SIAM J. SCI. COMPUT.
Vol. 30, No. 2, pp. 572–596

A PRIMAL-DUAL ACTIVE SET ALGORITHM FOR THREE-DIMENSIONAL CONTACT PROBLEMS WITH COULOMB FRICTION*

S. HÜEBER†, G. STADLER‡, AND B. I. WOHLMUTH†

Abstract. In this paper, efficient algorithms for contact problems with Tresca and Coulomb friction in three dimensions are presented and analyzed. The numerical approximation is based on mortar methods for nonconforming meshes with dual Lagrange multipliers. Using a nonsmooth complementarity function for the three-dimensional friction conditions, a primal-dual active set algorithm is derived. The method determines active contact and friction nodes and, at the same time, resolves the additional nonlinearity originating from sliding nodes. No regularization and no penalization are applied, and superlinear convergence can be observed locally. In combination with a multigrid method, it defines a robust and fast strategy for contact problems with Tresca or Coulomb friction. The efficiency and flexibility of the method is illustrated by several numerical examples.

Key words. 3D Coulomb friction, contact problems, dual Lagrange multipliers, inexact primal-dual active set strategy, semismooth Newton methods, nonlinear multigrid method

AMS subject classifications. 65K10, 65C20, 65N30, 65N55, 74M15

DOI. 10.1137/060671061

1. Introduction.

Solving contact problems with friction in three dimensions is a challenging task in mechanics and of crucial importance in various applications. The main difficulty lies in the conditions for contact and friction which are inherently nonlinear and complicate the theoretical analysis and the design of an efficient numerical algorithm. For a general introduction on contact problems with and without friction, we refer to [9, 14, 23, 27, 34, 36].

Very often, laws named after Tresca and Coulomb are used to model friction. It is well known that contact with Tresca friction leads to a classical variational inequality. For Coulomb friction, the friction bound depends on the solution, and this more realistic law results in a quasi-variational inequality.

A widely used approach for contact problems with Coulomb friction is to apply a sequence of Tresca friction problems together with a fixed point iteration (see, e.g., [10, 13, 16, 24, 28, 29]). Thus, a crucial component for the solver is a fast and robust algorithm for Tresca frictional contact problems. While in two dimensions, Tresca friction corresponds to linear pointwise inequality constraints for the boundary stresses, the situation in three dimensions is more involved due to the quadratic inequality constraint with which one has to deal.

Contributions to theoretically sound numerical algorithms for friction in three dimensions are quite rare. We refer to the recent papers [8, 16], where finite element tearing and interconnecting (FETI) domain decomposition techniques are combined with quadratic programming methods. The quadratic constraints are approximated

*Received by the editors September 28, 2006; accepted for publication (in revised form) June 19, 2007; published electronically February 14, 2008. This work was supported in part by the Deutsche Forschungsgemeinschaft, SFB 404, B8, and SPP 1146.

http://www.siam.org/journals/sisc/30-2/67106.html

†Institut für Angewandte Analysis und Numerische Simulation (IANS), Universität Stuttgart, Pfaffenwaldring 57, D-70569 Stuttgart, Germany (hueeber@ians.uni-stuttgart.de, wohlmuth@ians.uni-stuttgart.de).

‡Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, 1 University Station, C0200, Austin, TX 78712 (georgst@ices.utexas.edu).

as intersections of rotated squares in order to make the application of optimization algorithms possible. Improvements are proposed in [25, 26]. A different idea is followed in [10, 22], where monotone multigrid methods are used to construct a globally convergent solver. The implementation of these methods relies on a multilevel hierarchy of spaces and requires the use of modified coarse grid basis functions and suitable coarse grid constraints.

Already in an early paper [1], Newton-type methods for contact problems with friction are used. Similar methods are also studied in the more recent contributions [6,7], where the performance of generalized Newton-type methods for frictional contact problems is shown to be superior to interior point methods. The methods presented in [1,6,7] rely on the reformulation of the contact and friction conditions using nonsmooth equations and on generalized differentiability concepts. Here we apply these strategies but use a different complementarity function. Additionally, we study and exploit the structures arising in Newton-type steps and relate them to primal-dual active set strategies. We also apply our algorithm to two-body contact problems which are discretized in terms of mortar techniques and introduce different stabilizations.

Recently, the application of mortar methods as a discretization for the continuous problem has gained a considerable amount of interest (see, e.g., [3,11,12,18,21,27,30]). By using a dual basis function for the Lagrange multiplier, the weak form of the constraints can be written as independent constraints for each node on the slave side, resulting in a quasi-variational inequality. Using a suitable local basis transformation, only the coefficients on the slave side are affected by the nonpenetration condition and the friction constraints. Thus, the algebraic structure of a two-body contact problem is exactly the same as for a one-body contact problem. For details we refer to [21, 35]. In the mixed formulation which is motivated by the mortar approach, the Lagrange multipliers are treated as independent variables. Thus primal-dual active set strategies which use both types of variables are of special interest. We refer to [15], where the interpretation of the active set strategy as a semismooth Newton method is given. These methods have been successfully extended and applied to contact problems without friction [17, 21] and in two dimensions to friction [19, 24, 33]. However, the application to friction problems in three dimensions is not at all straightforward. In this paper, we derive a primal-dual active set method by applying a semismooth Newton method to a nonlinear complementarity function that expresses the three-dimensional (3D) Tresca friction conditions. In each step of the algorithm, we solve a linear problem with Robin boundary conditions on a suitable subset of the friction boundary and with Dirichlet conditions on the complement. This linear system can be solved using any efficient iterative or fast direct solver. Here we apply a multigrid method as an iterative solver. By solving these systems only approximately, we obtain an inexact primal-dual active set strategy. The resulting algorithm can be regarded as a nonlinear multigrid method and yields a reliable and fast convergence. Finally, we present and study a full Newton approach for the Coulomb friction problem. Our implementation is based on the finite element toolbox UG (see [4]), and the direct solver PARDISO (see [32]) is used for the full Newton approach.

The outline of this paper is as follows: The two-body Coulomb frictional contact problem is stated in section 2, and the discretization based on mortar methods is briefly discussed. In section 3, we develop and analyze our algorithm for 3D friction problems. For simplicity of presentation, we restrict ourselves to the case of a one-body contact problem with Tresca friction. In section 4, we study the performance

of the algorithm for a test problem with Tresca friction. Using fixed point ideas, this approach is extended to contact problems with Coulomb friction in section 5. In section 6, we apply our algorithms to two-body contact problems with Coulomb friction in three dimensions. Finally in section 7, we present a full Newton approach for Coulomb friction in three dimensions and compare its performance with the discussed fixed point-based methods.

2. Contact problem with Coulomb friction in three dimensions. In this section, we formulate the contact problem with Coulomb friction in linear elasticity. In addition, we briefly describe our nonconforming discretization based on mortar methods and the applied basis transformation. It guarantees that the algebraic structure of the two-body system is exactly the same as in the case of a one-body contact problem.

2.1. Problem statement. Let $\Omega^l \subset \mathbb{R}^3$, $l \in \{m, s\}$, denote two elastic bodies. The superscript "l" will refer to s for the slave or to m for the master body, as is common in the mortar setting. We assume that the boundaries $\partial\Omega^l$ are divided into three disjoint measurable parts Γ_D^l, Γ_N^l, and Γ_C^l with $\mathrm{meas}(\Gamma_D^l) \neq 0$. We impose homogeneous Dirichlet conditions on Γ_D^l and Neumann data, i.e., a surface traction $\boldsymbol{p}^l \in (L^2(\Gamma_N^l))^3$ on Γ_N^l. Moreover, we denote by $\boldsymbol{f}^l \in (L^2(\Omega^l))^3$ the volume forces acting on Ω^l. We use the constitutive law for linear elasticity, namely,

$$\boldsymbol{\sigma}^l := \lambda^l \mathrm{tr}(\boldsymbol{\varepsilon}^l) Id + 2\mu^l \boldsymbol{\varepsilon}^l \quad \text{in} \quad \Omega^l,$$

where $\boldsymbol{\sigma}^l$ denotes the stress field and $\boldsymbol{\varepsilon}^l := 1/2(\nabla \boldsymbol{u}^l + \nabla \boldsymbol{u}^{l\top})$ the linearized strain tensor that both depend on the displacement $\boldsymbol{u}^l$. The Lamé constants $\lambda^l, \mu^l > 0$ are given by $\mu^l := E^l/(2(1+\nu^l))$ and $\lambda^l := (E^l\nu^l)/((1+\nu^l)(1-2\nu^l))$ with Young's modulus $E^l > 0$ and Poisson ratio $\nu^l \in (0, 0.5)$. Moreover, tr denotes the matrix trace operator and Id the identity matrix. Then the displacement $\boldsymbol{u}^l$ satisfies the following equations:

$$\mathrm{Div}\boldsymbol{\sigma}^l + \boldsymbol{f}^l = 0 \ \text{ in } \Omega^l, \tag{2.1a}$$

$$\boldsymbol{u}^l = 0 \ \text{ on } \Gamma_D^l, \tag{2.1b}$$

$$\boldsymbol{\sigma}^l \boldsymbol{n}^l - \boldsymbol{p}^l = 0 \ \text{ on } \Gamma_N^l, \tag{2.1c}$$

where $\boldsymbol{n}^l$ denotes the unit normal outward vector on the boundary $\partial\Omega^l$. To state the contact and friction conditions, we introduce for each point of Γ_C^s the vectors $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$ that span the tangential plane and use $\boldsymbol{n} := \boldsymbol{n}^s$. We assume that $\{\boldsymbol{n}, \boldsymbol{\tau}_1, \boldsymbol{\tau}_2\}$ is an orthonormal basis in $\mathbb{R}^3$ for each point of Γ_C^s. In order to formulate the nonpenetration condition of the two bodies, we use a predefined relation between the points of the possible contact zones Γ_C^l. This relation is realized by a smooth mapping $R : \Gamma_C^s \to \Gamma_C^m$ satisfying $R(\Gamma_C^s) \subset \Gamma_C^m$. We assume that the mapping R is well-defined and maps any $x \in \Gamma_C^s$ to the intersection of the normal on Γ_C^s at x with Γ_C^m. Then the contact conditions on Γ_C^s are given by

$$[u]_n - d \leq 0, \quad \sigma_n \leq 0, \quad ([u]_n - d)\sigma_n = 0 \quad \text{on } \Gamma_C^s, \tag{2.2}$$

where $d \geq 0$ denotes the gap between the two elastic bodies, $\sigma_n := \boldsymbol{n}^\top \boldsymbol{\sigma}^s \boldsymbol{n}$ the normal component of the boundary stress, $[\boldsymbol{u}] := (\boldsymbol{u}^s(x) - \boldsymbol{u}^m(R(x)))$ the jump and $[u]_n := [\boldsymbol{u}]\boldsymbol{n}$ the jump in the normal direction. For the boundary stress in the normal direction on the possible contact part, we have to satisfy the condition

$$\sigma_n = \boldsymbol{n}^\top \boldsymbol{\sigma}^s(x)\boldsymbol{n} = \boldsymbol{n}^\top \boldsymbol{\sigma}^m(R(x))\boldsymbol{n} \quad \text{on} \quad \Gamma_C^s. \tag{2.3}$$

Additionally, we have to ensure $\boldsymbol{\tau}_k^\top \boldsymbol{\sigma}^s(x)\boldsymbol{\tau}_k = \boldsymbol{\tau}_k^\top \boldsymbol{\sigma}^m(R(x))\boldsymbol{\tau}_k$, $k = 1, 2$. Finally, the Coulomb friction law states that

$$\begin{cases} \|\boldsymbol{\sigma}_\tau\| \le \mathfrak{F}|\sigma_n|, \\ \|\boldsymbol{\sigma}_\tau\| < \mathfrak{F}|\sigma_n| \Rightarrow [\boldsymbol{u}]_\tau = 0, & \text{on } \Gamma_C^s. \\ \|\boldsymbol{\sigma}_\tau\| = \mathfrak{F}|\sigma_n| \Rightarrow \exists \beta \ge 0 : \boldsymbol{\sigma}_\tau = -\beta[\boldsymbol{u}]_\tau, \end{cases} \tag{2.4}$$

Here $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^3$, $\boldsymbol{\sigma}_\tau := \boldsymbol{\sigma}^s\boldsymbol{n} - \sigma_n\boldsymbol{n}$ the tangential component of the boundary stress, and $[\boldsymbol{u}]_\tau$ the relative tangential displacement given by $[\boldsymbol{u}]_\tau := [\boldsymbol{u}] - [\boldsymbol{u}]_n\boldsymbol{n}$. Furthermore, $\mathfrak{F} : \Gamma_C^s \to \mathbb{R}$, $\mathfrak{F} \ge 0$, is the friction coefficient.

2.2. Mortar discretization. By using dual Lagrange multipliers, we can apply locally a suitable basis transformation for the finite element basis. The shape functions on the master side are modified by adding a linear combination of nodal shape functions on the slave side. In the new basis the shape functions on the master side satisfy a weak continuity condition; i.e., the jump of these basis functions tested with a Lagrange multiplier being defined on the slave side is zero. Thus, the coefficients in this new basis on the slave side of the contact zone describe the relative displacement between the contact interfaces. In this basis, the two-body contact problem has the same structure as in the one-body case, since all constraints at the contact zone are restricted to the degrees of freedom on the slave side. For details, see [21, 35]. Motivated by these considerations, in what follows we will write $\boldsymbol{u}_h$ for the coefficient vector with respect to the new constrained basis. Therefore, the following discussion holds for both the one-body and the two-body cases. We denote the multiplier corresponding to the discretization of $-\boldsymbol{\sigma}^s\boldsymbol{n}$ on the slave side by $\boldsymbol{\lambda}_h$. In the new basis, (2.1) is satisfied in the discrete form

$$A_h\boldsymbol{u}_h + B_h\boldsymbol{\lambda}_h = \boldsymbol{f}_h, \tag{2.5}$$

with B_h of the form $(0, D)^\top$, where due to the use of dual Lagrange multipliers the matrix D is diagonal. Let us denote by $\mathcal{S}$ the set of all nodes of the finite element mesh belonging to Γ_C^s and by $\mathcal{N}$ the set containing all other nodes. Then, for each $p \in \mathcal{S}$, the entry of the matrix D is given by

$$D[p, q] = Id_3 \int_{\Gamma_C} \phi_p\psi_q \, ds = Id_3 \int_{\Gamma_C} \phi_p \, ds \, \delta_{pq} =: \delta_{pq} \, D_p Id_3, \tag{2.6}$$

where Id_3 denotes the identity matrix in $\mathbb{R}^{3\times3}$ and ϕ_p and ψ_q the primal and dual basis function associated with the node p, respectively. We remark that in (2.6) the scaled biorthogonality of the sets $\{\phi_p\}_{p\in\mathcal{S}}$ and $\{\psi_p\}_{p\in\mathcal{S}}$ is exploited. If the displacements $\boldsymbol{u}_h$ are known, the Lagrange multiplier can be computed directly from (2.5) as

$$\boldsymbol{\lambda}_h = D^{-1}(\boldsymbol{f}_h - A_h\boldsymbol{u}_h)_{\mathcal{S}}, \tag{2.7}$$

where the subscript $\mathcal{S}$ on the right side indicates that we use only the entries of the vector corresponding to the nodes $p \in \mathcal{S}$.

We next introduce the scaled normal and tangential components of the multiplier $\boldsymbol{\lambda}_h$ and the normal and tangential components of the relative deformation $\boldsymbol{u}_h$. For a node $p \in \mathcal{S}$, let

$$\boldsymbol{\lambda}_{\tau,p,s} := \begin{pmatrix} \boldsymbol{\lambda}_p^\top D_p\boldsymbol{\tau}_{1p} \\ \boldsymbol{\lambda}_p^\top D_p\boldsymbol{\tau}_{2p} \end{pmatrix} \in \mathbb{R}^2 \quad \text{and} \quad \lambda_{n,p,s} := \boldsymbol{\lambda}_p^\top D_p\boldsymbol{n}_p \in \mathbb{R},$$

$$\boldsymbol{u}_{\tau,p} := \begin{pmatrix} \boldsymbol{u}_p^\top\boldsymbol{\tau}_{1p} \\ \boldsymbol{u}_p^\top\boldsymbol{\tau}_{2p} \end{pmatrix} \in \mathbb{R}^2 \quad \text{and} \quad u_{n,p} := \boldsymbol{u}_p^\top\boldsymbol{n}_p \in \mathbb{R}.$$

The discretized (and scaled) gap at the node $p \in \mathcal{S}$ is defined as

$$d_p := \frac{1}{D_p} \int_{\Gamma_C^s} d\, \psi_p \, ds.$$

Then the discrete conditions for the normal contact (2.2) are given by

$$u_{n,p} \le d_p, \quad \lambda_{n,p,s} \ge 0, \quad \lambda_{n,p,s}(u_{n,p} - d_p) = 0, \qquad p \in \mathcal{S}. \tag{2.8}$$

We note that we use different scaling factors for the primal and dual variables and that D_p is proportional to the local mesh size. This is motivated by the fact that the $H^{-1/2}$-norm for the Lagrange multiplier and the $H^{1/2}$-norm for the displacements have the same error reduction. We remark that the proposed scaling factors yield better numerical convergence rates for the inexact version of the algorithms. The discrete Coulomb friction conditions are given by

$$\begin{cases} \|\boldsymbol{\lambda}_{\tau,p,s}\| \le \mathfrak{F}|\lambda_{n,p,s}|, \\ \|\boldsymbol{\lambda}_{\tau,p,s}\| < \mathfrak{F}|\lambda_{n,p,s}| \Rightarrow \boldsymbol{u}_{\tau,p} = 0, & \text{for all } p \in \mathcal{S}. \\ \|\boldsymbol{\lambda}_{\tau,p,s}\| = \mathfrak{F}|\lambda_{n,p,s}| \Rightarrow \exists \beta \ge 0 : \boldsymbol{\lambda}_{\tau,p,s} = \beta \boldsymbol{u}_{\tau,p}, \end{cases} \tag{2.9}$$

Here, for simplicity, $\mathfrak{F}$ is assumed to be constant and independent of the solution. Differently from the Coulomb friction conditions (2.4), where the friction bound is $\mathfrak{F}|\sigma_n|$, for Tresca friction this bound is a given function $g \in H^{-1/2}(\Gamma_C^s)$, with $g \ge 0$. The corresponding discrete bound is

$$g_p := \int_{\Gamma_C^s} g\, \phi_p \, ds;$$

i.e., for Tresca friction, we replace the friction bound $\mathfrak{F}|\lambda_{n,p,s}|$ in (2.9) by g_p.

3. Tresca friction in three dimensions. In this section, we introduce our algorithm for Tresca friction in the 3D case. To simplify the presentation, we restrict ourselves to the case $u_{n,p} = 0$ for all nodes $p \in \mathcal{S}$. In our iterative algorithm, we have to solve in each iteration step a linear problem with boundary conditions of Dirichlet-, Neumann-, or Robin-type on the contact zone Γ_C. The detection of the zones for the different types of boundary conditions is based on a primal-dual approach. We recall that the Lagrange multiplier playing the role of the dual variable can be locally computed in a postprocess from the primal variable $\boldsymbol{u}_h$; see (2.7). This algorithm for 3D friction is closely related to the primal-dual active set method for 2D friction. Consequently, it perfectly fits into the abstract framework used in [19,21] and inherits the advantages of these methods, i.e., their simple implementation and their numerical efficiency. The method can be regarded as an active set strategy or alternatively as a semismooth Newton method guaranteeing fast local convergence; see Theorem 4.1.

3.1. Iterative solver. To start with, we review the conditions for 3D Tresca friction as given by (2.9). Since for $g_p = 0$ these conditions simplify to homogeneous Neumann boundary conditions in the tangential direction, we first assume that $g_p > 0$ and later comment on the case $g_p = 0$. By a straightforward calculation, it can be verified that (2.9) with g_p instead of $\mathfrak{F}|\lambda_{n,p,s}|$ is equivalent to $\mathcal{C}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s}) = 0$ for all $p \in \mathcal{S}$, $c_\tau > 0$, where the nonlinear complementarity function $\mathcal{C}(\cdot\,,\cdot)$ is defined by

$$\mathcal{C}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s}) := \max(g_p, \|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\|)\boldsymbol{\lambda}_{\tau,p,s} - g_p\,(\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p})\,. \tag{3.1}$$

In the following, we take (3.1) as the starting point for our algorithm. Our main idea is to apply a Newton-type algorithm for the solution of $\mathcal{C}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s}) = 0$. Unfortunately, both the Euclidean norm and the max-function are not smooth and not differentiable in the classical sense. However, they are semismooth in the sense of [15,31] which justifies the application of a semismooth Newton method. As generalized derivative for $f(b) := \max(a, b)$, we use $\mathcal{G}_f(b) = 0$ if $a \geq b$ and $\mathcal{G}_f(b) = 1$ if $a < b$. We remark that the arbitrary choice for the case $a = b$ does not influence the convergence rate of the proposed algorithm. Note that, in the first term of (3.1), the Euclidean norm appears for nonzero arguments only. This is due to the fact that if $\|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\| = 0$, we obtain $\max(g_p, \|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\|) = g_p$, and the Euclidean norm vanishes. Therefore, the only nondifferentiability that matters in (3.1) is the max-function. In the semismooth Newton step the derivative of the Euclidean norm occurs only for points that are differentiable in the classical sense.

We now compute, for $p \in \mathcal{S}$, the generalized derivative $\mathcal{G}_\mathcal{C}$ of $\mathcal{C}(\cdot\,,\cdot)$. For the variation $(\delta\boldsymbol{u}_{\tau,p}, \delta\boldsymbol{\lambda}_{\tau,p,s}) \in \mathbb{R}^2 \times \mathbb{R}^2$, we obtain

$$\begin{aligned}\mathcal{G}_\mathcal{C}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s})(\delta\boldsymbol{u}_{\tau,p}, \delta\boldsymbol{\lambda}_{\tau,p,s}) = {} & \max(g_p, \|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\|)\delta\boldsymbol{\lambda}_{\tau,p,s} \\ & + \chi_\mathcal{A} \frac{\boldsymbol{\lambda}_{\tau,p,s}\left(\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\right)^\top}{\|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\|} \left(\delta\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \delta\boldsymbol{u}_{\tau,p}\right) - g_p \left(\delta\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \delta\boldsymbol{u}_{\tau,p}\right).\end{aligned} \tag{3.2}$$

Here $\chi_\mathcal{A}$ denotes the characteristic function of the set $\mathcal{A}_\tau = \{p \in \mathcal{S} : \|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\| > g_p\}$, i.e., $\chi_\mathcal{A} = 1$ if $\|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\| > g_p$ and $\chi_\mathcal{A} = 0$ if $\|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\| \leq g_p$. We note that $\boldsymbol{\lambda}_{\tau,p,s}\left(\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\right)^\top$ is a 2×2-matrix, either zero or of rank one. Performing a semismooth Newton step at a current iterate $(\boldsymbol{u}^{k-1}_{\tau,p}, \boldsymbol{\lambda}^{k-1}_{\tau,p,s})$, one derives the new iterates $(\boldsymbol{u}^{k}_{\tau,p}, \boldsymbol{\lambda}^{k}_{\tau,p,s})$ from

$$\begin{aligned}&\text{Solve} \quad \mathcal{G}_\mathcal{C}(\boldsymbol{u}^{k-1}_{\tau,p}, \boldsymbol{\lambda}^{k-1}_{\tau,p,s})(\delta\boldsymbol{u}^{k}_{\tau,p}, \delta\boldsymbol{\lambda}^{k}_{\tau,p,s}) = -\mathcal{C}(\boldsymbol{u}^{k-1}_{\tau,p}, \boldsymbol{\lambda}^{k-1}_{\tau,p,s}) \\ &\text{and update} \qquad\qquad (\boldsymbol{u}^{k}_{\tau,p}, \boldsymbol{\lambda}^{k}_{\tau,p,s}) = (\boldsymbol{u}^{k-1}_{\tau,p}, \boldsymbol{\lambda}^{k-1}_{\tau,p,s}) + (\delta\boldsymbol{u}^{k}_{\tau,p}, \delta\boldsymbol{\lambda}^{k}_{\tau,p,s}).\end{aligned} \tag{3.3}$$

The characteristic function $\chi_\mathcal{A}$ in (3.2) separates the nodes of $\mathcal{S}$ into the inactive set $\mathcal{I}^k_\tau$ and the active $\mathcal{A}^k_\tau$ set according to

$$\mathcal{I}^k_\tau := \{p \in \mathcal{S} : \|\boldsymbol{\lambda}^{k-1}_{\tau,p,s} + c_\tau \boldsymbol{u}^{k-1}_{\tau,p}\| - g_p \leq 0\}, \tag{3.4a}$$

$$\mathcal{A}^k_\tau := \{p \in \mathcal{S} : \|\boldsymbol{\lambda}^{k-1}_{\tau,p,s} + c_\tau \boldsymbol{u}^{k-1}_{\tau,p}\| - g_p > 0\}. \tag{3.4b}$$

Using this notation and (3.2) in (3.3), a straightforward computation shows that the new iterate $(\boldsymbol{u}^{k}_{\tau,p}, \boldsymbol{\lambda}^{k}_{\tau,p,s})$ satisfies

$$\boldsymbol{u}^k_{\tau,p} = 0 \qquad \text{for } p \in \mathcal{I}^k_\tau, \tag{3.5a}$$

$$\left(Id_2 - M^{k-1}_p\right)\boldsymbol{\lambda}^k_{\tau,p,s} - c_\tau M^{k-1}_p \boldsymbol{u}^k_{\tau,p} = \boldsymbol{h}^{k-1}_p \qquad \text{for } p \in \mathcal{A}^k_\tau, \tag{3.5b}$$

where $M^{k-1}_p := e^{k-1}_p (Id_2 - F^{k-1}_p)$, with the scalar value e^{k-1}_p and the 2×2-matrix F^{k-1}_p given by, respectively,

$$e^{k-1}_p := \frac{g_p}{\|\boldsymbol{\lambda}^{k-1}_{\tau,p,s} + c_\tau \boldsymbol{u}^{k-1}_{\tau,p}\|}, \quad F^{k-1}_p := \frac{\boldsymbol{\lambda}^{k-1}_{\tau,p,s}\left(\boldsymbol{\lambda}^{k-1}_{\tau,p,s} + c_\tau \boldsymbol{u}^{k-1}_{\tau,p}\right)^\top}{g_p \|\boldsymbol{\lambda}^{k-1}_{\tau,p,s} + c_\tau \boldsymbol{u}^{k-1}_{\tau,p}\|}, \tag{3.6}$$

and the vector $\boldsymbol{h}^{k-1}_p \in \mathbb{R}^2$

$$\boldsymbol{h}^{k-1}_p := e^{k-1}_p F^{k-1}_p (\boldsymbol{\lambda}^{k-1}_{\tau,p,s} + c_\tau \boldsymbol{u}^{k-1}_{\tau,p}) = \boldsymbol{\lambda}^{k-1}_p. \tag{3.7}$$

Note that, while on the inactive set $\mathcal{I}_\tau^k$ Dirichlet conditions are imposed, the condition on the active set $\mathcal{A}_\tau^k$ is of Robin type since it involves the displacement $\boldsymbol{u}_{\tau,p}^k$ and the surface traction $\boldsymbol{\lambda}_{\tau,p,s}^k$. The sets $\mathcal{A}_\tau^k$ and $\mathcal{I}_\tau^k$ approximate the sets of slippy and of sticky nodes, respectively. In the more general setting of Signorini conditions, we have also to set the boundary condition in the normal direction. The Robin condition (3.5b) can be easily handled if we rewrite (3.5b) as

$$-\boldsymbol{\lambda}_{\tau,p,s}^k + L_p^{k-1}\boldsymbol{u}_{\tau,p}^k = \boldsymbol{r}_p^{k-1}. \tag{3.8}$$

We note that $\boldsymbol{r}_p^{k-1}$ enters in the right-hand side of the linear system and L_p^{k-1} gives a contribution to the system matrix. Comparing (3.8) with (3.5b) and (3.7), we get, under the assumption that $Id_2 - M_p^{k-1}$ is regular,

$$L_p^{k-1} := c_\tau (Id_2 - M_p^{k-1})^{-1} M_p^{k-1} = c_\tau \big((Id_2 - M_p^{k-1})^{-1} - Id_2 \big), \tag{3.9a}$$

$$\boldsymbol{r}_p^{k-1} := -(Id_2 - M_p^{k-1})^{-1} \boldsymbol{h}_p^{k-1} = \gamma_p^{k-1} \boldsymbol{\lambda}_p^{k-1}, \tag{3.9b}$$

where γ_p^{k-1} is a suitable scaling factor. We mention that this Robin condition only guarantees positive definiteness of the system matrix if L_p^{k-1} is positive definite. Note that for $L_p^{k-1} = 0$ we find a pure Neumann condition. The degeneration of Robin-type to Dirichlet-type boundary conditions is not included in the form (3.9). However, this is not required in our situation, since nodes p with a Dirichlet condition belong to the set $\mathcal{A}_\tau^k$ and therefore are not handled by (3.8). One can easily see that defining the iteration process $Id_2 - M_p^{k-1}$ is not necessarily regular and therefore not invertible. However, in the case of convergence, $Id_2 - M_p^{k-1}$ tends to a positive definite and symmetric matrix. This observation motivates the introduction of three possible modifications of the Robin system (3.5b) such that a regular matrix $Id_2 - M_p^{k-1}$ is obtained. Two of these modifications give a positive definite and symmetric matrix L_p^{k-1}. We remark that all modifications converge in the limit case to the original system (3.5b) and thus preserve the local convergence properties of the algorithm.

3.2. Modifications of the Robin system. To obtain a robust and convergent scheme, we have to replace the matrix F_p^{k-1} by a scaled matrix $\tilde{F}_{p,l}^{k-1}$, $l = 1, 2, 3$. The index l stands for one of the three possibilities considered in this paper. Our numerical results show that the scaling is essential for the robustness of the iteration scheme. According to the definition of M_p^{k-1} and (3.7), we replace in (3.5b) M_p^{k-1} by $\tilde{M}_{p,l}^{k-1}$ and $\boldsymbol{h}_p^{k-1}$ by $\tilde{\boldsymbol{h}}_{p,l}^{k-1}$ given by

$$\tilde{M}_{p,l}^{k-1} := e_p^{k-1}(Id_2 - \tilde{F}_{p,l}^{k-1}), \quad \tilde{\boldsymbol{h}}_{p,l}^{k-1} := e_p^{k-1} \tilde{F}_{p,l}^{k-1} (\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}), \quad l = 1, 2, 3.$$

In a second step, we replace the matrix $Id_2 - M_p^{k-1}$ by $Id_2 - \beta_{p,l}^{k-1} \tilde{M}_{p,l}^{k-1}$ with a scaling factor $\beta_{p,l}^{k-1} > 0$, such that the resulting matrix is regular. Then the form (3.8) of the Robin boundary conditions is written as $-\boldsymbol{\lambda}_{\tau,p,s}^k + \tilde{L}_{p,l}^{k-1} \boldsymbol{u}_{\tau,p}^k = \tilde{\boldsymbol{r}}_{p,l}^{k-1}$, with

$$\tilde{L}_{p,l}^{k-1} := c_\tau \big((Id_2 - \beta_{p,l}^{k-1} \tilde{M}_{p,l}^{k-1})^{-1} - Id_2 \big), \quad \tilde{\boldsymbol{r}}_{p,l}^{k-1} := -(Id_2 - \beta_{p,l}^{k-1} \tilde{M}_{p,l}^{k-1})^{-1} \tilde{\boldsymbol{h}}_{p,l}^{k-1} \tag{3.10}$$

for $l = 1, 2, 3$. We mention that similar modifications as done for the matrix F_p^{k-1} are used in primal-dual algorithms for the minimization of functionals involving Euclidean norms; see, e.g., [2, 5, 20].

Since for all modifications $\tilde{F}_{p,l}^{k-1} \to F_p^{k-1}$ and $\beta_{p,l}^{k-1} \to 1$ as $(\boldsymbol{u}_{\tau,p}^k, \boldsymbol{\lambda}_{\tau,p,s}^k)$ converges to the solution, the modifications do not degrade the local superlinear convergence. Next we present the three possible modifications used in this paper.

First modification. We use a parameter $\beta_{p,1}^{k-1} \neq 1$ in $Id_2 - \beta_{p,1}^{k-1} M_p$ in (3.10) only if $(\boldsymbol{\lambda}_{\tau,p,s}^{k-1})^\top (\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}) < 0$. This condition is equivalent to the fact that the angle between the two vectors is greater than 90 degrees. Since in the limit case both vectors are parallel, the modification applies only when the iterates are far away from the solution. We define the scaled matrix

$$\tilde{F}_{p,1}^{k-1} := \frac{\boldsymbol{\lambda}_{\tau,p,s}^{k-1} \big(\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\big)^\top}{\max\big(g_p, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\|\big) \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\|}. \tag{3.11}$$

Note that $\tilde{F}_{p,1}^{k-1}$ only differs from F_p^{k-1} if $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\| > g_p$, i.e., if the Lagrange multiplier is not in the feasible set given in (2.9).

$$\alpha_p^{k-1} := \frac{(\boldsymbol{\lambda}_{\tau,p,s}^{k-1})^\top \big(\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}\big)}{\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\| \, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}\|}, \qquad \delta_p^{k-1} := \min\left\{ \frac{\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\|}{g_p}, 1 \right\},$$

and it is easy to see that $\gamma_{\tilde{F}_{p,1}}^1 = \alpha_p^{k-1} \delta_p^{k-1}$ is an eigenvalue of $\tilde{F}_{p,1}^{k-1}$ with $\boldsymbol{\lambda}_{\tau,p,s}^{k-1}$ as an eigenvector. The second eigenvalue is given by $\gamma_{\tilde{F}_{p,1}}^2 = 0$, where the corresponding eigenvector is the vector which is orthogonal to $\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}$. Therefore, the eigenvalues of $\tilde{M}_{p,1}^{k-1}$ are $\gamma_{\tilde{M}_{p,1}}^1 = e_p^{k-1}(1 - \alpha_p^{k-1} \delta_p^{k-1})$ and $\gamma_{\tilde{M}_{p,1}}^2 = e_p^{k-1}$. Since $-1 \leq \alpha_p^{k-1} \leq 1$ and $0 \leq \delta_p^{k-1} \leq 1$, we get due to $0 < e_p^{k-1} < 1$ the relation $0 \leq \gamma_{\tilde{M}_{p,1}}^1 < 2$ and $0 < \gamma_{\tilde{M}_{p,1}}^2 < 1$. Using (3.10) with

$$\beta_{p,1}^{k-1} := \begin{cases} \frac{1}{1 - \alpha_p^{k-1} \delta_p^{k-1}} & \text{if } \alpha_p^{k-1} < 0, \\ 1 & \text{otherwise} \end{cases}$$

yields a unsymmetric matrix $\tilde{L}_{p,1}^{k-1}$ with positive eigenvalues.

Second modification. In contrast to the first modification, we use a symmetric matrix $\tilde{F}_{p,2}^{k-1}$. Here we need a parameter $\beta_{p,2}^{k-1} \neq 1$ in (3.10) for all cases with $(\boldsymbol{\lambda}_{\tau,p,s}^{k-1})^\top (\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}) \neq 0$. We replace F_p^{k-1} by the symmetrization of (3.11), namely,

$$\tilde{F}_{p,2}^{k-1} := \frac{\boldsymbol{\lambda}_{\tau,p,s}^{k-1} \big(\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\big)^\top + \big(\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\big) \big(\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\big)^\top}{2 \max\big(g_p, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\|\big) \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\|}. \tag{3.12}$$

One can prove that the eigenvalues of $\tilde{F}_{p,2}^{k-1}$ are $\gamma_{\tilde{F}_{p,2}}^{1,2} = \frac{1}{2}(\alpha_p^{k-1} \pm 1)\delta_p^{k-1} \in [-1, 1]$, and therefore the eigenvalues of the matrix $\tilde{M}_{p,2}^{k-1}$ are $\gamma_{\tilde{M}_{p,2}}^{1,2} = e_p^{k-1}(2 - (\alpha_p^{k-1} \pm 1)\delta_p^{k-1})/2$. Using the same arguments as before, we get $0 \leq \gamma_{\tilde{M}_{p,2}}^{1,2} < 2$. Setting

$$\beta_{p,2}^{k-1} := \frac{2}{2 - (\alpha_p^{k-1} - 1)\delta_p^{k-1}}$$

results in a symmetric and positive definite matrix $\tilde{L}_{p,2}^{k-1}$.

Third modification. In the third modification, we use the matrix

$$\tilde{F}_{p,3}^{k-1} := \frac{\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\big(\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\big)^{\top}}{\max\big(g_p, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\|\big)^2} \tag{3.13}$$

instead of F_p^{k-1}. Obviously this matrix is symmetric and positive semidefinite with eigenvalues $\gamma^1_{\tilde{F}_{p,3}} = 0$ and $\gamma^2_{\tilde{F}_{p,3}} = (\delta_p^{k-1})^2$. Therefore we get for the eigenvalues of the matrix $\tilde{M}_{p,3}^{k-1}$ due to $0 < e_p^{k-1} < 1$ the relation $0 \le \gamma_{\tilde{M}_{p,3}} < 1$, and the matrix $\tilde{L}_{p,3}^{k-1}$ defined by (3.10) with $\beta_{p,3}^{k-1} = 1$ is symmetric and positive definite. We remark that the matrix $\tilde{F}_{p,3}^{k-1}$ converges in the limit case to the matrix F_p^{k-1} since, in the solution for a node $p \in \mathcal{A}_\tau$, we have

$$\frac{\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}}{\|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\|} = \frac{\boldsymbol{\lambda}_{\tau,p,s}}{g_p}.$$

Furthermore as for the unmodified case (3.7) and (3.9b), we observe that $\boldsymbol{h}_p^{k-1}$ and $\boldsymbol{r}_p^{k-1}$ are proportional to $\boldsymbol{\lambda}_p^{k-1}$ for the first and the third modification.

3.3. Extension to the case $g_p = 0$. In the above discussion, we assumed $g_p > 0$ for all $p \in \mathcal{S}$. The reason for this assumption is that for $g_p = 0$ one cannot stringently deduce $\boldsymbol{\lambda}_{\tau,p,s} = 0$ (which follows directly from (2.9)) from $\mathcal{C}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s}) = 0$. However, if a Tresca friction combined with fixed point ideas is used to solve Coulomb friction problems, $g_p = 0$ naturally occurs for all noncontact points, which makes the case $g_p = 0$ rather important. Fortunately nodes p with $g_p = 0$ can also be handled using (3.5) within the setting (3.11)–(3.13).

In the following, we consider $p \in \mathcal{S}$ with $g_p = 0$. First, we assume $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\| > 0$. Then $p \in \mathcal{A}_\tau^k$, since $g_p = 0$. In the case $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\| = 0$ we have to set $p \in \mathcal{A}_\tau^k$. In both cases, we set $\tilde{M}_{p,l}^{k-1} = 0$, and (3.5b) leads to the desired homogeneous Neumann condition $\boldsymbol{\lambda}_{\tau,p,s}^{k} = 0$. We mention that, in the case $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\| > 0$ and $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\| > 0$, we get due to (3.4b) and (3.6) automatically $p \in \mathcal{A}_\tau^k$ and $e_p^{k-1} = 0$, and therefore $\tilde{M}_{p,l}^{k-1} = 0$. In particular, these matrices are well-defined for this case. So only for the cases $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}\| = 0$ or $\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}\| = 0$ do we have to enforce the node p to be in $\mathcal{A}_\tau^{k-1}$.

3.4. General remarks. Computing generalized derivatives of nonsmooth functionals is a delicate issue. While in [1] mainly intuitive arguments are used, the related papers [6, 7] use the concept of Bouligand differentiability. This concept allows the use of globalization (e.g., linesearch) strategies, but calculating the search direction requires one to solve a nonlinear system in each Newton step. In these papers, this problem is circumvented by substituting this nonlinear system by a linear one. Despite this heuristic step, the authors report on good numerical results. The concept of semismoothness [15, 31] used in the present paper has the advantage that the search direction can be found by solving a linear system. Nevertheless, one can also prove local superlinear convergence of the iterates; see Theorem 4.1.

Furthermore, we remark that there is some freedom in choosing the nonlinear complementarity function to express the complementarity conditions for the Tresca friction law. For $g_p > 0$, we can also work with (3.1) replaced by

$$\bar{\mathcal{C}}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s}) := \boldsymbol{\lambda}_{\tau,p,s} - g_p \frac{(\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p})}{\max(g_p, \|\boldsymbol{\lambda}_{\tau,p,s} + c_\tau \boldsymbol{u}_{\tau,p}\|)}. \tag{3.14}$$

Complementarity functions closely related to (3.14) for dealing with friction conditions have been used in [1, 6, 7]. A semismooth Newton iteration for the solution of $\bar{\mathcal{C}}(\boldsymbol{u}_{\tau,p}, \boldsymbol{\lambda}_{\tau,p,s}) = 0$ results in an iteration rule that also uses the active and inactive sets defined in (3.4) but results in a modified iteration step on $\mathcal{A}_\tau^k$. Our numerical experience yields that algorithms based on (3.1) perform more robustly compared to those based on (3.14).

3.5. Algebraic representation. Now we give the matrix representation of the algebraic system we have to solve in each iteration step. As mentioned above, we restrict ourselves to the case $u_{n,p} = 0$. The tangential conditions are either Dirichlet conditions for the inactive nodes $p \in \mathcal{I}_\tau^k$ or Robin conditions for the active nodes $p \in \mathcal{A}_\tau^k$. To rotate the stiffness matrix A arising from standard linear elasticity and introduced in (2.5), we define the matrix with the normal vectors by

$$N := \begin{pmatrix} & & \ddots & 0 & 0 & 0 & & & \\ 0 & \cdots & 0 & & \boldsymbol{n}_p^\top & & 0 & \cdots & 0 \\ & & & 0 & 0 & 0 & \ddots & & \end{pmatrix} \in \mathbb{R}^{|\mathcal{S}|\times 3|\mathcal{S}|}$$

and the matrix with tangential vectors by

$$T := \begin{pmatrix} & & \ddots & 0 & 0 & 0 & & & \\ 0 & \cdots & 0 & & \boldsymbol{\tau}_{1p}^\top & & 0 & \cdots & 0 \\ 0 & \cdots & 0 & & \boldsymbol{\tau}_{2p}^\top & & 0 & \cdots & 0 \\ & & & 0 & 0 & 0 & \ddots & & \end{pmatrix} \in \mathbb{R}^{2|\mathcal{S}|\times 3|\mathcal{S}|}.$$

As above, for each $p \in \mathcal{S}$, $\boldsymbol{n}_p$, $\boldsymbol{\tau}_{1p}$, and $\boldsymbol{\tau}_{2p}$ denote the unit outward normal and the tangential vectors, respectively. Moreover, for $p \in \mathcal{A}_\tau^k$ the matrices $\tilde{L}_{p,l}^{k-1}$ are assembled into

$$L_{\mathcal{A}_\tau^k} := \operatorname{diag}\{\tilde{L}_{p,l}^{k-1}\}_{p\in\mathcal{A}_\tau^k} \in \mathbb{R}^{2|\mathcal{A}_\tau^k|\times 2|\mathcal{A}_\tau^k|}.$$

Similarly, we define the vector $\boldsymbol{r}_{\mathcal{A}_\tau^k}$ by the entries $\tilde{\boldsymbol{r}}_{p,l}^{k-1}$ for $p \in \mathcal{A}_\tau^k$. We now use the decomposition of $\mathcal{S}$ into $\mathcal{A}_\tau^k$ and $\mathcal{I}_\tau^k$. The vector $\boldsymbol{u}_h$ can then be partitioned into $(\boldsymbol{u}_{\mathcal{N}}^\top, \boldsymbol{u}_{\mathcal{I}_\tau^k}^\top, \boldsymbol{u}_{\mathcal{A}_\tau^k}^\top)^\top$, the multiplier $\boldsymbol{\lambda}_h$ into $(\boldsymbol{\lambda}_{\mathcal{I}_\tau^k}^\top, \boldsymbol{\lambda}_{\mathcal{A}_\tau^k}^\top)^\top$, and the right-hand side $\boldsymbol{f}_h$ into $(\boldsymbol{f}_{\mathcal{N}}^\top, \boldsymbol{f}_{\mathcal{I}_\tau^k}^\top, \boldsymbol{f}_{\mathcal{A}_\tau^k}^\top)^\top$. Correspondingly, we can decompose the matrices A_h, N, and T. Then the arising linear system has the form

$$\begin{pmatrix} A_{\mathcal{N}\mathcal{N}} & A_{\mathcal{N}\mathcal{I}_\tau^k} & A_{\mathcal{N}\mathcal{A}_\tau^k} \\ 0 & Id_{\mathcal{I}_\tau^k} & 0 \\ 0 & 0 & N_{\mathcal{A}_\tau^k} \\ T_{\mathcal{A}_\tau^k} A_{\mathcal{A}_\tau^k\mathcal{N}} & T_{\mathcal{A}_\tau^k} A_{\mathcal{A}_\tau^k\mathcal{I}_\tau^k} & T_{\mathcal{A}_\tau^k} A_{\mathcal{A}_\tau^k\mathcal{A}_\tau^k} + L_{\mathcal{A}_\tau^k} T_{\mathcal{A}_\tau^k} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_{\mathcal{N}}^k \\ \boldsymbol{u}_{\mathcal{I}_\tau^k}^k \\ \boldsymbol{u}_{\mathcal{A}_\tau^k}^k \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_{\mathcal{N}} \\ 0 \\ 0 \\ T_{\mathcal{A}_\tau^k} \boldsymbol{f}_{\mathcal{A}_\tau^k} + \boldsymbol{r}_{\mathcal{A}_\tau^k} \end{pmatrix}. \tag{3.15}$$

We remark that the Dirichlet boundary condition (3.5a) on $\mathcal{I}_\tau^k$ is reflected in the second row and the Robin-type conditions for $p \in \mathcal{A}_\tau^k$ (see (3.5b) or (3.8)) are included in the fourth row. In the case of the more general Signorini condition also used in the next section, we have to replace the second and third rows by more complex ones; see [21].

4. Algorithm and numerical examples for Tresca friction. In this section, we present an efficient primal-dual algorithm based on a multigrid method to solve the resulting linear problems. Furthermore, we study numerically the performance of our algorithm and compare the three modifications of section 3.2. While in the previous section, we have derived the iteration rule for Tresca friction without taking into account Signorini contact, the applied algorithm can handle Tresca friction and Signorini contact (2.2). The discrete Signorini conditions (2.8) are realized in terms of the nonlinear complementarity function

$$\lambda_{n,p,s} - \max(0, \lambda_{n,p,s} + c_n(u_{n,p} - d_p)) = 0, \tag{4.1}$$

with $c_n > 0$. The strategy uses the active and inactive sets

$$\begin{aligned} \mathcal{A}_n^k &:= \{p \in \mathcal{S} : \lambda_{n,p,s}^{k-1} + c_n(u_{n,p}^{k-1} - d_p) > 0\}, \\ \mathcal{I}_n^k &:= \{p \in \mathcal{S} : \lambda_{n,p,s}^{k-1} + c_n(u_{n,p}^{k-1} - d_p) \le 0\}. \end{aligned}$$

We next state the **i**nexact **p**rimal-**d**ual **a**ctive **s**et (IPDAS) strategy for contact with Tresca friction.

Algorithm 1: IPDAS for contact with 3D Tresca friction

(0) Set $k = 1$, choose $c_n > 0$, $c_\tau > 0$, and $m \in \mathbb{N}$.
Initialize $\boldsymbol{u}_h^{0,0}$ and $\boldsymbol{\lambda}_h^0$ as an initial solution.

(1) Define the active and inactive sets by

$$\begin{aligned} \mathcal{A}_n^k &:= \left\{p \in \mathcal{S} \,:\, \lambda_{n,p,s}^{k-1} + c_n\big(u_{n,p}^{k-1,m} - d_p\big) > 0\right\}, \\ \mathcal{I}_n^k &:= \left\{p \in \mathcal{S} \,:\, \lambda_{n,p,s}^{k-1} + c_n\big(u_{n,p}^{k-1,m} - d_p\big) \le 0\right\}, \\ \mathcal{A}_\tau^k &:= \left\{p \in \mathcal{S} \,:\, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1,m}\| - g_p > 0\right\}, \\ \mathcal{I}_\tau^k &:= \left\{p \in \mathcal{S} \,:\, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1,m}\| - g_p \le 0\right\}. \end{aligned}$$

(2) For $i = 1, \ldots, m$, compute

$$\boldsymbol{u}_h^{k,i} = MG(\boldsymbol{u}_h^{k,i-1}, \mathcal{A}_n^k, \mathcal{I}_n^k, \mathcal{A}_\tau^k, \mathcal{I}_\tau^k, \boldsymbol{u}_h^{k-1,m}, \boldsymbol{\lambda}_h^{k-1}).$$

(3) If $\|\boldsymbol{u}_h^{k,m} - \boldsymbol{u}_h^{k,0}\| \,/\, \|\boldsymbol{u}_h^{k,m}\| < \varepsilon_u$, stop.

(4) Compute the Lagrange multiplier as

$$\boldsymbol{\lambda}_h^k = D^{-1}\big(\boldsymbol{f}_{\mathcal{S}} - A_{\mathcal{S}\star}\boldsymbol{u}_h^{k,m}\big).$$

(5) Set $\boldsymbol{u}_h^{k+1,0} = \boldsymbol{u}_h^{k,m}$, $k = k + 1$, and go to step (1).

Above, we denote by $\boldsymbol{u}_h^{k,i} = MG(\boldsymbol{u}_h^{k,i-1}, \mathcal{A}_n^k, \mathcal{I}_n^k, \mathcal{A}_\tau^k, \mathcal{I}_\tau^k, \boldsymbol{u}_h^{k-1,m}, \boldsymbol{\lambda}_h^{k-1})$ the iterate after one multigrid cycle with the actual computational refinement level of the mesh as the top level and refinement level zero as the bottom level for the linear system (3.15). We mention that we start with a given mesh on level 0. To obtain the next finer mesh on level $l + 1$, we decompose each element on level l into 8 subelements. For all examples presented in this paper, we use a $\mathcal{W}$-cycle with three pre- and postsmoothing steps for the multigrid. As a smoother, a symmetric Gauß–Seidel iteration is applied. By $A_{\mathcal{S}\star}$, we denote the rows of the stiffness matrix A corresponding to the nodes in the set $\mathcal{S}$. In the case $m \in \mathbb{N}$, we solve the linear system approximately; i.e., we update the active and inactive sets after m multigrid steps. For $m =$ "∞," we get the

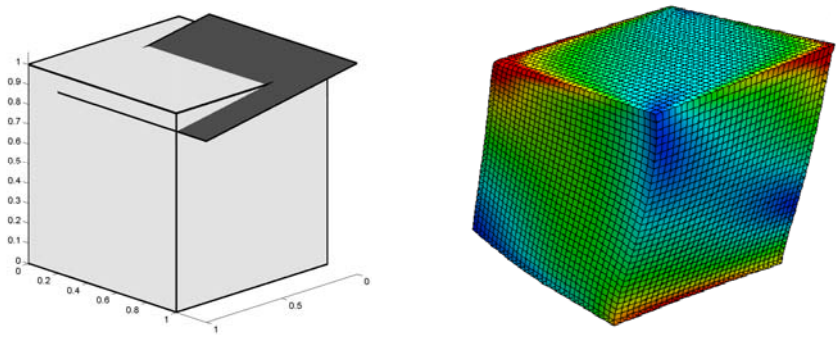

FIG. 4.1. *Problem definition with prescribed Dirichlet distortion (left) and distorted body with effective von Mises stress σ_{eff}; the lower surface in the plot is subject to unilateral contact and Tresca friction.*

exact version of our algorithm. Furthermore we mention that in each Newton step we have to change only the lines belonging to the contact nodes as described by (3.15). Therefore there is no need of reassemble the whole stiffness matrix in each iteration step if we store the original part $A_{\mathcal{S}\star}$. Recalling the derivation of our iteration rule as a semismooth Newton method, we obtain the following local convergence result; see [15, 31].

THEOREM 4.1. *For $m =$ "∞," Algorithm 1 converges locally superlinear.*

Our numerical experience reveals that Algorithm 1 also converges globally, and thus, there is no need for a globalization strategy. This has also been observed for the application of primal-dual active set strategies for related problems in, e.g., [17, 19, 21, 24]. A rigorous analysis for a simper model can be found in [15].

4.1. Example. As a first example for a 3D contact problem with Tresca friction, we consider a one-body contact problem.

The problem setting. We consider a linearly elastic cube $\Omega = [0,1]^3$ with material parameters $E = 200$ and $\nu = 0.3$ and take the xy-plane as the rigid obstacle which implies that the initial gap is $d = 0$. The cube is subject to the Dirichlet distortion $\boldsymbol{u}^\top = (0,\, 0.2,\, 0.06 - 0.15x)$ on its upper surface $[0,1]^2 \times \{1\}$; see the left part of Figure 4.1. For Tresca friction, the friction bound g is given a priori and does not depend on the distortion as in the case of Coulomb friction. For this problem, we choose $g = 800xy(1-x)(1-y)$. In the right of Figure 4.1, we show the distorted body with the effective von Mises stress σ_{eff} given by $\sigma_{\text{eff}}^2 := \sum_{i,j=1}^3 |\boldsymbol{\sigma}_{ij} - \delta_{ij} p|^2$, where the pressure p is given by $p := \frac{1}{3}\text{tr}(\boldsymbol{\sigma})$.

The Tresca friction law. To get a better understanding of the different types of nodes that occur for contact problems with Tresca friction, we show in Figure 4.2 the nodes of the contact surface on level 5. Different types of nodes are marked differently; see the legend in Figure 4.2. Note that for each node the displacement is parallel to the multiplier $\boldsymbol{\lambda}_\tau$ as required. We remark that, using the Tresca friction law, nodes can stick in the tangential direction without being in contact with the obstacle.

Performance of Algorithm 1. To investigate the performance of the algorithm, we first solve the linear system arising in each iteration step exactly. We note that, in each step, our algorithm updates the contact/noncontact sets $\mathcal{A}_n^k$ and $\mathcal{I}_n^k$ as well as the slip/stick sets $\mathcal{A}_\tau^k$ and $\mathcal{I}_\tau^k$ and at the same time performs a Newton step to adopt the distortion's direction $\boldsymbol{u}_\tau$ to the direction of $\boldsymbol{\lambda}_\tau$ for sliding nodes.

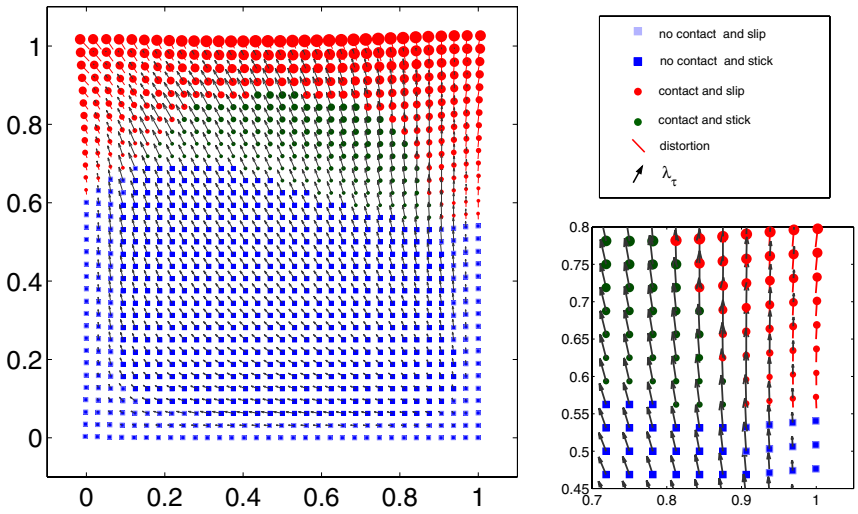

FIG. 4.2. *Visualization of the solution at the nodes $p \in \mathcal{S}$ for Tresca friction on level 5 (left), legend (upper right), and cutout (lower right).*

TABLE 4.1
Performance of Algorithm 1 for the exact strategy ($m =$ "∞"). Initialization: $\boldsymbol{u}_h^{0,0} = \mathbf{0}$, $\boldsymbol{\lambda}_h^0 = \mathbf{0}$ for all levels. Tolerance $\varepsilon_u = 10^{-9}$ with the modification (3.11).

l	$\lvert\mathcal{A}_n^k\rvert \,/\, \lvert\mathcal{A}_\tau^k\rvert$ for $k = 2,3,4,5,6,7,8,9$							
0	2/0	2/0						
1	6/4	5/8	5/8	5/8	5/8	5/8		
2	15/16	12/16	12/16	12/16	12/16			
3	49/32	38/37	33/39	33/39	33/39	33/39		
4	171/66	132/102	116/118	112/122	112/123	112/123	112/123	
5	651/252	506/357	443/411	412/435	406/444	406/446	406/446	406/446

For our tests, we initialize the algorithm on each level with $\boldsymbol{u}_h^{0,0} = \mathbf{0}$ and $\boldsymbol{\lambda}_h^0 = \mathbf{0}$. This leads to $\mathcal{A}_n^1 = \mathcal{A}_\tau^1 = \emptyset$. We terminate the IPDAS iteration if the relative change in the solution is less than 10^{-9}. In the complementarity function, we use $c_n = c_\tau = 100$; Algorithm 1 yields a fast and stable converge on all levels. Table 4.1 shows the number of iterations needed on different refinement levels and the number of nodes belonging to the active sets $\mathcal{A}_n^k$ and $\mathcal{A}_\tau^k$ for the first modification (3.11). We remark that for the second (3.12) and the third modification (3.13) only minor differences occur. Note that in each iteration step one linear system has to be solved. The number of iterations increases only weakly on finer levels; it seems to depend linearly on the level. Usually, after the exact active sets for both friction and contact conditions are found, the method requires about 3–4 more iterations to converge. In these steps, the algorithm adjusts, for $p \in \mathcal{S}$, the direction of the tangential traction $\boldsymbol{\lambda}_{\tau,p}$ to the tangential displacement $\boldsymbol{u}_{\tau,p}$.

Comparison between the modifications. Now we compare the convergence and the behavior of the factors α_p^{k-1} and $\beta_{p,l}^{k-1}$ for the three modifications. In the left of Figure 4.3, we show the errors $\|\boldsymbol{\lambda}_h^k - \boldsymbol{\lambda}_h^\star\|$ on level 5 in a logarithmic plot for the initialization $\boldsymbol{u}_h^{0,0} = (1,\, 1,\, 0)^\top$ and $\boldsymbol{\lambda}_h^0 = (-1,\, -1,\, 0)^\top$. Here $\boldsymbol{\lambda}_h^\star$ denotes the Lagrange multiplier of the solution. We remark that additional iteration steps, compared to Table 4.1, are needed due to the different initialization and the smaller tolerance

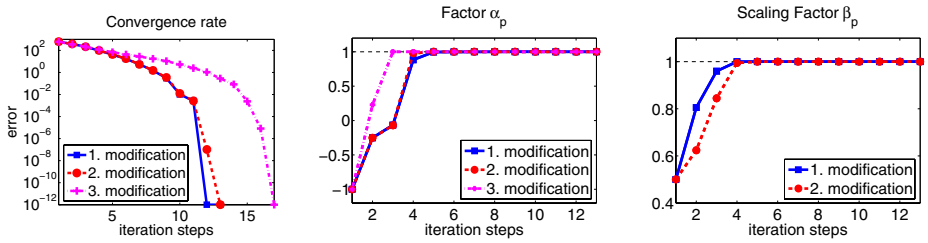

FIG. 4.3. *Convergence of* $\boldsymbol{\lambda}_h^k$ *on level* 5, $c_n = c_\tau = 100$, *and* $\varepsilon_u = 10^{-14}$ *(left); behavior of* $\alpha_{p,l}^{k-1}$ *(middle) and* $\beta_{p,l}^{k-1}$ *(right) at the node* $(0.0625, 1, 0)^\top \in \mathcal{S}$; $\boldsymbol{u}_h^{0,0} = (1, 1, 0)^\top$, $\boldsymbol{\lambda}_h^0 = (-1, -1, 0)^\top$.

TABLE 4.2

Comparison between the exact ($m =$ "∞") *and the inexact* ($m = 1$) *strategy with* $c_n = c_\tau = 100$ *using* (3.11). *Tolerance for IPDAS strategy:* $\varepsilon_u = 10^{-10}$.

Strategy		Exact		Inexact		Nested	
Level l	DOF	K_l	MG steps	M_l	MG steps	M_l	MG steps
1	27	3	41	3	11	2	10
2	125	3	44	3	13	2	12
3	729	4	59	4	14	5	13
4	4913	6	65	6	14	7	12
5	35937	7	85	8	17	7	14

$\varepsilon_u = 10^{-14}$. In the middle of Figure 4.3, we present the cosinus α_p^{k-1} of the angle between the vectors $\boldsymbol{\lambda}_{\tau,p,s}^{k-1}$ and $\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1}$ and on the right the behavior of the scaling factor $\beta_{p,l}^{k-1}$, $l = 1, 2$, for the node $p = (0.0625, 1, 0)^\top \in \mathcal{A}_\tau^k$ for all k. We observe for all modifications a superlinear convergence. Comparing the behavior of the factor $\alpha_{p,l}^{k-1}$, the first and the second modifications show almost the same behavior. Although the factor $\alpha_{p,l}^{k-1}$ tends faster towards 1.0 for the third modification, we observe a slower convergence. For the factor $\beta_{p,l}^{k-1}$ we observe a better behavior for the first modification than for the second one. From now on, we use modification one for all computations.

Exact versus inexact method. We next compare the exact version, i.e., $m =$ "∞," of Algorithm 1 with the inexact version. In the inexact version, we use $m = 1$; i.e., we update the active and inactive sets after each multigrid step. We denote by K_l the iteration step in which the correct active and inactive sets are found for the first time and kept afterwards. For the inexact approach, we denote this step by M_l. Table 4.2 shows the numbers K_l and M_l on each level and the necessary numbers of multigrid iterations to solve the full nonlinear problem on level l. We observe that the numbers K_l and M_l are almost the same. They seem to depend linearly on the level l. Therefore, there is no need to solve the linear system exactly. Furthermore, we compare the inexact approach, where we start with $\boldsymbol{u}_h^{0,0} = \mathbf{0}$ and $\boldsymbol{\lambda}_h^0 = \mathbf{0}$ on each level, with the nested approach in which we inherit $\boldsymbol{u}_h^{0,0}$ and $\boldsymbol{\lambda}_h^0$ on level $l+1$ from level l. The values M_l and the necessary numbers of multigrid (MG) iteration steps are shown in the last column of Table 4.2. We note that the inexact primal-dual active set strategy can be interpreted as a nonlinear multigrid method.

Influence of the parameter c_τ. In a last test, we investigate the influence of the parameter c_τ on our algorithm. Recall that c_τ can be seen as weight for the tangential distortion $\boldsymbol{u}_{\tau,p}$ in the sum with the tangential component $\boldsymbol{\lambda}_{\tau,p,s}$ of the Lagrange multiplier. Thus, it plays a similar role for the tangential component as c_n

TABLE 4.3
Comparison between different values for c_τ for the inexact strategy ($m = 1$) using $c_n = 100$ and $\varepsilon_u = 10^{-10}$. Initialization: $\boldsymbol{u}_h^{0,0} = \mathbf{0}$, $\boldsymbol{\lambda}_h^0 = \mathbf{0}$ for all levels l.

c_τ	1		10		100		1000		10000	
Level l	K_l	MG	K_l	MG	K_l	MG	M_l	MG	M_l	MG
1	3	12	3	11	3	11	3	11	3	11
2	3	12	3	12	3	13	3	13	4	13
3	4	14	4	14	4	14	4	14	4	14
4	6	15	6	14	6	14	6	14	6	14
5	–	–	8	17	8	17	9	17	9	17

for the normal component. In Table 4.3, we compare the numbers K_l and the numbers MG of necessary multigrid steps for different values of c_τ, where we fix $c_n = 100$ and use the inexact approach $m = 1$. As can be seen, the algorithm behaves quite stably and independently of c_τ, if c_τ is large enough. For this example, we find $c_\tau \geq 10$. For $c_\tau = 1$ the algorithm does not converge on level 5. A very similar behavior is observed in [21] with respect to the parameter c_n. Thus, in general it appears advantageous to choose both c_τ and c_n in order to balance the different scales of the distortion $\boldsymbol{u}_h$ and the Lagrange multiplier $\boldsymbol{\lambda}_h$. A closer look reveals that the scaling parameters c_τ and c_n should reflect the material parameter.

5. Fixed point algorithm and numerical examples for Coulomb friction. In this section, we extend Algorithm 1 to contact problems with Coulomb friction and give a numerical example. For Coulomb friction, the friction bound $g_p = \mathfrak{F}|\lambda_{n,p,s}|$ needs to be iteratively adjusted using the normal component of the Lagrange multiplier. Therefore, we get an additional outer loop for the update of the friction bound.

Algorithm 2: Fixed point IPDAS for contact with 3D Coulomb friction

(0) Set $k = 1$ and choose $c_n > 0$, $c_\tau > 0$, $m \in \mathbb{N}$, and $k_f \in \mathbb{N}$.
Initialize $\boldsymbol{u}_h^{0,0}$ and $\boldsymbol{\lambda}_h^0$.

(1) If $\mathrm{mod}_{k_f}(k-1) = 0$, set $k_c = k - 1$ and update the friction bound by

$$g_p^{k_c} = \mathfrak{F} \max\left\{0,\, \lambda_{n,p,s}^{k_c}\right\}, \qquad p \in \mathcal{S}.$$

(2) Define the active and inactive sets by

$$\begin{aligned}
\mathcal{A}_n^k &:= \left\{p \in \mathcal{S} \,:\, \lambda_{n,p,s}^{k-1} + c_n\big(u_{n,p}^{k-1,m} - d_p\big) > 0\right\},\\
\mathcal{I}_n^k &:= \left\{p \in \mathcal{S} \,:\, \lambda_{n,p,s}^{k-1} + c_n\big(u_{n,p}^{k-1,m} - d_p\big) \leq 0\right\},\\
\mathcal{A}_\tau^k &:= \left\{p \in \mathcal{S} \,:\, \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1,m}\| - g_p^{k_c} > 0\right\},\\
\mathcal{I}_\tau^k &:= \left\{p \in \mathcal{S} \,:\, \|\boldsymbol{\lambda}_{\tau,p,s}^{k} + c_\tau \boldsymbol{u}_{\tau,p}^{k-1,m}\| - g_p^{k_c} \leq 0\right\}.
\end{aligned}$$

(3) For $i = 1, \ldots, m$, compute

$$\boldsymbol{u}_h^{k,i} = MG(\boldsymbol{u}_h^{k,i-1}, \mathcal{A}_n^k, \mathcal{I}_n^k, \mathcal{A}_\tau^k, \mathcal{I}_\tau^k, \boldsymbol{u}_h^{k-1,m}, \boldsymbol{\lambda}_h^{k-1}).$$

(4) Compute the Lagrange multiplier as

$$\boldsymbol{\lambda}_h^k = D^{-1}\big(\boldsymbol{f}_{\mathcal{S}} - A_{\mathcal{S}\star}\boldsymbol{u}_h^{k,m}\big).$$

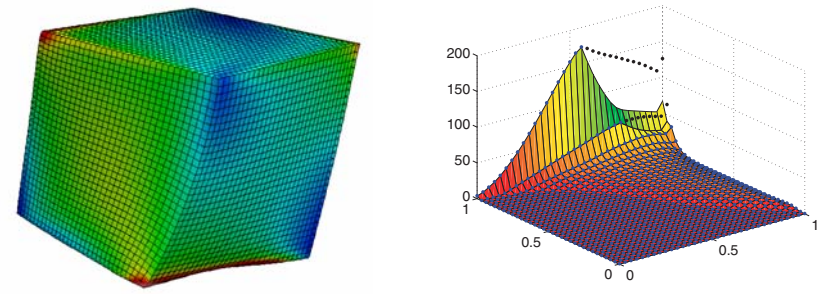

FIG. 5.1. *Distorted body with effective von Mises stress* σ_{eff} *(left); the lower surface in the plot is subject to contact and Coulomb friction. Visualization of the friction bound* $\mathfrak{F}|\lambda_{n,p}|$ *(small balls) and of* $\|\boldsymbol{\lambda}_{\tau,p}\|$ *(right).*

(5) If $k > k_f$ and $\|\boldsymbol{u}_h^{k,m} - \boldsymbol{u}_h^{k_c,m}\| / \|\boldsymbol{u}_h^{k,m}\| < \varepsilon_u$, stop.

(6) Set $\boldsymbol{u}_h^{k+1,0} = \boldsymbol{u}_h^{k,m}$ and $k = k + 1$, and go to step (1).

In this algorithm, we denote by mod$_*$∗ the modula-operator. Comparing this algorithm with Algorithm 1 for Tresca friction, we remark that here we update the friction bound after k_f steps of the (inexact) active set strategy. This update is done in step (1). Since we do not solve the resulting linear problems exactly, it is not guaranteed that $\lambda_{n,p,s}^k \geq 0$ for all $p \in \mathcal{S}$. Therefore, we set $g_p^k = \mathfrak{F} \max\{0, \lambda_{n,p,s}^k\}$. For the choice $m = k_f = 1$, the friction bound and the active and inactive sets are updated after each multigrid step. As the stopping criterion, we use the relative error between the actual solution $\boldsymbol{u}_h^{k,m}$ and the solution for the last friction bound $\boldsymbol{u}_h^{k_c,m}$. For the choice $m = k_f =$ "∞," we get the exact version of the algorithm. In this case, we solve the resulting Tresca friction problem exactly for each friction bound. Obviously, this approach is rather costly. However, for a small friction coefficient $\mathfrak{F}$ it can be shown that this discrete fixed point mapping is contractive and thus converges; see [29].

5.1. Example. In this section, we study the performance of Algorithm 2 for Coulomb friction. This friction model is physically more realistic than the Tresca model, since only points that are in contact with the obstacle are points for which friction occurs. Points on the contact boundary with a positive distance to the obstacle are traction-free; i.e., we apply homogeneous Neumann conditions.

Recall that, in the Coulomb law, the friction bound becomes $g_p = \mathfrak{F}|\lambda_{n,p,s}|$ and thus depends on the actual distortion. In the following tests, we consider the same geometry and data as for the example of section 4.1 and choose the friction coefficient $\mathfrak{F} \equiv 1$. The distorted cube can be seen in the left plot of Figure 5.1. Note that its distortion is significantly different from the one obtained with the Tresca law (Figure 4.1). In the right of Figure 5.1, we visualize the constraint $\|\boldsymbol{\lambda}_{\tau,p,s}\| \leq \mathfrak{F}|\lambda_{n,p,s}|$ that holds for all $p \in \mathcal{S}$. The few nodes where this inequality holds in a strict sense are nodes in contact with the obstacle that, at the same time, stick to this obstacle, i.e., $u_{n,p} = d_p = 0$ and $\boldsymbol{u}_{\tau,p} = \mathbf{0}$. We remark that the solution has a singularity at the node $(1, 1, 0)^\top$.

The Coulomb friction law. As for Tresca friction (Figure 4.2), we also visualize the different types of contact nodes for the Coulomb law. Note that, in contrast to

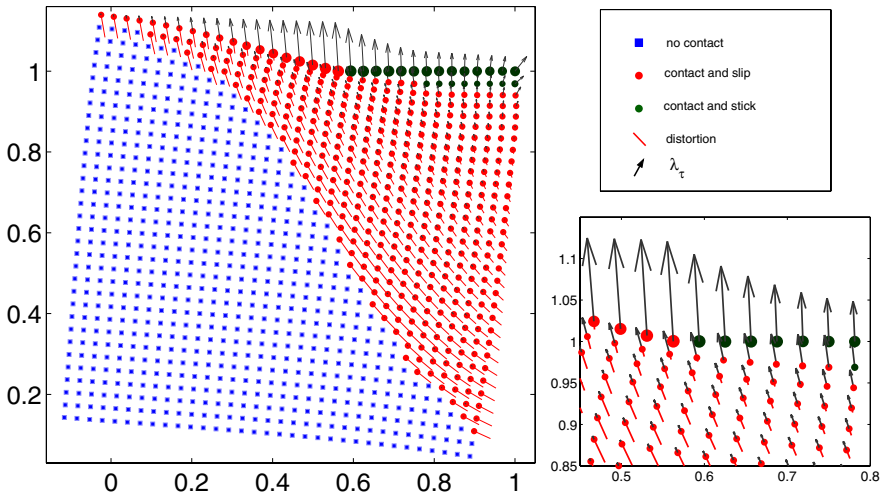

FIG. 5.2. *Visualization of the solution at the nodes $p \in \mathcal{S}$ for Coulomb friction on level* 5 *(left), legend (upper right), and cutout (lower right).*

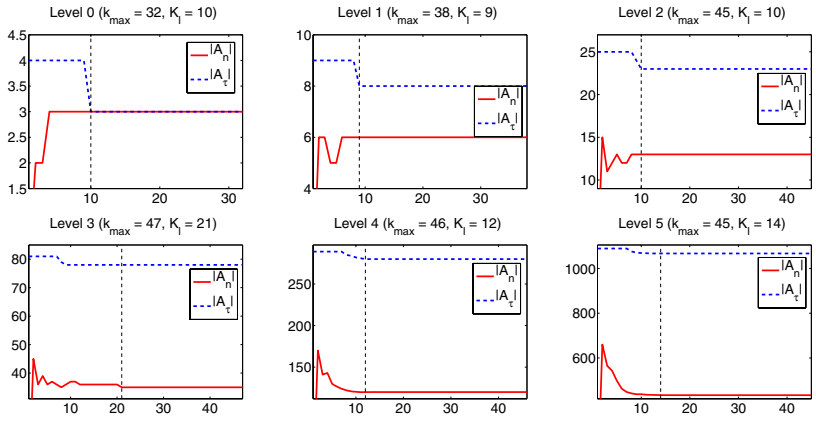

FIG. 5.3. *Behavior of Algorithm* 2*: Numbers $|\mathcal{A}_n^k|$ and $|\mathcal{A}_\tau^k|$ of active nodes in each iteration step k on each level l. We used the parameters $c_n = c_\tau = 100$, $m = k_f = 1$, and $\varepsilon_u = 10^{-10}$. Initialization: $\boldsymbol{u}_h^{0,0} = \mathbf{0}$ and $\boldsymbol{\lambda}_h^0 = \mathbf{0}$.*

the Tresca model, the Coulomb law allows only three types of nodes, since nodes that are not in contact with the obstacle are not subject to any (friction) constraints. The size of the nodes in Figure 5.2 is proportional to the normal contact force $|\lambda_{n,p}|$.

Performance of the algorithm. We apply the inexact version of Algorithm 2 with $m = k_f = 1$ to solve the contact problem with Coulomb friction. We use $c_n = c_\tau = 100$, $\varepsilon_u = 10^{-10}$ and initialize the iteration with $\boldsymbol{u}_h^{0,0} = \mathbf{0}$ and $\boldsymbol{\lambda}_h^0 = \mathbf{0}$ on each level. Figure 5.3 shows the behavior of our algorithm on various levels. The number of nodes contained in the active sets are plotted over the iteration steps k. We denote by $k_{\max}$ the necessary number of iteration steps and by K_l the iteration step in which the correct active sets are found for the first time and do not change afterwards. The number K_l is marked by a dashed vertical line in Figure 5.3. We observe that both $k_{\max}$ and K_l appear to be almost independent of the level l. Looking at Figure 5.3 more closely, we note that on each level there are only minor changes of the active nodes after $k = 10$. Table 5.1 shows a comparison between the initialization

TABLE 5.1
Behavior of Algorithm 2: K_l and $k_{\max}$ (see text for explanation) on each level l for $c_n = c_\tau = 100$, $m = k_f = 1$, and $\varepsilon_u = 10^{-10}$.

Level l	DOF	$\boldsymbol{u}_h^{0,0} = \mathbf{0}$, $\boldsymbol{\lambda}_h^0 = \mathbf{0}$		Nested approach	
		K_l	$k_{\max}$	K_l	$k_{\max}$
0	8	10	32	10	32
1	27	9	38	3	34
2	125	10	45	5	32
3	729	21	47	3	37
4	4913	12	46	4	38
5	35937	14	45	5	36

$\boldsymbol{u}_h^{0,0} = \mathbf{0}$ and $\boldsymbol{\lambda}_h^0 = \mathbf{0}$ on each level and the nested approach. As expected, the correct active sets are found earlier, and as a result fewer iterations are required for the nested approach.

6. Numerical examples for two-body contact with Coulomb friction. Now we consider a curved contact interface subjected to Coulomb friction. A two-dimensional cross section of our geometry is shown in Figure 6.1. The lower domain Ω^m, assumed to be the master side, models a spherical shell that is fixed at the outer boundary. Against this shell, we press the body modeled by the domain Ω^s, which is assumed to be the slave side. At the top surface of Ω^s, we apply the Dirichlet data $(0, 0, -0.2)^\top$. The geometry is given by $r_i = 0.7$, $r_a = 1.0$, $r = 0.6$, $h = 0.5$, and $d = 0.3$. In Ω^s, we use a Young modulus $E^s = 300$ and a Poisson ratio $\nu^s = 0.3$, while in Ω^m we have $E^m = 400$ and $\nu^m = 0.3$.

Discussion of the results and comparison for various friction coefficients. The results for the friction coefficient $\mathfrak{F} = 0.5$ are shown in Figures 6.1–6.2. Figure 6.1 shows the deformed body with the effective von Mises stress σ_{eff} on level 3 and a two-dimensional cross section on level 2. In Figure 6.2, we show the nodes being in contact, their relative tangential slip $[\boldsymbol{u}_{\tau,p}]$ (lines), and the tangential contact pressure $\boldsymbol{\lambda}_{\tau,p}$ (arrows). We remark that the nodes in the middle with only an arrow are the sticky nodes; the others are slippy. In the two pictures on the right, the normal and the tangential part of the Lagrange multiplier together with the friction bound are plotted. The comparison of the result for various friction coefficients $\mathfrak{F}$ is shown in Figures 6.3–6.4. In the first one, the visualization of the contact nodes together with the relative tangential slip and the tangential contact pressure are shown. Figure 6.4 shows a two-dimensional plot of the normal and tangential part of the Lagrange multiplier and the friction bound for all nodes $p \in \mathcal{S}$ over their distance to the midpoint of the

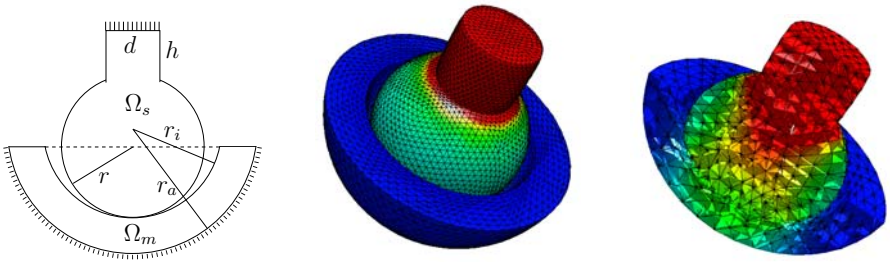

FIG. 6.1. *Problem definition (left), deformed mesh with effective von Mises stress σ_{eff} on level 3 (middle), and two-dimensional cross section of the deformed mesh on level 2 (right) for $\mathfrak{F} = 0.5$.*

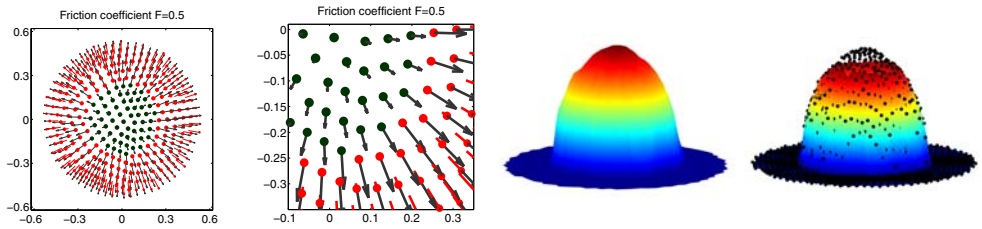

FIG. 6.2. *Visualization of the nodes being in contact together with the deformation and stress vectors on level* 3 *for* $\mathfrak{F} = 0.5$ *(first) and a cutout (second)—for the legend we refer to Figure* 5.2*—and visualization of* λ_n *(third) and* $\|\boldsymbol{\lambda}_\tau\|$ *(fourth) with the friction bound* $0.5|\lambda_n|$ *(dotted line).*

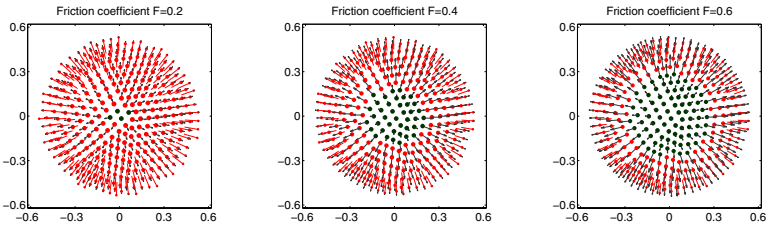

FIG. 6.3. *Visualization of the nodes being in contact together with the deformation and stress vector on level* 3 *for* $\mathfrak{F} = 0.2, 0.4, 0.6$*; for the legend we refer to Figure* 5.2.

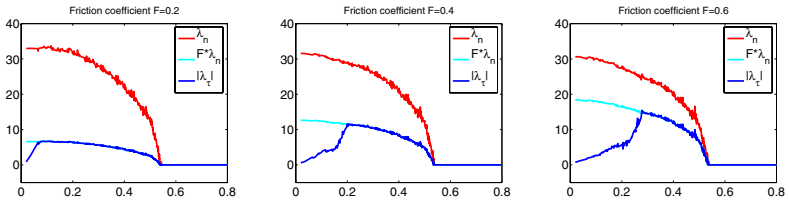

FIG. 6.4. *A two-dimensional visualization of the Lagrange multipliers for* $\mathfrak{F} = 0.2, 0.4, 0.6$.

contact zone on the surface of Ω^s. The small oscillations in the plot occur due to the fact that all nodes $p \in \mathcal{S}$ over the whole contact zone are projected onto a straight radial line and that we work with an unstructured mesh. We observe that the number of sticky nodes decreases for smaller $\mathfrak{F}$.

Performance of the algorithm. We consider the performance of Algorithm 2 for the friction coefficient $\mathfrak{F} = 0.5$. For the initialization, we set $\boldsymbol{u}_p^{0,0} = 0.1\boldsymbol{n}_p$ for $p \in \mathcal{S}$, $\boldsymbol{u}_p^{0,0} = \mathbf{0}$ for $p \notin \mathcal{S}$, and $\boldsymbol{\lambda}_p^0 = 0.0001\boldsymbol{n}_p$ on each level. Using the parameters $c_n = c_\tau = 100$, $m = k_f = 1$, and $\varepsilon_u = 10^{-9}$, we get the performance of Algorithm 2 shown in Figure 6.5. Here the number of nodes in $|\mathcal{A}_n^k|$ and $|\mathcal{A}_\tau^k|$ is shown for levels 1–3. Again, the dashed vertical line marks the step K_l in which the correct active sets are found for the first time and remain unchanged. A comparison between this approach and the nested approach for various friction coefficients is shown in Table 6.1. Although for $\mathfrak{F} = 0.2$ the convergence rates of the multigrid method for levels 1 and 2 are worse, the results obtained show qualitatively the same behavior as for the cube; see section 5.1.

Problem with nonsymmetric boundary traction. For the next example, we use the same data and geometry as above (see Figure 6.1) but use a nonsymmetric boundary traction at the top of Ω^s instead of the Dirichlet data; namely, we apply the surface

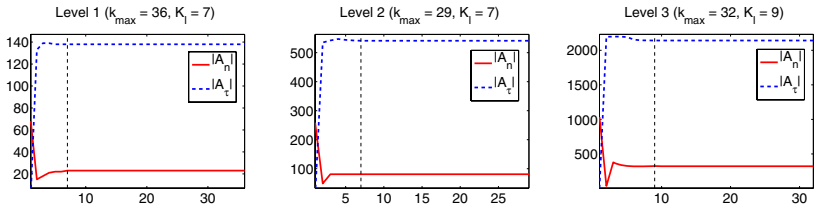

FIG. 6.5. *Behavior of Algorithm 2: Numbers* $|\mathcal{A}_n^k|$ *and* $|\mathcal{A}_\tau^k|$ *in each iteration step* k *on levels* $1, 2, 3$ *for* $\mathfrak{F} = 0.5$. *We use the parameters* $c_n = c_\tau = 100$, $m = k_f = 1$, *and* $\varepsilon_u = 10^{-9}$.

TABLE 6.1

Behavior of Algorithm 2: Necessary numbers $k_{\max}$ *and* K_l *on each level for* $c_n = c_\tau = 100$, $m = k_f = 1$, $\varepsilon_u = 10^{-9}$ *and* $\mathfrak{F} = 0.2, 0.4, 0.6$.

Level	DOF	$\boldsymbol{u}_p^{0,0} = 0.1\,\boldsymbol{n}_p$, $\boldsymbol{\lambda}_p^0 = 0.0001\,\boldsymbol{n}_p$						Nested approach					
		K_l			$k_{\max}$			K_l			$k_{\max}$		
		0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6	0.2	0.4	0.6
0	104	4	4	4	12	16	21	4	4	4	12	16	21
1	541	10	5	5	111	20	33	1	3	2	91	20	25
2	3384	5	7	6	67	27	26	2	3	2	62	21	22
3	23694	9	11	9	33	32	32	3	3	2	29	21	21

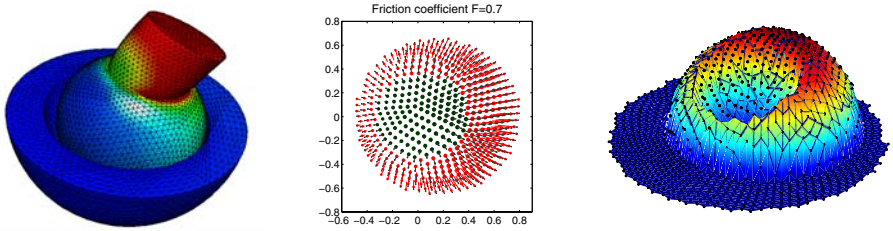

FIG. 6.6. *Nonsymmetric boundary data: Deformed mesh with effective von Mises stress* σ_{eff} *on level 3 (left); visualization of the nodes being in contact together with the deformation and the stress vectors (middle) and of* $\|\boldsymbol{\lambda}_\tau\|$ *(right) with the friction bound* $0.7|\lambda_n|$ *(dotted line).*

traction $(15,\ 0,\ -150\exp(-100r^2))^\top$, where r denotes the distance to the midpoint of the top surface of Ω^s. The results for $\mathfrak{F} = 0.7$ are presented in Figure 6.6.

7. Full Newton approach for multibody contact with Coulomb friction. While our algorithm in section 5 for the solution of the Coulomb friction problem is based on fixed point ideas, we now present a full Newton approach. The main advantage of this approach is its fast convergence, which is due to the fact that the friction bound is updated in the Newton iteration and not via a fixed point loop. In this section, we apply a fast direct solver [32] to solve the linear system in each Newton step. To derive the full Newton iteration, one replaces g_p in (3.1) either by $g_p(u_{n,p}, \lambda_{n,p,s}) := \mathfrak{F}\max(0, \lambda_{n,p,s})$ (see also [6,7]) or by $g_p(u_{n,p}, \lambda_{n,p,s}) := \mathfrak{F}\max(0, \lambda_{n,p,s} + c_n(u_{n,p} - d_p))$ before deriving the Newton iteration step. The equivalence of these two choices follows from (4.1). In what follows, we use the latter replacement, since then the Newton-type iteration automatically takes the form of an active set method that in each iteration estimates the three relevant sets for Coulomb friction (no contact, contact and stick,

contact and slip). The resulting nonlinear complementarity function is

$$\begin{aligned}\mathcal{D}(\boldsymbol{u}_h,\boldsymbol{\lambda}_h) := &\max\left(\mathfrak{F}(\lambda_{n,p,s}+c_n(u_{n,p}-d_p)),\|\boldsymbol{\lambda}_{\tau,p,s}+c_\tau\boldsymbol{u}_{\tau,p}\|\right)\boldsymbol{\lambda}_{\tau,p,s}\\ &-\mathfrak{F}\max\left(0,\lambda_{n,p,s}+c_n(u_{n,p}-d_p)\right)\left(\boldsymbol{\lambda}_{\tau,p,s}+c_\tau\boldsymbol{u}_{\tau,p}\right).\end{aligned}\tag{7.1}$$

Similarly as in section 3, we derive a semismooth Newton step for the solution of $\mathcal{D}(\boldsymbol{u}_h,\boldsymbol{\lambda}_h)=0$ and (4.1). Using the notation $g_p^{k-1} := \mathfrak{F}(\lambda_{n,p,s}^{k-1}+c_n(u_{n,p}^{k-1}-d_p))$, we obtain the following settings.

- On $\mathcal{I}_n^k := \{p\in\mathcal{S} : g_p^{k-1}\le 0\}$ (estimation for a set of nodes not in contact):

$$\lambda_{n,p,s}^k = 0 \quad \text{and} \quad \boldsymbol{\lambda}_{\tau,p,s}^k = 0.$$

  Note that $\mathcal{I}_n^k\subset\mathcal{A}_\tau^k$, where $\mathcal{A}_\tau^k := \{p\in\mathcal{S} : \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1}\|-g_p^{k-1}\ge 0\}$. We remark that the setting for $\boldsymbol{\lambda}_{\tau,p,s}^k$ is derived directly from (7.1) for that case, namely, from $\|\boldsymbol{\lambda}_{\tau,p,s}+c_\tau\boldsymbol{u}_{\tau,p}\|\boldsymbol{\lambda}_{\tau,p,s}=0$.

- On $\mathcal{I}_\tau^k := \{p\in\mathcal{S} : \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1}\|-g_p^{k-1}<0\}$: (estimation for a set of sticky nodes):

$$u_{n,p}^k = d_p \quad \text{and} \quad \boldsymbol{u}_{\tau,p}^k + \left(\mathfrak{F}\boldsymbol{u}_{\tau,p}^{k-1}/g_p^{k-1}\right)\lambda_{n,p,s}^k = \boldsymbol{u}_{\tau,p}^{k-1}.\tag{7.2}$$

  Note that $\mathcal{I}_\tau^k\subset\mathcal{A}_n^k$, where $\mathcal{A}_n^k := \{p\in\mathcal{S} : g_p^{k-1}>0\}$.

- On $\mathcal{A}_{\tau n}^k := \mathcal{A}_\tau^k\cap\mathcal{A}_n^k = \{p\in\mathcal{S} : \|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1}\|\ge g_p^{k-1}>0\}$ (estimation for a set of slippy nodes):

$$\begin{aligned}u_{n,p}^k &= d_p \quad \text{and}\\ -\boldsymbol{\lambda}_{\tau,p,s}^k + L_{p,l}^{k-1}\boldsymbol{u}_{\tau,p}^k + \mathfrak{F}\boldsymbol{v}_{p,l}^{k-1}\lambda_{n,p,s}^k &= \boldsymbol{r}_{p,l}^{k-1} + g_p^{k-1}\boldsymbol{v}_{p,l}^{k-1},\end{aligned}\tag{7.3}$$

  where $\boldsymbol{v}_p^{k-1} := (Id_2 - M_p^{k-1})^{-1}(\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1})/\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1}\|\in\mathbb{R}^2$ and g_p^{k-1} is used for the friction bound g_p in the matrices M_p^{k-1} and L_p^{k-1}.

Note that $\mathcal{I}_n^k$, $\mathcal{I}_\tau^k$, and $\mathcal{A}_{\tau n}^k$ represent a disjoint splitting of $\mathcal{S}$. Comparing (7.2) and (7.3) with (3.5a) and (3.5b), respectively, we observe that the main difference is the term involving $\lambda_{n,p,s}^k$ on the left-hand side of (7.2) and (7.3). Again we apply the modifications stated in subsection 3.2 and replace the matrices M_p^{k-1} and L_p^{k-1} by $\tilde{M}_{p,l}^{k-1}$ and $\tilde{L}_{p,l}^{k-1}$, respectively. Therefore, we use instead of $\boldsymbol{v}_p^{k-1}$ the vector $\tilde{\boldsymbol{v}}_{p,l}^{k-1} := (Id_2-\beta_{p,l}^{k-1}\tilde{M}_{p,l}^{k-1})^{-1}(\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1})/\|\boldsymbol{\lambda}_{\tau,p,s}^{k-1}+c_\tau\boldsymbol{u}_{\tau,p}^{k-1}\|$. Now we briefly state the algebraic representation of the above system, where we use in addition to the notation introduced in section 3

$$G := \operatorname{diag}\left\{g_p^{k-1}Id_2\right\}_{p\in\mathcal{S}} \in\mathbb{R}^{2|\mathcal{S}|\times 2|\mathcal{S}|}.$$

and

$$U := \operatorname{diag}\left\{\boldsymbol{u}_{\tau,p}^{k-1}\right\}_{p\in\mathcal{S}}, \qquad V := \operatorname{diag}\left\{\tilde{\boldsymbol{v}}_{p,l}^{k-1}\right\}_{p\in\mathcal{S}} \in\mathbb{R}^{2|\mathcal{S}|\times|\mathcal{S}|}.$$

We use a subblock notation for the matrices, e.g., $G_{\mathcal{I}_\tau^n}\ldots$. Defining $\tilde{\mathcal{N}} := \mathcal{N}\cup\mathcal{I}_n^k$, we obtain, after eliminating the Lagrange multiplier, the following linear system to be

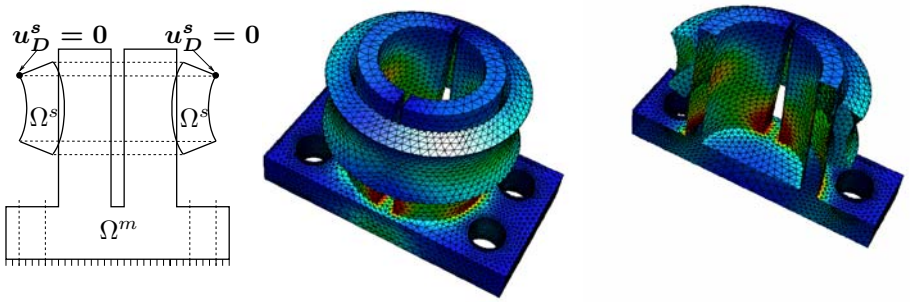

FIG. 7.1. *Problem definition (left), deformed mesh with effective von Mises stress σ_{eff} (middle), and two-dimensional cross section of the deformed mesh (right) for $\mathfrak{F} = 0.7$.*

solved in each full Newton step for the Coulomb problem as

$$\begin{pmatrix} A_{\tilde{\mathcal{N}}\tilde{\mathcal{N}}} & A_{\tilde{\mathcal{N}}\mathcal{I}_\tau^k} & A_{\tilde{\mathcal{N}}\mathcal{A}_{\tau m}^k} \\ 0 & N_{\mathcal{I}_\tau^k} & 0 \\ 0 & 0 & N_{\mathcal{A}_{\tau m}^k} \\ K_{\mathcal{I}_\tau^k} A_{\mathcal{I}_\tau^k \tilde{\mathcal{N}}} & K_{\mathcal{I}_\tau^k} A_{\mathcal{I}_\tau^k \mathcal{I}_\tau^k} - G_{\mathcal{I}_\tau^k} T_{\mathcal{I}_\tau^k} & K_{\mathcal{I}_\tau^k} A_{\mathcal{I}_\tau^k \mathcal{A}_{\tau m}^k} \\ T'_{\mathcal{A}_{\tau m}^k} A_{\mathcal{A}_{\tau m}^k \tilde{\mathcal{N}}} & T'_{\mathcal{A}_{\tau m}^k} A_{\mathcal{A}_{\tau m}^k \mathcal{I}_\tau^k} & T'_{\mathcal{A}_{\tau m}^k} A_{\mathcal{A}_{\tau m}^k \mathcal{A}_{\tau m}^k} + L_{\mathcal{A}_{\tau m}^k} T_{\mathcal{A}_{\tau m}^k} \end{pmatrix} \begin{pmatrix} \boldsymbol{u}_{\tilde{\mathcal{N}}}^k \\ \boldsymbol{u}_{\mathcal{I}_\tau^k}^k \\ \boldsymbol{u}_{\mathcal{A}_{\tau m}^k}^k \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}_{\tilde{\mathcal{N}}} \\ d_{\mathcal{I}_\tau^k} \\ d_{\mathcal{A}_{\tau m}^k} \\ K_{\mathcal{I}_\tau^k} \boldsymbol{f}_{\mathcal{I}_\tau^k} - \boldsymbol{j}_{\mathcal{I}_\tau^k} \\ T'_{\mathcal{A}_{\tau m}^k} \boldsymbol{f}_{\mathcal{A}_{\tau m}^k} + \boldsymbol{r}'_{\mathcal{A}_{\tau m}^k} \end{pmatrix},$$

with the notation

$$T'_{\mathcal{A}_{\tau m}^k} := T_{\mathcal{A}_{\tau m}^k} - \mathfrak{F} V_{\mathcal{A}_{\tau m}^k} N_{\mathcal{A}_{\tau m}^k}, \qquad K_{\mathcal{I}_\tau^k} := \mathfrak{F} U_{\mathcal{I}_\tau^k} N_{\mathcal{I}_\tau^k},$$

and

$$\boldsymbol{r}'_{\mathcal{A}_{\tau m}^k} := \boldsymbol{r}_{\mathcal{A}_{\tau m}^k} + G_{\mathcal{A}_{\tau m}^k} V_{\mathcal{A}_{\tau m}^k}, \qquad \boldsymbol{j}_{\mathcal{I}_\tau^k} := G_{\mathcal{I}_\tau^k} U_{\mathcal{I}_\tau^k}.$$

Comparing this system with (3.15), we observe that here the normal and tangential components are pointwise coupled by lines four and five.

Numerical example. As example, we consider the situation presented in the left of Figure 7.1, where a two-dimensional cross section of the problem definition is shown. The ring is fixed on its upper outer edge and the tool on its bottom. Note that the bodies penetrate in their reference configuration. We use Young modulus $E^m = 8.13 \times 10^8$ and a Poisson ratio $\nu^m = 0.3$ for the inner tool modeled by Ω^m and $E^s = 9 \times 10^7$ and $\nu^s = 0.3$ for the outer ring being the slave domain Ω^s. The friction coefficient is $\mathfrak{F} = 0.7$. The deformed mesh with effective von Mises stress σ_{eff} is shown in the middle and the right of Figure 7.1. The first two pictures of Figure 7.2 show the possible contact nodes on the ring being the slave side. The nodes without a line are nodes not being in contact with the inner tool.

Comparison between the fixed point Newton and the full Newton approaches. To show the behavior of the full Newton approach, we compare the convergence rates of the Lagrange multiplier with those obtained from the fixed point Newton approach given by Algorithm 2 with $m =$ "∞" and $k_f = 1$. Our algorithm is initialized with $\boldsymbol{u}_h^{0,0} = \mathbf{0}$ and $\boldsymbol{\lambda}_h^0 = \mathbf{0}$, and the parameters $c_n = c_\tau = 10^8$ and the tolerance $\varepsilon_u = 10^{-9}$ are used. The finite element mesh consists of 66.600 degrees of freedom. In the right of Figure 7.2, the relative error $\|\boldsymbol{\lambda}_h^k - \boldsymbol{\lambda}_h^*\|/\|\boldsymbol{\lambda}_h^*\|$ is shown for the fixed point and the full Newton approach in a logarithmic scale. We observe superlinear convergence of the full Newton approach, and only half of the iteration steps compared with the fixed point approach are required. For the fixed point approach the convergence rate

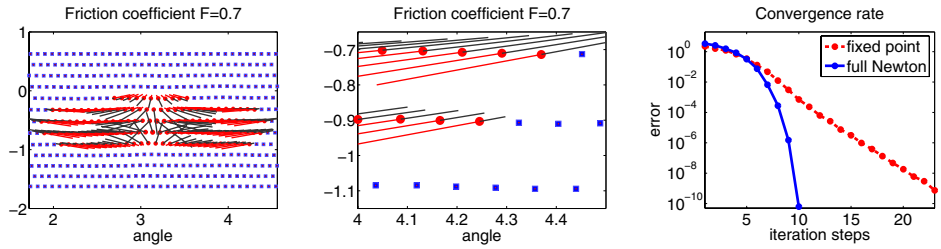

FIG. 7.2. *Visualization of the nodes being in contact, together with the deformation and stress vectors for* $\mathfrak{F} = 0.7$ *(left) and a cutout (middle)—for the legend we refer to Figure* 5.2*—and convergence rates for the fixed point Newton and the full Newton approaches (right).*

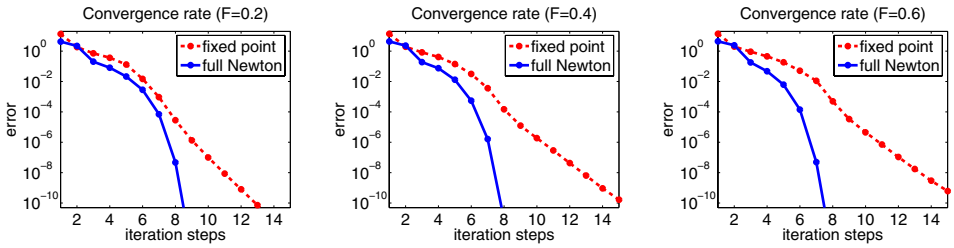

FIG. 7.3. *Convergence rates for the example of Figure* 6.1 *for the fixed point Newton and the full Newton approaches for* $\mathfrak{F} = 0.2, 0.4, 0.6$.

$\|\boldsymbol{\lambda}_h^{k+1} - \boldsymbol{\lambda}_h^*\| / \|\boldsymbol{\lambda}_h^k - \boldsymbol{\lambda}_h^*\|$ tends to approximately 0.3. In a last test we investigate the influence of the friction coefficients on the full Newton approach. The convergence rates for the example presented in Figure 6.1 for the friction coefficients $\mathfrak{F} = 0.2, 0.4, 0.6$ are shown in Figure 7.3. As initial values we use the same as in Figure 6.5. Again we observe the same behavior as before for all friction coefficients. Therefore the behavior of the full Newton method is independent of the size of the friction coefficient.

REFERENCES

[1] P. ALART AND A. CURNIER, *A mixed formulation for frictional contact problems prone to Newton like solution methods*, Comput. Methods Appl. Mech. Engrg., 92 (1991), pp. 353–375.

[2] K. ANDERSEN, E. CHRISTIANSEN, A. CONN, AND M. OVERTON, *An efficient primal-dual interior-point method for minimizing a sum of Euclidean norms*, SIAM J. Sci. Comput., 22 (2000), pp. 243–262.

[3] F. BEN BELGACEM, *Numerical simulation of some variational inequalities arisen from unilateral contact problems by the finite element methods*, SIAM J. Numer. Anal., 37 (2000), pp. 1198–1216.

[4] P. BASTIAN, K. BIRKEN, K. JOHANNSEN, S. LANG, N. NEUSS, H. RENTZ–REICHERT, AND C. WIENERS, *UG – A flexible software toolbox for solving partial differential equations*, Comput. Vis. Sci., 1 (1997), pp. 27–40.

[5] T.F. CHAN, G.H. GOLUB, AND P. MULET, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comput., 20 (1999), pp. 1964–1977.

[6] P. CHRISTENSEN, A. KLARBRING, J. PANG, AND N. STRÖMBERG, *Formulation and comparison of algorithms for frictional contact problems*, Internat. J. Numer. Methods Engrg., 42 (1998), 145–173.

[7] P. CHRISTENSEN AND J. PANG, *Frictional contact algorithms based on semismooth Newton methods*, in Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Appl. Optim. 22, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1999, pp. 81–116.

[8] Z. Dostál, D. Horák, R. Kučera, V. Vondrák, J. Haslinger, J. Dobiáš, and S. Pták, *FETI based algorithms for contact problems: Scalability, large displacements and 3D Coulomb friction*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 395–409.

[9] C. Eck, J. Jarusek, and M. Krbec, *Unilateral Contact Problems. Variational Methods and Existence Theorems*, Chapman and Hall/CRC, Boca Raton, FL, 2005.

[10] K. Fackeldey and R. Krause, *Solving frictional contact problems with multigrid efficiency*, in Domain Decomposition Methods in Science and Engineering XVI, Lect. Notes Comput. Sci. Eng. 55, Springer-Verlag, Berlin, 2007, pp. 547–554.

[11] K. Fischer and P. Wriggers, *Frictionless 2D Contact formulations for finite deformations based on the mortar method*, Comput. Mech., 36 (2005), pp. 226–244.

[12] K. Fischer and P. Wriggers, *Mortar based frictional contact formulation for higher order interpolations using the moving friction cone*, Comput. Methods Appl. Mech. Engrg., 195 (2006), pp. 5020–5036.

[13] J. Haslinger, Z. Dostál, and R. Kučera, *On a splitting type algorithm for the numerical realization of contact problems with Coulomb friction*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 2261–2281.

[14] I. Hlaváček, J. Haslinger, J. Nečas, and J. Lovíšek, *Solution of Variational Inequalities in Mechanics*, Appl. Math. Sci. 66, Springer, New York, 1988.

[15] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.

[16] J. Haslinger, R. Kučera, and Z. Dostál, *An algorithm for the numerical realization of 3D contact problems with Coulomb friction*, J. Comput. Appl. Math., 164/165 (2004), pp. 387–408.

[17] M. Hintermüller, V. Kovtunenko, and K. Kunisch, *Semismooth Newton methods for a class of unilaterally constrained variational problems*, Adv. Math. Sci. Appl., 14 (2004), pp. 513–535.

[18] P. Hild and P. Laborde, *Quadratic finite element methods for unilateral contact problems*, Appl. Numer. Math., 41 (2002), pp. 410–421.

[19] S. Hüeber, A. Matei, and B. Wohlmuth, *Efficient algorithms for problems with friction*, SIAM J. Sci. Comput., 29 (2007), pp. 70–92.

[20] M. Hintermüller and G. Stadler, *An infeasible primal-dual algorithm for total variation-based inf-convolution-type image restoration*, SIAM J. Sci. Comput., 28 (2006), pp. 1–23.

[21] S. Hüeber and B. Wohlmuth, *A primal-dual active set strategy for non–linear multibody contact problems*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 3147–3166.

[22] R. Kornhuber and R. Krause, *Adaptive multigrid methods for Signorini's problem in linear elasticity*, Comput. Vis. Sci., 4 (2001), pp. 9–20.

[23] N. Kikuchi and J.T. Oden, *Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods*, SIAM Studies in Applied Mathematics 8, SIAM, Philadelphia, PA, 1988.

[24] K. Kunisch and G. Stadler, *Generalized Newton methods for the 2D-Signorini contact problem with friction in function space*, M2AN Math. Model. Numer. Anal., 39 (2005), pp. 827–854.

[25] R. Kučera, J. Haslinger, and Z. Dostál, *A new FETI-based algorithm for solving 3D contact problems with Coulomb friction*, in Domain Decomposition Methods in Science and Engineering XVI, Lect. Notes Comput. Sci. Eng. 55, Springer-Verlag, Berlin, 2007, pp. 643–650.

[26] R. Kučera, *Minimizing quadratic functions with separable quadratic constraints*, Optim. Methods Softw., 22 (2007), pp. 453–467.

[27] T. Laursen, *Computational Contact and Impact Mechanics*, Springer, Berlin, Heidelberg, 2002.

[28] C. Licht, E. Pratt, and M. Raous, *Remarks on a numerical method for unilateral contact including friction*, in Unilateral Problems in Structural Analysis, IV (Capri, 1989), Internat. Ser. Numer. Math. 101, Birkhäuser, Basel, 1991, pp. 129–144.

[29] J. Nečas, J. Jarušek, and J. Haslinger, *On the solution of the variational inequality to the Signorini problem with small friction*, Boll. Unione Mat. Ital., 5 (1980), pp. 796–811.

[30] M. Puso and T. Laursen, *A mortar segment-to-segment frictional contact method for large deformations*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 4891–4913.

[31] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.

[32] O. Schenk, K. Gärtner, and W. Fichtner, *Efficient sparse LU factorization with left-right looking strategy on shared memory multiprocessors*, BIT, 40 (2000), pp. 158–176.

[33] G. Stadler, *Semismooth Newton and augmented Lagrangian methods for a simplified friction problem*, SIAM J. Optim., 15 (2004), pp. 39–62.
[34] K. Willner, *Kontinuums- und Kontaktmechanik*, Springer, New York, 2003.
[35] B. Wohlmuth and R. Krause, *Monotone multigrid methods on nonmatching grids for nonlinear multibody contact problems*, SIAM J. Sci. Comput., 25 (2003), pp. 324–347.
[36] P. Wriggers, *Computational Contact Mechanics*, John Wiley & Sons, New York, 2002.