak{e}}} \\
&\lesssim h_{\mathfrak{s}}^{-d+1} \Bigg(\| r_1(\boldsymbol{u}_{\mathfrak{m}}) \|_{\mathfrak{e}} \| \phi_{p,\mathfrak{e}} \|_{\mathfrak{e}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^I} \| [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})] \| \| \phi_{p,\mathfrak{e}} \|_{\mathfrak{s}} \\
&\quad + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^N} \| [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) - \pi_1] \| \| \phi_{p,\mathfrak{e}} \|_{\mathfrak{s}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^C} \| \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \| \| \phi_{p,\mathfrak{e}} \|_{\mathfrak{s}} \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \| [u_{\mathfrak{m},1}] \|_{\mathfrak{s}} \| \| \phi_{p,\mathfrak{e}} \|_{\mathfrak{s}} + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \| (u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^D) \|_{\mathfrak{s}} \| \| \phi_{p,\mathfrak{e}} \|_{\mathfrak{s}} \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \| [\boldsymbol{u}_{\mathfrak{m}}] \|_{\mathfrak{s}} \| \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}} \boldsymbol{e}_1) \|_{\mathfrak{s}} + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \| (\boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D) \|_{\mathfrak{s}} \| \hat{\boldsymbol{\sigma}}(\phi_{p,\mathfrak{e}} \boldsymbol{e}_1) \|_{\mathfrak{s}} \Bigg) \\
&\lesssim h_{\mathfrak{e}}^{-d+1} \Bigg(\| r_1(\boldsymbol{u}_{\mathfrak{m}}) \|_{\mathfrak{e}} h_{\mathfrak{e}}^{\frac{d}{2}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^I} \| [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}})] \| h_{\mathfrak{e}}^{\frac{d-1}{2}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^N} \| [\hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) - \pi_1] \| h_{\mathfrak{e}}^{\frac{d-1}{2}} \\
&\quad + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^C} \| \hat{\sigma}_1(\boldsymbol{u}_{\mathfrak{m}}) \| h_{\mathfrak{e}}^{\frac{d-1}{2}} + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \| [u_{\mathfrak{m},1}] \|_{\mathfrak{s}} h_{\mathfrak{e}}^{\frac{d-1}{2}} + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \frac{\hat{\eta}}{h_{\mathfrak{s}}} \| u_{\mathfrak{m},1} - u_{\mathfrak{m},1}^D \|_{\mathfrak{s}} h_{\mathfrak{e}}^{\frac{d-1}{2}} \\
&\quad + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^I} \| [\boldsymbol{u}_{\mathfrak{m}}] \|_{\mathfrak{s}} h_{\mathfrak{e}}^{-1} h_{\mathfrak{e}}^{\frac{d-1}{2}} + \sum_{\mathfrak{s}\in\mathfrak{S}_{\mathfrak{e}}^D} \| \boldsymbol{u}_{\mathfrak{m}} - \boldsymbol{u}_{\mathfrak{m}}^D \|_{\mathfrak{s}} h_{\mathfrak{e}}^{-1} h_{\mathfrak{e}}^{\frac{d-1}{2}} \Bigg) \\
&\lesssim h_{\mathfrak{e}}^{\frac{-d}{2}} \left(\eta_{1,\mathfrak{e}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^I} \eta_{2,\mathfrak{s}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^N} \eta_{3,\mathfrak{s}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^C} \eta_{6,\mathfrak{s}} + \sum_{\mathfrak{s}\in(\mathfrak{S}_{\mathfrak{e}}^I \cup \mathfrak{S}_{\mathfrak{e}}^D)} \eta_{4,\mathfrak{s}} \right). \qquad (32)
\end{aligned}
$$

Combining (31) and (32) and exploiting the local lower bounds in terms of $\eta_{1,\mathfrak{e}}, \eta_{2,\mathfrak{s}}, \eta_{3,\mathfrak{s}}, \eta_{6,\mathfrak{s}}$ and the fact that $\eta_{4,\mathfrak{s}}$ is part of the norm

$$\begin{aligned}
\eta_7^2 &= \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} s_{p,\mathfrak{e}} d_{p,\mathfrak{e}} \\
&\lesssim \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}^C} \Bigg(\eta_{1,\mathfrak{e}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^I} \eta_{2,\mathfrak{s}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^N} \eta_{3,\mathfrak{s}} + \sum_{\mathfrak{s}\subset\mathfrak{S}_{\mathfrak{e}}^C} \eta_{6,\mathfrak{s}} + \sum_{\mathfrak{s}\in(\mathfrak{S}_{\mathfrak{e}}^I\cup\mathfrak{S}_{\mathfrak{e}}^D)} \eta_{4,\mathfrak{s}} \\
&\qquad + \sum_{\mathfrak{s}\in\gamma_p^I} \eta_{2,\mathfrak{s}} + \sum_{q\in\omega_p} \sum_{\mathfrak{s}\subset\gamma_q^I} \eta_{4,\mathfrak{s}} + h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} \Bigg)^2 \\
&\lesssim \Bigg(\|\boldsymbol{u}-\boldsymbol{u}_{\mathfrak{m}}\|_{a,\mathsf{sip}} + \|\boldsymbol{\lambda}-\boldsymbol{\lambda}_{\mathfrak{m}}\|_{-1} + \sum_{\mathfrak{e}\in\mathfrak{M}} h_{\mathfrak{e}} \|\boldsymbol{f}-\bar{\boldsymbol{f}}\|_{\mathfrak{e}} + \sum_{\mathfrak{e}\in\mathfrak{M}^{sC}} \sum_{p\in\mathfrak{N}_{\mathfrak{e}}} h_{\mathfrak{e}}^{\frac{1}{2}} \| [\hat{\boldsymbol{\sigma}} \begin{pmatrix} \bar{g}_{\mathfrak{m}} \\ 0 \end{pmatrix}] \|_{\gamma_p^I} \Bigg)^2 .
\end{aligned}$$

6. Numerical results

The implementation has been carried out in MATLAB. For the adaptive mesh generation we have taken from [20, Chapter 5] the refinement strategy *refineNVB.m* for simplicial meshes and we extended the *provideGeometricData.m*. As solver for the variational inequalities we implemented a primal-dual-active set method similar to [21, Chapter 5.3.1].

In this section we have a look at different contact problems in $2D$. We examine the structure of the refined simplicial meshes and the rate of convergence of the new estimator for adaptive und uniform refinement.

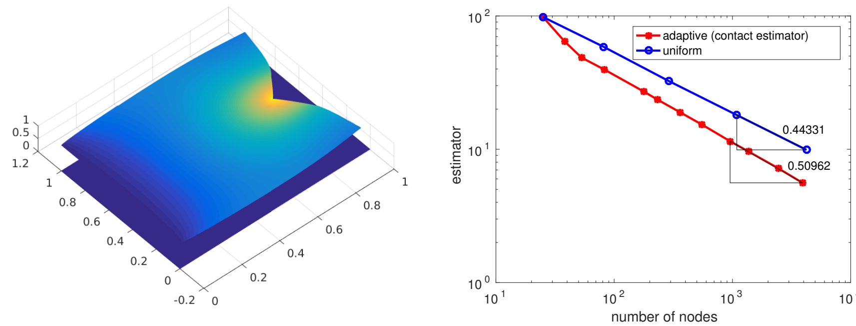
The starting grid has been two times uniformly refined by means of newest vertex bisection, compare [20, Chapter 5] and thus consists of 32 elements. As marking strategy for the adaptive process we use the maximum strategy, i.e. an element is marked for refinement if its local element estimator is bigger than 0.6 times the maximum of all element estimators.

In all our experiments Dirichlet conditions are applied to push the linear elastic body against a rigid obstacle while the force density $\boldsymbol{f}$ and the Neumann values $\boldsymbol{\pi}$ are set to zero. The Poisson ratio is $\nu = 0.3$ and the Young's modulus is $E = 500\frac{kN}{mm^2}$.

6.1. Contact with a rigid wedge

In the first example we simulate the deformation of a linear elastic unit square which is moved in x-direction towards the obstacle $g(y) = -0.2 + 0.5 \cdot |y-0.5|$ which describes a wedge with a semi-angle $\alpha \approx 63$. The Dirichlet values on the left side at $x = 0$ are $u_{\mathfrak{m},1}^D = 0.1$.

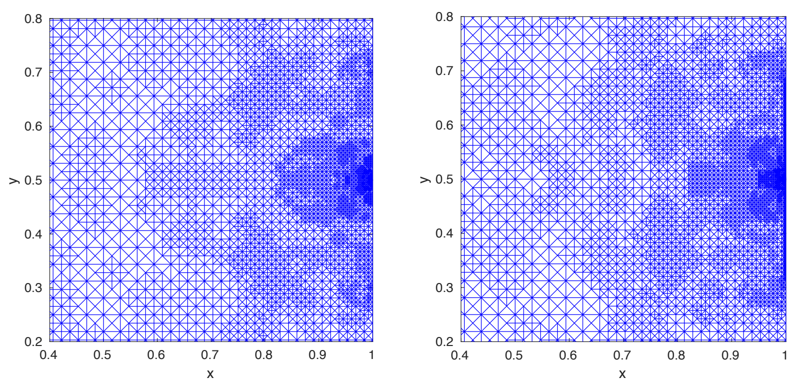
Figure 1 (a) illustrates the deformed unit square above the undeformed geometry. The convergence of the new error estimator for contact on adaptively and uniformly refined grids is shown in Figure 1 (b) with logarithmic scales on



(a) *deformation*

(b) *convergence of estimator*

Figure 1: Wedge example



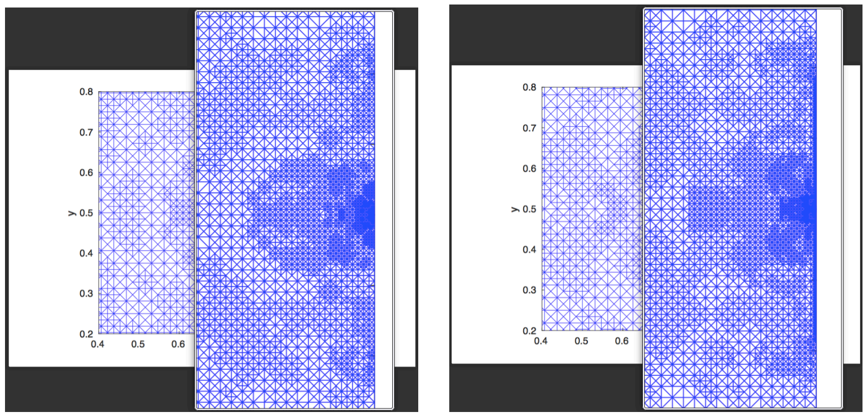
(a) *new estimator for contact*

(b) *standard estimator*

Figure 2: Wedge example: Part of the adaptively refined grid

both axes. The blue line (with circles) refers to uniform refinement. The experimental order of convergence is about 0.443. And the red line (with stars) refers to adaptive refinement steered by our new error estimator for contact problems. The experimental order of convergence is about 0.509.

In Figure 2 the adaptively refined grid in the area $[0.4, 1] \times [0.2, 0.8]$ is shown. The adaptively refined grid steered by the presented estimator for contact (Figure 2 (a)) is refined strongly around the corner caused by the tip of the wedge and at the free boundary. In contrast, the adaptively refined grid steered by the standard residual error estimator for linear elliptic problems (Figure 2(b)) is refined strongly at the whole contact boundary. This can be seen even clearer in Figure 3 where a magnifying glass shows the area around the contact boundary for both methods.



(a) *new estimator for contact* (b) *standard estimator*

Figure 3: Wedge example: Zoom of grid at the contact boundary

6.2. Contact with a rigid block

In the second example we simulate the deformation of a linear elastic unit square which is moved in x-direction towards the obstacle which is a block, coming into contact in the interval $[0.25, 0.75]$. The Dirichlet values on the left side at $x = 0$ are $u^D_{\mathfrak{m},1} = 0.1$.

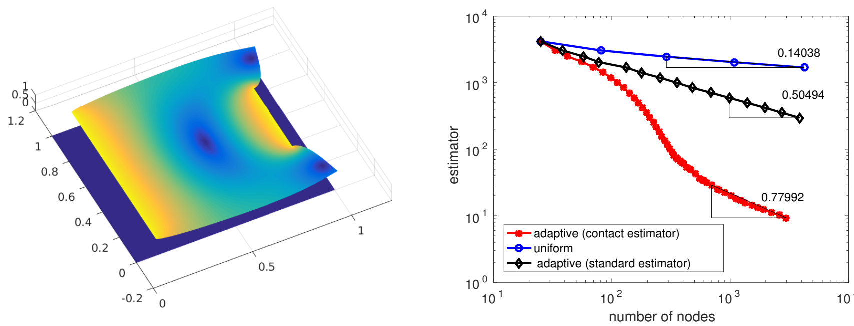
Figure 4 (a) illustrates the deformed unit square above the undeformed geometry. The convergence of the new error estimator for contact on adaptively and uniformly refined grids is shown in Figure 4 (b) with logarithmic scales on both axes. The blue line (with circles) refers to uniform refinement. The black line (with diamonds) refers to adaptive refinement steered by the standard residual error estimator for linear elliptic problems without constraints. And the red line (with stars) refers to adaptive refinement steered by our new error estimator for contact problems.

The adaptively refined grid, steered by the presented estimator for contact (Figure 5(a)) and steered by the standard estimator (Figure 5(b)) can be seen. It is obvious that we have an over-refinement of the contact boundary in Figure 5(b) due to the standard residual error estimator for linear elliptic problems. In Figure 6 we see the area around the contact boundary through a magnifying glass.

6.3. Contact with a parable

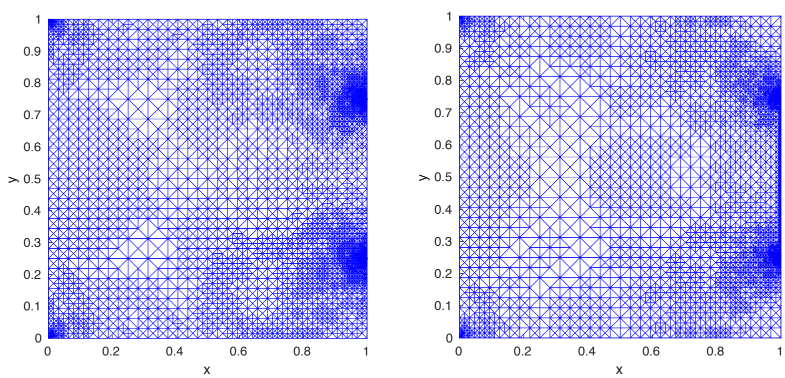
In the third example we simulate the deformation of a linear elastic unit square which is moved in x-direction towards the obstacle which has the form of a parable $g(y) = (y - 0.5)^2 - 0.1$. The Dirichlet values on the left side at $x = 0$ are $u^D_{\mathfrak{m},1} = 0.1$.

Figure 7 (a) illustrates the deformed unit square above the undeformed geometry. The convergence of the new error estimator for contact on adaptively



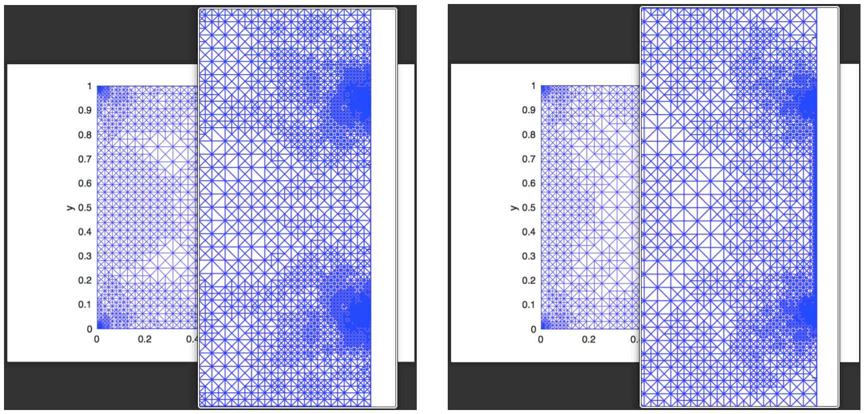
(a) *deformation* (b) *convergence of estimator*

Figure 4: Block example



(a) *new estimator for contact* (b) *standard estimator*

Figure 5: Block example: Adaptively refined mesh

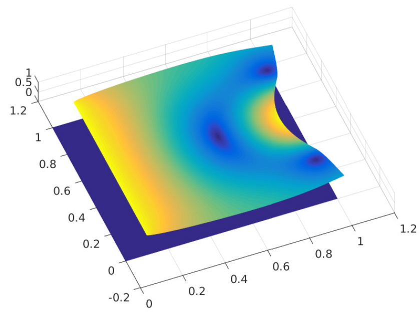


(a) *new estimator for contact* (b) *standard estimator*

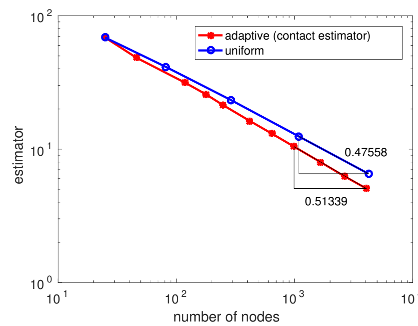
Figure 6: Block example: Zoom of grid at the contact boundary

and uniformly refined grids is shown in Figure 7 (b) with logarithmic scales on both axes. The blue line (with circles) refers to uniform refinement. The experimental order of convergence is about 0.475. And the red line (with stars) refers to adaptive refinement steered by our new error estimator for contact problems. The experimental order of convergence is about 0.513.

In Figure 8 the adaptively refined grid in the area $[0.2, 1] \times [0.1, 0.9]$ is shown. In Figure 8 (a) the adaptive refinement is steered by the new error estimator for contact and in Figure 8 (b) the adaptive refinement is steered by the standard residual error estimator for linear elliptic problems. While in Figure 8 (b) a thin line around the whole contact boundary is highly refined in Figure 8 (a) the contact boundary is not entirely strongly refined. But there are well refined regions close to the boundary which are symmetrically distributed. This refinement comes mainly from the jumps in the stresses and in the solution due to the deformation. The different structure of the grids can be seen even clearer in Figure 9. We denote the number of nodes of a mesh by $\#\mathfrak{N}$ and the number of nodes in the discontinuous finite element setting, meaning the number of pairs $(p, \mathfrak{e})$, is denoted by $\#\mathfrak{N}_{\mathsf{dg}}$. The number of nodes at the contact boundary in the grid shown in Figure 8 (a) amounts to $\#\mathfrak{N}^C = 123$ which corresponds to $\#\mathfrak{N}_{\mathsf{dg}}^C = 244$. The whole number of nodes in the grid represented in Figure 8 (a) is $\#\mathfrak{N}_{\mathsf{dg}} = 39366$. In comparison, the grid shown in Figure 8 (b) consists of $\#\mathfrak{N}_{\mathsf{dg}}^C = 1968$ contact nodes where the whole number of nodes $\#\mathfrak{N}_{\mathsf{dg}} = 38850$ is approximately the same as in Figure 8 (a). In other words we have approximately 8 times less contact boundary nodes using the new a posteriori error estimator for contact which reflects the fact that in the full-contact area the solution is fixed and cannot be improved by adaptive refinement.

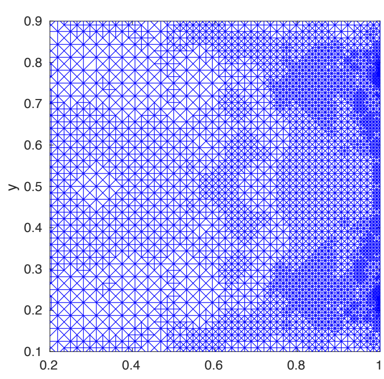


(a) *deformation*

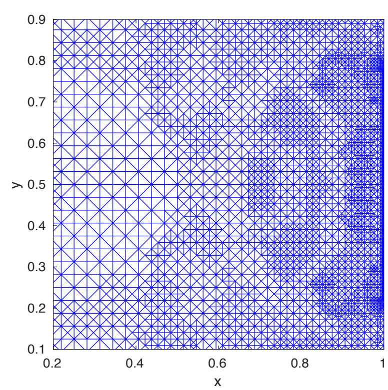


(b) *convergence of estimator*

Figure 7: Parable example

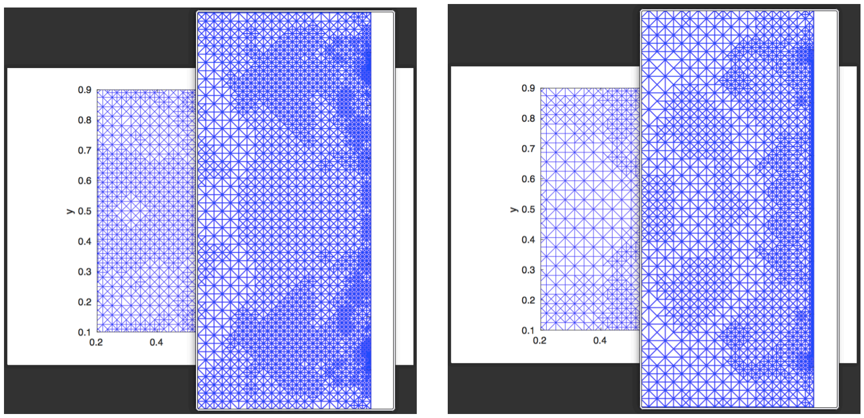


(a) *new estimator for contact*



(b) *standard estimator*

Figure 8: Parable example: Part of the adaptively refined mesh



(a) *new estimator for contact* (b) *standard estimator*

Figure 9: Parable example: Zoom of grid at the contact boundary

References

[1] N. Kikuchi, J. T. Oden, Contact problems in elasticity: a study of variational inequalities and finite element methods, vol. 8 of *SIAM Studies in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, ISBN 0-89871-202-5, doi: 10.1137/1.9781611970845, 1988.

[2] D. A. Di Pietro, A. Ern, Mathematical aspects of discontinuous Galerkin methods, vol. 69 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*, Springer, Heidelberg, ISBN 978-3-642-22979-4, doi: 10.1007/978-3-642-22980-0, 2012.

[3] F. Wang, W. Han, X. Cheng, Discontinuous Galerkin methods for solving the Signorini problem, IMA J. Numer. Anal. 31 (4) (2011) 1754–1772, ISSN 0272-4979, doi:10.1093/imanum/drr010.

[4] R. Krause, A. Veeser, M. Walloth, An efficient and reliable residual-type a posteriori error estimator for the Signorini problem, Numer. Math. 130 (1) (2015) 151–197, ISSN 0029-599X, doi:10.1007/s00211-014-0655-8.

[5] A. Veeser, Efficient and reliable a posteriori error estimators for elliptic obstacle problems, SIAM J. Numer. Anal. 39 (1) (2001) 146–167, ISSN 0036-1429, doi:10.1137/S0036142900370812.

[6] R. H. Nochetto, K. G. Siebert, A. Veeser, Pointwise a posteriori error control for elliptic obstacle problems, Numer. Math. 95 (1) (2003) 163–195, ISSN 0029-599X, doi:10.1007/s00211-002-0411-3.

[7] P. Hild, S. Nicaise, Residual a posteriori error estimators for contact problems in elasticity, M2AN Math. Model. Numer. Anal. 41 (5) (2007) 897–923, ISSN 0764-583X, doi:10.1051/m2an:2007045.

[8] R. H. Nochetto, K. G. Siebert, A. Veeser, Fully localized a posteriori error estimators and barrier sets for contact problems, SIAM J. Numer. Anal. 42 (5) (2005) 2118–2135, ISSN 0036-1429, doi:10.1137/S0036142903424404.

[9] K.-S. Moon, R. H. Nochetto, T. von Petersdorff, C.-s. Zhang, A posteriori error analysis for parabolic variational inequalities, M2AN Math. Model. Numer. Anal. 41 (3) (2007) 485–511, ISSN 0764-583X, doi:10.1051/m2an:2007029.

[10] T. Gudi, K. Porwal, A posteriori error control of discontinuous Galerkin methods for elliptic obstacle problems, Math. Comp. 83 (286) (2014) 579–602, ISSN 0025-5718, doi:10.1090/S0025-5718-2013-02728-7.

[11] T. Gudi, K. Porwal, A posteriori error estimates of discontinuous Galerkin methods for the Signorini problem, J. Comput. Appl. Math. 292 (2016) 257–278, ISSN 0377-0427, doi:10.1016/j.cam.2015.07.008.

[12] F. Fierro, A. Veeser, A posteriori error estimators for regularized total variation of characteristic functions, SIAM J. Numer. Anal. 41 (6) (2003) 2032–2055, ISSN 0036-1429, doi:10.1137/S0036142902408283.

[13] C. Eck, Existenz und Regularität der Lösungen für Kontaktprobleme mit Reibung, Mathematisches Institut A der Universität Stuttgart, Stuttgart, dissertation, Universität Stuttgart, Stuttgart, 1996, 1996.

[14] O. A. Karakashian, F. Pascal, A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems, SIAM J. Numer. Anal. 41 (6) (2003) 2374–2399, ISSN 0036-1429, doi:10.1137/S0036142902405217.

[15] K. Y. Kim, A posteriori error estimators for locally conservative methods of nonlinear elliptic problems, Appl. Numer. Math. 57 (9) (2007) 1065–1080, ISSN 0168-9274, doi:10.1016/j.apnum.2006.09.010.

[16] A. Veeser, R. Verfürth, Explicit upper bounds for dual norms of residuals, SIAM J. Numer. Anal. 47 (3) (2009) 2387–2405, ISSN 0036-1429, doi:10.1137/080738283.

[17] K. G. Siebert, A. Veeser, A unilaterally constrained quadratic minimization with adaptive finite elements, SIAM J. Optim. 18 (1) (2007) 260–289, ISSN 1052-6234, doi:10.1137/05064597X.

[18] O. Steinbach, Numerical approximation methods for elliptic boundary value problems, Springer, New York, ISBN 978-0-387-31312-2, doi:10.1007/978-0-387-68805-3, finite and boundary elements, Translated from the 2003 German original, 2008.

[19] R. Verfürth, A posteriori error estimation techniques for finite element methods, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, ISBN 978-0-19-967942-3, doi: 10.1093/acprof:oso/9780199679423.001.0001, 2013.

[20] S. Funken, D. Praetorius, P. Wissgott, Efficient implementation of adaptive P1-FEM in Matlab, Comput. Methods Appl. Math. 11 (4) (2011) 460–490, ISSN 1609-4840, doi:10.2478/cmam-2011-0026.

[21] S. Bartels, Numerical methods for nonlinear partial differential equations, vol. 47 of *Springer Series in Computational Mathematics*, Springer, Cham, ISBN 978-3-319-13796-4; 978-3-319-13797-1, doi:10.1007/978-3-319-13797-1, 2015.