Arch Computat Methods Eng (2017) 24:37–87
DOI 10.1007/s11831-015-9159-7

ORIGINAL PAPER

Surface Smoothing Procedures in Computational Contact Mechanics

D. M. Neto[1] · M. C. Oliveira[1] · L. F. Menezes[1]

Received: 19 October 2015 / Accepted: 28 October 2015 / Published online: 11 November 2015

Abstract This work addresses the problems arising in the finite element simulation of contact problems undergoing large deformation. The frictional contact problem is formulated in the continuum framework, introducing the interface laws for the normal and tangential stress components in the contact area. The variational formulation is presented, considering different methods to enforce the contact constraints. The spatial discretization within the finite element method is applied, as well as the temporal discretization required to solve the three sources of nonlinearities: geometric, material and frictional contact. The discretization of contact surfaces is discussed in detail, including different surface smoothing procedures. This numerical strategy allows to solve the difficulties associated with the discontinuities in the contact surface geometry introduced by finite element discretization, which leads to nonphysical oscillations of the contact force for large sliding problems. The geometrical accuracy of different interpolation methods is evaluated, paying particular attention to the Nagata patch interpolation recently proposed. In this framework, the Node-to-Nagata contact elements are developed using the augmented Lagrangian method to regularize the variational frictional contact problem. The techniques used to search for contact in case of large deformations are discussed, including self-contact phenomena. Several numerical examples are presented, comprising both the contact between deformable and rigid obstacles and the contact between deformable bodies. The results show that the accuracy and robustness of the numerical simulations is improved when the contact surface is smoothed with Nagata patches.

✉ D. M. Neto
diogo.neto@dem.uc.pt

[1] CEMUC, Department of Mechanical Engineering, University of Coimbra, Polo II, Rua Luís Reis Santos, Pinhal de Marrocos, 3030-788 Coimbra, Portugal

1 Introduction

Any mechanical load results from the contact interaction between two separate bodies or parts of a single body. Thus, almost all mechanical systems comprise contact interactions, which from the engineering point of view can be intentional, such as in the sheet metal forming processes (Fig. 1), or undesired, as in a car crash (Fig. 2). In order to increase the efficiency in intentional contact interactions and decrease the adverse effects in not intended contact interactions, the full understanding of the contact interaction process is mandatory. By nature, mechanical contact always involves friction phenomena. Nevertheless, it is neglected in some situations (sufficiently small frictional forces) to simplify the analysis. Despite the importance of contact mechanics in several engineering applications, contact effects are rarely taken into account in conventional structural analysis due to the complexity of the contact phenomena (multi-physical nature) [1]. In fact, mathematical models of contact problems are inherently nonlinear since the contacting surface on which the loads are transferred from one body to another is unknown a priori. Furthermore, the accurate modelling of friction is very difficult because it is dependent of several factors (contact pressure, surface roughness, temperature, etc.) [2].

Concerning the equilibrium formulation, the frictional contact problems can be categorized into static, *quasi*-static and dynamic. Typically, the terminology contact is adopted in static and *quasi*-static formulations, while the term impact is used to outline dynamic contact problems [1].

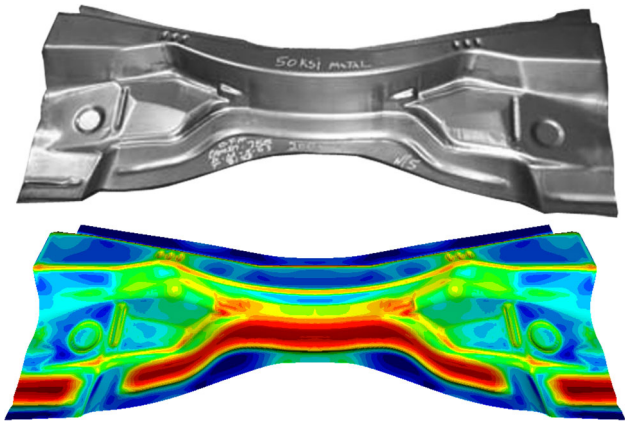

Fig. 1 Sheet metal forming of the automotive underbody cross member panel proposed as benchmark at the Numisheet 2005 conference: experimental (*top*) and numerical simulation result (*bottom*)

Fig. 2 Frontal impact of a car with initial speed of 64 km/h against a deformable barrier: experimental (*top*) and numerical simulation result (*bottom*) (courtesy of Daimler Chrysler AG)

The contact arising in building frames can be modelled as a static problem, while the contact in sheet metal forming processes (see Fig. 1) can be treated as a *quasi*-static phenomenon. In both cases the inertial effects are negligible. On the other hand, in contact problems such as automobile crashworthiness (see Fig. 2) the inertial effects need to be taken into account. In this review paper, the presented formulation and computational strategies will be restricted mainly to static and *quasi*-static frictional contact problems. For a detailed treatment of explicit transient dynamics the reader should consult the literature, e.g. the book by Zhong [1] presenting several contact algorithms for highly transient applications.

The history of contact mechanics, which began hundreds of years ago in ancient Egypt with the movement of large stone blocks, can be divided in three distinct research periods [1, 3]. The first period of investigation occurred in the eighteenth century, where the Newton's third law and the Coulomb's friction law can be stated as the two principal contributions. The contact bodies were assumed rigid to keep the formulation simple, being the analysis restricted to the total contact forces (global phenomena). The prediction of deformations and stresses inside the body defines the beginning of the second research period (nineteenth century). The key point of this period was dictated by the work carried out by Hertz [4, 5], establishing the analytical solution for the elastic frictionless contact between two ellipsoidal bodies. The prediction of the pressure distribution in the contact area represents a milestone in the field of modern contact mechanics. In order to overcome the limitations of the Hertz theory, several researchers (mainly mathematicians) studied contact problems using different shapes and considering different conditions. Nevertheless, these approaches are obviously very restrictive and can be only applied to a few simple problems. The third research period began in the twentieth century with the application of numerical methods to solve contact problems, where the finite element method has been the most widely used [6]. The general formulation for frictionless contact between an elastic body and a rigid foundation was formulated by Signorini [7], subsequently called Signorini problem. The numerical approximation of this problem was described in detail in the book of Kikuchi and Oden [2], where the existence and uniqueness of the solution is provided. The inclusion of frictional effects in the Signorini problem was formulated by Campos et al. [8] assuming the Coulomb's friction law at the contact interface. The extension of this mathematical framework to inelastic materials involving large deformations was presented by Wriggers et al. [9]. The general formulation for the finite deformation frictional contact between deformable bodies was developed by Laursen and Simo [10] in the continuum setting.

Since many frictional contact problems of industrial importance cannot be solved analytically, various commercial and research codes have been developed [11]. In fact, the fast development of the finite element method in the last decades is directly connected with the growth of the computer power [12]. The first attempts to solve contact problems using the finite element method were published in the seventies, namely the work carried out by Wilson and Parsons [13] and the study presented by Chan and Tuba [14]. The huge importance of the frictional contact problems in engineering and the rapid improvement of modern computer technology motivates the development of efficient and robust numerical algorithms to improve the accuracy of the numerical simulations [15]. Nowadays, some commercial finite element packages include the possibility to solve frictional contact problems undergoing

finite deformations (e.g. ANSYS, ABAQUS, LS-DYNA). Nonetheless, in spite of important progresses achieved in computational mechanics, the finite element simulation of contact problems continues to be nowadays a very complex task, mainly due to the strong nonlinearities involved, including geometrical effects, contact nonlinearity itself and possibly nonlinear material behaviour [16].

The low-order finite element discretization method is the most widely used approach in computational contact mechanics, leading to non-smooth contact surfaces, as illustrated in Fig. 3. For the general case of frictional contact problems involving large relative sliding between the contacting deformable bodies, the Node-to-Segment (NTS) contact formulation is frequently adopted [9, 11]. This approach enforces contact conditions at the finite element nodes, preventing the nodes on one contact surface (slave) from penetrating the contact segments on the counterpart contact surface (master) [17]. Although implemented in several finite element codes, the NTS approach presents some well-known drawbacks. First, the single pass NTS algorithm fails the contact patch test, which was introduced by Taylor and Papadopoulos [18] to check the ability of a contact formulation to exactly transmit constant normal stresses between two flat contacting surfaces. Furthermore, when a slave node slides between two adjacent master segments, the sudden normal change induced by the faceted surface representation with linear elements (see Fig. 3) can lead to non-physical oscillations in the computed contact force. Indeed, the normal and tangential contact force components arising in a slave node are aligned with the normal and tangent directions of the closest master segment, respectively [19]. Therefore, the aforementioned numerical jumps may cause serious convergence problems when using iterative solution techniques. In order to overcome this issue, various contact smoothing procedures have been proposed based on Hermite, Bézier, Spline and NURBS descriptions [20–30]. These techniques consider the smoothing of the master surface in order to increase the accuracy archived in the evaluation of the contact kinematic and static variables, leaving the bulk description of the interacting deformable bodies unchanged [31].

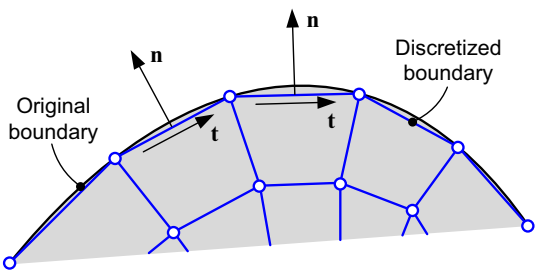

Fig. 3 Faceted contact surface resulting from the low-order finite element discretization

The paper is organized as follows. Section 2 deals with the formulation of the frictional contact problem, defining the kinematic and static contact variables used to define the unilateral contact law and classical Coulomb's friction law. The variational formulation of the problem under analysis is presented in Sect. 3, including the description of the methods widely used in contact mechanics to enforce of contact constraints. Besides, the incremental procedure is presented for the implicit time integration scheme, using the Newton–Raphson method to solve the nonlinear system of equations arising in each time step. The discretization of the contact interfaces is described in Sect. 4, separating the description of rigid obstacles with the Node-to-Surface discretization technique from the discretization of the contact interface between deformable bodies (i.e. Node-to-Node, Node-to-Segment and Segment-to-Segment). Section 5 deals with the smoothing of the contact surface to improve both the geometrical accuracy and the robustness of the numerical methods. The surface interpolations through Bézier and NURBS parameterizations are presented, highlighting the Nagata patch interpolation method. The accuracy of different interpolation methods is evaluated, namely comparing faceted and smoothed surface description methods. The definition of the contact elements is introduced in Sect. 5, comprising both the contact search algorithm and the formulation of the residual vectors and Jacobian matrices for the contact element derived, using the augmented Lagrangian method. The paper ends with several numerical examples and concluding remarks concerning the surface smoothing procedures used in computational contact mechanics.

2 Frictional Contact Problems

Several important engineering applications comprise contact phenomena, which can be classified in three categories: (i) contact between deformable and rigid bodies, (ii) contact between deformable bodies and (iii) self-contact (deformable body contacting itself) [32]. From the mathematical point of view, the contact problem is a physical system subject to a governing variational inequality [2], which can be defined as an initial boundary value problem within a constrained solution space [33]. The difficulties involved in the formulation of frictional contact problems arise due to several reasons: (i) nonlinear nature of contact mechanics, because the contact area is unknown a priori; (ii) both the contact and friction laws are defined by non-smooth and multivalued relations, and (iii) geometrical and material nonlinearities resulting from the large deformation analysis with inelastic material models [16].

The description of the mechanical behaviour of the deformable bodies coming into contact will not be addressed. For a detailed treatment of this subject the reader should consult the literature, e.g. the book of Bonet and Wood [34]. This paper presents a comprehensive overview on the behaviour of the contact interface, in the presence of finite deformations and significant relative sliding. The formulation of frictional contact problems involves two fundamental conditions: the principle of impenetrability of one body by another and the friction law on their common interface [23]. Thus, the unilateral contact law and the classical Coulomb's friction law are defined in the continuum point of view, by means of kinematic and static relations [10, 35, 36].

2.1 Kinematic and Static Variables

For simplicity purposes, without any loss of generality, attention is restricted to contact between two deformable bodies with a single contact zone, as shown schematically in Fig. 4. Since the bodies $\mathcal{B}^1$ and $\mathcal{B}^2$ undergo a finite deformation process, it is necessary to distinguish between reference and current configurations. Thus, the bodies in the reference configuration are represented by the open sets $\Omega_0^1 \subset \Re^3$ and $\Omega_0^2 \subset \Re^3$, while their boundaries are denoted by $\partial\Omega_0^1 = \Gamma^1$ and $\partial\Omega_0^2 = \Gamma^2$. The current configuration of the bodies is obtained by applying the deformation mappings $\boldsymbol{\varphi}^1$ and $\boldsymbol{\varphi}^2$, where the open sets Ω^1 and Ω^2 represent the bodies in the current configuration and their surfaces are denoted by γ^1 and γ^2. Accordingly, for each body, $\mathbf{X}^i \in \Omega_0^i$ and $\mathbf{x}^i \in \Omega^i$ denote the position vector of the same material point in the reference and current configuration, respectively. The motion of the two deformable bodies in continuum mechanics is based on the displacement vector $\mathbf{u}^i = \mathbf{x}^i - \mathbf{X}^i$, as shown in Fig. 4. Since two distinct points in the initial configuration can occupy the same position in the current configuration, contact conditions are formulated with respect to the current configuration.

For each deformable body, its boundary surface in the current configuration γ^i is subdivided in three disjoint subsets, such that:

$$\begin{aligned} &\gamma^i = \gamma_{\mathrm{u}}^i \cup \gamma_{\sigma}^i \cup \gamma_{\mathrm{c}}^i, \\ &\gamma_{\mathrm{u}}^i \cap \gamma_{\sigma}^i = \gamma_{\mathrm{u}}^i \cap \gamma_{\mathrm{c}}^i = \gamma_{\sigma}^i \cap \gamma_{\mathrm{c}}^i = \emptyset, \end{aligned} \tag{1}$$

where γ_{u}^i is the Dirichlet boundary (prescribed motion), γ_{σ}^i is the Neumann boundary (prescribed traction) and γ_{c}^i represents the potential contact surface. Since the frictional contact constraints are defined only in the potential contact surface, all material points on the boundary with possibility to establish contact are included in the potential contact boundary surface. Moreover, the potential contact surface can be divided into two nonintersecting sets: active contact surface $\bar{\gamma}_{\mathrm{c}}^i \subset \gamma_{\mathrm{c}}^i$ (points in contact) and inactive contact surface $\gamma_{\mathrm{c}}^i \backslash \bar{\gamma}_{\mathrm{c}}^i$ (points not in contact). Nevertheless, the active contact surface is unknown a priori (can change over time). Thus, it is determined as part of the nonlinear solution procedure.

In order to define the kinematics of the body surfaces and contact constraints, it is useful to consider a parameterisation of the contact surfaces by means of a local coordinate system [35], shown schematically in Fig. 5 for body 2. The potential contact boundary is a smooth surface for a 3D problem. Thus, all points on the contact surfaces can be parameterized by mappings $\boldsymbol{\psi}^i$, such that $\boldsymbol{\psi}^i: \boldsymbol{\mathcal{A}}^i \rightarrow \Re^3$, where the parametric domain $\boldsymbol{\mathcal{A}}^i \subset \Re^2$ defines the set of points located in the potential contact

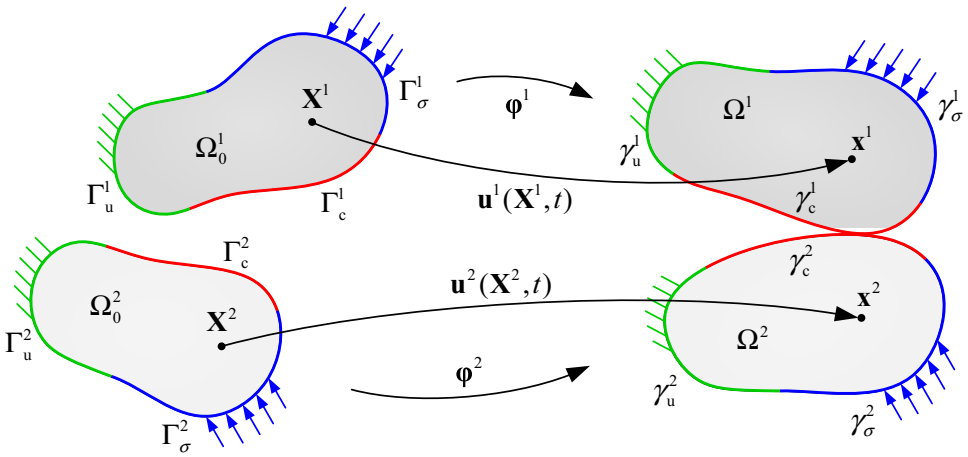

Fig. 4 Schematic description of the two body finite deformation contact problem (reference and current configurations)

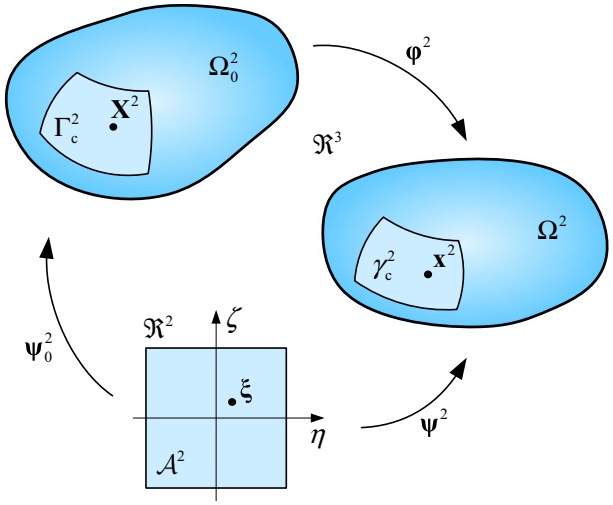

Fig. 5 Parameterization of the contact surface for body 2

surface. The parameterization of the contact surfaces to the reference and current configurations is obtained through $\Gamma_c^i = \boldsymbol{\psi}_0^i(\mathcal{A}^i)$ and $\gamma_c^i = \boldsymbol{\psi}^i(\mathcal{A}^i)$, respectively. Considering any point on the potential contact surface of the body, its vector position in the reference and current configuration can be obtained respectively by $\mathbf{X}^2 = \boldsymbol{\psi}_0^2(\boldsymbol{\xi})$ and $\mathbf{x}^2 = \boldsymbol{\psi}^2(\boldsymbol{\xi})$ for some point $\boldsymbol{\xi} \in \mathcal{A}^2$, as shown in Fig. 5. Furthermore, it is assumed that the parameterization ensures sufficient smoothness, such that the required derivatives can be evaluated [10].

In the framework of the continuum contact problem the classification of the bodies as master or slave is somewhat arbitrary, although the choice becomes important in the discrete setting. In the following, the bodies $\mathcal{B}^1$ and $\mathcal{B}^2$ are conveniently referred as the slave and master body, respectively. Accordingly, the contact surface γ_c^1 is denoted as the slave surface and γ_c^2 is denoted as the master surface. In order to distinguish between the material points located in the interior of the bodies and the points placed on the contact surfaces, $\mathbf{x}^s \in \gamma_c^1$ and $\mathbf{x}^m \in \gamma_c^2$ refer to the slave and master points, respectively. The potential contact surface of the master body is parameterized using a local coordinate system, which is formed by two surface tangent vectors and the normal vector. The surface tangent vectors are given as partial derivatives of the configuration mapping $\boldsymbol{\psi}^2$ with respect to the surface parameterization (covariant basis vectors), defined in the current configuration by:

$$\boldsymbol{\tau}_1(\eta,\zeta) = \frac{\partial \boldsymbol{\psi}^2(\eta,\zeta)}{\partial \eta}; \quad \boldsymbol{\tau}_2(\eta,\zeta) = \frac{\partial \boldsymbol{\psi}^2(\eta,\zeta)}{\partial \zeta}, \tag{2}$$

where $\boldsymbol{\xi} = (\eta, \zeta)$ defines the local parameterization of the master surface via convective coordinates, as illustrated in Fig. 5. The surface normal vector can be computed using the vector product of the tangent basis.

The kinematic variables used to measure the relative motion of two bodies coming into frictional contact are the normal gap function and the tangential relative sliding. The contact interaction between the bodies requires a one-to-one correspondence between each point of the slave surface and its counterpart point on the master surface, even if the bodies are not in contact. Typically the contact point $\mathbf{x}^m \in \gamma_c^2$ is determined according to the closest point projection of the slave point $\mathbf{x}^s$ onto the master surface γ_c^2, defined as:

$$\bar{\mathbf{x}}^m(\mathbf{x}^s) = \arg \min_{\mathbf{x}^m \in \gamma_c^2} \|\mathbf{x}^s - \mathbf{x}^m(\boldsymbol{\xi})\|, \tag{3}$$

where the terminology $\bar{\mathbf{x}}^m(\mathbf{x}^s)$ indicates a one-to-one correspondence between points $\mathbf{x}^s$ and $\mathbf{x}^m$, resulting from the minimization problem presented in (3). Under certain restrictions (at least locally convex region), the closest point problem (3) results in the orthogonal projection of the slave point onto the master surface, where the normal gap function is defined as:

$$g_n(\mathbf{x}^s) = (\mathbf{x}^s - \bar{\mathbf{x}}^m) \cdot \bar{\mathbf{n}}, \tag{4}$$

where $\bar{\mathbf{n}}$ denotes the outward unit normal vector to γ_c^2 at the point $\bar{\mathbf{x}}^m$. The orthogonal projection of a given slave point $\mathbf{x}^s$ onto the master surface γ_c^2 is schematically presented in Fig. 6. The resulting projection point and other quantities evaluated at the solution point are denoted by a bar over the quantity, as illustrated in Fig. 6. The normal gap function g_n can be expressed as the signed distance between the points $\mathbf{x}^s$ and $\bar{\mathbf{x}}^m$, measured in the normal direction to the master surface. Note that applying the definition of normal gap function presented in (4), this quantity is positive if the slave point is outside of the master body, otherwise it will be negative. Two possible geometrical situations of a slave

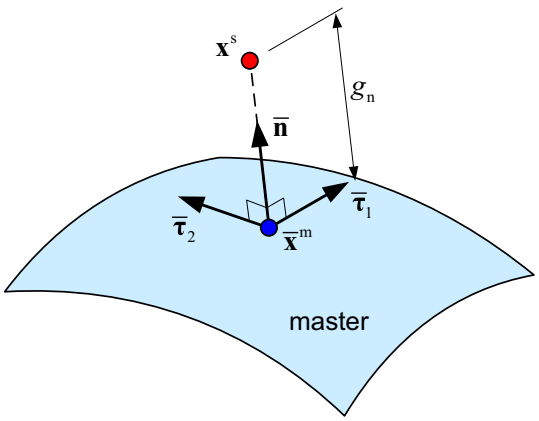

Fig. 6 Orthogonal projection of the slave point onto the master surface

point with respect to the master body are schematically indicated in Fig. 7.

The change of the closest point projection describes the tangential relative sliding between surfaces, which is necessary for modelling friction effects. It is connected to the change of the point $\mathbf{x}^{\mathrm{s}}$ relative to the projection $\bar{\mathbf{x}}^{\mathrm{m}}$. This means that the solution point $\bar{\xi} = (\bar{\eta}, \bar{\zeta})$, which has been obtained via the minimal distance problem (3), will move on the master surface. Then, the tangential slip increment of a slave point on the contact master surface can be defined in the incremental form as:

$$\Delta \mathbf{g}_{\mathrm{t}} = \bar{\tau}_1 \Delta \bar{\eta} + \bar{\tau}_2 \Delta \bar{\zeta}, \tag{5}$$

where $\bar{\tau}_1$ and $\bar{\tau}_2$ are the covariant tangential basis vectors, defined by (2), at the projection point. Nevertheless, when the initial and final positions of the slip path belong to different local parameterization domains, as a consequence of the finite element discretization, the frictional time integration becomes meaningless [35]. This problem can be avoided using the history information, as described by Agelet de Saracibar [37], where the slip path length is evaluated through the position of the slave point at the beginning and the end of a time increment. Hence, quantities of the previous and current time steps will be denoted as ${}^{n}(\cdot)$ and ${}^{n+1}(\cdot)$, respectively. In order to define the incremental slip with respect to the current configuration, the variables from the last converged configuration ${}^{n}(\cdot)$ are mapped forward to the current configuration using the notation ${}^{n}(\tilde{\cdot})$. This means that these variables are evaluated in the current configuration using the convective coordinates of the projection point in the last converged configuration ${}^{n}\bar{\xi} = ({}^{n}\bar{\eta}, {}^{n}\bar{\zeta})$, as illustrated in Fig. 8.

The simplest approximation for the slip path is given by the vector connecting the solution point in the current configuration and the projection point calculated in the last converged configuration, mapped forward to the current configuration (see Fig. 8). It is expressed by:

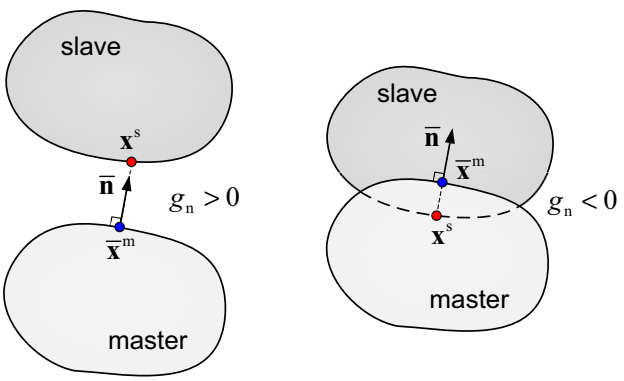

Fig. 7 Geometrical condition of the slave point with respect to the master body: (*left*) separated and (*right*) penetration

$$ {}^{n+1}\Delta \mathbf{g} = {}^{n+1}\bar{\mathbf{x}}^{\mathrm{m}}({}^{n+1}\bar{\eta}, {}^{n+1}\bar{\zeta}) - {}^{n}\tilde{\mathbf{x}}^{\mathrm{m}}({}^{n}\bar{\eta}, {}^{n}\bar{\zeta}), \tag{6}$$

where ${}^{n+1}\bar{\mathbf{x}}^{\mathrm{m}}({}^{n+1}\bar{\eta}, {}^{n+1}\bar{\zeta})$ denotes the position vector of the projection point in the current configuration and ${}^{n}\tilde{\mathbf{x}}^{\mathrm{m}}({}^{n}\bar{\eta}, {}^{n}\bar{\zeta})$ represents the position vector of the projection point in the last converged configuration, mapped into the current configuration. Since, in general, the slip vector (6) is not lying in the tangential plane of the contact surface (see Fig. 8), the tangential slip vector is given by:

$$ {}^{n+1}\mathbf{g}_{\mathrm{t}} = (\mathbf{I} - {}^{n+1}\bar{\mathbf{n}} \otimes {}^{n+1}\bar{\mathbf{n}})\, {}^{n+1}\Delta \mathbf{g}, \tag{7}$$

where ${}^{n+1}\bar{\mathbf{n}}$ denotes the master surface normal vector at the solution point, evaluated in the current time step, as illustrated in Fig. 8. Note that the tangential slip vector defines the direction of the frictional force in case of slip status.

In order to prevent one body from penetrating the other, contact forces arise at the contact interface, which are the static variables used to model the frictional contact interactions. Since adhesive stresses will not be allowed in the contact interface, the contact pressure is assumed negative in compression and zero in inactive contact zones. In continuum mechanics, the contact traction is expressed by the Cauchy stress vector $\mathbf{t} = \mathbf{t}^2 = -\mathbf{t}^1$, which satisfy the action–reaction principle in the contact point. Hence, the contact traction is decomposed into the normal and tangential components:

$$\begin{aligned} \mathbf{t} &= p_{\mathrm{n}}\mathbf{n} + \mathbf{t}_{\mathrm{t}}, \\ p_{\mathrm{n}} &= \mathbf{t} \cdot \mathbf{n}, \\ \mathbf{t}_{\mathrm{t}} &= (\mathbf{I} - \mathbf{n} \otimes \mathbf{n})\mathbf{t}, \end{aligned} \tag{8}$$

which physically represents the force exerted by the slave point on the master surface. Note that the tangential component is zero in the case of frictionless contact. The mechanical formulation of frictional contact problems involves constraints related with impenetrability and friction conditions. These constraints are expressed considering relationships between the previously presented kinematic and static variables. The contact traction in the normal and in the tangential directions are coupled with the normal distance and the tangential slip increment, respectively.

2.2 Unilateral Contact Law

The unilateral contact condition defines the physical requirements of impenetrability and compressive interaction between the bodies. Due to the sign convention chosen for the gap function defined in (4), a positive value $g_{\mathrm{n}} > 0$ defines points not in contact, while a negative value $g_{\mathrm{n}} < 0$ denotes penetration between bodies (see Fig. 7), which is physically not admissible. Therefore, the normal contact conditions on the contact interface can be formulated using

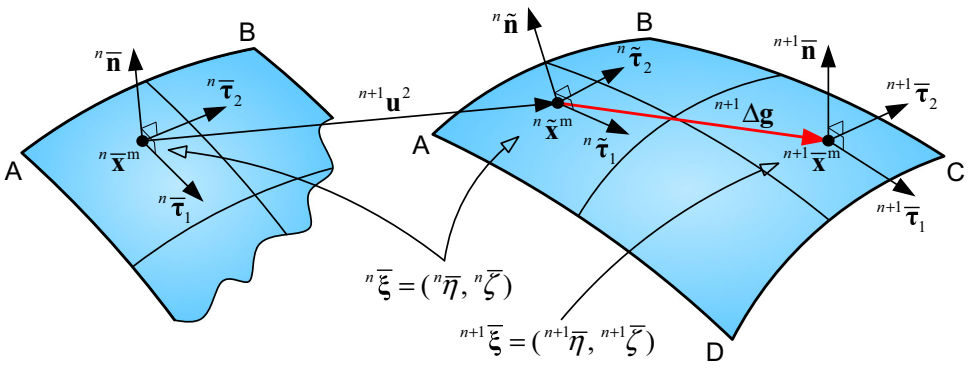

Fig. 8 Definition of the tangential slip vector using the mapping of the closest point projection from the previous time step forward to the current time step

the classical set of Karush–Kuhn–Tucker (KKT) conditions for optimality, being stated as:

$$g_{\mathrm{n}} \geq 0, \quad p_{\mathrm{n}} \leq 0, \quad p_{\mathrm{n}} g_{\mathrm{n}} = 0, \tag{9}$$

which must hold for all points on the slave contact surface. These conditions are commonly called Hertz–Signorini–Moreau conditions for frictionless contact mechanics. The first condition in (9) simply states the geometric impenetrability condition, whereas the second one refers that the contact pressure must be compressive (no adhesive stresses are allowed) in the contact zone. The last condition in (9) states the complementarity condition, which forces the gap to be zero if compressive tractions occur and the pressure to be zero if the gap function is positive. Moreover, the unilateral contact conditions can be decomposed into two parts: active and inactive contact zones, defined as:

$$\begin{cases} g_{\mathrm{n}} = 0, & p_{\mathrm{n}} < 0, \quad \text{at } \bar{\gamma}_{\mathrm{c}}^{1} \\ g_{\mathrm{n}} > 0, & p_{\mathrm{n}} = 0, \quad \text{at } \gamma_{\mathrm{c}}^{1} \backslash \bar{\gamma}_{\mathrm{c}}^{1} \end{cases}, \tag{10}$$

where the two possible geometrical situations denoted as contact and gap are illustrated in Fig. 9.

The set of inequality conditions stated in (9) leads to a non-smooth contact law for the normal contact pressure, but also the relation between the contact pressure and the normal gap is multivalued at $g_{\mathrm{n}} = 0$, as shown in Fig. 9. This means that the contact law is non-differentiable and can take an infinite number of values at the origin. This singularity can be physically interpreted as a result from the fact that the contact pressure is a reaction force, which cannot be calculated by the unilateral contact law (9). It is a result of the equilibrium between the bodies in contact. This difficulty usually arises also in optimization problems subjected to inequality constraints. Thus, standard solution techniques from optimization theory can be adapted for contact mechanics. In spite of these peculiarities, the unilateral contact law (9) can be expressed by sub-gradients of non-differentiable convex potentials, following Moreau's works on convex analysis. This formalism applied to contact problems is stated in the works of Alart and Curnier [38] and Heegaard and Curnier [39].

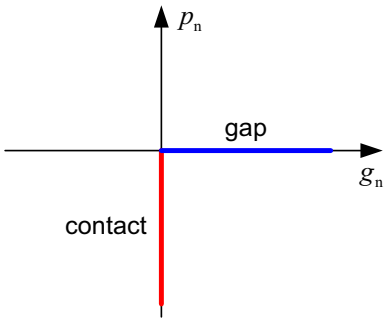

Fig. 9 Unilateral contact law defined by the Karush–Kuhn–Tucker conditions

Since the present analysis is restricted to a macroscopic description of normal contact compliance, the contact constraints defined in (9) for the contact interface are based on a purely geometrical perspective. Consequently, the micromechanical behaviour is not taken into account [40], which in general depends upon several parameters such as hardness and surface roughness [3]. Nevertheless, some contact problems require the knowledge of the micromechanical interaction for a proper treatment of the physical phenomena. Hence, some constitutive equations for normal contact have been developed based on experiments, which take into account the micro deformation of the contact bodies [41].

2.3 Friction Law

The friction laws used in the contact interface of engineering problems can be divided in two groups: (i) the classic Coulomb's friction law [42–44] and (ii) other non-associated friction laws [45, 46]. In this study, the friction response is formulated through the classical non-associated Coulomb's friction law. This law establishes that the frictional force at the contact interface depends on the contact

pressure. Since the contact pressure is only known at the solution, the constraints imposed by the friction law are solution dependent, leading to additional difficulties in the formulation of the frictional contact problem. The Coulomb's non-associated friction law can be described by the following three conditions:

$$\begin{aligned} &\|\mathbf{t}_{\mathrm{t}}\| - \mu|p_{\mathrm{n}}| \leq 0, \quad \mathbf{t}_{\mathrm{t}} - \mu|p_{\mathrm{n}}| \frac{\mathbf{g}_{\mathrm{t}}}{\|\mathbf{g}_{\mathrm{t}}\|} \leq \mathbf{0}, \\ &\|\mathbf{g}_{\mathrm{t}}\|(\|\mathbf{t}_{\mathrm{t}}\| - \mu|p_{\mathrm{n}}|) = 0, \end{aligned} \tag{11}$$

where μ is the friction coefficient, which is assumed constant in the classical Coulomb law. Nevertheless, in general the friction coefficient depends on several parameters, such as normal pressure, relative tangential velocity, surface roughness and temperature [47]. The first condition in (11), usually referred as Coulomb friction condition, imposes that the magnitude of the frictional force does not exceed a threshold value, defined as the product of the friction coefficient by the contact pressure modulus (see Fig. 10). The second condition in (11), also referred as slip rule, defines that the frictional force vector is collinear with the tangential relative sliding velocity. Besides, the frictional force arising at the slave point is opposite to its slip direction. The last equation in (11) is a complementarity condition, which distinguishes two different contact situations, usually referred as stick and slip statuses.

If the frictional force has not reached the Coulomb threshold, the contact point in contact is not allowed to move in the tangential direction (null velocity), assigned with a commonly called stick status. On the other hand, when the tangential traction reaches the Coulomb limit, the contact point moves in the tangential direction of the contact interface, the point is assigned with the slip status. The conditions (11) can be represented graphically, as shown in Fig. 10, where stick and slip statuses are depicted.

Such as in the unilateral contact law, the Coulomb's friction law also yields a non-smooth functional at the onset of sliding (see Fig. 10a), creating difficulties from the mathematical and numerical point of view. However, the friction law can also be expressed by means of convex analysis in the form of sub-gradients of non-differentiable, convex *quasi*-potentials, as presented by Alart and Curnier [38] and Pietrzak and Curnier [23]. The cone depicted in Fig. 10b is called the Coulomb's cone, which relates the contact pressure and the components of the contact tangential stress vector, defined in the first condition of (11). Thus, any admissible contact stress vector corresponds to a unique point either in the interior of the cone (stick status) or on its closure (slip status). The change of position in the interior of Coulomb's cone does not result in relative tangential displacements, while the relative sliding implies that the point is located on Coulomb's cone surface.

The frictional contact conditions, which involve both the unilateral contact law and the Coulomb's friction law, can be decomposed in different constraints applied in three zones of the potential contact surface:

$$\begin{cases} g_{\mathrm{n}} = 0, \quad \mathbf{g}_{\mathrm{t}} = \mathbf{0}, \quad p_{\mathrm{n}} < 0, \quad \|\mathbf{t}_{\mathrm{t}}\| < \mu|p_{\mathrm{n}}|, & \text{in the stick zone } \gamma_{\mathrm{c}}^{1\cdot} \\ g_{\mathrm{n}} = 0, \quad p_{\mathrm{n}} < 0, \quad \mathbf{t}_{\mathrm{t}} = \mu\|p_{\mathrm{n}}\| \dfrac{\mathbf{g}_{\mathrm{t}}}{\|\mathbf{g}_{\mathrm{t}}\|}, & \text{in the slip zone } \gamma_{\mathrm{c}}^{1*} \\ g_{\mathrm{n}} > 0, \quad p_{\mathrm{n}} = 0, \quad \mathbf{t}_{\mathrm{t}} = \mathbf{0}, & \text{in the inactive contact zone } \gamma_{\mathrm{c}}^{1} \backslash \bar{\gamma}_{\mathrm{c}}^{1} \end{cases}, \tag{12}$$

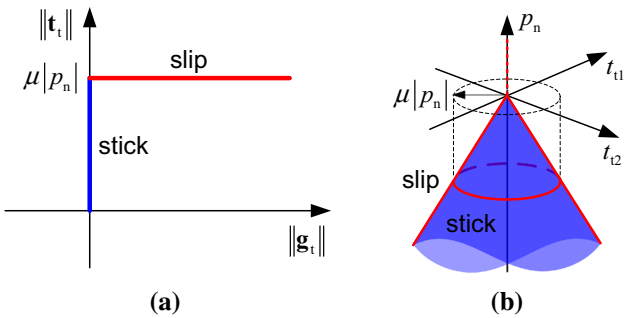

Fig. 10 Graphical representation of Coulomb's friction law: **a** relation between the norm of the tangential velocity and the norm of the friction force; **b** relation between the contact pressure and the friction force (Coulomb's frictional cone)

where the active contact zone $\bar{\gamma}_{\mathrm{c}}^{1}$ is split into a stick $\gamma_{\mathrm{c}}^{1\cdot}$ and a slip γ_{c}^{1*} zone, such that $\gamma_{\mathrm{c}}^{1\cdot} \cup \gamma_{\mathrm{c}}^{1*} = \bar{\gamma}_{\mathrm{c}}^{1}$ and $\gamma_{\mathrm{c}}^{1\cdot} \cap \gamma_{\mathrm{c}}^{1*} = \emptyset$. On the other hand, the inactive contact zone does not presents any imposed boundary condition, as expressed in (12). The source of nonlinearity in all contact problems comes from the presence of the unknown active contact zone, which is divided into stick and slip zones.

3 Variational Equality Formulation

The frictional contact problems can be solved employing variational inequality or variational equality formulations [33]. The variational inequality formulation provides the

principle of virtual work in the inequality form due to the contact and friction constraints (12). The frictionless elastostatic contact problem was formulated by Kikuchi and Oden [2] using variational inequalities, establishing the existence and uniqueness of the solution using standard optimization methods. Nevertheless, the variational inequality for frictional contact problems poses several formulation difficulties [38] because the friction force depends on the normal contact force, while both are unknown quantities. In fact, the variational inequality containing the non-convex and non-differentiable term describing the virtual work of frictional forces cannot be modelled as a standard optimization problem [48]. If the contact surface is assumed as known in the variational equality formulation (virtual work equation written in the equality form), it is possible establish an equivalent optimization problem. Simo and Laursen [42] formulate the frictional contact problem using the variational equality approach, adopting the augmented Lagrangian method to enforce the contact and friction conditions. The problem was formulated in the context of contact between a deformable body and a rigid obstacle. The generalisation for contact between deformable bodies undergoing large deformation was formulated by Laursen and Simo [10]. The finite element solution of thermo-mechanical frictional contact problems based on the equality formulation was presented by Agelet de Saracibar [49]. The variational equality formulation has been much more successful than the variational inequality formulation with respect to applicability to large scale industrial frictional contact problems, since it can be easily incorporated into the existing finite element codes [33].

The nonlinear boundary value problem (BVP) for the frictional contact system undergoing finite deformation, shown in Fig. 4, is stated as follows:

$$\begin{cases} \operatorname{div}(\boldsymbol{\sigma}^i) + \mathbf{b}^i = \mathbf{0}, & \text{in } \Omega^i \\ \mathbf{t}^i = \boldsymbol{\sigma}^i \mathbf{n}^i = \bar{\mathbf{t}}^i, & \text{on } \gamma^i_\sigma \\ \mathbf{u}^i = \bar{\mathbf{u}}^i, & \text{on } \gamma^i_{\mathrm{u}} \end{cases}, \tag{13}$$

where the inertia terms are neglected. The balance of linear and angular momentum for each body $\mathcal{B}^i$ $(i = 1, 2)$ coming into contact is described by the first equation, where $\boldsymbol{\sigma}^i$ is the Cauchy stress tensor acting in the body i, and $\mathbf{b}^i$ are the volume or body forces (e.g. due to gravitation). The notation $\bar{\mathbf{t}}^i$ represents a prescribed Cauchy traction (Neumann boundary conditions) and $\bar{\mathbf{u}}^i$ denotes a prescribed displacement (Dirichlet boundary conditions), as shown schematically in Fig. 4. Furthermore, the inequality constraints related with frictional contact conditions given in (12) need to be included in the problem formulation, in order to restricts the solution space.

The numerical solution of the nonlinear BVP using the finite element method requires the weak form, often called variational form of the local field equations. Both bodies must satisfy the balance of virtual work, expressed in the current configuration to account for the frictional contact conditions. The principle of virtual work for the two body system is obtained by adding the weak form of each body. The derivation of the weak form starts from the local equilibrium equation, which is multiplied by a virtual displacement vector (variations on the solution displacement field) before the integration over the volume under consideration, yielding:

$$\int_\Omega \boldsymbol{\sigma} : \delta\nabla\mathbf{u} d\Omega - \int_\Omega \mathbf{b} \cdot \delta\mathbf{u} d\Omega - \int_\gamma (\boldsymbol{\sigma}\mathbf{n}) \cdot \delta\mathbf{u} d\gamma = 0, \tag{14}$$

where the union of two bodies domains is denoted by $\Omega = \Omega^1 \cup \Omega^2$ and $\gamma = \gamma^1 \cup \gamma^2$. The Cauchy stress vector $\boldsymbol{\sigma}\mathbf{n}$ is non-null in the boundary on which tractions are prescribed $\gamma_\sigma = \gamma^1_\sigma \cup \gamma^2_\sigma$, on the surface boundary where the displacements are prescribed $\gamma_{\mathrm{u}} = \gamma^1_{\mathrm{u}} \cup \gamma^2_{\mathrm{u}}$ and in the active contact zone. Thus, the contact contribution to the variational principle is included in the last term of (14). The virtual work related with the surface forces acting on the bodies can be written as:

$$\int_\gamma (\boldsymbol{\sigma}\mathbf{n}) \cdot \delta\mathbf{u} d\gamma = \int_{\gamma_\sigma} \bar{\mathbf{t}} \cdot \delta\mathbf{u} d\gamma - \int_{\gamma^1_{\mathrm{c}}} \mathbf{t} \cdot \delta(\mathbf{x}^{\mathrm{s}} - \bar{\mathbf{x}}^{\mathrm{m}}) d\gamma, \tag{15}$$

which can be expressed as an equality within an active set strategy, i.e. the active contact zone is known within an incremental solution step. Due to the balance of linear momentum at the contact interface, the contact traction vector satisfies the action–reaction principle in the contact point, allowing to define the contact contribution using a single integral over the slave surface. Considering the variations of the kinematic contact variables (normal gap and tangential slip increment) associated with the variations of the solution displacement field $\delta\mathbf{u}$, the first order variations of kinematic quantities are given by [10, 23]:

$$\delta g_{\mathrm{n}} = (\delta\mathbf{x}^{\mathrm{s}} - \delta\bar{\mathbf{x}}^{\mathrm{m}}) \cdot \bar{\mathbf{n}}, \quad \delta\mathbf{g}_{\mathrm{t}} = \delta\bar{\eta}\bar{\tau}_1 + \delta\bar{\zeta}\bar{\tau}_2, \tag{16}$$

which are obtained from (4) and (5), respectively. Using the static contact variables, the virtual work balance for the frictional contact problem of the two body system may be summarized as:

$$\begin{aligned} &\int_\Omega \boldsymbol{\sigma} : \delta\nabla\mathbf{u} d\Omega - \int_\Omega \mathbf{b} \cdot \delta\mathbf{u} d\Omega - \int_{\gamma_\sigma} \bar{\mathbf{t}} \cdot \delta\mathbf{u} d\gamma \\ &\quad + \int_{\gamma^1_{\mathrm{c}}} (p_{\mathrm{n}} \delta g_{\mathrm{n}} + \mathbf{t}_{\mathrm{t}} \cdot \delta\mathbf{g}_{\mathrm{t}}) d\gamma = 0, \end{aligned} \tag{17}$$

where the normal contact pressure p_{n} and the frictional force vector $\mathbf{t}_{\mathrm{t}}$ have definitions that depend on the regularisation method adopted for the contact constraints, as discussed in the next section.

3.1 Contact Constraint Enforcement Methods

Assuming that the active contact zone is known, the variational inequality is replaced by a variational equality with an additional contact term (17), resulting in an unconstrained or partly unconstrained optimization problem [15]. This allows to apply methods well known from optimization theory to solve contact mechanics problems [50]. Due to the requirement of a known contact zone, the active contact zone defined by the frictional contact constraints should be checked and updated in each solution step. The methods widely used in contact mechanics to enforce the contact constraints are the penalty method, the Lagrange multiplier method and the augmented Lagrangian method [51]. Each one presents its own advantages and drawbacks, which are discussed in detail in the following.

3.1.1 Penalty Method

The penalty method is one of the most widely used for treating contact problems, both in commercial and scientific finite element codes [51]. The basic idea behind this method is to remove the contact constraints explicitly from the variational equality formulation by means of a penalization of the constraint violations. The magnitude of the penalization increases according to how severely the constraint is violated [2]. Therefore, it is known to be simple and can be physical interpreted as a series of springs in the contact interface with zero initial length, as represented in Fig. 11. Nevertheless, the contact conditions are fulfilled exactly only in case the penalty parameter is infinite, which results in ill-conditioning of the numerical problem, i.e. high condition number of the tangent matrix [42].

Since the contact constraints are explicitly removed from the variational formulation using the penalty method, the contact problem is formulated as an unconstrained optimization problem, where the unknowns are only the displacement variables. The penalty regularization dictates that the non-penetration condition stated in (9) is only approximately fulfilled (see Fig. 11). The contact pressure is assumed to be a continuous function dependent of the penetration, given by:

$$p_{\mathrm{n}}(g_{\mathrm{n}}) = -\varepsilon_{\mathrm{n}}\langle -g_{\mathrm{n}}\rangle = \begin{cases} \varepsilon_{\mathrm{n}} g_{\mathrm{n}}, & g_{\mathrm{n}} \leq 0 \\ 0, & g_{\mathrm{n}} > 0 \end{cases}, \tag{18}$$

where ε_{n} is the penalty parameter and the notation $\langle\cdot\rangle$ is used to denote the Macaulay bracket, which simply gives the positive part of its operand. The regularization of the unilateral contact law is illustrated schematically in Fig. 12a, where the contact condition is strictly fulfilled for non-negative gaps (compare with Fig. 9). On the other hand, the linear relation between contact pressure and the gap function dictates that the contact arises only for negative gap values, allowing penetration of the bodies. Therefore, the impenetrability condition (9) is only perfectly represented using an infinite value for the penalty parameter.

Analogously, the classical Coulomb's friction law (11) can be approximately fulfilled using a penalty function. The regularization of the Coulomb friction law considers the frictional force as a function of the tangential sliding limited by the Coulomb's cone surface, such as:

$$\mathbf{t}_{\mathrm{t}}(\mathbf{g}_{\mathrm{t}}) = \begin{cases} \varepsilon_{\mathrm{t}}\mathbf{g}_{\mathrm{t}}, & \varepsilon_{\mathrm{t}}\|\mathbf{g}_{\mathrm{t}}\| < \mu|p_{\mathrm{n}}| \quad \textit{stick} \\ \mu|p_{\mathrm{n}}|\dfrac{\mathbf{g}_{\mathrm{t}}}{\|\mathbf{g}_{\mathrm{t}}\|}, & \varepsilon_{\mathrm{t}}\|\mathbf{g}_{\mathrm{t}}\| \geq \mu|p_{\mathrm{n}}| \quad \textit{slip} \end{cases}, \tag{19}$$

where ε_{t} denotes the tangential penalty parameter, which is not necessarily equal to the normal penalty parameter ε_{n}. The multivalued nature of the Coulomb friction law (see Fig. 10) is removed through the penalty regularization, providing the frictional force single-valued function of the tangential displacement, as shown schematically in Fig. 12b. The perfect representation of the Coulomb law is only recovered using an infinite value for the penalty parameter. The penalty method allows some tangential movement at the contact interface for the stick status (see Fig. 12b), situation inadmissible in the unregularized Coulomb law (11). Therefore, if the selected penalty parameter value is too low, this can produce non-physical behaviour at the interface. Nevertheless, there is significant experimental evidence indicating that the sharp

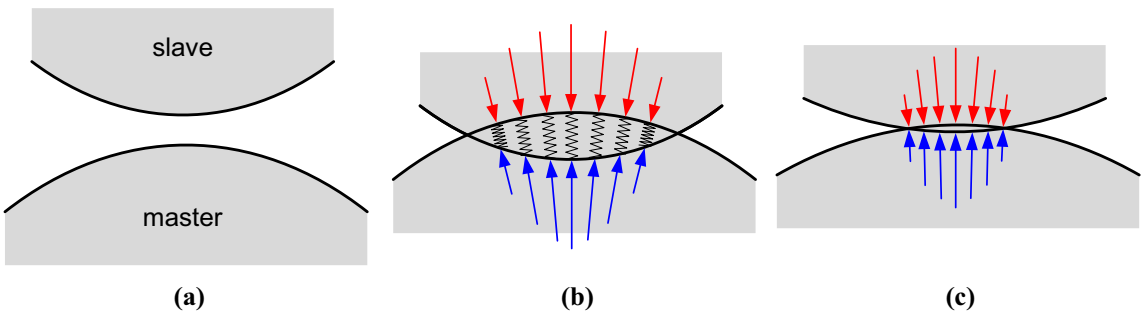

Fig. 11 Physical interpretation of the penalty method: **a** initial configuration; **b** configuration after penetration; **c** equilibrium state

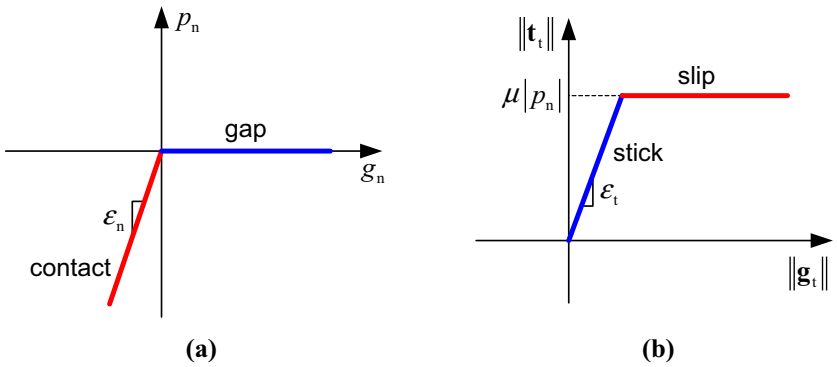

Fig. 12 Application of the penalty method to the frictional contact problem: **a** regularized unilateral contact law; **b** regularized Coulomb's friction law

transition between slick and slip friction (Fig. 10a) does not exist in many systems due to the elastic deformation of the contact surface asperities [52]. Thus, in some cases the penalty regularization can be used with the purpose to better model the friction behaviour [3].

Since the penalty term is only added for active constraints, normal and tangential conditions are formulated separately, distinguishing between stick and slip conditions. In the case of slip status the frictional force is evaluated through the contact pressure and the slip direction. Nevertheless, when only stick status occurs in the contact interface there is no need to distinguish between the normal and tangential directions, dictating equal penalty parameters for both directions [3]. Therefore, the frictional contact virtual work defined according to the penalty method is given by:

$$G^{c}(\mathbf{u},\delta\mathbf{u}) = \int_{\bar{\gamma}_c^{1*}} -\varepsilon_n\langle -g_n\rangle(\delta g_n - \mu\frac{\mathbf{g}_t}{\|\mathbf{g}_t\|}\cdot\delta\mathbf{g}_t)d\gamma_c^1 + \int_{\bar{\gamma}_c^{1\cdot}} -\varepsilon_n\langle -g_n\rangle\delta g_n + \varepsilon_t\mathbf{g}_t\cdot\delta\mathbf{g}_t)d\gamma_c^1, \tag{20}$$

where $\bar{\gamma}_c^{1*}$ and $\bar{\gamma}_c^{1\cdot}$ are the slip and the stick active contact zones on the master surface, respectively.

3.1.2 Lagrange Multiplier Method

The Lagrange multiplier method can also be used to include the frictional contact constraints in the virtual work balance. This method is commonly used in optimization problems to find the maximum/minimum of a functional subjected to equality constraints. The idea behind the Lagrange multiplier method is to introduce a vector of additional unknowns $\boldsymbol{\lambda}$, called Lagrange multipliers, replacing the original constrained optimization problem by a functional stationary in the solution. Nevertheless, the Lagrange multiplier method does not converts a minimization problem with inequality constraints into a fully unconstrained one, since the constraint λ_n has still to be fulfilled [51]. The Lagrange multiplier method allows an exact fulfilment of the frictional contact constrains (9) and (11) introducing additional degrees of freedom (Lagrange multipliers), which dictates an increase of the number of primary unknowns in comparison with the penalty method. The disadvantage of an increased system size can be avoided by employing the primal–dual active set strategy, which allows integrate all nonlinearities into one single nonlinear iteration scheme [53, 54].

The contribution of the frictional contact to the virtual work balance, using the Lagrange multiplier method to enforce the contact constraints, is given by:

$$\delta\mathcal{L}(\mathbf{u},\lambda_n,\boldsymbol{\lambda}_t) = \int_{\bar{\gamma}_c^{1\cdot}} \mathbf{g}(\mathbf{u})\cdot\delta\boldsymbol{\lambda} + \boldsymbol{\lambda}\cdot\delta\mathbf{g}(\mathbf{u})d\gamma_c^1 + \int_{\bar{\gamma}_c^{1*}} g_n(\mathbf{u})\delta\lambda_n + \lambda_n\delta g_n(\mathbf{u}) + \mu|\lambda_n|\frac{\mathbf{g}_t}{\|\mathbf{g}_t\|}\cdot\delta\mathbf{g}_t(\mathbf{u})d\gamma_c^1, \tag{21}$$

where the Lagrange multiplier vector $\boldsymbol{\lambda}$ represents the contact force vector needed to fulfil the frictional contact constraints. Since the contact virtual work is integrated only in the points where the contact constraints are active (union of stick and slip zones), the solution involves an iterative procedure to determine the active contact zone, which is typically performed with an active set strategy [50]. For a more rigorous formulation of Lagrange multiplier method for contact problems the reader is referred to the book by Kikuchi and Oden [2].

3.1.3 Augmented Lagrangian Method

Another approach widely used to solve minimization problems under constraints is the augmented Lagrangian approach, originally proposed by Hestenes [55] and Powell [56] to solve constrained optimization problems. The application of this method to frictional contact problems involving large displacements was firstly presented by Alart and Curnier [38]. The main idea of this method is to

combine the advantages of both the penalty and the Lagrange multiplier methods, allowing an exact enforcement of contact constraints for a finite value of the penalty parameter, while providing a smooth functional [23]. The augmented Lagrangian method can be expressed as a Lagrange multiplier formulation regularized by penalty functions [51]. It yields a C^1-differentiable energy functional, which is described in detail in [23] and provides a saddle point solution (minimize primal variables and maximize dual variables) fully unconstrained. This advantage is not verified in the Lagrange multiplier method since the condition of non-positivity of the Lagrange multipliers must be satisfied. Thus, the augmented Lagrangian method is better from a numerical point of view.

The implementation of the augmented Lagrangian method to solve frictional contact problems can be carried out in two different ways. The first approach considers the augmented Lagrangian method with Uzawa's algorithm, referred as nested augmented Lagrangian algorithm, which was firstly reported by Simo and Laursen in [42]. In that case, the value of the Lagrange multipliers is assumed know and removed from the unknowns. Therefore, the system to be solved is composed only by displacements, presenting the same dimension as the problem without contact. Nevertheless, this leads to a double loop algorithm, where the Lagrange multiplier is held constant only during an iteration loop to solve the weak form in the inner loop, where Lagrange multiplier is updated in the outer loop. The rate of convergence of this method, where primal and dual variables are updated independently, is linear [56]. The advantage is that the resulting functional is smooth enough to apply a standard Newton's technique, in order to obtain the solution problem. The second approach is the coupled augmented Lagrangian method, which has been undertaken by Alart and Curnier [38]. In that case, the Lagrange multipliers are retained as independent variables in the coupled problem, increasing the computational effort, while both variables are updated simultaneously [57]. Similar to the Lagrange multiplier method, a vector of additional unknowns is introduced, which physically represent the contact forces [39]. Nevertheless, since the functional is not sufficiently smooth to apply the standard Newton's technique, the generalized Newton method is adopted to deal with non-smooth functions [58]. Since the penalty parameter value is a crucial factor for the convergence rate of the augmented Lagrangian method, a technique for penalty parameter updating has been proposed by Mijar and Arora [59, 60].

The application of the augmented Lagrangian method to solve frictional contact problems is given in detail by Pietrzak [23]. The formulation is based in the formalism used by Alart and Curnier [38] and Heegaard and Curnier [39], which follows the Moreau's convex analysis. Following the cited authors, the Hertz–Signorini–Moreau conditions that corresponds to the unilateral contact law (9) can be equivalently written as the sub-differential inclusion:

$$p_{\mathrm{n}} \in \partial\psi^{+}(g_{\mathrm{n}}), \tag{22}$$

where ψ^{+} denotes the indicator function of the positive half-line $\Re^{+}$ and $\partial\psi^{+}$ represents its sub-differential [39]. Similarly, the conditions arising from the Coulomb's friction law (11) can be reformulated as a sub-differential inclusion, where the frictional contact force vector is expressed as:

$$\mathbf{t}_{\mathrm{t}} \in \partial\psi^{*}_{C(p_{\mathrm{n}})}(\mathbf{g}_{\mathrm{t}}), \tag{23}$$

where $\psi^{*}_{C(p_{\mathrm{n}})}$ denotes the conjugate function of the disk indicator function $\psi_{C(p_{\mathrm{n}})}$, while $C(p_{\mathrm{n}})$ represents the convex disk of radius $\mu|p_{\mathrm{n}}|$ centred at the origin (section of the Coulomb's cone), which is function of the unknown contact pressure.

The augmented Lagrangian functional is formulated including the framework of sub-differential inclusions, where the frictional contact contribution is defined by:

$$\mathcal{L}^{\mathrm{a}}(\mathbf{u}, \lambda_{\mathrm{n}}, \boldsymbol{\lambda}_{\mathrm{t}}, \hat{p}_{\mathrm{n}}) = \int_{\gamma_{\mathrm{c}}^{1}} l_{\mathrm{n}}(g_{\mathrm{n}}, \lambda_{\mathrm{n}}) + l_{\mathrm{t}}(\mathbf{g}_{\mathrm{t}}, \boldsymbol{\lambda}_{\mathrm{t}}, \hat{p}_{\mathrm{n}}) d\gamma_{\mathrm{c}}^{1}, \tag{24}$$

where l_{n} and l_{t} denotes the augmented Lagrangian related to normal and frictional contact, which represent the regularized functions (22) and (22), respectively. The Lagrange multipliers λ_{n} and $\boldsymbol{\lambda}_{\mathrm{t}}$ represent the normal contact force and the friction force, respectively. Since the contact pressure is independent of the friction force but the friction force depends on the contact pressure (non-associated character of the Coulomb's friction law), the contact pressure at the solution is denoted by $\hat{p}_{\mathrm{n}}$.

The closed form of the augmented Lagrangian functional related with the unilateral contact law takes the form [23]:

$$l_{\mathrm{n}}(g_{\mathrm{n}}, \lambda_{\mathrm{n}}) = \begin{cases} g_{\mathrm{n}}\hat{\lambda}_{\mathrm{n}} - \dfrac{\varepsilon_{\mathrm{n}}}{2} g_{\mathrm{n}}^{2}, & \hat{\lambda}_{\mathrm{n}} \leq 0, \quad \textit{contact} \\ -\dfrac{1}{2\varepsilon_{\mathrm{n}}} \lambda_{\mathrm{n}}^{2}, & \hat{\lambda}_{\mathrm{n}} > 0, \quad \textit{non-contact} \end{cases}, \tag{25}$$

where the augmented Lagrange multiplier is denoted by a hat:

$$\hat{\lambda}_{\mathrm{n}} = \lambda_{\mathrm{n}} + \varepsilon_{\mathrm{n}} g_{\mathrm{n}}. \tag{26}$$

The augmented Lagrangian functional due to Coulomb's friction law is written as:

$$l_{\mathrm{t}}(\mathbf{g}_{\mathrm{t}}, \boldsymbol{\lambda}_{\mathrm{t}}, \hat{p}_{\mathrm{n}}) = \begin{cases} \boldsymbol{\lambda}_{\mathrm{t}} \cdot \mathbf{g}_{\mathrm{t}} - \dfrac{\varepsilon_{\mathrm{n}}}{2} \mathbf{g}_{\mathrm{t}} \cdot \mathbf{g}_{\mathrm{t}}, & \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| \leq -\mu \hat{p}_{\mathrm{n}}, \quad \textit{stick} \\ -\dfrac{1}{2\varepsilon_{\mathrm{t}}} (\boldsymbol{\lambda}_{\mathrm{t}} \cdot \boldsymbol{\lambda}_{\mathrm{t}} + 2\mu \hat{p}_{\mathrm{n}} \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| + \mu^2 \hat{p}_{\mathrm{n}}^2), & \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| > -\mu \hat{p}_{\mathrm{n}}, \quad \textit{slip} \\ -\dfrac{1}{2\varepsilon_{\mathrm{t}}} \boldsymbol{\lambda}_{\mathrm{t}} \cdot \boldsymbol{\lambda}_{\mathrm{t}}, & \hat{p}_{\mathrm{n}} > 0, \quad \textit{non-contact} \end{cases}, \tag{27}$$

where $\hat{p}_{\mathrm{n}}$ is a regularized contact pressure at solution:

$$\hat{p}_{\mathrm{n}} = p_{\mathrm{n}} + \varepsilon_{\mathrm{n}} g_{\mathrm{n}}. \tag{28}$$

The augmented Lagrangian method converts the constrained minimization problem into a fully unconstrained problem, contrary to the Lagrange multiplier method which requires fulfilment of the constraint related to the Lagrange multiplier $\lambda_{\mathrm{n}} \leq 0$ [51]. Thus, the tangential functional l_{t} is extended to the non-contact domain $\hat{p}_{\mathrm{n}} > 0$, resulting in the prolongation of the Coulomb's cone for positive values of the normal contact pressure, as represented in Fig. 10b by a dashed red line. The augmented Lagrange multiplier used in (27) is defined as:

$$\hat{\boldsymbol{\lambda}}_{\mathrm{t}} = \boldsymbol{\lambda}_{\mathrm{t}} + \varepsilon_{\mathrm{t}} \mathbf{g}_{\mathrm{t}}. \tag{29}$$

The integration of (25) and (27) over the master surface leads to the following contribution of the contact conditions to the energy of the system:

$$\delta \mathcal{L}^{\mathrm{a}}(\mathbf{u}, \lambda_{\mathrm{n}}, \boldsymbol{\lambda}_{\mathrm{t}}, \hat{p}_{\mathrm{n}}) = \int_{\gamma_{\mathrm{c}}^1} \frac{\partial l_{\mathrm{n}}}{\partial g_{\mathrm{n}}} \delta g_{\mathrm{n}} + \frac{\partial l_{\mathrm{n}}}{\partial \lambda_{\mathrm{n}}} \delta \lambda_{\mathrm{n}} + \frac{\partial l_{\mathrm{t}}}{\partial \mathbf{g}_{\mathrm{t}}} \cdot \delta \mathbf{g}_{\mathrm{t}} + \frac{\partial l_{\mathrm{t}}}{\partial \boldsymbol{\lambda}_{\mathrm{t}}} \cdot \delta \boldsymbol{\lambda}_{\mathrm{t}} d\gamma_{\mathrm{c}}^1, \tag{30}$$

where the contact pressure $\hat{p}_{\mathrm{n}}$ is not subjected to the variation since it is assumed to be the known contact pressure at the solution.

In order to evaluate the derivatives contained in (30) it is useful to divide the potential contact surface into three non-intersecting zones, resulting from the three possible contact statuses (stick, slip and non-contact):

$$\frac{\partial l_{\mathrm{n}}(g_{\mathrm{n}}, \lambda_{\mathrm{n}})}{\partial g_{\mathrm{n}}} = \begin{cases} \hat{\lambda}_{\mathrm{n}}, & \hat{\lambda}_{\mathrm{n}} \leq 0, \quad \textit{contact} \\ 0, & \hat{\lambda}_{\mathrm{n}} > 0, \quad \textit{non-contact} \end{cases}, \tag{31}$$

$$\frac{\partial l_{\mathrm{n}}(g_{\mathrm{n}}, \lambda_{\mathrm{n}})}{\partial \lambda_{\mathrm{n}}} = \begin{cases} g_{\mathrm{n}}, & \hat{\lambda}_{\mathrm{n}} \leq 0, \quad \textit{contact} \\ -\dfrac{1}{\varepsilon_{\mathrm{n}}} \lambda_{\mathrm{n}}, & \hat{\lambda}_{\mathrm{n}} > 0, \quad \textit{non-contact} \end{cases}, \tag{32}$$

$$\frac{\partial l_{\mathrm{t}}(\mathbf{g}_{\mathrm{t}}, \boldsymbol{\lambda}_{\mathrm{t}})}{\partial \mathbf{g}_{\mathrm{t}}} = \begin{cases} \hat{\boldsymbol{\lambda}}_{\mathrm{t}}, & \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| \leq -\mu \hat{p}_{\mathrm{n}}, \quad \textit{stick} \\ -\mu \hat{p}_{\mathrm{n}} \dfrac{\hat{\boldsymbol{\lambda}}_{\mathrm{t}}}{\left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\|}, & \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| > -\mu \hat{p}_{\mathrm{n}}, \quad \textit{slip} \\ 0, & \hat{p}_{\mathrm{n}} > 0, \quad \textit{non-contact} \end{cases}, \tag{33}$$

$$\frac{\partial l_{\mathrm{t}}(\mathbf{g}_{\mathrm{t}}, \boldsymbol{\lambda}_{\mathrm{t}})}{\partial \boldsymbol{\lambda}_{\mathrm{t}}} = \begin{cases} \mathbf{g}_{\mathrm{t}}, & \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| \leq -\mu \hat{p}_{\mathrm{n}}, \quad \textit{stick} \\ -\dfrac{1}{\varepsilon_{\mathrm{t}}} \left(\boldsymbol{\lambda}_{\mathrm{t}} + \mu \hat{p}_{\mathrm{n}} \dfrac{\hat{\boldsymbol{\lambda}}_{\mathrm{t}}}{\left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\|} \right), & \left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\| > -\mu \hat{p}_{\mathrm{n}}, \quad \textit{slip} \\ -\dfrac{1}{\varepsilon_{\mathrm{t}}} \boldsymbol{\lambda}_{\mathrm{t}}, & \hat{p}_{\mathrm{n}} > 0, \quad \textit{non-contact} \end{cases}. \tag{34}$$

Considering the derivatives (31)–(34), the contribution of the contact conditions to the virtual work balance, using the augmented Lagrangian method, takes the following form:

$$\begin{aligned} \delta \mathcal{L}^{\mathrm{a}}(\mathbf{u}, \lambda_{\mathrm{n}}, \boldsymbol{\lambda}_{\mathrm{t}}, \hat{p}_{\mathrm{n}}) = & \int_{\bar{\gamma}_{\mathrm{c}}^1} \hat{\lambda}_{\mathrm{n}} \delta g_{\mathrm{n}} + g_{\mathrm{n}} \delta \lambda_{\mathrm{n}} + \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \cdot \delta \mathbf{g}_{\mathrm{t}} + \mathbf{g}_{\mathrm{t}} \cdot \delta \boldsymbol{\lambda}_{\mathrm{t}} d\gamma_{\mathrm{c}}^1 \\ & + \int_{\bar{\gamma}_{\mathrm{c}}^{1*}} \hat{\lambda}_{\mathrm{n}} \delta g_{\mathrm{n}} + g_{\mathrm{n}} \delta \lambda_{\mathrm{n}} - \mu \hat{p}_{\mathrm{n}} \frac{\hat{\boldsymbol{\lambda}}_{\mathrm{t}}}{\left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\|} \\ & \cdot \delta \mathbf{g}_{\mathrm{t}} - \frac{1}{\varepsilon_{\mathrm{t}}} \left(\boldsymbol{\lambda}_{\mathrm{t}} + \mu \hat{p}_{\mathrm{n}} \frac{\hat{\boldsymbol{\lambda}}_{\mathrm{t}}}{\left\| \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \right\|} \right) \cdot \delta \boldsymbol{\lambda}_{\mathrm{t}} d\gamma_{\mathrm{c}}^1 \\ & + \int_{\gamma_{\mathrm{c}}^1 \setminus \bar{\gamma}_{\mathrm{c}}^1} -\frac{1}{\varepsilon_{\mathrm{n}}} \lambda_{\mathrm{n}} \delta \lambda_{\mathrm{n}} - \frac{1}{\varepsilon_{\mathrm{t}}} \boldsymbol{\lambda}_{\mathrm{t}} \cdot \delta \boldsymbol{\lambda}_{\mathrm{t}} d\gamma_{\mathrm{c}}^1 \end{aligned} \tag{35}$$

where $\bar{\gamma}_c^{1\cdot}$ denoted the stick zone, $\bar{\gamma}_c^{1*}$ denoted the slip zone and $\bar{\gamma}_c^1$ represents the active contact zone and, consequently, $\gamma_c^1 \backslash \bar{\gamma}_c^1$ is the non-contact zone.

3.2 Incremental Solution Procedure

The application of the *quasi*-static formulation is appropriate when the inertial forces are negligible in comparison with the internal and applied forces. The numerical solution of contact problems in nonlinear solid mechanics involves three sources of nonlinearities: geometric (large deformation of bodies), material (constitutive behaviour) and the frictional contact phenomena governed by nonlinear and non-smooth laws. Many materials widely used in mechanical applications present a nonlinear behaviour and path-dependent (behaviour dependent of the deformation history). A large class of nonlinear materials can be described by the assumption of elastoplastic behaviour, such as the metallic ones. In that case, the mechanical model need to take into account large elastoplastic strains and rotations [61, 62]. Moreover, when the friction phenomenon is taken into account, the solution becomes path-dependent due to the dissipative effect of the friction process [19].

In order to capture all time or history dependent effects (e.g. plastic material behaviour and friction), the time interval of interest is subdivided into a set of subintervals (temporal discretization). Therefore, an incremental solution procedure is required to solve the *quasi*-static nonlinear problem involving large sliding between the curved contacting surfaces. The motion of the bodies coming into contact is described by a fully implicit time integration method, which is unconditionally stable, allowing for relatively large time step sizes in comparison with explicit schemes [63]. Nevertheless, the implicit time integration schemes require the solution of a nonlinear system of equations at each time step. Typically, the Newton–Raphson method is employed, which involves the evaluation of the tangent matrix for the global system. Although this method exhibits quadratic convergence near the solution point, some convergence problems can arise, particularly due to the frictional contact constraints.

3.2.1 Generalized Newton Method

The finite element spatial discretization transforms the original continuum boundary value problem, expressed by the principle of virtual work (17), into a nonlinear system of algebraic equations. Since these three sources of nonlinearities are embedded in the discrete system of equations, an iterative scheme has to be employed. The Newton's method (also called Newton–Raphson) is a very efficient algorithm to solve the nonlinear problems occurring in the finite element method, since it exhibits quadratic convergence near the solution. The main idea of the Newton–Raphson method is to replace the nonlinear problem by a series of linear problems, which are directly solvable by standard methods of linear algebra. The solution of a linear system of equations in each iteration yields a high computational cost, which depends quadratically on the number of degrees of freedom [63].

The Newton–Raphson method provides quadratic rate of convergence near the solution if the conditions of convexity and smoothness are fulfilled. However, the virtual work balance resulting from the augmented Lagrangian method is only piecewise smooth, presenting first derivative discontinuity across the gap–contact status line. Due to this lack of differentiability, most studies based on the augmented Lagrangian in contact mechanics apply the Uzawa's algorithm [42], leading to stable procedures with poor convergence rates due to the alternate treatment of the primal (displacements) and the dual (Lagrange multipliers representing the contact forces) variables. On the other hand, the simultaneous treatment of both variables by the Newton–Raphson method converges much faster. The extension to non-differentiable problems arising from contact mechanics was investigated by Alart and Curnier [38, 58] and Heegaard and Curnier [39], developing the generalized Newton method (GNM).

The main idea of the generalized Newton method is to split into two parts the system of nonlinear equations, i.e. a differentiable structural part $\mathcal{F}^s$ and a non-differentiable contact part $\mathcal{F}^c$, such that:

$$\mathcal{F}(\mathbf{u},\boldsymbol{\lambda}) = \mathcal{F}^s(\mathbf{u}) + \mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}) = \mathbf{0}, \tag{36}$$

where $\mathcal{F}^s$ represents the virtual work of the two body system in absence of contact and $\mathcal{F}^c$ denotes the virtual work due to the frictional contact forces. Then, the generalized Newton method is stated as:

$$\left[\begin{array}{c} \nabla_{\mathbf{u}}\mathcal{F}^s(\mathbf{u}) + \nabla_{\mathbf{u}}\mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}) \\ \nabla_{\boldsymbol{\lambda}}\mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}) \end{array}\right]\Bigg|_{\mathbf{u}_i,\boldsymbol{\lambda}_i} \left\{\begin{array}{c} \Delta\mathbf{u}_i \\ \Delta\boldsymbol{\lambda}_i \end{array}\right\} = \\ = -\{\mathcal{F}^s(\mathbf{u}_i) + \mathcal{F}^c(\mathbf{u}_i,\boldsymbol{\lambda}_i)\}, \tag{37}$$

where i is the iteration index and $\nabla_{\mathbf{u}}\mathcal{F}^s$ denotes the tangent matrix of the contacting bodies. The sub-gradients $\nabla_{\mathbf{u}}\mathcal{F}^c$ and $\nabla_{\boldsymbol{\lambda}}\mathcal{F}^c$ are components of the generalized Jacobians for primal and dual variables:

$$\begin{array}{l} \nabla_{\mathbf{u}}\mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}) \in \partial_{\mathbf{u}}\mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}), \\ \nabla_{\boldsymbol{\lambda}}\mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}) \in \partial_{\boldsymbol{\lambda}}\mathcal{F}^c(\mathbf{u},\boldsymbol{\lambda}). \end{array} \tag{38}$$

In practice, the generalized Jacobian is evaluated in each node as a classical Jacobian at the current iteration. Thus, a different Jacobian matrix is derived for each contact status. In the general case of frictional contact, the resulting tangent matrices contained in (37) are non-symmetric, non-positive definite and present zero values on the diagonal [64]. The convergence properties of the augmented

Lagrangian method for frictional contact problem were investigated in detail by Alart [58]. The results shown a good convergence for frictionless contact, both in case of small and large slip. In case of frictional problems, some conditions are imposed in the selection of the penalty parameter value to avoid infinite cycling, ensuring convergence of the GNM.

The solution technique employed to solve the linear systems of Eqs. (37) in each iteration step has a significant effect on the efficiency of the nonlinear finite element solution. In case of large finite element models, very large systems of equation arise in the iterative solution. Besides, the finite element discretization dictates a sparse matrix structure (most of the entries are zero) with an evident band structure, which results from the locality of the shape functions. There are many different methods to solve sparse linear systems of equations, which can be classified as direct and iterative methods. The direct methods are based on an LU decomposition of the underlying system matrix, leading to two triangular matrices easily invertible. The main advantage of direct solvers is that the solution is always achieved, allowing to solve ill-conditioned and non-positive definite systems of equations, as long as round-off does not affect the solution [65]. Modern sparse solvers have less memory requirements, allowing to solve problems of several million of unknowns [66, 67].

On the other hand, iterative solvers are advantageous when large systems of equations have to be solved, since the memory requirement and the total number of operations is less, when compared to direct solvers. However, preconditioning techniques are a prerequisite for the success of the iterative solvers, since its rate of convergence is largely influenced by the condition number. The main idea of the preconditioning is to convert the original linear system into an equivalent system with lower condition number. The preconditioned iterative solution of sparse linear systems is usually achieved through so-called Krylov subspace methods, where the conjugate gradient method and the generalized minimal residual method are the most popular [65]. The Jacobi and the Gauss–Seidel preconditioners are the simplest ones, which are derived from the iterative methods with the same designation. A broad class of preconditioners is based on incomplete factorizations of the coefficient matrix, leading to the so-called incomplete LU factorization techniques, which are the most popular and efficient to solve large sparse linear systems [68]. However, the selection and evaluation of a good preconditioner for an iterative method can be computationally more expensive than using a direct method [67].

Although direct methods usually need more operations and larger memory requirements than iterative ones, the Direct Sparse Solver (DSS) from Intel® Math Kernel Library (Intel® MKL) is adopted in the present study for solving the large sparse linear systems [69]. This modern library is highly optimized for scientific and engineering applications that require solving large problems. Moreover, it is optimized for the latest Intel processors, including processors with multiple cores, which are currently standard in personal computers. The typical invoking sequence of the DSS interface routines is depicted in Fig. 13, which is divided into six phases. The names of the routines corresponding to each phase are presented in the box.

The first phase (dss_create) initializes the solver and creates the basic data structures necessary for the solver. The purpose of the second phase (dss_define_structure) is to define the locations of the nonzero entries of the matrix, i.e. the sparse matrix pattern. The general nonzero structure of the matrix: symmetrically structured, symmetric and non-symmetric is defined. Typically, the matrices arising from the finite element method are symmetrically structured, i.e. the pattern of nonzero entries is symmetric. The nonzero entries of the sparse matrix are stored in a linear array in order to improve the computational efficiency. The location of the nonzero entries is performed by means of two arrays, one provides the column number containing the entry, while the other gives the location of the first nonzero entry within each matrix row. The next phase of the DSS interface (dss_reorder) comprises the permutation of rows and columns in order to minimize the fill-in during the factorization phase. The factorization phase (dss_factor_real) computes the LU factorization of the sparse matrix. In the following phase (dss_solve_real), the solution vector is computed based on the factorization computed in the previous phase. The last phase (dss_delete) deletes all data structures created during the solving process. In some finite element problems (e.g. contact between deformable and rigid bodies) the nonzero pattern of the matrix is unchanged during the incremental solution procedure. Thus, the phase related with the LU factorization (dss_factor_real) is repeated for each matrix, as represented by the dashed line in Fig. 13. For more details about each routine see the Intel MKL reference manual [70].

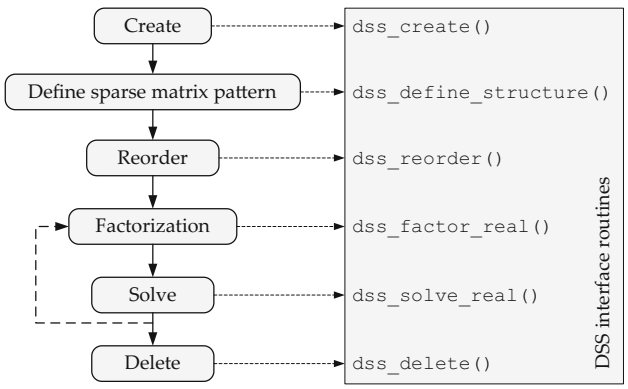

Fig. 13 Typical sequence for invoking DSS interface routines

4 Contact Surface Discretization

The finite element method approximates the real geometry by a finite element mesh, which is composed by nodes, connected to form elements, which are coupled together into a structure. In general, the spatial discretization of the contacting bodies using a finite element mesh originates differences between the continuous and the discretized boundary geometry, as highlighted in Fig. 3. In fact, the real surface of the body is approximated by the finite element mesh, producing piecewise smooth contacting surfaces, which introduce some numerical difficulties in the contact treatment.

The finite element analysis of contact problems can be classified in two distinct categories: (i) contact between discretized deformable bodies and (ii) contact between one discretized deformable body and a rigid surface. In fact, several engineering problems can be included in the second category, such as: metal forming processes, tyre on road, rubber seals and indentation tests [3, 19]. In this type of problems it is possible to consider that the contact occurs between two bodies with significantly different stiffness, i.e. the stiffer body is approximated by a rigid surface while the other body is modelled as deformable. This simplification is very useful since there are no calculations over the rigid body (no additional degrees of freedom involved), leading to a contact problem computationally more efficient. The contact forces arise only on the slave surface due to the violation of geometrical constraints and friction law conditions.

4.1 Rigid Bodies

Concerning the numerical simulation of the contact of a deformable body with a rigid obstacle, the geometrical penetration and the friction law are evaluated only at the nodes of the discretized body (slave). Since the discretization of the master surface is not required, this contact discretization technique is usually called the Node-to-Surface, where the slave nodes are prohibited to penetrate the master surface. Typically, this approach is used in conjunction with lower order finite elements due to the non-uniform distribution of nodal forces associated with higher order shape functions in the presence of a uniform pressure (1/4/1 distribution for a line with quadratic shape function) [71]. Several approaches have been developed to handle with the contact surface description of rigid bodies, which can be divided in three groups: (i) analytical functions; (ii) finite element meshes (Fig. 14a) and (iii) parametric patches (Fig. 14b).

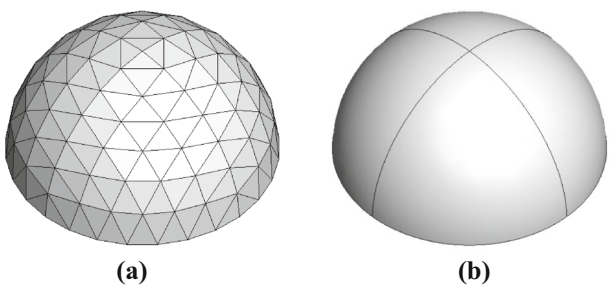

Fig. 14 Description of a rigid contact surface (half spherical shell) using: **a** finite element mesh; **b** parametric patches

The first description scheme is restricted to simple geometries, where the contact surface is composed by an assembly of simple analytical shapes (planes, cylinders, spheres and tori) [72]. This method is usually adopted in the axisymmetric contact problems, due to the exact description of the contact geometry and associated good convergence, dictated by the smooth surface. On the other hand, for arbitrarily-shaped contact surfaces, the finite element mesh scheme is commonly employed due to its ability and simplicity. However, this approximation can lead to large errors in the geometry, requiring an extremely fine mesh in curved regions to attain a sufficiently accurate surface description [73]. Typically, in order to use the minimum number of finite elements for a proper description, small elements are used in curved areas and large elements are applied in flat regions. Nevertheless, the discontinuities in the contact surface normal field (faceted surfaces) introduced by the finite element discretization (see Fig. 14a) leads to artificial oscillations in the contact force when sliding is significant, generating convergence problems in the iterative procedure [74].

These problems related with the discontinuity of the surface normal vector field can be avoided using parametric patches, which can be obtained directly from the Computer-Aided Design (CAD) model. This surface description method allows creating complex geometries ensuring a high level of continuity with a small number of patches, as shown in Fig. 14b, leading to a more robust behaviour of the iterative solution algorithms [20]. Various parameterizations, originally developed for CAD models, can be used to define 3D rigid contact surfaces, such as Bézier patches [21], Hermite patches [22] and trimmed NURBS patches [75, 76]. However, they are characterized by high order interpolation and complex algorithms, which leads to high computational cost in the contact search algorithms [77]. Furthermore, CAD models are known to be plagued by geometrical or topological errors and inconsistencies (gaps/overlaps between abutting surfaces), which result from the lack of a robust solution for the surface intersection problem [78]. Therefore, before using the surface model within a finite element environment it is always necessary to perform some laborious manual intervention such as geometry repair, clean-up and

preparation [79]. Since this manual treatment is incompatible with complex models involving hundreds of patches, the faceted finite element mesh is still the contact surface description method most used in commercial finite element codes.

The parametric patches can also be employed to smooth the discretized rigid contact surfaces, improving both the accuracy of the surface representation and the convergence behaviour [80]. Nevertheless, this strategy is more frequently used to smooth the master surface in contact problems between deformable bodies, where the adoption of a CAD model to describe deformable bodies is impossible. Since the smooth parametric patches are created based on the information from several adjacent finite elements, its application is frequently restricted to structured meshes composed by regular quadrilateral elements (surrounded by eight neighbouring elements). In fact, the extension of the surface smoothing method to arbitrary surface meshes involve several difficulties [29, 81].

4.2 Deformable Bodies

The finite element analysis of frictional contact problem between deformable bodies requires the transmission of the contact forces through the contact interface, which increases the complexity of the problem. The contact forces arising in the slave surface due to the violation of geometrical constraints and friction law conditions are transferred to the master surface through contact elements. Contact elements can be geometrically interpreted as "bridge elements" between separated surfaces, which share components (nodes or segments) of both surfaces. Typically, the discretization of the contact interface is classified into three main types: (i) Node-to-Node (NTN); (ii) Node-to-Segment (NTS) and (iii) Segment-to-Segment (STS) [51].

4.2.1 Node-to-Node Discretization

The simplest discretization technique adopted in contact problems is the so-called Node-to-Node, which incorporates the contact constraints directly on the nodal pairs [82], as shown in Fig. 15. Therefore, this contact discretization can be established only in case of small deformations, not allowing the finite sliding. Furthermore, restrictions on mesh generation are introduced due to the requirement of guaranteeing conforming meshes along the contacting surfaces [83], i.e. each node of one contacting surface has a single corresponding node on the opposite surface (see Fig. 15). Thus, the contact detection procedure simply establish contact pairs of nodes (one slave and one master), which do not change during the solution steps. On the other hand, due to the imposed conforming contact interface, the NTN discretization passes the contact patch test proposed by Taylor and Papadopoulos [18] (uniform pressure transferred correctly through the contact interface).

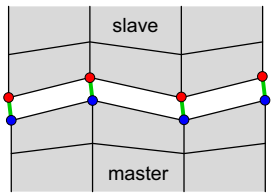

Fig. 15 Schematic illustration of the Node-to-Node contact discretization, including the associated pairs of nodes

4.2.2 Node-to-Segment Discretization

For the general case of contact including large deformation and large sliding, the so-called Node-to-Segment discretization technique is quite popular and widely used [9, 11]. This contact discretization is valid for non-conforming meshes (nodes at the contact interface located at dissimilar positions), which can arise from the sliding of the contact interface or when the finite element meshes are generated independently in each body. The contact pairs are composed by a node of the slave surface and a corresponding segment of the master surface, as shown in Fig. 16, where the contact segments are lines (in 2D problems) joining adjacent master nodes [84]. However, in case of non-conforming meshes the NTS discretization fails the contact patch test [18]. Nevertheless, it is often implemented in commercial finite element codes due to its simplicity and flexibility [17]. The first step of the NTS discretization comprises the selection of one contacting surface as slave and the other one as master, leading to an asymmetry in the contact problem treatment because the contact surfaces are treated differently [85]. The impenetrability conditions are enforced only in a finite number of points on the slave surface, preventing the slave nodes from penetrating on the contact master surface [11]. However, the master nodes are allowed to penetrate into the slave surface.

Each contact element (not structural) is composed by a slave node and the closest segment (element edge/facet) on the master surface (see Fig. 16), which is selected through

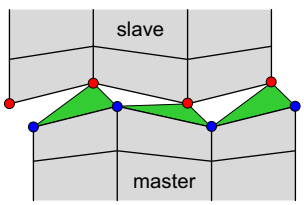

Fig. 16 Schematic illustration of the Node-to-Segment contact discretization with three contact elements spanned on three slave nodes

the projection of the slave node onto the master surface. Nevertheless, particularly when the NTS approach is applied with low order finite elements [71], the identification of the master segment is either ambiguous or impossible, which may result in slow convergence or even in divergence of the numerical solution. Some strategies have been specially developed to deal with such problems in 2D frictionless contact problems [17].

The proper selection of the master and slave surfaces is fundamental to the success of the NTS contact discretization. After identifying the pair of surfaces that will interact, one surface is assigned as master and the other one as slave. The selection of the master and slave surfaces should be carried out taking into account that the slave nodes cannot penetrate into the master surfaces, but the master nodes are free to penetrate the slave surface. Hence, the main guidelines for choosing the master and slave surfaces are listed below:

- The contact surface that presents the finer mesh should be the slave surface, while the surface with the coarser mesh is the master surface;
- When the stiffness between the contacting bodies is different, the contact surface of the stiffer body should be the master surface and the other should be assigned as slave surface;
- When the contact occurs between a deformable body and a rigid obstacle, the surfaces of the rigid obstacle must be specified as master surfaces;
- In the case of contact between a convex surface with a flat or concave surface, the master surface should be the flat/concave surface;
- If one body slides over another with a contact surface considerably larger, the larger surface should be the master surface in order to minimize the creation/deletion of the contact elements.

In order to highlight the importance of the master/slave surface selection, the contact patch test example is presented. Two elastic cubes with identical geometry (each edge with 10 mm) and the same material properties ($E = 100$ MPa and $\nu = 0.3$) are pressed against each other under frictionless conditions. Each cube is discretized independently with 8-node hexahedral finite elements, as shown in Fig. 17a. The bottom surface of the lower cube is constrained against vertical displacements and the four lateral surfaces of the cubes are constrained against displacements in its normal direction. The uniform pressure is imposed by applying a vertical displacement of 1 mm on the top surface of the upper cube.

The distribution of the vertical stress component in the cubes is depicted in Fig. 17b, employing the NTS contact discretization with the upper cube (finer mesh) defined as master. The obtained results do not satisfy the contact patch test due to the non-conforming meshes at the contact interface. Since the finer mesh is assigned as master (incorrect choice), the penetration of some master nodes into the slave cube can be considered excessive (see Fig. 17b), which leads to high deviations in the predicted contact stress, i.e. inaccurate transmission of constant normal stresses between the two contacting surfaces. On the other hand, by exchanging the master and slave surfaces definition, the resulting distribution of vertical stress is shown in Fig. 17c. The noise in the contact stress is considerably reduced when the coarse mesh is assigned as master surface, leading to a contact surface approximately flat after loading. A modification of the NTS discretization has been proposed by Zavarise and De Lorenzis [86], which passes the contact patch test in 2D frictionless contact problems using the penalty method to enforce the contact constrains. The basic idea of this algorithm is to create two virtual slave nodes located at the quarter points of each slave

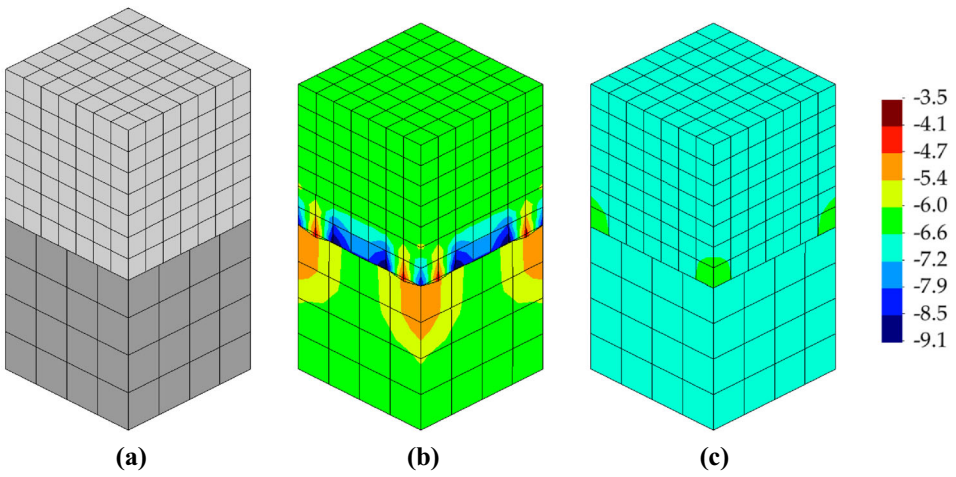

Fig. 17 Distribution of the vertical stress component for different choices of the master and slave surfaces in the NTS contact discretization: **a** finite element mesh of contacting cubes; **b** upper cube defined as master and lower as slave; **c** upper cube defined as slave and lower as master

segment, improving the contact contribution to the stiffness matrix and to the internal force vector.

The master–slave formulation used in the NTS discretization is inherently asymmetric. Therefore, the two-pass contact (also called symmetric contact) approach was developed to try to overcome this problem. The main idea of the two-pass approach is the definition of the each contact surface as master and slave simultaneously, performing a double definition of the contact pair, exchanging the master and slave surfaces. This approach precludes penetration of the slave nodes into the master segments (first pass), while the master nodes are restricted from penetrating on the slave segments, when the master and slave surfaces are exchanged in the second pass. Therefore, this strategy allows eliminating the geometric asymmetry by reversing the role of master and slave surfaces and repeating the same process performed in the single-pass algorithm. Since the number of contact elements created is higher, the two-pass NTS algorithm is computationally more expensive. The NTS discretization associated with the two-pass contact algorithm passes the contact patch test in 2D and in some 3D mesh configurations (with sufficient symmetry) for low order finite elements [18]. Nevertheless, this discretization technique based on the two-pass contact algorithm has the recognised deficiency of locking, due to the overconstrained system of equations [87]. Indeed, if any two nodes on both contact surfaces have identical locations, the corresponding contact constraint is created in duplicate during the two-pass algorithm, which results in a rank deficient matrix (linearly dependent rows and columns) [88].

The comparison between the single-pass and the two-pass NTS contact algorithm is presented in Fig. 18 for the discretization employed previously. For the single-pass NTS contact algorithm, the cube with the coarse mesh is defined as master, while the selection of the master and slave surfaces in the two-pass algorithm is arbitrary due to its exchange in the second pass. The distribution of the vertical stress component obtained with the single-pass algorithm is presented in Fig. 18a, which are the same results shown in Fig. 17c using a different scale. In fact, the single-pass algorithm does not satisfies the contact patch test. On the other hand, the two-pass contact algorithm exactly transmits the constant normal stresses between the contacting surfaces, as shown in Fig. 18b. The patch test is passed since the expected contact surface is perfectly horizontal (flat).

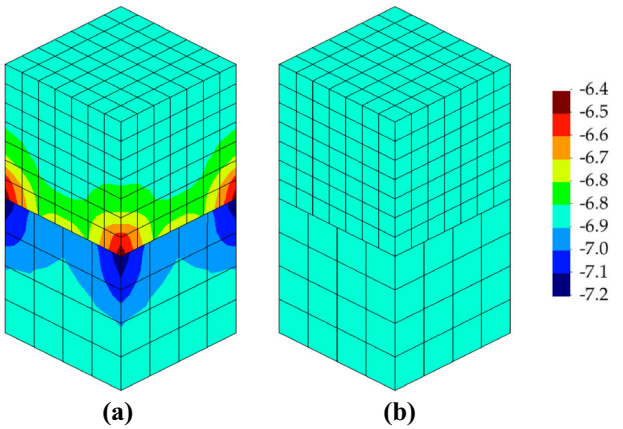

Fig. 18 Distribution of the vertical stress component in the contact patch test using: **a** the single-pass NTS algorithm with the lower cube defined as master; **b** the two-pass NTS algorithm

4.2.3 Segment-to-Segment Discretization

Alternative methods based in the Segment-to-Segment contact discretization have been first proposed for 2D problems [89]. Recently, this discretization technique has been successfully coupled with the mortar method, which was originally developed in the context of domain decomposition techniques for nonconforming meshes [90, 91]. The extension of the mortar method to contact problems between elastic bodies was formulated by Belgacem et al. [92] and its application for 2D frictional contact problems, subjected to small deformations, was firstly established by McDevitt and Laursen [93]. The extension of the STS contact discretization using the mortar method to 3D contact problems involving large deformation and large sliding has been developed during the last few years [87, 94–96]. Nowadays, the application of mortar methods in computational contact mechanics is still one of the most active research topics [97, 98].

The principal property of the mortar method is the incorporation of the contact constrains in a weak form (see Fig. 19) instead of point-wise constraints, allowing an accurate exchange of the contact forces between interface surfaces. Therefore, this technique is stable and passes the contact patch test for non-conforming meshes [87]. Moreover, using the dual spaces for the Lagrange multipliers it is possible to eliminate the Lagrange multipliers from the set of linear equations, thus the unknowns are only the nodal displacements [98–100]. Although the mortar-based

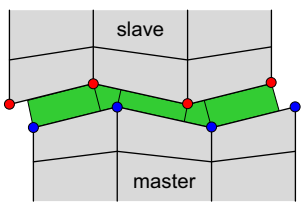

Fig. 19 Schematic illustration of the Segment-to-Segment contact discretization

STS discretization is more robust and accurate than the classic NTS counterpart, its implementation presents a great challenge. Indeed, for the general 3D case, the mortar interface coupling and numerical integration for arbitrary surface meshes is a complex task due to the surface segmentation.

5 Surface Smoothing Procedures

The finite element discretization of the contact surfaces may cause two types of difficulties in the solution of contact problems. The first one is associated with the local searching procedure based on the normal projection method, which can provide multiple solutions and blind spots in the projection domain due to the C^0 continuity of the surface. The second difficulty is created by the sudden changes in the surface normal field between adjacent master elements (see Fig. 3), producing a contact surface only piecewise differentiable. Therefore, combining the Node-to-Segment contact discretization technique with faceted master surface description leads to non-physical jumps in the contact forces, when a slave node slides over two adjacent master elements [101]. Moreover, these jumps induced by spatial discretization can cause convergence problems and possible divergence of the numerical solution [15]. In order to overcome the above mentioned drawbacks, various surface smoothing procedures have been developed, trying to achieve a smooth description of the master surface.

The idea behind the contact smoothing procedures is to define an accurate and continuous contact master surface using higher order interpolations, while the bulk is discretized with the classical linear finite elements. This procedure ensures a continuous projection of the slave nodes on the master surface, improving the accuracy obtained in the evaluation of the normal gap function. When the contact surface is itself deformable, the smoothing method needs to be applied in each iteration in order to evaluate accurately the kinematic contact variables (Sect. 2.1). The surface smoothing methods adopted in 2D contact problems are typically based in Hermite [24], Bézier [25], Spline [26, 102] or NURBS [28, 31] interpolation of the mesh, attaining at least C^1 continuity in the resulting smoothed master surface. In fact, the technique proposed by Stadler et al. [28] allows obtaining an arbitrary level of continuity in the master surface representation.

However, the straightforward extension of such interpolation methods to 3D contact problems is limited to structured quadrilateral finite element meshes [23, 24]. In fact, the development of smoothing techniques for deformable contact surfaces discretized with an arbitrary mesh topology is significantly more complex, presenting an actual challenge and research topic in computational contact mechanics [3]. Therefore, only two different approaches are currently available to address this issue. One is the surface smoothing method proposed by Puso and Laursen [29], which uses Gregory patches to interpolate both structured and unstructured meshes of quadrilateral elements. However, this method only ensures G^1 continuity at the patch boundaries of the surface representation, i.e. the direction of the tangent vector is continuous (not its magnitude). On the other hand, this smoothed contact algorithm is roughly twelve times computationally more expensive than the non-smoothed implementation. The other approach, proposed by Krstulović-Opara et al. [27], employs quartic triangular Bézier patches in the surface smoothing of tetrahedral finite element meshes, using the nodes and the centroid of the exterior triangular elements. This method allows achieving *quasi*-C^1 continuity surfaces, producing a smooth transition between edges of triangular elements, except at the nodes. However, since this smoothed contact surface passes through the centroids of the finite elements, which are not necessarily points lying on the original surface, the obtained geometrical accuracy can be unsatisfactory.

Alternative techniques to the classic surface smoothing method have been developed, such as the subdivision scheme presented by Stadler and Holzapfel [103], which can be applied for both quadrilateral and triangular meshes of arbitrary topology. However, this technique requires special treatment of the nodes where the mesh is unstructured. Another alternative is the smoothing procedure based on a meshfree technique, as proposed by Chamoret et al. [80], which allows dealing with hybrid surface meshes (tetrahedral and hexahedral elements). This approach generates a smooth contact surface using a least-squares approximation. A similar approach was suggested by Belytschko et al. [81], where the smoothing is performed implicitly by constructing smooth signed gap functions for the bodies. However, the contact surface provided by these methods does not pass through the nodes of the master surface discretization, which introduces some imprecisions in the geometry of the contacting bodies [104].

5.1 Interpolation Methods

The surface interpolations through Bézier and NURBS parameterizations are usually adopted in CAD systems, allowing high level of continuity using a small number of patches. These interpolation methods have been used to describe the geometry of the contact surfaces, leading to high accuracy in the contact integral evaluation. The simpler description of the Bézier in comparison with the NURBS may significantly reduce the computational cost [105]. On the other hand, NURBS can reproduce exactly several analytical geometries while other spline functions only provide approximations [106]. The isogeometric

analysis (IGA) was first introduced by Hughes et al. [107] with the objective to tightly integrate CAD with finite element analysis, where the NURBS are used as basis functions for finite elements. Thus, several computational contact formulations within the IGA framework have been developed in the past few years [108].

The tensor product method is widely used in geometric modelling to create a bidirectional curve scheme. Thus, a Bézier patch of deegre n in the u direction and degree m in the v direction can be defined as the tensorial product of two Bézier curves, given by:

$$\mathbf{S}(u,v) = \sum_{i=0}^{n}\sum_{j=0}^{m} B_{i,n}(u)B_{j,m}(v)\mathbf{P}_{i,j}, \quad 0 \le u,v \le 1, \tag{39}$$

where $\mathbf{P}_{i,j}$ are the position vector of the vertices defining the polygonal control net, composed by $(n+1)(m+1)$ control points. The Bernstein basis functions $B_{i,n}$ and $B_{j,m}$ are defined for each direction through the local coordinates u and v. The Bernstein polynomial of degree n is defined by:

$$B_{i,n}(t) = \binom{n}{i} t^i (1-t)^{n-i}, \quad \text{for} \quad i = 0,\ 1,\ \ldots,\ n, \tag{40}$$

where the binomial coefficient is given explicitly by:

$$\binom{n}{i} = \begin{cases} \dfrac{n!}{i!(n-i)!} & \text{if} \quad 0 \le i \le n \\ 0 & \text{otherwise} \end{cases}, \tag{41}$$

where $n!$ denotes the factorial of n. The Bézier patches can also be described using the monomial form, also called power basis form. The main advantage of this form is that insures lower computational time for operations such as calculations of point coordinates or derivatives. However, it is numerically less stable, mainly for patches with high degree [105].

A NURBS curve is defined by a set of weighted control points and a knot vector. It is constructed from B-spline basis functions (see Fig. 20) using the weighted position vector of the control points as coefficients of the basis functions [109]. The shape of the curve is dictated mainly by the location of the control points. However, the curve can be pulled or pushed towards each control point through its associated weight. Mathematically, a NURBS curve of degree l is a piecewise rational function defined by a set of $h+1$ control points along with their weights, expressed by:

$$\mathbf{C}(t) = \frac{\sum_{i=0}^{h} N_{i,l}(t) w_i \mathbf{P}_i}{\sum_{i=0}^{h} N_{i,l}(t) w_i}, \quad 0 \le t \le 1, \tag{42}$$

where w_i are the so-called weights, $\mathbf{P}_i$ are the position vectors of the control points and $N_{i,l}$ are the normalized B-spline basis functions of degree l. Typically, the B-spline basis functions are defined recursively starting with zero order basis function ($l = 0$). The ith B-spline basis function of degree l is defined by:

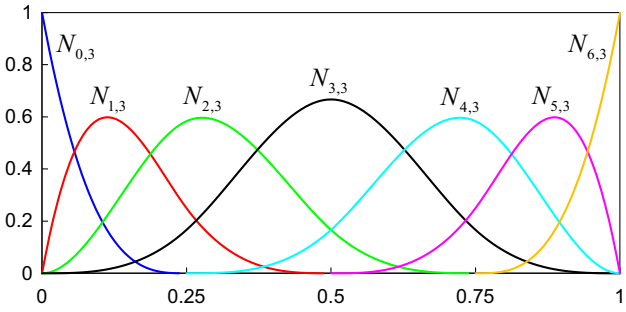

Fig. 20 Cubic B-spline basis functions for open and non-uniform knot vector

$$N_{i,0}(t) = \begin{cases} 1 & \text{if } t_i \le t < t_{i+1} \\ 0 & \text{otherwise} \end{cases},$$
$$N_{i,l}(t) = \frac{t - t_i}{t_{i+l} - t_i} N_{i,l-1}(t) + \frac{t_{i+l+1} - t}{t_{i+l+1} - t_{i+1}} N_{i+1,l-1}(t), \tag{43}$$

which is referred to as the Cox–de Boor recursion formula [110, 111]. Note that $N_{i,0}(t)$ is a step function, equal to zero everywhere except on the half-open interval $t \in [t_i, t_{i+1})$. Throughout the calculation of functions (43), ratios of the form 0/0 are defined as zero by convention.

The computation of the set of basis functions (43) requires the specification of a knot vector, which is a non-decreasing sequence of coordinates in the parametric space, defined in one dimension by:

$$\mathbf{T} = \left\{ \underbrace{0, \ldots, 0}_{l+1}, t_{l+1}, \ldots, t_{g-l-1}, \underbrace{1, \ldots, 1}_{l+1} \right\}, \tag{44}$$

where t_i are the so-called knots composing the knot vector of dimension $g+1$. The curve degree l, the number of knots $g+1$ and the number of control points $h+1$ are related through the relationship $g = h + l + 1$. The knot vector is referred as uniform when the knots are equally spaced in the parametric space, otherwise it is defined as non-uniform. Since consecutive knots can have the same value (repeated knots), a knot vector is assumed to be open if its first and last knots are repeated with multiplicity $l+1$, leading to a curve with start and end in a control point. The normalized B-spline basis functions of degree 3 for the open, non-uniform knot vector

$$\mathbf{T} = \{0,\ 0,\ 0,\ 0,\ 0.25,\ 0.5,\ 0.75,\ 1,\ 1,\ 1,\ 1\}$$

are shown in Fig. 20.

A NURBS surface is obtained as the tensor product of two NURBS curves. The detailed mathematical description can be found in the literature, e.g. [106]. Hence, a NURBS surface of degree p in the u direction and degree q in the v direction has the following form:

$$\mathbf{S}(u,v) = \frac{\sum_{i=0}^{n}\sum_{j=0}^{m} N_{i,p}(u)N_{j,q}(v)w_{i,j}\mathbf{P}_{i,j}}{\sum_{i=0}^{n}\sum_{j=0}^{m} N_{i,p}(u)N_{j,q}(v)w_{i,j}}, \quad 0 \le u,v \le 1, \tag{45}$$

where $\mathbf{P}_{i,j}$ are the position vectors of the control points that form a bidirectional control net (see Fig. 21a) and $w_{i,j}$ are the weights of the control points. The number of control points is defined as $n+1$ in the u direction and $m+1$ in the v direction. The $N_{i,p}$ and $N_{j,q}$ are the B-spline basis functions (43) of degree p and q, respectively, defined on the following knot vectors:

$$\begin{aligned} \mathbf{U} &= \left\{ \underbrace{0,\ldots,0}_{p+1}, u_{p+1},\ldots,u_{r-p-1}, \underbrace{1,\ldots,1}_{p+1} \right\}, \\ \mathbf{V} &= \left\{ \underbrace{0,\ldots,0}_{q+1}, v_{q+1},\ldots,v_{s-q-1}, \underbrace{1,\ldots,1}_{q+1} \right\}, \end{aligned} \tag{46}$$

where $r = n+p+1$ and $s = m+q+1$ express the size of each knot vector less one. Open knot vectors are commonly used in the definition of NURBS surfaces, i.e. the first and last knots are repeated with multiplicity $p+1$ and $q+1$, respectively, as represented in (46).

Figure 21 presents an example of a NURBS surface using equal weights for all control points. The control net is composed by 20 control points (red dots in Fig. 21a), which are distributed into a rectangular grid with 5 and 4 points in each direction. The surface passes at the four corner control points $\mathbf{P}_{0,0}$, $\mathbf{P}_{4,0}$, $\mathbf{P}_{0,3}$, $\mathbf{P}_{4,3}$ since the two knot vectors adopted in the surface definition are open.

5.2 Nagata Patch Interpolation

The Nagata patch interpolation was proposed by Nagata [112] for interpolating discretized surfaces in order to recover the original geometry with good accuracy. Its central idea is the quadratic interpolation, requiring only the position and normal vectors at the nodes of the surface mesh. Moreover, it can be applied to general finite element meshes with arbitrary topology. This interpolation method has been recently applied in the smoothing of (rigid) contact surfaces involved in metal forming processes [73, 74, 113]. Although the formulation can account for discontinuity of normal vectors (sharp edges and singular points), in the present study the interpolation method is restricted to smooth contact surfaces.

The Nagata patch formulation is described in the following, starting with the 2D interpolation applied to an edge, where $\mathbf{x}_0$ and $\mathbf{x}_1$ denote the position vectors of the edge ends (element nodes). The interpolation of this edge is replaced by a Nagata curve in the form:

$$\mathcal{C}(\xi) = \mathbf{x}_0 + (\mathbf{d} - \mathbf{c})\xi + \mathbf{c}\,\xi^2, \tag{47}$$

where ξ is the local coordinate that satisfies the condition $0 \le \xi \le 1$, and $\mathbf{d} = \mathbf{x}_1 - \mathbf{x}_0$ is the vector joining the end points of the edge. The coefficient vector $\mathbf{c}$, called the curvature parameter, adds the curvature to the edge. Assuming that the Nagata curve is orthogonal to the unit normal vectors $\mathbf{n}_0$ and $\mathbf{n}_1$, the curvature parameter $\mathbf{c}$ is determined minimizing its norm, as follows:

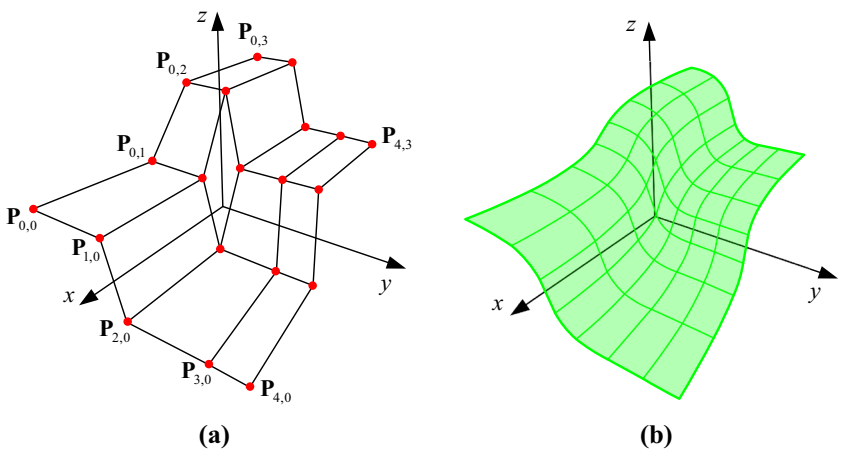

Fig. 21 Example of a NURBS surface: **a** control points denoted by *red dots* (forming a control net); **b** NURBS surface. (Color figure online)

$$\mathbf{c}(\mathbf{d},\,\mathbf{n}_0,\,\mathbf{n}_1)=\begin{cases}\dfrac{[\mathbf{n}_0,\,\mathbf{n}_1]}{1-a^2}\begin{bmatrix}1 & -a\\ -a & 1\end{bmatrix}\begin{Bmatrix}\mathbf{n}_0\cdot\mathbf{d}\\ -\mathbf{n}_1\cdot\mathbf{d}\end{Bmatrix} & (a\neq\pm1)\\ \dfrac{[\mathbf{n}_0,\,\pm\mathbf{n}_0]}{2}\begin{Bmatrix}\mathbf{n}_0\cdot\mathbf{d}\\ \mp\mathbf{n}_0\cdot\mathbf{d}\end{Bmatrix}=\mathbf{0} & (a=\pm1)\end{cases}, \tag{48}$$

where $a=\mathbf{n}_0\cdot\mathbf{n}_1$ denotes the cosine of the angle between the two normal vectors and $[\cdot,\cdot]$ represents a matrix composed by two vectors. When the normal vectors are parallel $(a=\pm1)$, the curvature parameter vanishes and the Nagata curve degenerates into a linear segment. Some modifications in the curvature parameter definition were proposed by Sekine and Obikawa [114] in order to apply the interpolation method in the tool path generation for ball end milling. The modified formulation does not satisfy exactly the boundary conditions, i.e. the surface is not mandatorily orthogonal to the normal vectors given at the nodes. Nevertheless, the very sharp surfaces with inverted orientation that can arise in the original formulation, are avoided as highlighted by Boschiroli et al. [115]. The above interpolation algorithm expressed by (47) and (48) is the basis to apply the Nagata interpolation to general n-sided patches, such as triangular and quadrilateral patches. The idea is to first interpolate independently each edge of the finite element through the quadratic curve (47). Afterwards, the Nagata patch is defined by its trace on the quadratic curves.

In the case of a quadrilateral finite element, schematically presented in Fig. 22, the Nagata patch is defined by interpolating each edge of the element using (47) and then the interior is filled. The input data necessary in the vertices v_1, v_2, v_3 and v_4 are the position vectors $\mathbf{x}_{00}$, $\mathbf{x}_{10}$, $\mathbf{x}_{11}$ and $\mathbf{x}_{01}$, and the unit normal vectors $\mathbf{n}_{00}$, $\mathbf{n}_{10}$, $\mathbf{n}_{11}$ and $\mathbf{n}_{01}$ (see Fig. 22). In case of a quadrilateral patch, the interpolated surface is approximated by the following quadratic polynomial:

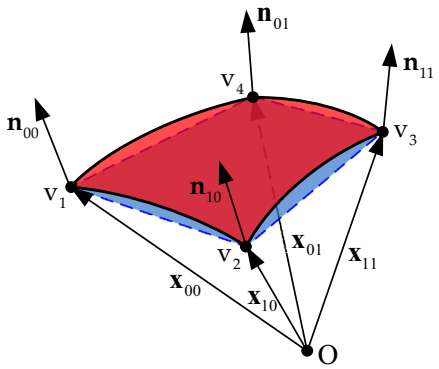

Fig. 22 Sketch of the quadrilateral Nagata patch interpolation using nodal normal vectors

$$\begin{aligned}\mathcal{P}(\eta,\,\zeta)=\;&\mathbf{c}_{00}+\mathbf{c}_{10}\eta+\mathbf{c}_{01}\zeta+\mathbf{c}_{11}\eta\zeta+\mathbf{c}_{20}\eta^2+\mathbf{c}_{02}\zeta^2\\&+\mathbf{c}_{21}\eta^2\zeta+\mathbf{c}_{12}\eta\zeta^2,\end{aligned} \tag{49}$$

where the patch domain in the local coordinates is defined by $0\leq\eta,\zeta\leq1$. The coefficient vectors for the quadrilateral Nagata patch are given by:

$$\begin{aligned}\mathbf{c}_{00}&=\mathbf{x}_{00},\\ \mathbf{c}_{10}&=\mathbf{x}_{10}-\mathbf{x}_{00}-\mathbf{c}_1,\\ \mathbf{c}_{01}&=\mathbf{x}_{01}-\mathbf{x}_{00}-\mathbf{c}_4,\\ \mathbf{c}_{11}&=\mathbf{x}_{11}-\mathbf{x}_{10}-\mathbf{x}_{01}+\mathbf{x}_{00}+\mathbf{c}_1-\mathbf{c}_2-\mathbf{c}_3+\mathbf{c}_4,\\ \mathbf{c}_{20}&=\mathbf{c}_1,\\ \mathbf{c}_{02}&=\mathbf{c}_4,\\ \mathbf{c}_{21}&=\mathbf{c}_3-\mathbf{c}_1,\\ \mathbf{c}_{12}&=\mathbf{c}_2-\mathbf{c}_4,\end{aligned} \tag{50}$$

where $\mathbf{c}_1$, $\mathbf{c}_2$, $\mathbf{c}_3$ and $\mathbf{c}_4$ are the curvature parameters defined for the finite element edges $(\mathbf{x}_{00},\,\mathbf{x}_{10})$, $(\mathbf{x}_{10},\,\mathbf{x}_{11})$, $(\mathbf{x}_{01},\,\mathbf{x}_{11})$ and $(\mathbf{x}_{00},\,\mathbf{x}_{01})$, respectively. These vectors are determined by applying (48) to each of these edges considering:

$$\begin{aligned}\mathbf{c}_1&\equiv\mathbf{c}(\mathbf{d}_1,\mathbf{n}_{00},\mathbf{n}_{10}),\\ \mathbf{c}_2&\equiv\mathbf{c}(\mathbf{d}_2,\mathbf{n}_{10},\mathbf{n}_{11}),\\ \mathbf{c}_3&\equiv\mathbf{c}(\mathbf{d}_3,\mathbf{n}_{01},\mathbf{n}_{11}),\\ \mathbf{c}_4&\equiv\mathbf{c}(\mathbf{d}_4,\mathbf{n}_{00},\mathbf{n}_{01}).\end{aligned} \tag{51}$$

The bilinear parametric representations of the quadrilateral finite elements is recovered as a particular case when the curvature parameters defined in (48) are set as zero vectors. The triangular Nagata patch is obtained in a similar way as the quadrilateral patch [116].

5.3 Approximation of the Normal Vectors

Since the analytical evaluation of the nodal normal vectors is restricted to simple geometries, the application of the Nagata interpolation to general surfaces requires the approximation of the normal vectors. Typically, the finite element mesh of a surface is generated from the model created in a CAD software package. Thus, the information available in the CAD model can be used to evaluate the nodal normal vectors, using the neutral file format Initial Graphics Exchange Specification (IGES) [117]. The IGES file format contains all the information required for the mathematical definition of the surface geometry (trimmed NURBS) and it is organised in a structured manner, following a standard specification [118]. The flowchart presented in Fig. 23 shows the surface smoothing procedure using Nagata patches with normal vectors evaluated from the information contained in the CAD model. The

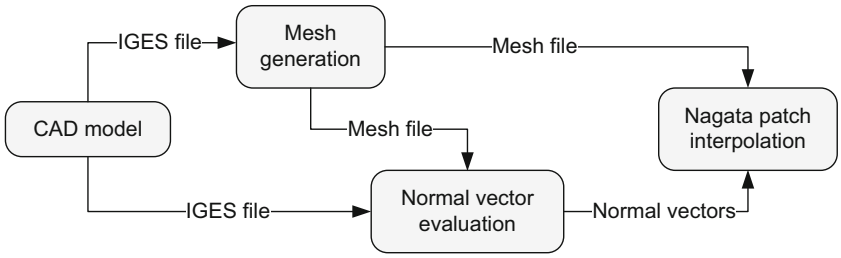

Fig. 23 Procedure followed to evaluate the surface normal vectors from the CAD model information

procedure can be divided in three steps: (i) finite element mesh generation; (ii) nodal normal vectors evaluation and (iii) Nagata patch interpolation.

In the general case of contact between two deformable bodies, the shape of the contact surface changes due to the interaction between the bodies. Therefore, the finite element discretization of the contact surfaces is the only information available to evaluate the required nodal normal vectors. Two different approaches have been developed to approximate the surface normal vectors in the nodes of a finite element mesh. The first approach is based on the fitting of a smooth parametric surface using the neighbouring mesh nodes and subsequent evaluation of the normal vector using the generated surface [119–121]. However, the accuracy of this approach is highly dependent on the distribution of the neighbouring nodes position, as well as on the degree of the interpolated surface.

The idea of the second approach is the evaluation of the surface normal vector using the weighted average of the normal vectors of facets adjacent to the node [122–124]. This approach is simple (see Fig. 24) and presents low computational cost. All variants share the concept of weighting adjacent facet normal vectors, but they differ substantially in the weighting factor adopted [123, 124], which takes into account topological information provided by the finite element mesh. The normal vector at a generic node of a finite element mesh, surrounded by n_f facets, is obtained from the weighted sum of the normal vector of each facet, expressed by:

$$\mathbf{n}_\mathrm{approx} = \frac{\sum_{i=1}^{n_\mathrm{f}} w_i \mathbf{n}_i^\mathrm{facet}}{\left\| \sum_{i=1}^{n_\mathrm{f}} w_i \mathbf{n}_i^\mathrm{facet} \right\|}, \tag{52}$$

where $\mathbf{n}_\mathrm{approx}$ is the approximated unit normal vector of the unknown surface. The unit normal vector of the ith finite element (facet) surrounding the node is denoted by $\mathbf{n}_i^\mathrm{facet}$, while its weight in the average is indicated by w_i. The graphical representation of (52) is illustrated in Fig. 24 for a node surrounded by 5 triangular elements. When quadrilateral finite elements (generally non-coplanar) are adopted in the surface description, the normal vector of each facet required for (52) is evaluated using the two edges that share the node. The approximated nodal normal vector provided by the weighted average may lead to inaccurate results in some locations, particularly in the intersection between two surface geometries, as well as in the surface boundaries. Therefore, these situations should be identified a priori in the finite element mesh, in order to perform the adjustment of the nodal normal vector. In case of intersection between flat and curved geometries, the nodal normal vector of the nodes located in the transition zone should present the normal vector of the flat surface. Regarding the normal vectors for the nodes located in symmetry planes, the reduced amount of information available dictates the mirroring of the existing geometry [125].

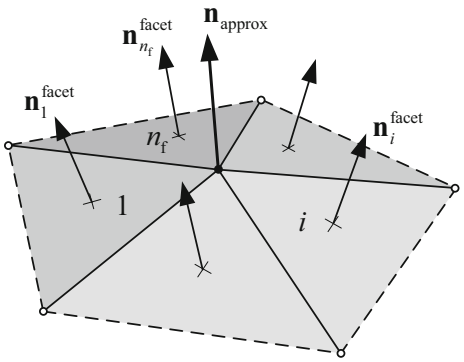

Fig. 24 Schematic representation of the nodal normal vector evaluated through the normal vectors of the surrounding facets

5.4 Geometrical Accuracy

The accuracy of the surface representation is evaluated through two distinct types of error: shape of the interpolated surface and deviation in the surface normal vector [74, 101]. The first one dictates the accuracy in the computation of the normal gap function (4), while the second one is related with the non-physical oscillations arising in the contact force for large sliding contact problems. Since the nodal normal vectors are required for the Nagata interpolation, only geometries defined by analytical

functions are evaluated, namely the circular arc and the sphere. The shape accuracy of the interpolation is evaluated by means of the radial error. Considering a circular arc of radius r, the radial error associated with the interpolation is defined by:

$$\delta_r(\eta,\zeta) = \frac{(\boldsymbol{\mathcal{P}}(\eta,\zeta) - \mathbf{o}) \cdot \mathbf{n}_{\text{analytical}} - r}{r}, \tag{53}$$

where $\boldsymbol{\mathcal{P}}(\eta,\zeta)$ denotes the position vector of a generic point on the interpolated surface, $\mathbf{o}$ is the position vector of the circle centre and $\mathbf{n}_{\text{analytical}}$ is the unit normal vector of the circular arc defined by the analytic function. The radial error represents the dimensionless distance between the interpolated curve and the circular arc defined by the analytical function, measured in the radial direction. The second type of error studied is the normal vector error, which is defined by:

$$|\delta_n(\eta,\zeta)| = \cos^{-1}(\mathbf{n}_{\text{interpol}}(\eta,\zeta) \cdot \mathbf{n}_{\text{analytical}}) \ [^\circ], \tag{54}$$

where $\mathbf{n}_{\text{interpol}}$ is the unit normal vector of the interpolated surface, which is defined by the cross product of the derivatives. The modulus of the normal vector error expresses the angle between the normal vector of the interpolated surface and the analytical normal vector evaluated in the projection point. This error is directly associated with the discontinuity of the normal vector orientation in the contact surface (see Fig. 3), which is a key point for the robustness of the solution procedure adopted for solving contact problems. In fact, if the error in the normal vector is zero in the transition between neighbouring patches, the contact surface is smooth.

The first geometry analysed is the circular arc (2D geometry), which is described by linear segments, Nagata and Bézier curves. Although it is a simple geometry, constant radius fillet surfaces and surfaces of revolution are frequently found in the description of complex surfaces. The circular arc can be represented exactly through a NURBS curve of order 3 using three control points [106]. Concerning the other interpolation methods, the quarter of circle is divided in a different number of divisions in order to achieve a similar value of radial error for all methods. The radial error distribution in the circular arc is presented in Fig. 25 for different interpolation methods. Adopting the linear interpolation, the radial error is negative (inside interpolation), attaining its largest value in the middle of the segment. The division of the circular arc in 34 equal divisions (35 points) leads to the error presented in Fig. 25. Both the Bézier and Nagata interpolation methods produce a curve outside the analytical one. The Bézier interpolation uses four control points, which are strategically positioned to obtain zero radial error at the midpoint of the Bézier curve [126]. On the other hand, the Nagata interpolation is performed with the discretization of the circular arc using 4 equal divisions (5 nodes). The maximum value of radial error is slightly higher in the Bézier interpolation than adopting the Nagata interpolation, as shown in Fig. 25.

The radial error distribution obtained with linear and Nagata interpolation of a circular arc discretized by two neighbouring finite elements is presented in Fig. 26. The central angle β is the parameter selected to quantity the mesh refinement, which gives the length of each discretized circular arc. Three different values of central angle are analysed, corresponding to the division of half circumference into 8, 10 and 12 elements. As expected, the radial error decreases with the mesh refinement for both interpolation methods, as shown in Fig. 26. Although the evolution of the radial error is roughly similar for both interpolation methods (negative in linear and positive in Nagata interpolation), the order of magnitude of the results is completely different.

The normal vector error distribution obtained with linear and Nagata interpolation of a circular arc discretized by two neighbouring finite elements is presented in Fig. 27. Such as for the radial error, the normal vector error decreases with the increasing of the number of elements used to describe the circular arc (mesh refinement). The sudden changes in the normal vector direction created by the linear interpolation can be observed through the discontinuity of the normal vector error distribution across element boundaries, as illustrated in Fig. 27a. Considering

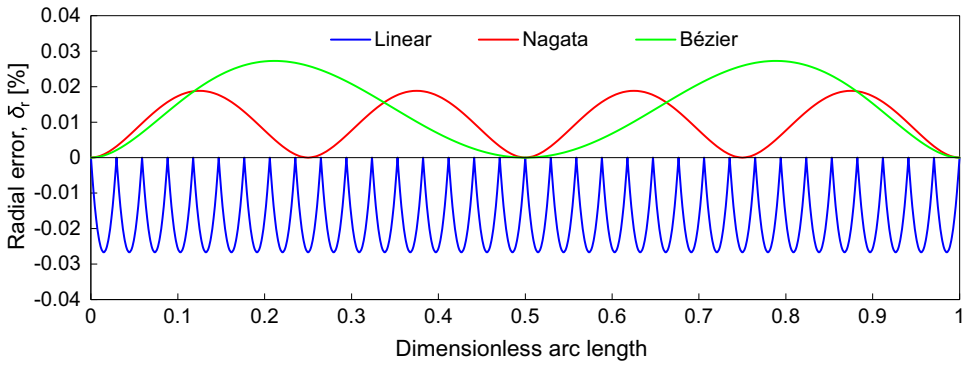

Fig. 25 Radial error distribution in the circular arc described by linear segments, Nagata and Bézier curves

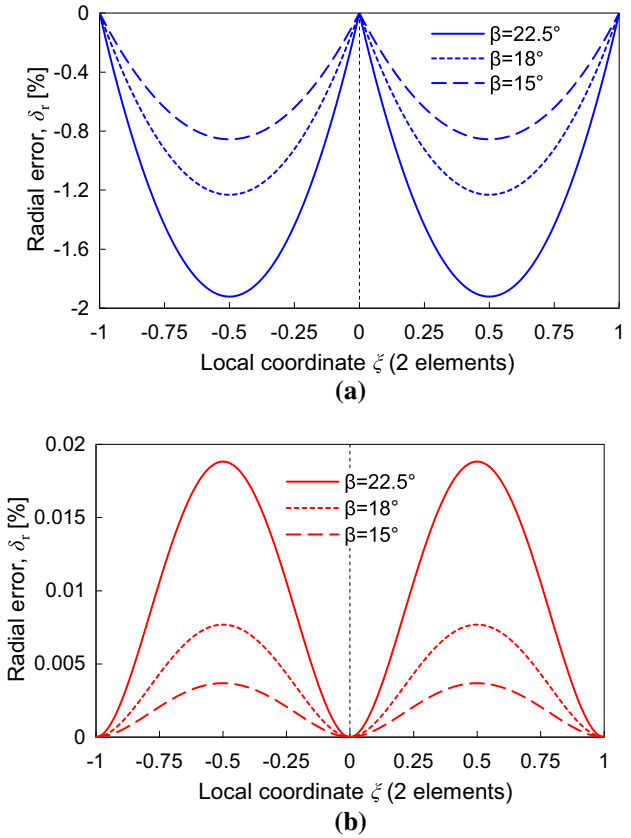

Fig. 26 Radial error distribution in a circular arc defined by two elements: **a** linear interpolation; **b** Nagata interpolation

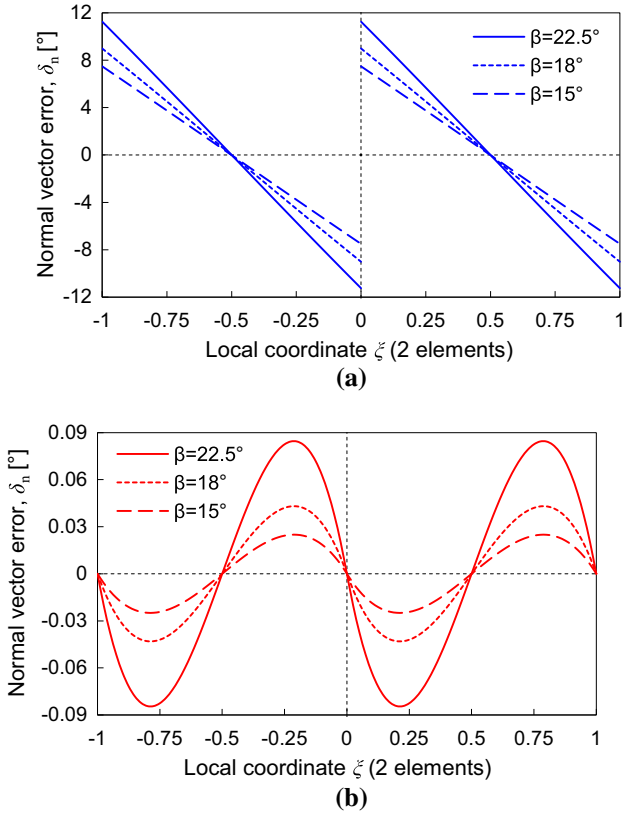

Fig. 27 Normal vector error distribution in a circular arc defined by two elements: **a** linear interpolation; **b** Nagata interpolation

the linear interpolation, the normal vector error is zero where the radial error attains its maximum value (negative), while the maximum value of the normal vector error (half value of the arc central angle) occurs at the nodes. On the other hand, the Nagata interpolation assures G^1 continuity in the curve, providing a continuous distribution of the normal vector error, as shown in Fig. 27b.

In order to study the influence of the discretization on the interpolation error, different lengths of circular arc are studied. The range considered for the normalized arc length is from 1.571 until 0.079, which corresponds to dividing a quarter of circle from 1 to 20 equal elements, respectively. The normalized arc length expresses the division of the arc length by the radius r of the circular arc. Figure 28a presents the maximum norm of radial error in function of the normalized arc length (ℓ/r), for both the linear and the Nagata interpolation. The order of convergence in the radial error provided by the linear interpolation is quadratic, while when applying Nagata interpolation the convergence rate is quartic [113]. Figure 28b shows the maximum normal vector error modulus as function of the normalized arc length, for both surface description methods. The error decreases linearly when the linear interpolation is adopted, while the Nagata interpolation method provides a cubic convergence rate. In fact, considering the wide range of normalized arc length analysed, the maximum normal vector error modulus is always larger than 2° for the linear interpolation, while the Nagata interpolation gives an error always inferior to 1°, with the exception of the coarser mesh with only one element (Fig. 28b).

The spherical surface of unit radius is selected to evaluate the accuracy of both interpolation methods, using structured meshes composed either by triangular or quadrilateral finite elements. The modulus of the maximum radial error attained in the description of the spherical surface is presented in Fig. 29a, considering both faceted elements and Nagata patches. The radial error is significantly lower when the Nagata interpolation is applied. The maximum value attained in the Nagata interpolation decreases quartically with the square root of the maximum element area normalized by the sphere radius, while quadratic convergence rate is observed in the faceted surface description. For both surface description methods, the discretization with quadrilateral finite elements provides a maximum value of radial error always inferior to the one obtained with triangular elements, considering the same finite element area (see Fig. 29a). The convergence rate of the maximum normal vector error modulus attained in the

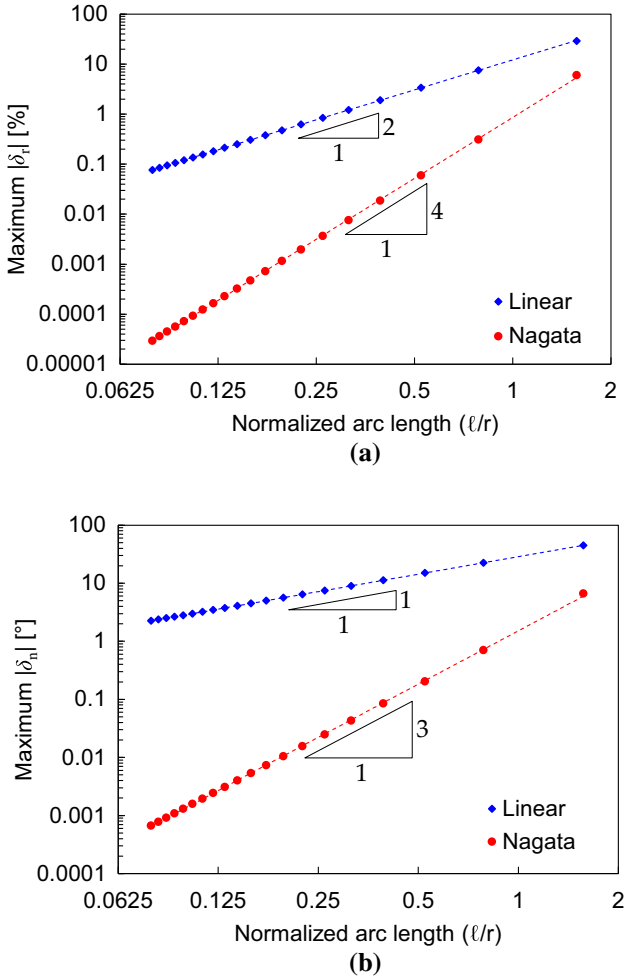

Fig. 28 Comparison between linear and Nagata interpolation accuracy applied to a circular arc: **a** maximum radial error modulus; **b** maximum normal vector error modulus

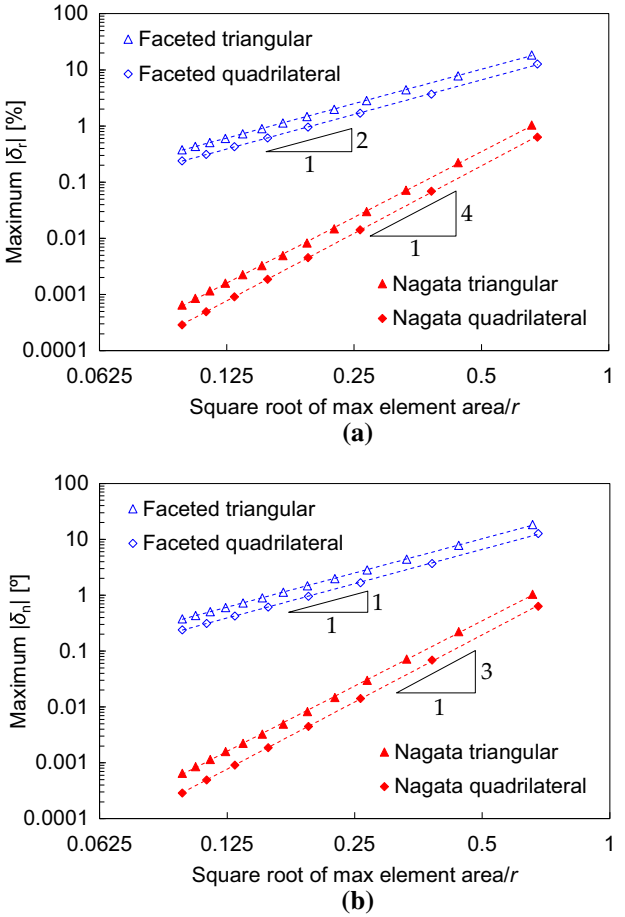

Fig. 29 Comparison between linear and Nagata interpolation accuracy applied to a spherical surface: **a** maximum radial error modulus; **b** maximum normal vector error modulus

description of the spherical surface is presented in Fig. 29b. The maximum error value decreases linearly adopting the faceted surface description, while the Nagata patches interpolation provides a cubic convergence rate. Therefore, this convergence rate allows considering that the smoothing method with Nagata patches ensures *quasi-*G^1 continuity in the patch boundaries [113].

6 Contact Elements Definition

6.1 Contact Search Algorithms

The contact search is the step preceding the creation of the contact elements, which aims to determine the contacting pairs on the discretized surfaces. Consequently, the contact search algorithm is strongly connected with the contact surface discretization scheme, as well as the type of contact, i.e. contact between two separated bodies or self-contact (unknown master/slave pair). Moreover, its robustness determines the accuracy of the entire resolution scheme [127], highlighting its importance to the numerical treatment of contact problems. Nowadays, the complexity demanded by computational contact problems includes large deformation, large sliding and self-contact, requiring the development of efficient contact search algorithms for large size models [128]. Since the contact zones can change considerably during the deformation process, the contact search has to be performed in every time step of the numerical simulation, presenting a bottleneck for an efficient treatment of contact problems [51, 129]. On the other hand, in case of small deformations (Node-to-Node discretization) the contact pairs of nodes are established at the beginning, and do not change during the solution steps (no slip occurs), requiring only one execution of the search procedure [38]. The presented contact search algorithm was specifically developed for the NTS discretization.

The contact search algorithms are typically decomposed into two distinct phases [3]: (i) global search and (ii) local

search. The global contact search procedure is related to purely geometrical considerations and its purpose is to determine potential contacting bodies/surfaces. A hierarchical structure is created to find out which contact surfaces are able to come into contact in a given time step [1]. The aim of the local contact search process is to find for each slave node the closest point on the associated master segment, defining the contact element [130]. Nevertheless, within implicit integration scheme, the possible penetration has to be known at the beginning of the time step, in order to incorporate properly the contact residual vector and the stiffness matrix in the resolution stage [129].

6.1.1 Global Search

Several algorithms have been developed to identify all possible candidate contact partners. The simplest search algorithm is the all-to-all detection, sometimes referred to as the "brute force" approach. In that case, all slave nodes are projected on all master segments in order to determine for each slave node the closest master segment with existent projection. Since the evaluation of the projection point requires the solution of a nonlinear system of equations, this approach leads to unsupportable computational cost for most situations [51]. Nevertheless, the global search is highly dependent of the contact surface description, namely the regular or irregular finite element mesh discretization. In fact, the global search is quite straightforward and efficient using a regular mesh for the contact definition [72], while the irregular discretization of the contact surfaces (general case) leads to complex and less efficient contact search algorithms. The most efficient global contact search algorithms are the Hierarchy-Territory algorithm (HITA) [131], the typical bucket sorting algorithm [132] and the position code algorithm [133].

Considering the contact search algorithm HITA, proposed by Zhong [1], the contact system is decomposed into four contact hierarchies: bodies, surfaces, segments and nodes. The search procedure is based in the definition of a territory for each contact element (cubic box containing the element). The territories at the same hierarchical level are compared to detect common territory, where the lower levels of hierarchy are tested. A node and a segment are a candidate pair for contact when the distance between the node and the segment is sufficiently small. The main idea of the bucket sorting algorithm proposed by Benson and Hallquist [132] is the subdivision of the space into equally cells or buckets, assigning each node/segment of the contact surface to a bucket number. Then, the list of segments in each bucket is used to find the candidate master segments for each slave node, allowing to reduce locally the area of search (inside the bucket). In the position code algorithm, developed by Oldenburg and Nilsson [133], each master segment is checked for the presence of slave nodes situated within the segment territory, which is defined by the smallest cubic box holding the contact territory. The algorithm to detect the contact nodes within the segment territories is based on sorting and searching in one dimension. The mapping from three dimensions to one dimension is achieved by the definition of a discrete position code, which is the number of the box resulting from the division of the space into cubic boxes. The study conducted by Oldenburg and Nilsson [133] concluded that the HITA and the position code algorithms are superior in terms of computational efficiency. While the bucket sorting algorithm performs sorting and searching in three dimensions in a nested manner, the position code reduces the three-dimensional space into a one dimensional searching problem, hence decreasing the number of operations involved, due to the binary search procedure used [134].

The starting point of the global search is the definition of contact pairs, i.e. the assignment of slave and master surfaces by the user. In contact problems involving rigid surfaces, they are automatically specified as master surfaces. Regarding the contact problems between separated deformable bodies, the contacting surface pairs are created based on the expected interaction between the bodies. The discretized master contact surface is defined from the assembly of solid element faces looking outside the body. For deformable bodies discretized with linear solid finite elements, the contact faces are defined by quadrilateral elements, in meshes composed by 8-node hexahedral elements, and triangular elements, in meshes composed by 4-node tetrahedral elements. Although detection methods based on the search for the closest master node and adjacent segments have been widely used [132], they are known for not being robust. In fact, the assumption that the slave node is in contact with the master segment sharing the closest master node is not always correct, as shown in Fig. 30. Thus, this detection method may fail for unstructured meshes. Figure 30 presents two examples of flat master surfaces described by irregular finite element meshes, highlighting a slave node that does not contact a master segment connected to the closest master node.

Applying the two-pass NTS contact algorithm to deal with the contact constraints, the classification of the master and slave surfaces is arbitrary due to its exchange role in the second pass. In fact, for some mechanical problems, the a priori assignment of the contact pairs is impossible or presents a big challenge. The analysis of self-contact problems, which arise generally in the post-buckling of thin walled structures, is a representative example. The bucket sorting method is frequently applied in these problems, which is known as the single surface contact algorithm, where an additional history variable is created to track which side of the contact surface has been penetrated

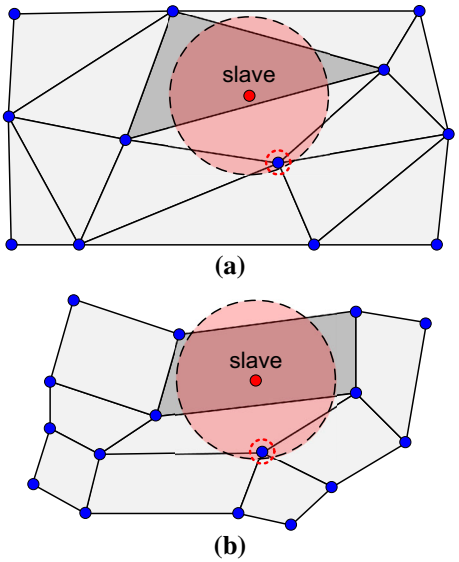

Fig. 30 Failure of the global search algorithm based on the closest master node: **a** master surface described by triangular elements; **b** master surface described by quadrilateral elements

[132]. Alternative approaches based in geometric informations (i.e. curvature criterion) have been recently proposed to deal with self-contact problems [95]. Since the outward normal vectors of the contacting surfaces at the contact point are always opposed (pointing towards each other), only the nodes with opposite normal vectors are considered.

Only the two-pass NTS contact formulation is allowed in self-contact problems, because the contact pair is specified by a single surface instead of two different surfaces. Thus, each node of the surface acts as a slave node and every finite element face composing the surface is defined as master surface. Since one or two dimensions are much smaller than the others in thin walled structures, some difficulties arise in the contact search due to the contact with the reverse side. Figure 31 presents a thin solid with two sided contact zones (red and blue), which can be treated independently. Adopting the contact search algorithm based in the closest master node, the square points can be incorrectly identified as being penetrating the surface with the same colour due to the proximity and opposite surface normal vectors. The proposed solution takes into account the normal vectors at the nodes to determine potential contacting surfaces, establishing a maximal detection distance in the normal direction smaller than two times the wall thickness (see the rectangular boxes in Fig. 31b).

6.1.2 Local Search

The purpose of the local contact search algorithm is to find the point on the master surface closest to each slave node. In order to reduce the computational time associated to the contact detection, the local search is performed only within the set of master segments identified in the global contact search. However, the computational cost of the local search is typically higher than the one required by the global search procedure, since it is repeated in each iteration of every time step. The methods commonly adopted in the local search are the closest point projection algorithm [11, 135], the pinball algorithm proposed by Belytschko and Neal [136] and the inside–outside algorithm suggested by Wang and Nakamachi [77]. For the closest point projection procedure, the minimum distance between each slave node and the master surface is calculated based on the normal projection of the node onto the surface. The Newton–Raphson method is typically used to solve this problem and find the contact point coordinates. On the other hand, the pinball algorithm is very efficient when combined with the penalty method, since it is based on simple checks, thus eliminating any iterative procedure. However, some inaccuracies concerning the real geometry of the contacting bodies are introduced because the penetration between their surfaces is assumed as the interpenetration of two spherical balls. The inside–outside algorithm is based in the status of the projected point of the slave node along the mesh normal direction. Only two states are allowed, i.e. the projection point is located either inside or outside the

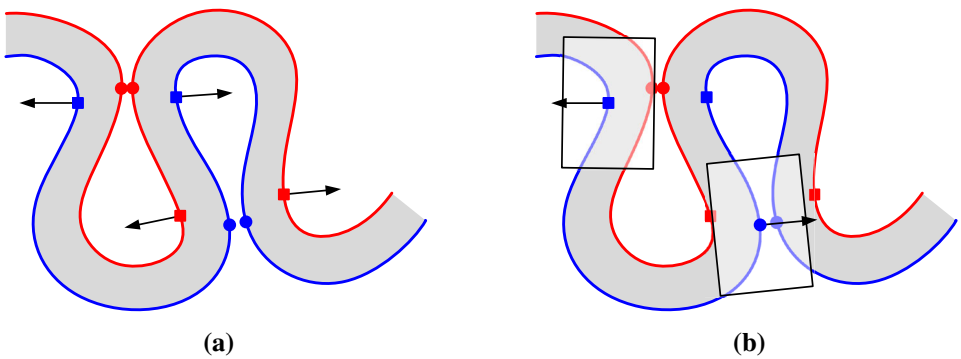

Fig. 31 Self-contact in a thin walled structure: **a** difficulties in the contact with the reverse side; **b** global search taking into account the nodal normal vectors

master segment. This algorithm is fast, robust and does not requires any iterative procedure to perform the search (closed-form expression) if faceted finite elements are used in the contact surfaces description [77].

The normal gap previously defined in (4) is strongly connected with the closest point projection used in the local contact detection procedure [135]. The aim of the closest point projection is to find, for each slave node, the point belonging to the contact master surface that is closest to the slave node. The coordinates of a generic slave node $\mathbf{x}^{\mathrm{s}}$ can be correlated with a vector describing any point on the master surface $\mathbf{x}^{\mathrm{m}}$ through the normal gap value, as given by the following equation:

$$\mathbf{F}^{\mathrm{proj}}(\eta, \zeta, g_{\mathrm{n}}) = \mathbf{x}^{\mathrm{m}}(\eta, \zeta) + g_{\mathrm{n}}\mathbf{n}(\eta, \zeta) - \mathbf{x}^{\mathrm{s}}, \tag{55}$$

where $\mathbf{n}$ denotes the unit normal vector of the master surface, while the value of the normal gap function is the third coordinate of the surface coordinate system [137]. The solution of $\mathbf{F}^{\mathrm{proj}} = \mathbf{0}$ provides the local coordinates of the contact point and assures that the vector connecting this point to the slave node is collinear with the normal vector (see Fig. 6). The Newton–Raphson method is frequently used for solving the nonlinear system of equations, which can be summarized as follows:

$$\mathbf{s}_{i+1} = \mathbf{s}_i - \left[\nabla \mathbf{F}^{\mathrm{proj}}(\mathbf{s}_i)\right]^{-1} \mathbf{F}^{\mathrm{proj}}(\mathbf{s}_i), \tag{56}$$

where $\mathbf{s}_i = [\eta, \zeta, g_{\mathrm{n}}]_i^{\mathrm{T}}$ contains the solution vector at iteration i and $\nabla \mathbf{F}^{\mathrm{proj}}$ denotes the Jacobian matrix of the system of equations. Note that the solution yields simultaneously the normal gap g_{n} and the local coordinates of the contact point on the master segment $(\bar{\eta}, \bar{\zeta})$ [21].

The finite element approximation of the master surface leads to a non-smooth surface representation between adjacent finite elements, which arises mostly in the bilinear parametric representations. In fact, this situation leads to several difficulties in finding the projection of the slave node on the master segment. Each master segment presents its "normal projection" zone, where the slave node can have at least one projection onto the master surface. However, sometimes the assembly of the "normal projection" zones does not fill the neighbouring space completely, creating *deadzones* where no normal projection exists, as shown in Fig. 32. When no projection is found, as shown in Fig. 32, serious numerical problems may arise [17]. Two types of blind spots can be distinguished: internal and external. Slave nodes situated in external blind spots are not detected before they penetrate the master surface, as shown in Fig. 32a. On the other hand, in the presence of internal blind spots the contact is predicted correctly (before penetration), but if the slave node penetrates the master surface it is not detected, as shown in Fig. 32b. If several projections are found, the projection

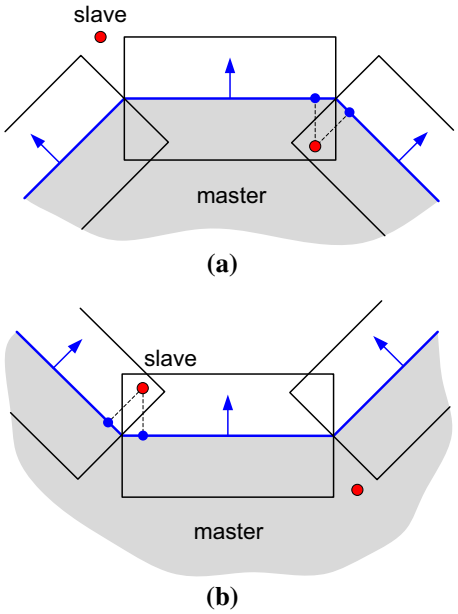

Fig. 32 Example of a slave node near a sharp corner/valley: **a** convex master surface; **b** concave master surface

point with minimum normal gap is selected to create a contact element [21].

The sudden changes in the surface normal field of the master surface, induced by the finite element discretization, cause serious convergence difficulties in the solution of contact problems involving sliding, namely in local search detection procedure. Thus, the master surface smoothing is typically adopted to overcome the above mentioned problems (blind spots). The behaviour of the closest point projection is evaluated through a simple example, comparing the faceted and the smoothed surface description methods. The setup is composed by a flat surface and a spherical convex surface, representing the slave and master surfaces, respectively. The spherical surface is discretized with 16 bilinear quadrilateral finite elements, while the square flat surface is defined by a fine grid of points (300 divisions in each direction). The relative position of the two surfaces is presented in Fig. 33, where two views are shown.

The local search procedure base in the normal projection of a slave point on the master surface fails when it is described by bilinear quadrilateral finite elements, specifically near the common edges of the finite elements. Figure 34a shows the colour map denoting the finite elements on which the slave points are projected. Some *deadzones* (white colour) arise, which are larger for points located more distant to the surface due to the pyramidal shape of the blind spots, as illustrated in Fig. 32a. The points of the slave surface located within the white regions do not have any normal projection with any master finite element. On

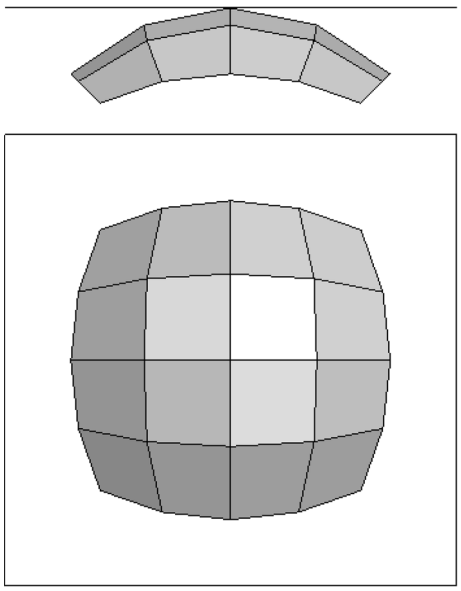

Fig. 33 Flat square surface close to a spherical convex surface (lateral and top views)

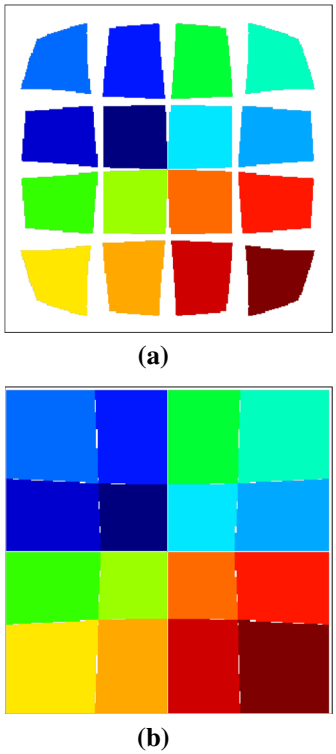

Fig. 34 Regions of flat surface with normal projection on the convex surface: **a** faceted spherical surface, **b** smoothed spherical surface. *Each colour* denotes a different finite element of the master surface. (Color figure online)

the other hand, the smoothing of the spherical surface with Nagata patches improves significantly the local contact search procedure. The colour map denoting the patches on which the slave points are projected is presented in Fig. 34b. The blind spots observed in the surface modelled by faceted elements are strongly reduced using a smoothed master surface. In fact, the zones of the slave surface without normal projection (white colour) are now located in a very narrow range near the edges of the patches.

6.2 Node-to-Nagata Contact Elements

The formulation of these contact elements is developed in the framework of the Node-to-Segment contact discretization. Due to the geometrical description of the master surface with Nagata patches, they are called Node-to-Nagata contact elements [125]. Since the frictional contact problem between a deformable body and a rigid obstacle is a simplification of the contact between two deformable bodies, only the general case is presented. The residual vectors and Jacobian matrices of the developed contact element are derived for the augmented Lagrangian method. Assuming that the master surface is rigid, the contact forces arise only in the slave nodes due to the violation of geometrical constraints and are not transferred to the master surface. On the other and, when two deformable bodies come in contact, the contact forces arising in the contact interface are transferred from one body to the other according to the impenetrability and friction conditions.

Adopting the NTS contact discretization technique, the slave surface is represented by a set of nodes while the master surface is defined by a collection of Nagata patches. Each contact element used to connect the two deformable bodies comprises one slave node and several master nodes attached to one structural finite element, as illustrated schematically in Fig. 16. The geometry of a contact element presents a pyramid like shape, where the slave node is the top vertex and the base is composed by master nodes, as shown in Fig. 35. The Nagata patch (triangular or

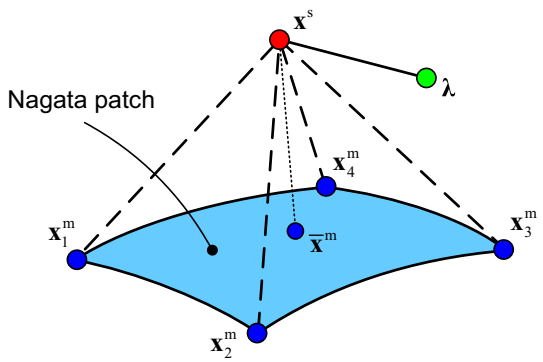

Fig. 35 Contact element of the type Node-to-Nagata using four master nodes (the artificial node corresponding to the Lagrange multipliers is marked in *green*). (Color figure online)

quadrilateral) connected with each slave node is selected according to the closest point projection (local search algorithm described in Sect. 6.1.2). Since the augmented Lagrangian method is adopted to formulate the frictional contact problem [23, 38], each contact element is complemented by an artificial node to store the contact force (Lagrange multipliers), as illustrated in Fig. 35. The finite element discretization of the contact problem leads to a system of nonlinear equations with both nodal displacements and contact forces as unknowns [138–140].

Following the augmented Lagrangian approach proposed by Alart and Curnier [38], the static equilibrium for the frictional contact problem between two discretized deformable bodies is governed by the following system of non-differentiable equations:

$$\begin{cases} (\mathbf{F}^{\text{int}}(\mathbf{u}^1) - \mathbf{F}^{\text{ext}})^1 + \mathbf{F}_{\text{s}}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda}) = \mathbf{0} \\ (\mathbf{F}^{\text{int}}(\mathbf{u}^2) - \mathbf{F}^{\text{ext}})^2 + \mathbf{F}_{\text{m}}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda}) = \mathbf{0} \\ -\dfrac{1}{\varepsilon}(\boldsymbol{\lambda} - \mathbf{F}_{\text{s}}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda})) = \mathbf{0} \end{cases}, \tag{57}$$

where $\mathbf{F}^{\text{int}}$ and $\mathbf{F}^{\text{ext}}$ denote the internal and the external force vectors, respectively. The first line is related with the equilibrium of the slave body while the second one expresses the equilibrium of the master body. The connection between both deformable bodies is expressed by the frictional contact operator $\mathbf{F}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda})$, which is responsible for the transmission of the frictional contact force from the slave node to the master body. The last line denotes the supplementary equations necessary to evaluate the contact force in the slave node. In the absence of contact, the static equilibrium is expressed independently for each body (two systems of equations), as illustrated in (57), where the last line vanishes as well as the frictional contact operator. The discretized frictional contact operator for a single contact element is defined by:

$$\mathbf{F}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda}) = \text{proj}_{\Re^-}(\hat{\lambda}_{\text{n}})\mathbf{n} + \text{proj}_{C^{\text{aug}}(\hat{\lambda}_{\text{n}})}(\hat{\boldsymbol{\lambda}}_{\text{t}}), \tag{58}$$

where $\hat{\lambda}_{\text{n}}$ and $\hat{\boldsymbol{\lambda}}_{\text{t}}$ are the augmented Lagrange multipliers defined in (26) and (29), respectively. The projection of x on $\Re^-$ is denoted by $\text{proj}_{\Re^-}(x)$, corresponding to the closest point to x belonging in $\Re^-$. The augmented convex involved in the last term of (58) is defined by:

$$C^{\text{aug}}(\hat{\lambda}_{\text{n}}) = C(\text{proj}_{\Re^-}(\hat{\lambda}_{\text{n}})), \tag{59}$$

which consists in the prolongation the multivalued convex set (friction cone) by the positive half-line. For positive values of the normal component of the augmented Lagrange multiplier, the radius of the convex disk is equal to zero and the Coulomb's cone shrinks to the vertical half-line.

In order to highlight the contribution of the frictional contact to the global system of equations, the structural part is separated from the contact part, where the former is derived from the variation of the augmented Lagrangian functional [38]. The nodal displacements and contact forces are evaluated only in the slave nodes, while the variables associated with the contact point $\bar{\mathbf{x}}^{\text{m}}$ on the master Nagata patch are dependent. Since the contact interaction is governed by the principle of action and reaction, assuming contact, the slave node presents the same position than the contact point $\bar{\mathbf{x}}^{\text{m}}$. Nevertheless, in general the contact point does not coincide with any master node (non-conforming meshes at contact interface or large sliding). Thus, the contact force in the slave node should be distributed on the nodes of the master patch, according to the moment equilibrium. According to the contact action–reaction principle, the frictional contact operator vector defined at the contact point (master) presents the same magnitude and opposite direction of the same operator defined in the slave node. Therefore, the frictional contact operator associated with each master node composing the contact element can be expressed by a weighting factor of the frictional contact operator defined for the slave node:

$$\mathbf{F}_{\text{m}i}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda}) = -w_i \mathbf{F}_{\text{s}}^{\text{c}}(\mathbf{u}, \boldsymbol{\lambda}), \tag{60}$$

where the weight for each master node is evaluated according to the local coordinates $(\bar{\eta}, \bar{\zeta})$ of the contact point in the Nagata patch domain. The summation of all weights always gives one, i.e. constitute a partition of unity. The weight associated with each master node can be evaluated by local coordinates of the contact point [125]. The area of the triangle/rectangle opposite to the node defines the associated weight, as highlighted in the Fig. 36. The expressions for the weight associated to each master node in case of quadrilateral Nagata patches are given by:

$$w_1 = (1-\bar{\eta})(1-\bar{\zeta}), \quad w_2 = \bar{\eta}(1-\bar{\zeta}), \quad w_3 = \bar{\eta}\bar{\zeta} \quad \text{and} \quad w_4 = \bar{\zeta}(1-\bar{\eta}), \tag{61}$$

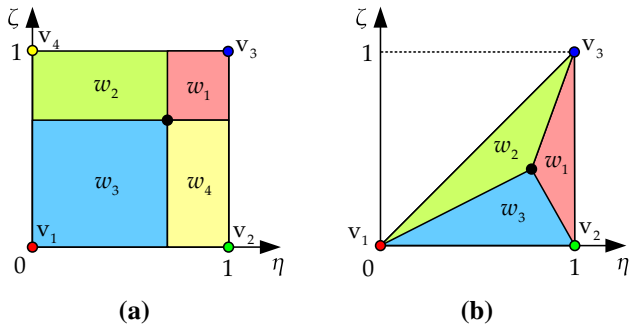

Fig. 36 Definition of the weight associated with each master node based in the opposite area: **a** triangular patch; **b** quadrilateral patch

which involve the local coordinates of the contact point, evaluated through the closest point projection method, described in Sect. 6.1.2.

6.2.1 Residual Vectors

Considering that the contact element is composed by a quadrilateral Nagata patch, the incorporation of the impenetrability and friction constraints into the static equilibrium problem is expressed by the mixed contact operator defined as follows:

$$\mathbf{F}^{*}(\mathbf{u},\boldsymbol{\lambda}) = \left\{ \begin{array}{c} \mathbf{F}_{\mathrm{s}}^{\mathrm{c}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_1\mathbf{F}_{\mathrm{s}}^{\mathrm{c}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_2\mathbf{F}_{\mathrm{s}}^{\mathrm{c}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_3\mathbf{F}_{\mathrm{s}}^{\mathrm{c}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_4\mathbf{F}_{\mathrm{s}}^{\mathrm{c}}(\mathbf{u},\boldsymbol{\lambda}) \\ -\dfrac{1}{\varepsilon}(\boldsymbol{\lambda} - \mathbf{F}_{\mathrm{s}}^{\mathrm{c}}(\mathbf{u},\boldsymbol{\lambda})) \end{array} \right\} = \left\{ \begin{array}{c} \mathbf{F}^{\mathrm{equi}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_1\mathbf{F}^{\mathrm{equi}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_2\mathbf{F}^{\mathrm{equi}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_3\mathbf{F}^{\mathrm{equi}}(\mathbf{u},\boldsymbol{\lambda}) \\ -w_4\mathbf{F}^{\mathrm{equi}}(\mathbf{u},\boldsymbol{\lambda}) \\ \mathbf{F}^{\mathrm{suppl}}(\mathbf{u},\boldsymbol{\lambda}) \end{array} \right\}, \tag{62}$$

where $\mathbf{F}^{\mathrm{equi}}(\mathbf{u},\boldsymbol{\lambda})$ denotes the constraints imposed by the obstacle in the equilibrium equation and $\mathbf{F}^{\mathrm{suppl}}(\mathbf{u},\boldsymbol{\lambda})$ expresses the supplementary equations required to evaluate the frictional contact forces [21]. The contact operator defined for the slave node is extended to the nodes of the master body, according to the relationship expressed in (60). Depending on the contact status (gap, stick and slip), the contribution of the contact element to the virtual work of the system is given by different integrals, as illustrated in (35). Thus, the mixed contact operator is derived for each contact status. In the framework of the augmented Lagrangian method, the contact status is determined by the linear combination of kinematic and static variables (primal and dual variables) [23]. In the absence of contact between the bodies (gap status), defined by the normal component of the augmented Lagrange multiplier $\hat{\lambda}_{\mathrm{n}} > 0$, the mixed contact operator is written as:

$$\mathbf{F}_{\mathrm{gap}}^{*}(\mathbf{u},\boldsymbol{\lambda}) = \left\{ \begin{array}{c} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -\dfrac{1}{\varepsilon}\boldsymbol{\lambda} \end{array} \right\}, \tag{63}$$

which does not present any connection between the slave node and the master nodes composing the contact element. It only removes (if it exists) the contact force in the slave node. The mixed contact operator for stick status, which is defined through the condition $\left\|\hat{\boldsymbol{\lambda}}_{\mathrm{t}}\right\| \leq -\mu\hat{\lambda}_{\mathrm{n}}$, is given by:

$$\mathbf{F}_{\mathrm{stick}}^{*}(\mathbf{u},\boldsymbol{\lambda}) = \left\{ \begin{array}{c} \hat{\lambda}_{\mathrm{n}}\mathbf{n} + \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \\ -w_1(\hat{\lambda}_{\mathrm{n}}\mathbf{n} + \hat{\boldsymbol{\lambda}}_{\mathrm{t}}) \\ -w_2(\hat{\lambda}_{\mathrm{n}}\mathbf{n} + \hat{\boldsymbol{\lambda}}_{\mathrm{t}}) \\ -w_3(\hat{\lambda}_{\mathrm{n}}\mathbf{n} + \hat{\boldsymbol{\lambda}}_{\mathrm{t}}) \\ -w_4(\hat{\lambda}_{\mathrm{n}}\mathbf{n} + \hat{\boldsymbol{\lambda}}_{\mathrm{t}}) \\ g_{\mathrm{n}}\mathbf{n} + \mathbf{g}_{\mathrm{t}} \end{array} \right\}, \tag{64}$$

where is highlighted the connection between the slave node (first line) and the four master nodes (next four lines), due to the contact action–reaction principle. This imposes zero relative displacement between the slave node and the master surface. Finally, for the slip status, defined by the condition $\left\|\hat{\boldsymbol{\lambda}}_{\mathrm{t}}\right\| > -\mu\hat{\lambda}_{\mathrm{n}}$, the mixed contact operator is given by:

$$\mathbf{F}_{\mathrm{slip}}^{*}(\mathbf{u},\boldsymbol{\lambda}) = \left\{ \begin{array}{c} \hat{\lambda}_{\mathrm{n}}(\mathbf{n} - \mu\mathbf{t}) \\ -w_1(\hat{\lambda}_{\mathrm{n}}(\mathbf{n} - \mu\mathbf{t})) \\ -w_2(\hat{\lambda}_{\mathrm{n}}(\mathbf{n} - \mu\mathbf{t})) \\ -w_3(\hat{\lambda}_{\mathrm{n}}(\mathbf{n} - \mu\mathbf{t})) \\ -w_4(\hat{\lambda}_{\mathrm{n}}(\mathbf{n} - \mu\mathbf{t})) \\ g_{\mathrm{n}}\mathbf{n} - \dfrac{1}{\varepsilon}(\boldsymbol{\lambda}_{\mathrm{t}} + \mu\hat{\lambda}_{\mathrm{n}}\mathbf{t}) \end{array} \right\}, \tag{65}$$

where the tangential slip direction unit vector is defined by:

$$\mathbf{t} = \hat{\boldsymbol{\lambda}}_{\mathrm{t}} \Big/ \left\|\hat{\boldsymbol{\lambda}}_{\mathrm{t}}\right\|, \tag{66}$$

which is evaluated in the contact point. The mixed contact operators (63)–(65) only comprise three lines for the master nodes in case of contact elements defined by a triangular Nagata patch.

6.2.2 Jacobian Matrices

The resulting system of nonlinear and partially non-differentiable equations is solved using the generalized Newton method. The convergence of the generalized Newton method applied to frictional contact problems using the augmented Lagrangian is discussed by Alart [58] and more recently by Renard [141] considering elasticity. In order to attain quadratic convergence in the iterative solution scheme, a consistent linearization is required. Therefore, in analogy to the tangent matrix of a structural

finite element, the elemental contact Jacobian matrix is defined from the mixed contact operator vector (62), and it is expressed by:

$$\mathbf{J}^{*}=\begin{bmatrix} \nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & \nabla_{\boldsymbol{\lambda}}\mathbf{F}^{\text{equi}} \\ -w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_1w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_1w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_1w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_1w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_1\nabla_{\boldsymbol{\lambda}}\mathbf{F}^{\text{equi}} \\ -w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_2w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_2w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_2w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_2w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_2\nabla_{\boldsymbol{\lambda}}\mathbf{F}^{\text{equi}} \\ -w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_3w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_3w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_3w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_3w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_3\nabla_{\boldsymbol{\lambda}}\mathbf{F}^{\text{equi}} \\ -w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_4w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_4w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_4w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & w_4w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{equi}} & -w_4\nabla_{\boldsymbol{\lambda}}\mathbf{F}^{\text{equi}} \\ \nabla_{\mathbf{u}}\mathbf{F}^{\text{suppl}} & -w_1\nabla_{\mathbf{u}}\mathbf{F}^{\text{suppl}} & -w_2\nabla_{\mathbf{u}}\mathbf{F}^{\text{suppl}} & -w_3\nabla_{\mathbf{u}}\mathbf{F}^{\text{suppl}} & -w_4\nabla_{\mathbf{u}}\mathbf{F}^{\text{suppl}} & \nabla_{\boldsymbol{\lambda}}\mathbf{F}^{\text{suppl}} \end{bmatrix}, \tag{67}$$

which is evaluated according with the contact status of the slave node belonging to the contact element. Note that the matrix only contains the gradients of the two sub-operators involved in (62). When the contact element is created using the triangular Nagata patch (three master nodes), the penultimate row and column are removed from the contact Jacobian matrix. Taking into account (67) and the mixed contact operators (63)–(65), the contact Jacobian matrices can be explicitly evaluated for each contact status (gap, stick and slip). The elemental Jacobian matrix for the gap status is obtained from (63) and it is written as:

$$\mathbf{J}^{*}_{\text{gap}}(\mathbf{u},\boldsymbol{\lambda})=\begin{bmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & -(1/\varepsilon)\mathbf{I} \end{bmatrix}, \tag{68}$$

$$\mathbf{J}^{*}_{\text{stick}}(\mathbf{u},\boldsymbol{\lambda})=\begin{bmatrix} \varepsilon\mathbf{I} & -w_1\varepsilon\mathbf{I} & -w_2\varepsilon\mathbf{I} & -w_3\varepsilon\mathbf{I} & -w_4\varepsilon\mathbf{I} & \mathbf{I} \\ -w_1\varepsilon\mathbf{I} & w_1w_1\varepsilon\mathbf{I} & w_1w_2\varepsilon\mathbf{I} & w_1w_3\varepsilon\mathbf{I} & w_1w_4\varepsilon\mathbf{I} & -w_1\mathbf{I} \\ -w_2\varepsilon\mathbf{I} & w_2w_1\varepsilon\mathbf{I} & w_2w_2\varepsilon\mathbf{I} & w_2w_3\varepsilon\mathbf{I} & w_2w_4\varepsilon\mathbf{I} & -w_2\mathbf{I} \\ -w_3\varepsilon\mathbf{I} & w_3w_1\varepsilon\mathbf{I} & w_3w_2\varepsilon\mathbf{I} & w_3w_3\varepsilon\mathbf{I} & w_3w_4\varepsilon\mathbf{I} & -w_3\mathbf{I} \\ -w_4\varepsilon\mathbf{I} & w_4w_1\varepsilon\mathbf{I} & w_4w_2\varepsilon\mathbf{I} & w_4w_3\varepsilon\mathbf{I} & w_4w_4\varepsilon\mathbf{I} & -w_4\mathbf{I} \\ \mathbf{I} & -w_1\mathbf{I} & -w_2\mathbf{I} & -w_3\mathbf{I} & -w_4\mathbf{I} & \mathbf{0} \end{bmatrix}, \tag{69}$$

which considers that the local frame defined on the master surface is fixed in all Newton iterations within an increment. This assumption is admissible since the solution obtained for the stick status is path-independent [142]. Finally, the elemental Jacobian matrix for the slip status, assuming a fixed local frame attached to the Nagata patch, is defined by:

$$\mathbf{J}^{*}_{\text{slip}}(\mathbf{u},\boldsymbol{\lambda})=\begin{bmatrix} \varepsilon\mathbf{M} & -w_1\varepsilon\mathbf{M} & -w_2\varepsilon\mathbf{M} & -w_3\varepsilon\mathbf{M} & -w_4\varepsilon\mathbf{M} & \mathbf{M} \\ -w_1\varepsilon\mathbf{M} & w_1w_1\varepsilon\mathbf{M} & w_1w_2\varepsilon\mathbf{M} & w_1w_3\varepsilon\mathbf{M} & w_1w_4\varepsilon\mathbf{M} & -w_1\mathbf{M} \\ -w_2\varepsilon\mathbf{M} & w_2w_1\varepsilon\mathbf{M} & w_2w_2\varepsilon\mathbf{M} & w_2w_3\varepsilon\mathbf{M} & w_2w_4\varepsilon\mathbf{M} & -w_2\mathbf{M} \\ -w_3\varepsilon\mathbf{M} & w_3w_1\varepsilon\mathbf{M} & w_3w_2\varepsilon\mathbf{M} & w_3w_3\varepsilon\mathbf{M} & w_3w_4\varepsilon\mathbf{M} & -w_3\mathbf{M} \\ -w_4\varepsilon\mathbf{M} & w_4w_1\varepsilon\mathbf{M} & w_4w_2\varepsilon\mathbf{M} & w_4w_3\varepsilon\mathbf{M} & w_4w_4\varepsilon\mathbf{M} & -w_4\mathbf{M} \\ \mathbf{M} & -w_1\mathbf{M} & -w_2\mathbf{M} & -w_3\mathbf{M} & -w_4\mathbf{M} & \dfrac{1}{\varepsilon}(\mathbf{M}-\mathbf{I}) \end{bmatrix}, \tag{70}$$

where the second-order tensor $\mathbf{M}$ is defined as:

$$\mathbf{M}=(\mathbf{n}-\mu\mathbf{t})\otimes\mathbf{n}+\rho(\mathbf{I}-\mathbf{n}\otimes\mathbf{n}-\mathbf{t}\otimes\mathbf{t}), \tag{71}$$

which is independent of the master surface curvature. Its definition includes the unit slip vector defined in (66), the normal vector of the master surface at the closest point and a projection scaling factor defined by:

$$\rho=-\mu\hat{\lambda}_{\text{n}}\Big/\left\|\hat{\boldsymbol{\lambda}}_{\text{t}}\right\|, \tag{72}$$

which is always in the range between zero and one. Although Jacobian matrix (70) considers a fixed local frame, it is updated in each iteration of the Newton's method. In opposite to the other contact status, the elemental Jacobian matrix defined for the slip status is not symmetric, excluding for the frictionless contact case. The Jacobian matrix for slip status that takes into account the contributions of geometric terms arising from the gradient of the normal vector was developed by Heege and Alart [21], for contact between deformable and rigid bodies. Nevertheless, since the local frame defined by the surface normal vector is updated within the iterative solution, the convergence rate is not degraded using the Jacobian matrix (70).

In order to exemplify the structure of the global tangent matrix after the incorporation of the contact elements, a simple 2D example is presented [129]. The mechanical system involves two discretized deformable bodies coming in frictional contact, as shown in Fig. 37. The upper body is assigned as slave body and the lower one as master, each one discretized by a single structural element. A constant pressure is imposed on the upper segment between nodes 1 and 3. Besides, nodes 6 and 7 are fixed, thus they are not included in the global system of equations. In absence of contact, the global tangent matrix and the residual vector for such configuration have the pattern presented in Fig. 38a. The contribution of each body to the tangent matrix is illustrated using two different colours, highlighting the geometrical independence of the bodies.

The frictional contact between the bodies is introduced by means of the contact element (green), which is defined by the slave node 2, the master segment joining nodes 4 and 5, and a complementary node 8 for the Lagrange multipliers. The selection of the master segment (nodes 4 and 5) that defines the contact element together with the salve node is performed by the contact search algorithm. Indeed, from the geometrical point of view, the slave node 2 may come in contact with the master segment located between nodes 4 and 5 (existence of the normal projection), as presented in Fig. 37b. The contribution of the contact element to the global tangent matrix and residual

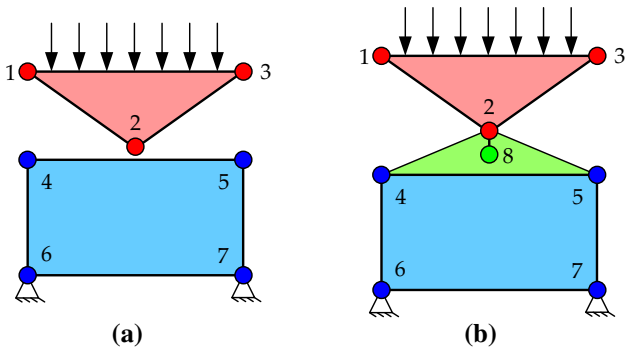

Fig. 37 Example of two discretized deformable bodies coming in contact for the cases: **a** absence of contact elements; **b** with a contact element

vector is shown in Fig. 38b by means of the green colour. Since the considered example is 2D, the contact element is composed by four nodes (including the artificial one), leading to a contact Jacobian matrix (67) with four rows/columns. The sixteen elements of the Jacobian matrix are added to the global tangent matrix taking into account the four nodes that define the contact element, as shown in Fig. 38b. The internal force vector corresponding to the contact element is introduced in the global residual vector, connecting the involved nodes, as shown in Fig. 38b. The pattern of nonzero entries arising in the global tangent matrix is symmetric (see Fig. 38b), allowing solve the linear system of equations considering the matrix symmetrically structured (DSS from Intel MKL library).

The large sliding between the bodies under frictional contact requires the update of the Nagata patch associated with the slave node according to the normal projection of the slave node on the master surface. This procedure increases the computational cost associated with the contact search and involves the update of the global tangent matrix pattern [51]. In fact, this comprises invoking each Intel MKL DSS interface routine once, as shown in Fig. 13, i.e. delete old and create new all data structures. Note that the global tangent matrix presents a fixed pattern when the master surface is assumed rigid, independently of the Nagata patch that defines the contact element. Since the

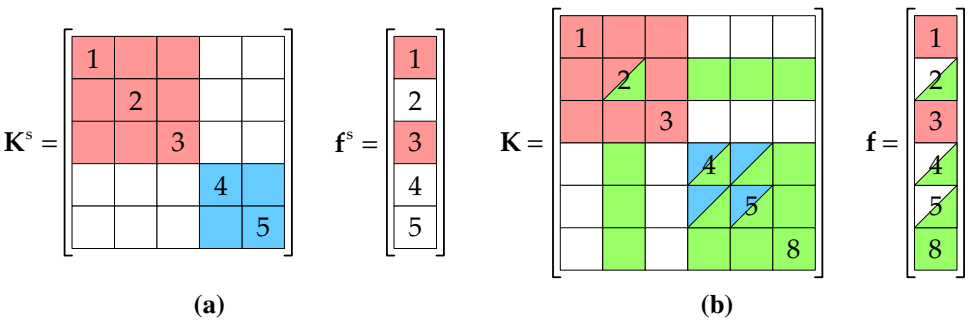

Fig. 38 Pattern of the global tangent matrix and residual vector of two discretized bodies coming in contact: **a** absence of contact elements; **b** with a contact element

update of the nonzero pattern of the global tangent matrix is computationally expensive, two distinct approaches can be applied in large sliding contact problems: (i) Extension of the parametric (local) domain of the Nagata patch (see Fig. 39a) and (ii) creation of multi-face contact elements in advance, i.e. incorporation of the adjacent Nagata patches into the contact element definition (see Fig. 39b). The first approach yields a continuous sliding of the slave node along the extension of the actual patch. The equilibrium is reached for the slave node located outside the patch domain, which results either in penetration, for a locally concave master surface, or in gap opening, in case of a convex surface. The second approach uses the multi-face contact elements proposed by Heegaard and Curnier [39], which consists in creating contact elements composed by one slave node and several master patches. This strategy avoids the frequent updating of the contact elements and allows dealing with large sliding in a single increment.

7 Numerical Examples

Several numerical examples are presented in order to validate the accuracy and effectiveness of the presented surface smoothing procedure. The selected problems are divided into three groups: (i) contact between deformable and rigid bodies; (ii) contact between deformable bodies and (iii) self-contact, all of them modelled in 3D framework. The accuracy, robustness and performance of the surface smoothing method is demonstrated by means of its comparison with the traditional piecewise bilinear finite element mesh representation. The numerical simulations are performed with the in-house static implicit finite element code DD3IMP [143], which has been specifically developed to simulate sheet metal forming processes [144]. The evolution of the deformation process is described by an updated Lagrangian scheme. Several advanced constitutive models and work hardening laws are currently implemented [145]. In order to improve the convergence in the iterative procedure, an explicit approach is used to calculate the trial solution, which is adjusted using a generalization of the $r_{\min}$ strategy [146]. In order to improve computational performance, some high performance computing techniques have been incorporated to take advantage of multi-core processors, namely OpenMP directives in the most time consuming branches of the code [69]. All numerical simulations are carried out on a computer machine equipped with an Intel Core^{TM} i7–2600 K Quad-Core processor (3.4 GHz), 8.0 GB RAM and the Windows 7 Professional (64-bit platform) operating system.

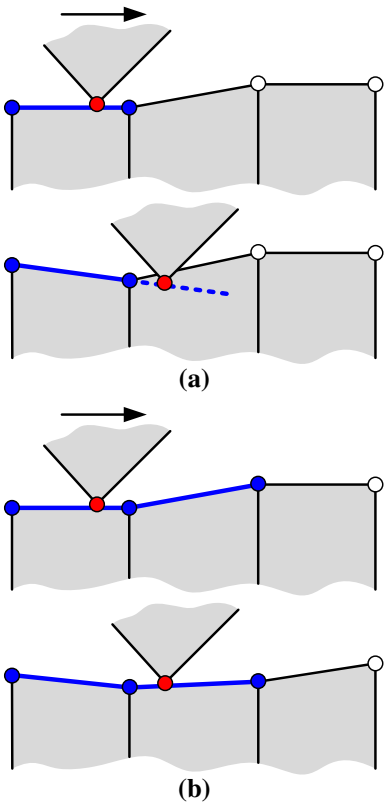

Fig. 39 Example of large sliding contact: **a** extension of the master segment domain; **b** multi-face contact element (adapted from [51])

7.1 Contact Between Deformable and Rigid Bodies

7.1.1 Frictional Ironing Problem

The first numerical example presented is the so-called frictional ironing problem, which was adapted from the one proposed by Tur et al. [147], where the contact arises between two deformable bodies using the mortar formulation. The present example comprises the sliding of a rigid indenter along an elastic slab, as illustrated in Fig. 40. The idea behind this example is to show the performance of the surface smoothing procedure in problems with large deformation and sliding. The dimensions of the bodies and the material properties of the slab are indicated in Fig. 40. The friction between the indenter and the elastic slab is

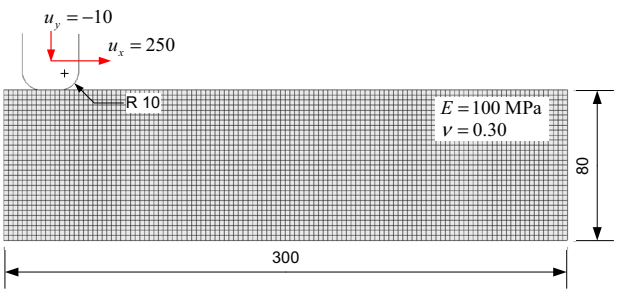

Fig. 40 Initial configuration of the frictional ironing problem with finite element mesh of the slab

modelled by the Coulomb's law considering $\mu = 0.3$. A downward displacement of $u_y = -10$ mm is applied to the rigid indenter, followed by a horizontal displacement $u_x = 250$ mm. Although the performed simulation is *quasi-static*, load steps are defined as a function of time for the sake of presentation of the results. From 0 to 1 s, the indenter is moved vertically towards the slab, while from 1 to 2 s, the indenter is displaced horizontally.

The deformable slab is discretized with 3600 hexahedral finite elements (see Fig. 40), while the surface of the rigid indenter is described by: (i) bilinear facets and (ii) Nagata patches, as shown in Fig. 41. Each circular arc resulting from the fillet operation is discretized with 4 finite elements. The faceted description of the rigid surface is defined by 64 quadrilateral facets, while the smoothed description is composed by 26 quadrilateral Nagata patches. Plane strain conditions are assumed for the numerical simulation, considering 1 mm of thickness for the elastic slab.

The evolution of the total horizontal and vertical force components between the contacting bodies during the simulated period is presented in Fig. 42, comparing the faceted and smoothed surface description methods. Concerning the description of the rigid surface using the classical faceted finite elements (Fig. 41a), the chatter effect in the predicted force during the sliding is related with the accuracy in the surface representation. The maximum value of radial error is approximately 1.9 % in the fillet radius (see Fig. 28a), while the discontinuity of the normal vector between adjacent facets is larger than 10° (see Fig. 28b). On the other hand, the application of the surface smoothing method (Fig. 41b) yields important improvements in the predicted force evolution, because both the radial and the normal vector errors are negligible (see Fig. 28). The reduction of the chatter effect in the horizontal force component is highlighted in Fig. 43. The periodic oscillations observed in the force evolution, when using the surface smoothing method, are exclusively related with the discretization of the deformable slab. The improvement in the rigid surface representation increases the vertical force value and reduces the horizontal force value, as shown in Fig. 42.

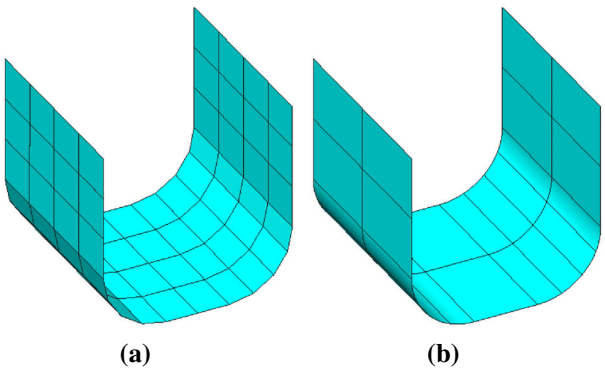

Fig. 41 Rigid indenter described by: **a** bilinear facets; **b** Nagata patches

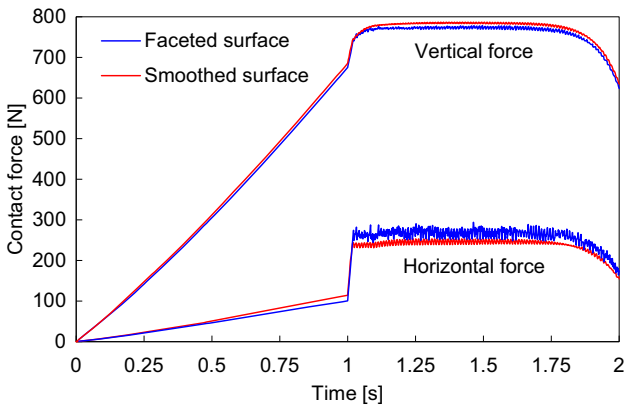

Fig. 42 Vertical and horizontal contact force components for the frictional ironing problem (rigid indenter)

7.1.2 *Deep Drawing of a Square Cup*

This example involves the numerical simulation of the deep drawing process used to produce a square cup, proposed as benchmark at the Numisheet 1993 conference [148]. The blank is a 150 × 150 × 0.78 mm sheet, of a mild steel which is modelled considering its elastoplastic mechanical behaviour, with the elastic and plastic properties listed in Table 1. The material follows Hooke's law in the elastic region, while in the plastic domain the hardening is described Swift law. The orthotropic plastic behaviour of the steel blank is described by Hill'48 quadratic yield criterion using the parameters presented in Table 1. The friction coefficient between the sheet and the rigid tools is taken from the benchmark specifications as $\mu = 0.144$. The

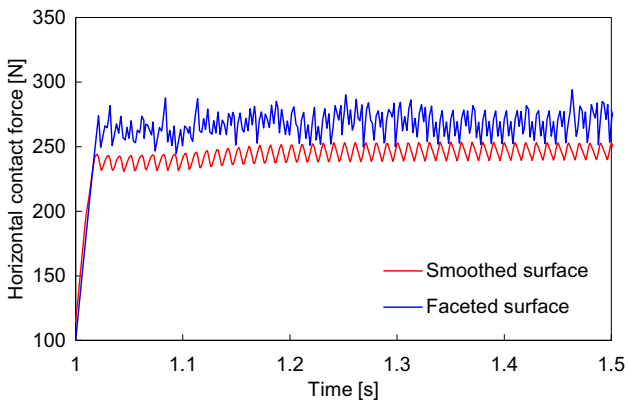

Fig. 43 Detail of the chatter effect observed in the horizontal contact force of the frictional ironing problem (rigid indenter)

Table 1 Elastic and plastic material properties of the mild steel used in the deep drawing of a square cup

Elastic properties		Swift hardening law			Hill'48 yield criterion			
E [MPa]	ν	σ_0 [MPa]	K [MPa]	n	F	G	H	N
206	0.3	157.12	565.32	0.2589	0.283	0.358	0.642	1.289

total punch displacement is 40 mm, while the blank holding force is 19.6 kN.

Considering the geometrical and material symmetry of the problem, only a quarter of the entire geometry is simulated, as shown in Fig. 44. The blank is discretized with 7200 hexahedral finite elements, using a structured mesh defined by 60 finite elements in each direction and 2 layers through the thickness. The surface of the forming tools is described using three different methods [149], as shown in Fig. 44. The faceted surface description uses 10 bilinear facets to define each circular arc, leading to a total of 1924 bilinear facets (see Fig. 44a). The model composed by Nagata patches uses only 2–3 patches to describe each circular arc, as illustrated in Fig. 44b, leading to a total of 492 Nagata patches, i.e. about 25 % of the total number of facets. The last approach uses 20 Bézier patches in the surface description, which are obtained with the aid of a specific CAD package [144]. Note that the required finite element meshes of the tool surfaces are generated from this CAD model, which is composed by 20 Bézier patches.

The comparison of the predicted punch force evolution using different tool surface description methods is presented in Fig. 45. The effect of the tool surface accuracy can be observed in the non-physical oscillations arising in the contact force. In fact, only the surface smoothing method with Nagata patches and the tools defined by Bézier patches lead to a force evolution with insignificant oscillations, as shown in the zoom view included in Fig. 45. On the other hand, the faceted tool surface description leads to significant oscillations in the contact force, due to the discontinuity of the surface normal vector field. The amplitude of the oscillations is larger after 20 mm of punch displacement since the relative sliding between the sheet and the tools is higher.

The evolution of the number of slave nodes in contact with the die surface for each tool surface description method is presented in Fig. 46. The number of nodes in contact predicted by model with the tool surfaces modelled by Bézier patches is globally higher due to the excellent geometrical accuracy in the surfaces definition (at least C^1 continuity). On the other hand, the piecewise bilinear representation of the tool surfaces provides a lower number of nodes in contact due to the artificial roughness induced in the contact surface, particularly between 5 and 20 mm of the punch displacement, where the contact occurs with the die radius involving small sliding. The evolution of the number of nodes in contact, adopting the smoothing procedure with Nagata patches, is identical to the ones provided by the model using Bézier patches, as shown in Fig. 46.

The computational performance of the numerical simulation is evaluated in this study through the number of increments and the computational time, which are presented in Table 2. The number of increments is lowest when the geometrical description of the tools with Bézier patches is adopted. Indeed, the smoothing with Nagata patches leads to a slight increase of the required number of increments. On the other hand, the faceted tool surface description method involves an increase of approximately 20 % for the required number of increments. Figure 47 presents the increment size evolution with the punch displacement. The reduction in the increment size defined by the $r_{\min}$ strategy [150] is more emphasised for the faceted tool surface description method, predominantly after 20 mm of punch displacement. Concerning the Bézier and the Nagata patch surface description methods, its increment size evolution is analogous, as shown in Fig. 47. The computational time is mainly dictated by the number of

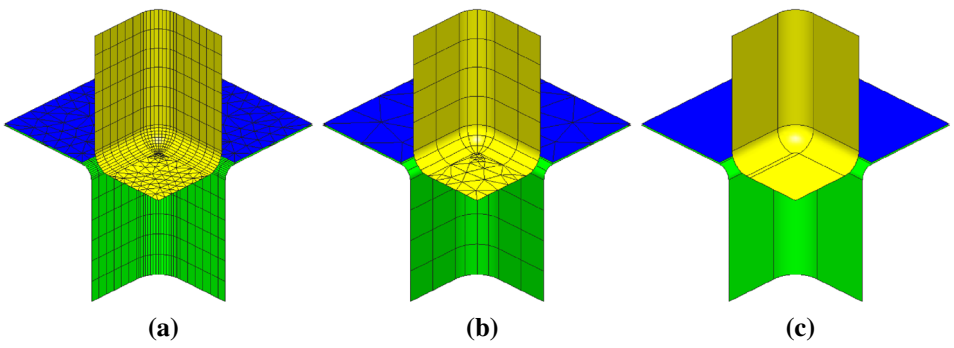

Fig. 44 Geometrical description of the forming tools using: **a** faceted mesh; **b** Nagata patches; **c** Bézier patches

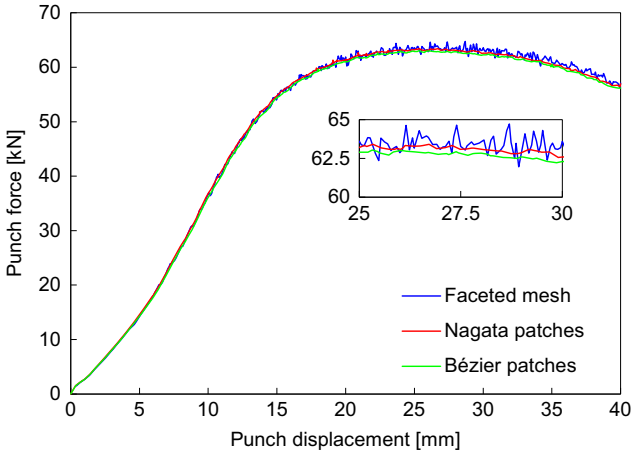

Fig. 45 Punch force evolution for different tool surface description methods and zoom of the chatter effect in the force

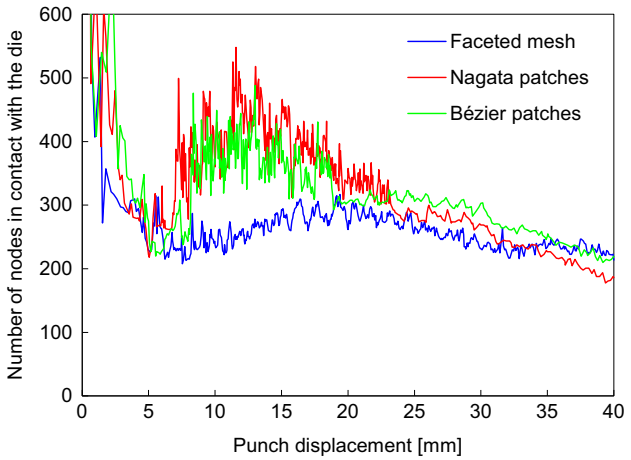

Fig. 46 Evolution of the number of slave nodes in contact with the die for different tool surface description methods

Table 2 Computational performance of the deep drawing problem using different tool surface description methods

	Faceted mesh	Nagata	Bézier
No. of increments	537	451	428
Computational time (s)	3020	2381	2354

increments required to perform the simulation. Therefore, the computational time required to carry out the numerical simulation using the faceted tool surface description is approximately 25 % higher than the one required for the smooth surface description, as shown in Table 2. Although the total number of Nagata patches used in tool surface description is higher than the number of Bézier patches (see Fig. 44), the computational time is identical for both cases.

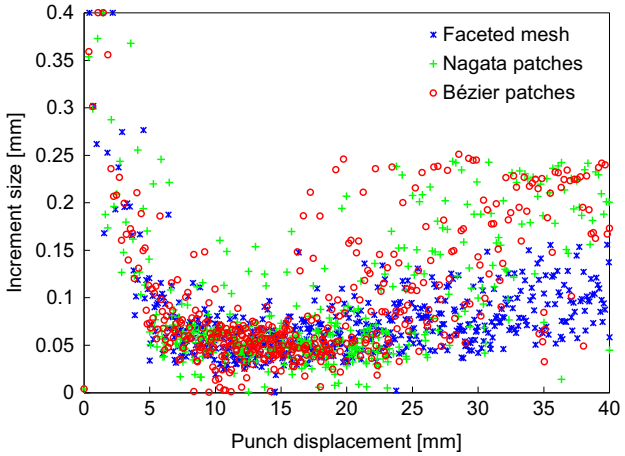

Fig. 47 Increment size evolution with the punch displacement

7.2 Contact Between Deformable Bodies

7.2.1 Disk Embedded in a Bored Plate

This example comprises an elastic disk and an elastic infinite plate with a hole of almost the same radius of the disk. Both the disk and the plate have unitary thickness. The disk is pressed by a concentrated load, situated in its centre, against the cylindrical bore drilled in the infinite plate, as illustrated in Fig. 48. The main dimensions of the two bodies are listed in the same figure and the Coulomb's friction law is $\mu = 0.4$. The material properties of the disk and the plate are identical ($E = 210$ GPa, $\nu = 0.3$). The value of the concentrated force is chosen such that the contact occurs at one third of the interface between the hole and the disc (half contact angle $\alpha = 60°$). The purpose of this example is to evaluate the friction shear stress distribution at the interface, as well to determine the angle for which the transition between the stick and the slip zones occurs.

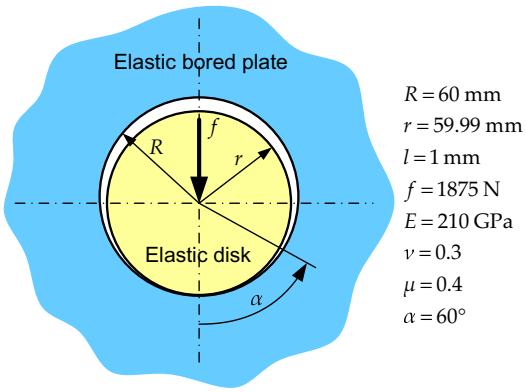

Fig. 48 Thin elastic disk embedded in a thin elastic infinite plane with a circular hole

The problem is modelled as 3D using plane strain conditions, i.e. the out-of-plane displacements are fixed on both faces, as previously performed by other authors [38]. Due to the symmetry conditions, only one half of the disk and the plate are modelled. The finite element model of the problem is shown in Fig. 49. The inner disk is discretized by 1896 finite elements using a single layer through the thickness, while the plate is defined by 3228 finite elements. The external radius of the bored plate used in the finite element model is 10 times higher than its internal radius, allowing taking into account the half space influence. Besides, the nodal displacements are fixed in all nodes of the periphery (Fig. 49). The finite element mesh is refined in the contact area to attain an accurate friction stress distribution. Moreover, the mesh is generated using the same number of finite elements in the circumferential direction of the disk and the hole of the plate, i.e. each slave node is located close to a master node. The interface between the contacting bodies is modelled by a set of Node-to-Nagata contact elements, where the disk and the bored plate are chosen as the slave and the master, respectively.

The semi-analytical solution of this problem, found by Klang [151], involves integrals that cannot be expressed in closed form and have to be evaluated numerically. The semi-analytical shear stress distribution for the selected material and geometrical data is presented in Fig. 50 (replicated from [38]). In the presence of friction, the contact surface is divided into an inner stick region and an outer slip region, easily identified in the shear contact stress distribution taking into account that the normal pressure distribution is nearly elliptical. The semi-analytical solution is independent of the loading rate. Nevertheless, the predicted friction stress distribution is very sensitive to the contact point location, provided by the local search algorithm. Therefore, previous studies shown that the numerical solution is strongly dependent on the number of

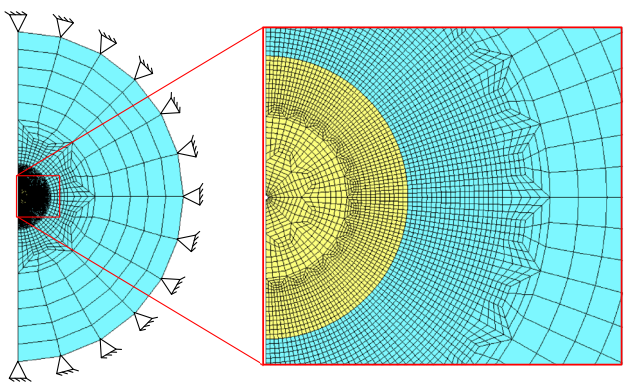

Fig. 49 Finite element mesh (10,626 nodes, 74 active slave nodes in the interface) and zoom of the contact region

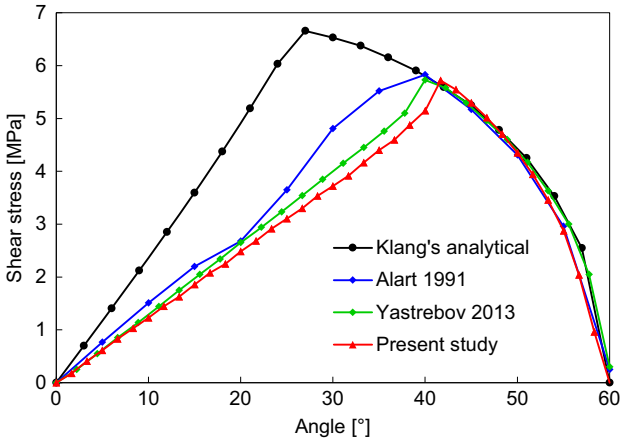

Fig. 50 Comparison between semi-analytical and numerical solutions for the shear stress distribution in the contact surface

increments adopted to apply the load, particularly the shear stress distribution in the stick zone [51]. The numerical solution reported by Yastrebov [51] was obtained using 100 increments, adopting the classical NTS contact discretization, while the solution provided by Alart and Curnier [38] adopts the NTN contact discretization. In the present model, the external load is applied in 100 increments with equal size.

The comparison between semi-analytical and numerical solutions for the shear stress distribution in the contact surface is presented in Fig. 50. The three presented numerical solutions are in good agreement, particularly for the estimate of the angle for which the division between stick and slip regions occurs (corresponding to the maximum shear stress value). Nevertheless, the value for this angle is substantially lower in the semi-analytical solution (approximately 27°), while the finite element solutions predict a value of about 40° (see Fig. 50). The shear stress distribution in the slip region is in good agreement with the semi-analytical solution, which indicates the accurate prediction of the normal contact pressure distribution. On the other hand, the shear stress estimate provided by the present finite element model in the stick region is slightly lower than the others numerical solutions. However, its linear distribution in the stick region is accurately predicted.

The numerical solution of this problem is strongly dependent on the number of increments due to the friction at the interface. The direction and magnitude of the friction force is dictated by the location of the contact point (on the master surface), which defines the sliding path. Since the sliding path of each slave node over the master surface is inferior to 0.0001 mm (very small sliding), the success of the simulation is defined by the precision achieved in the calculation of the contact point, which is evaluated by

means of the local search algorithm. The nodal contact forces distribution in the slave nodes is presented in Fig. 51. In order to obtain an accurate solution, the increase of the external load should be performed such that only one new contact element is activated in each load increment, i.e. the normal gap immediately before the contact should be significantly lower than the predicted penetration [23].

7.2.2 Frictional Ironing Problem

This example is the extension of the problem presented in Sect. 7.1.1, considering the elastic deformation of the indenter, as proposed for instance, in [147, 152]. The material parameters of the contacting bodies are given in Fig. 52, where the indenter is 10 times stiffer than the slab. The friction coefficient between the indenter and the slab is $\mu = 0.3$. The problem is solved under the plane strain assumption. The deformable indenter is discretized with 112 hexahedral finite elements, while the discretization of the slab is the one previously adopted in Sect. 7.1.1 (see Fig. 52). A downward displacement of $u_y = -10$ mm is applied to the upper line of the indenter, followed by a horizontal displacement $u_x = 250$ mm. Since the indenter is 10 times stiffer than the slab, it is chosen as master body.

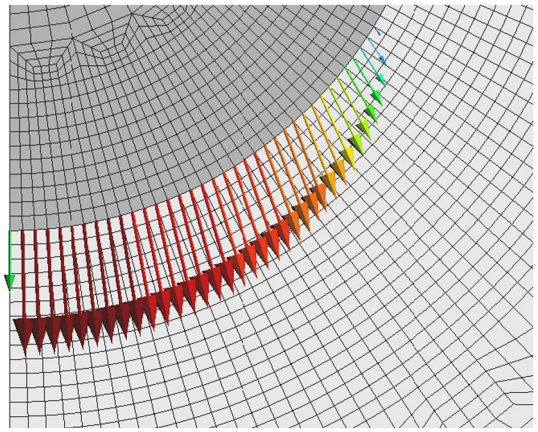

Fig. 51 Deformed configuration and nodal contact forces distribution in the slave nodes of the disk

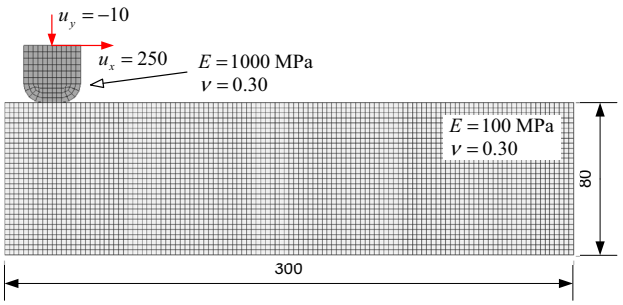

Fig. 52 Initial configuration of the frictional ironing problem with finite element discretization of the bodies

The predicted deformed geometry of the contacting bodies is presented in Fig. 53 for three different instants, considering the master surface (indenter) smoothed with Nagata patches. The first instant presented corresponds to the end of the prescribed vertical displacement, while the last instant corresponds to the end of the horizontal sliding. The contour plot of shear stress is presented in Fig. 53 for the same three different instants, showing the typical distribution [51]. The large sliding between the indenter and the slab requires the update of the Node-to-Nagata contact elements. Therefore, the nonzero pattern of the global tangent matrix need to be update during the incremental procedure. In order to reduce the amount of involved updates, the creation of multi-face contact elements is performed, creating contact elements composed by one slave node and several master patches. Besides, the parametric domain of the Nagata patches is slightly extended in order to eliminate the *deadzones* arising during the local search.

The evolution of the horizontal and vertical reaction force components, computed at the top of the indenter, is presented in Fig. 54, comparing the faceted and smoothed surface description methods. The chatter effect in the predicted force during the sliding is significantly reduced when adopting the surface smoothing method in the description of the master surface. In addition, the value of both force components is reduced through the smoothing procedure with Nagata patches, as shown in Fig. 54. On the

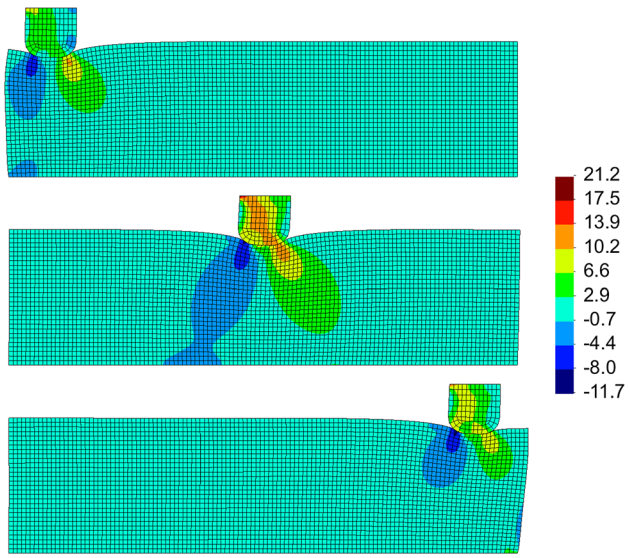

Fig. 53 Deformed configuration for the frictional ironing problem at three different instants: end of the prescribed vertical displacement (*top*) and end of the horizontal sliding (*bottom*). Contour plots of shear stress in MPa

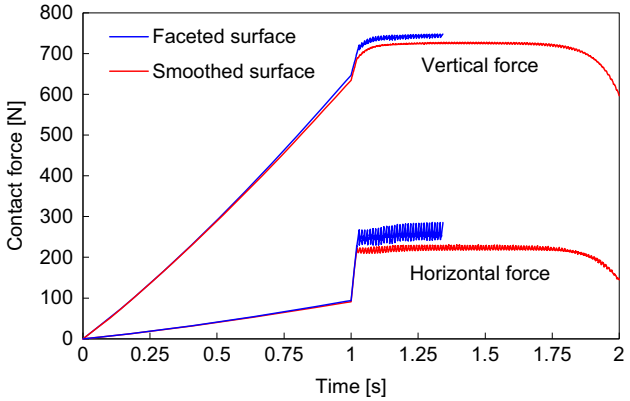

Fig. 54 Vertical and horizontal reaction force components for the frictional ironing problem

other hand, in case of rigid indenter (Sect. 7.1.1), the magnitude of the vertical force increases slightly adopting the surface smoothing procedure, as shown in Fig. 42. Considering the elastic deformation of the indenter, the numerical simulation using the faceted description of the master surface leads to severe convergence problems at 85 mm of horizontal prescribed displacement, causing the divergence of the iterative procedure (see Fig. 54).

Although the indenter (master) presents elastic deformation, the accuracy in its surface representation is similar to the one discussed in Sect. 7.1.1 concerning the rigid indenter. Therefore, the surface smoothing method yields important improvements in the predicted force evolution. The chatter effect in the horizontal force component is shown in Fig. 55. The comparison with the force evolution predicted considering the rigid indenter (see Fig. 43) allows concluding that the periodic oscillations observed in the force evolution using the smoothed surface are exclusively related with the discretization of the slab (slave body). The application of the surface smoothing procedure allows to reduce considerably the amplitude of the force oscillations and improves the convergence rate due to the smooth surface normal vector field.

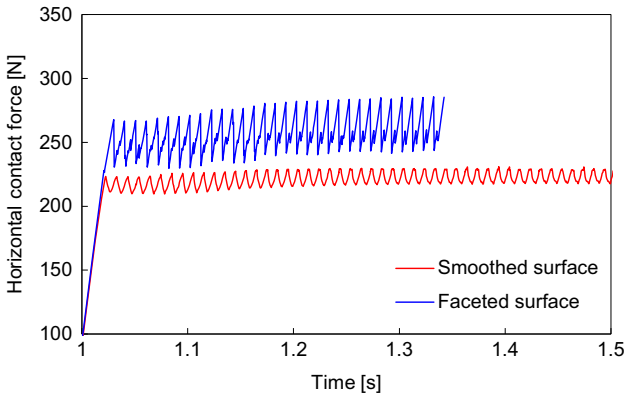

Fig. 55 Detail of the chatter effect observed in the horizontal contact force of the frictional ironing problem

The deformed geometry of the contacting bodies and the nodal contact forces are depicted in Fig. 56, for the time instant corresponding to 50 mm of horizontal prescribed displacement. The contact forces distribution is according to the common understanding of the system mechanics. Besides, the angle between the contact force vectors and the normal to the master surface seems to be very close to the friction angle of 17°, corresponding to the given friction coefficient. Due to the penetration of the master nodes into the slave body when using the faceted surface description (Fig. 56a), the magnitude of the predicted reaction force components is larger using this surface description, as shown in Fig. 54.

7.2.3 Extrusion of an Aluminium Billet

This example involves the extrusion of an aluminium billet into a deformable conical die. It has been previously studied by Padmanabhan and Laursen [24] in a 2D setting (axisymmetric model), to demonstrate the advantages of using a smoothing scheme on the discretized master contact surface. In the present study, the problem is modelled in a 3D setting, the contact interface is treated as a surface instead of a curve. The cylindrical billet is drawn through a conical die under frictionless contact, as shown in Fig. 57.

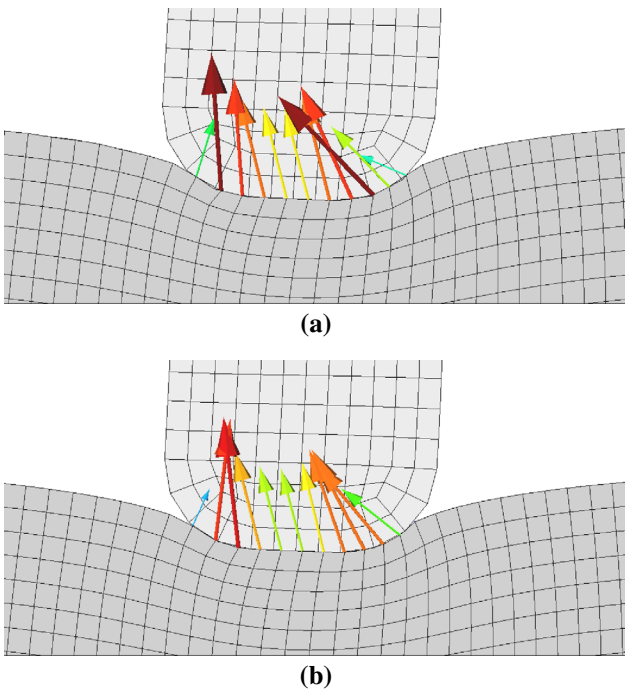

Fig. 56 Nodal contact forces for 50 mm of horizontal prescribed displacement (magnitude denoted by *arrow size* and *colour*) considering: **a** faceted surface; **b** smoothed surface. (Color figure online)

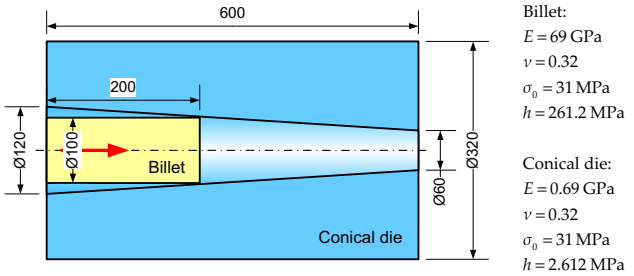

Fig. 57 Extrusion of an aluminium billet in a conical die including geometrical and material properties (dimensions in mm)

The backside of the billet is subjected to an axial displacement (150 mm) into the conical die, while the outer boundary (exterior radius) of the die is fixed in all directions. The dimensions of each body are indicated in Fig. 57. Both bodies are modelled with an elastoplastic material behaviour, adopting the linear isotropic hardening law given by $\sigma = 31 + 261.2\bar{\varepsilon}^{p}\ [MPa]$ for the cylindrical billet and the hardening law given by $\sigma = 31 + 2.612\bar{\varepsilon}^{p}\ [MPa]$ for the conical die. The elastic material properties were selected in order to ensure considerable deformation of both bodies [24].

Due to symmetry conditions, only one quarter of the problem is modelled. The finite element mesh of both deformable bodies is presented in Fig. 58, where the billet is discretized by 540 finite elements and the conical die is defined by 840 finite elements. The element size in the axial direction of the billet is half the size used to discretize the die. The billet is defined as slave body while the conical die is assigned as master. Since the master surface (conical surface of the die) is convex, some convergence problems can arise in the numerical solution when using the faceted description of the master surface, namely when the slave and master nodes are very close. Therefore, the finite element mesh of the contact surfaces is created using six elements in the circumferential direction of the billet and seven elements for the die, as shown in the detail of Fig. 58. This configuration allows avoiding the convergence problems related with the flip–flop effect in the circumferential direction. The displacements of the nodes located in the exterior radius of the die are fixed, while the axial displacement is applied incrementally in the backside of the billet.

The evolution of the axial force acting on the backside of the billet as a function of its displacement is presented in Fig. 59, for both descriptions of the master surface (faceted and smoothed). The abrupt oscillations in the axial force for the frictionless extrusion using the master surface described by bilinear facets are reduced through the smoothing procedure with Nagata patches. The chatter effect is produced by the sudden changes in the surface normal between adjacent master facets, when the slave nodes slide along the axial direction. The value of axial force predicted by the model with faceted master surfaces is slightly higher than using the smoothed description of the surface (see Fig. 59). This results from the selection of the convex surface as master, which dictates a higher value of contact pressure due to the artificial gap between the billet and the die surface.

The equivalent plastic strain distribution plotted in the deformed configuration is presented in Fig. 60, comparing faceted and smoothed master surface descriptions. The maximum value of equivalent plastic strain occurs in the

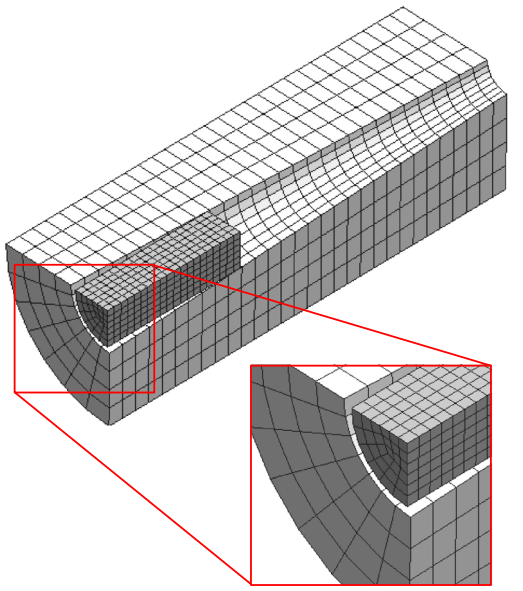

Fig. 58 Finite element mesh of the billet (777 nodes) and the conical die (1240 nodes) with detail of the discretization in the circumferential direction

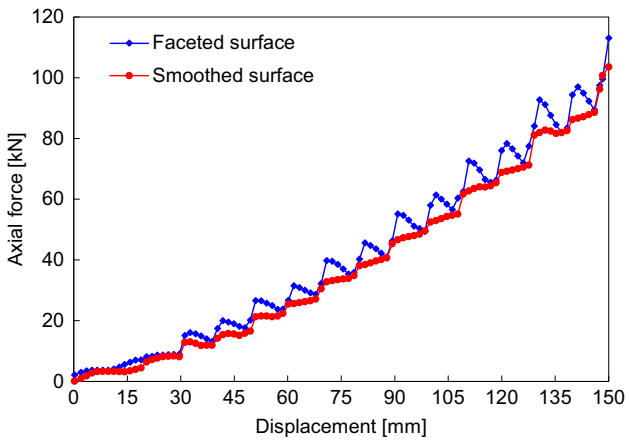

Fig. 59 Axial force evolution as function of the displacement of the billet for both faceted and smoothed master surface description methods

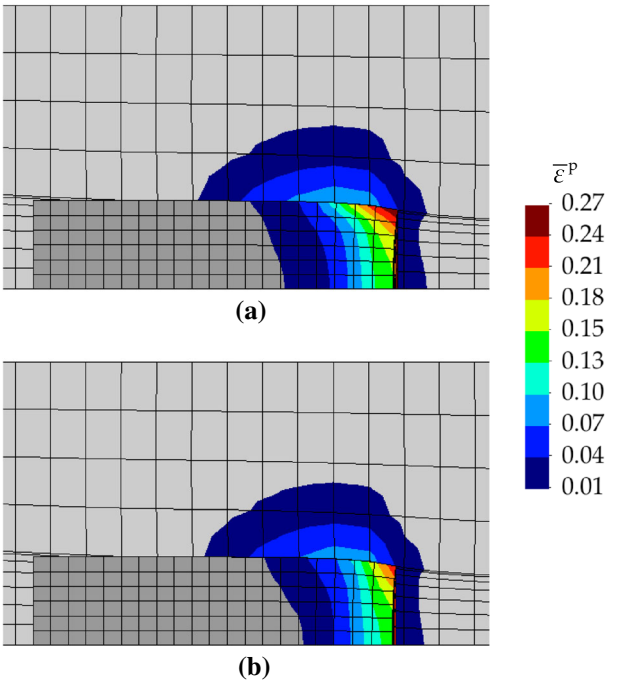

Fig. 60 Equivalent plastic strain distribution for the extrusion problem plotted in the fully deformed configuration: **a** faceted master surface; **b** smoothed master surface

periphery of the billet (forward side). Its value is approximately 27 % for the faceted description of the master surface and 23 % considering the surface smoothed with Nagata patches. The plastic strain in the conical die is only located near the contact interface (see Fig. 60), where the nodal contact forces attain its maximum value. The computational performance of the frictionless extrusion problem is identical for both surface description methods (faceted and smoothed). The total computational time is about 67 s, using 100 increments to impose the axial displacement. Although the sudden changes in the surface normal vector are reduced when using Nagata patches in the definition of the contact elements, the computational cost is not affected. Nevertheless, the accuracy achieved in the evaluation of the contact forces is clearly improved, as highlighted in Fig. 59.

7.3 Self-Contact

7.3.1 Post-buckling of a Thin Walled Tube

The *quasi*-static post-buckling of a thin walled tube is analysed. This axisymmetric example involves large plastic deformations, post-buckling and self-contact phenomena. This problem was firstly introduced by Laursen and Simo [10] in 2D setting (axisymmetric finite element model). More recently, a simplified version has been studied by Yang and Laursen [95] to demonstrate the effectiveness of the mortar based method to deal with self-contact phenomena, undergoing large deformations and sliding. The bottom surface of the thin walled tube is fixed in all directions, while a vertical displacement is applied to the top surface of the tube (40 mm), as shown in Fig. 61a. The tube dimensions are given in Fig. 61a, as well as the elastoplastic material properties (linear isotropic hardening). Frictionless response is assumed in the self-contacting buckle regions of this problem. Since the application of the prescribed axial displacement (40 mm) in the top surface of the tube is equivalent to apply 20 mm of displacement in each face of the tube (top and bottom), it allows to model only half length of the tube. Thus, due to the symmetry conditions, only one eighth of the tube is simulated, as shown in Fig. 61b, which is discretized with 240 finite elements, using a single layer through the thickness. This finite element mesh is identical to the one adopted in [95], which is denoted by coarse mesh due to the reduced number of elements.

In post-buckling problems it is difficult to predict in advance which portions of a surface will come into contact with each other. Therefore, the contact pairs required for the master–slave discretization are unknown a priori (a

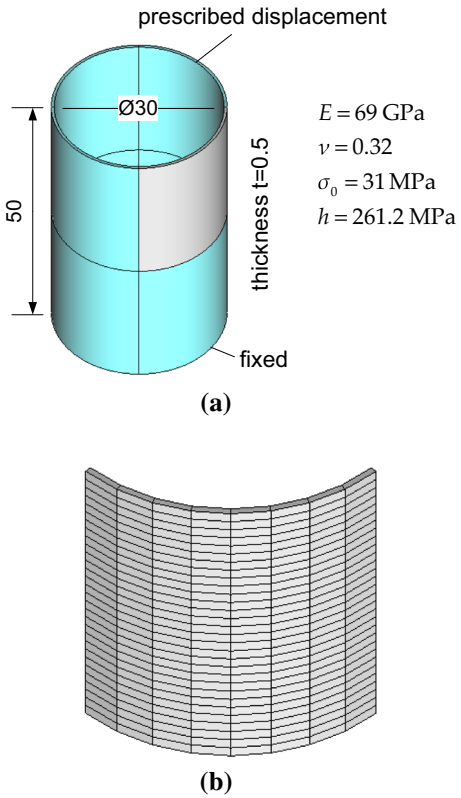

Fig. 61 Post-buckling of a thin walled tube: **a** geometrical and material properties (dimensions in mm); **b** finite element mesh of one eighth of the tube

series of buckles occur). Therefore, the two-pass Node-to-Segment approach is employed in this example, where each node of the contact surface is defined as slave node, while all Nagata patches composing the contact surface are assigned as master segments. Two independent self-contact surfaces are defined, i.e. the inner and outer surface of the tube. Thus, the nodes belonging to the interior surface of the tube are not allowed to penetrate their own surface, while the nodes in the exterior surface of the tube are not allowed to penetrate the exterior surface.

The post-buckling geometry and the corresponding equivalent plastic strain distribution (four instants) are shown in Fig. 62, for half height of the tube. Note that the buckling occurs without introducing any initial geometric imperfection. The presented results are obtained with the smoothed master surface description. A rigid contact surface is introduced in the horizontal plane of symmetry, which imposes the necessary impenetrability conditions for the last buckle, as illustrated in the bottom of Fig. 62d. The predicted number of buckles after the progressive folding of the tube is the same reported in [95] and the shape is very similar. The maximum value of plastic strain is reached at the final state, in the interior of the buckles, which is approximately 57 % for the last buckle. In fact, the entire tube comprises plastic deformation, particularly the buckles with large strains.

The total axial force as function of the total (top and bottom) axial displacement is presented in Fig. 63. The numerical results are compared with the ones presented in [95], which uses the coarse mesh to discretize the tube. The sequence of buckle cycles is clearly apparent in the curve, as well as the contribution of the self-contact. The buckling (indicated by the drop in force) occurs when a critical axial load is reached. The self-contact phenomena are illustrated

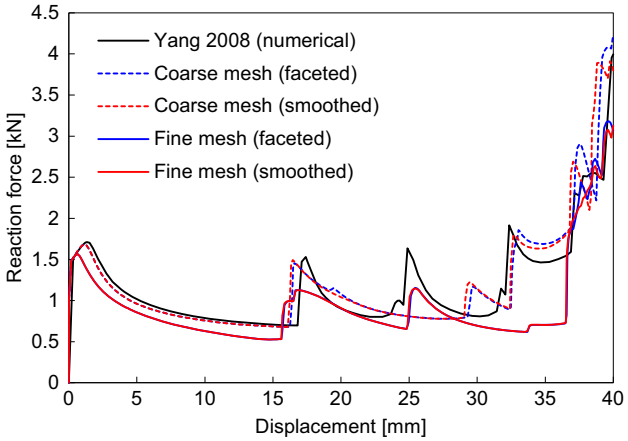

Fig. 63 Axial force evolution as a function of the displacement in the post-buckling problem, for both faceted and smoothed master surface descriptions

as a reversal of the force at the bottom of each buckle cycle, which provides the stiffening mechanism needed to activate the next buckle. The reaction force evolution predicted by the present model is identical to the one predicted in [95] for the first buckle cycle, as shown in Fig. 63. Nevertheless, the results reported in [95] present a different buckling sequence, i.e. the first buckle cycle occurs simultaneously in both extremities of the tube, while the next buckle cycles arise alternating between the two extremities of the tube. This behaviour cannot be reproduced using half length of the tube (present model), nevertheless, identical results are obtained when using the full model.

The buckling occurs slightly sooner when the master surface is smoothed with Nagata patches, in comparison with the piecewise bilinear representation. This offset in

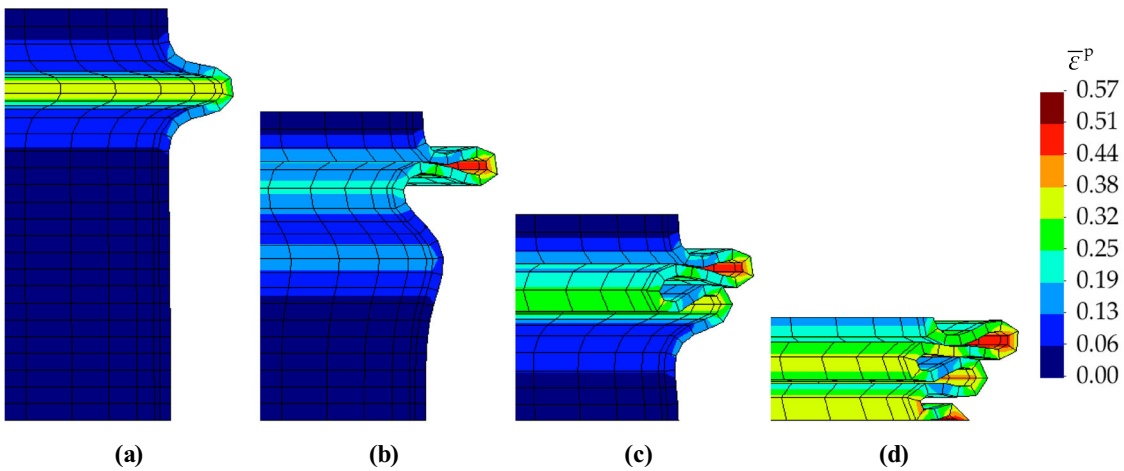

Fig. 62 Post-buckling geometry and equivalent plastic strain distribution (smoothed coarse mesh) for different values of total displacement: **a** 10 mm; **b** 20 mm; **c** 30 mm; **d** 40 mm

the curves due to the contact surface representation is more evident in the final stage. In order to evaluate the influence of the tube discretization in the numerical results, a fine mesh is also created, where the one eighth of the tube is discretized by 4800 finite elements (40 elements in the circumferential direction and 60 in the axial direction and two layers of elements through the thickness). The axial force evolution as a function of the total (top and bottom) axial displacement, obtained with the fine mesh, is shown in Fig. 63. Globally the force is lower and the instants at which self-contact occurs are different. The post-buckling geometry and the corresponding equivalent plastic strain distribution (four instants) are shown in Fig. 64, for the fine mesh with smoothed master surface. The sequence of post-buckling shapes obtained in the simulation with the fine mesh is substantially different, as shown through the comparison with Fig. 62. Typically, the finite element mesh refinement reduces the critical axial force necessary to produce buckling, as shown in Fig. 63. On the other hand, the application of the surface smoothing method presents a negligible influence in the numerical solution (force evolution curves are almost coincident). Globally, the equivalent plastic strain provided by the fine mesh is lower than the one obtained when adopting the coarse mesh, due to the better geometrical description of the folding of the tube (buckles). Although the number of buckles at the final state is the same for both finite element meshes (see Figs. 62d, 64d), the instants at which self-contact occurs are different, as well as the post-buckling shape.

Since the two-pass NTS contact algorithm is adopted in self-contact problems, the Node-to-Nagata contact elements are created in all nodes belonging to the potential contact surfaces. The nodal contact forces arising at the final state of the post-buckling problem are shown in Fig. 65. Since friction is not taken into account, the direction of the nodal contact force is given by the master surface normal vector at the contact point. Concerning the computational time, the numerical simulation is carried out in <1 min when adopting the coarse mesh, while the fine mesh requires about 17 min. In fact, the computational time is approximately the same for both surface description methods (faceted and smoothed) when the fine mesh is adopted. The same conclusion is extracted for the total number of iterations. On the other hand, the application of the surface smoothing in the coarse mesh yields a reduction of about 10 %, both in terms of the total number of iterations as well as on the computational time. The effectiveness of the surface smoothing method is more apparent in the coarse mesh than in the fine mesh.

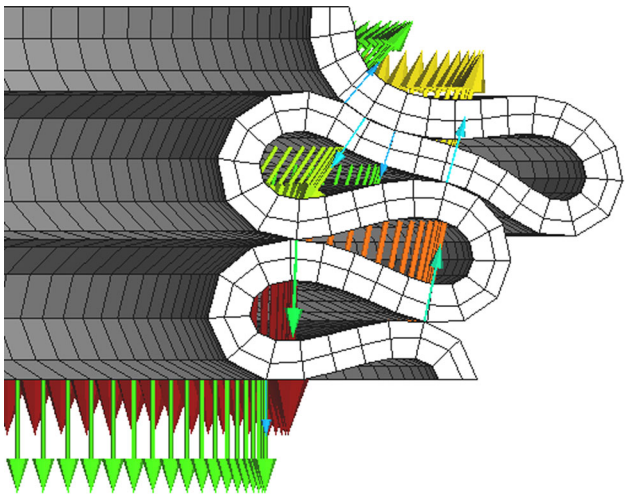

Fig. 65 Deformed configuration of the tube at the final state using the smoothed fine mesh, including the nodal contact forces

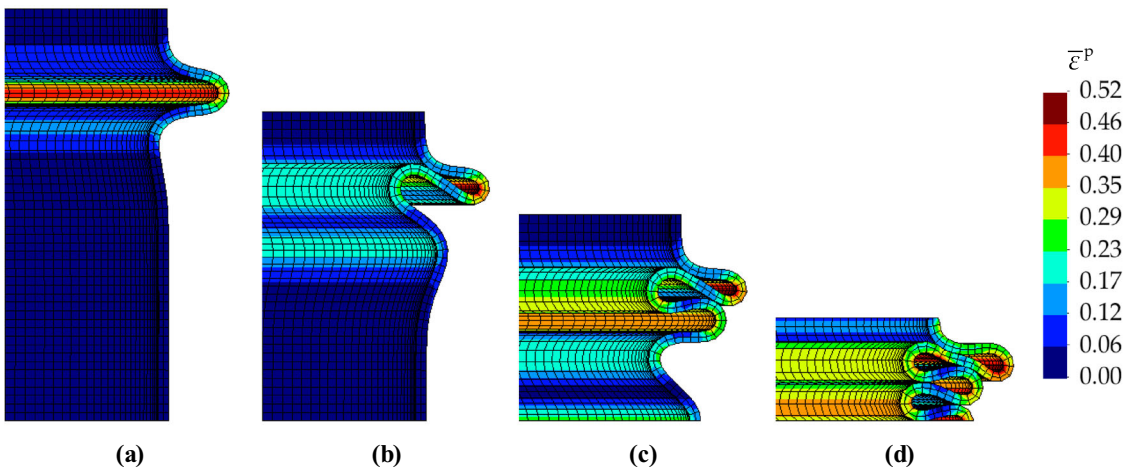

Fig. 64 Post-buckling geometry and equivalent plastic strain distribution (smoothed fine mesh) for different values of total displacement: **a** 10 mm; **b** 20 mm; **c** 30 mm; **d** 40 mm

8 Concluding Remarks

This paper presents the current research topics in computational contact mechanics. The finite element simulation of contact problems undergoing large deformation is addressed in detail, particularly the behaviour in the contact interface. The principle of impenetrability between bodies and the friction law on their common interface are formulated in the continuum setting. Nevertheless, the faceted contact surfaces introduced by finite element discretization induce nonphysical oscillations in the contact force, particularly in large sliding problems. The smoothing of the contact surfaces allows to solve the difficulties associated with the discontinuities in the contact surface geometry, improving the robustness of the numerical algorithms. The interpolation methods currently applied in the smoothing are based in parametric surfaces (e.g. Bézier, Spline, NURBS), which are computed from the position of the finite element nodes. The Nagata patch interpolation, recently proposed as smoothing procedure, is presented in detail, highlighting improvements in terms of geometrical representation of the contact surfaces in comparison with linear finite element meshes. Accordingly, the contact element called Node-to-Nagata is developed using the augmented Lagrangian method, allowing to solve problems with significant relative sliding between the contact surfaces. The results of the presented numerical examples show that both the accuracy and the robustness of the numerical simulations is improved when the contact surface is smoothed. The oscillations in the contact force are strongly reduced, while the performance of the numerical algorithms is enhanced.

Acknowledgments The authors gratefully acknowledge the financial support of the Portuguese Foundation for Science and Technology (FCT) via the project PTDC/EMS-TEC/1805/2012. The first author is also grateful to the FCT for the postdoctoral Grant SFRH/BPD/101334/2014.

References

1. Zhong Z-H (1993) Finite element procedures for contact-impact problems. Oxford University Press, Oxford
2. Kikuchi N, Oden JT (1988) Contact problems in elasticity: a study of variational inequalities and finite element methods. SIAM, Philadelphia
3. Wriggers P (2006) Computational contact mechanics. Springer, Berlin
4. Hertz H (1881) Über die Berührung fester elastische Körper. J für die reine und Angew Math 92:156–171. doi:10.1243/PIME_PROC_1982_196_039_02
5. Johnson KL (1982) One hundred years of hertz contact. Proc Inst Mech Eng 196:363–378. doi:10.1243/PIME_PROC_1982_196_039_02
6. Zienkiewicz OC, Taylor RL, Zhu JZ (2013) The finite element method: its basis and fundamentals. Finite Elem Method Basis Fundam. doi:10.1016/B978-1-85617-633-0.00019-8
7. Signorini A (1933) Sopra alcune questioni di elastostatica. Atti della Soc Ital per Prog delle Sci
8. Campos LT, Oden JT, Kikuchi N (1982) A numerical analysis of a class of contact problems with friction in elastostatics. Comput Methods Appl Mech Eng 34:821–845. doi:10.1016/0045-7825(82)90090-1
9. Wriggers P, Van Vu T, Stein E (1990) Finite element formulation of large deformation impact-contact problems with friction. Comput Struct 37:319–331. doi:10.1016/0045-7949(90)90324-U
10. Laursen TA, Simo JC (1993) A continuum-based finite element formulation for the implicit solution of multibody, large deformation-frictional contact problems. Int J Numer Methods Eng 36:3451–3485. doi:10.1002/nme.1620362005
11. Hallquist JO, Goudreau GL, Benson DJ (1985) Sliding interfaces with contact-impact in large-scale Lagrangian computations. Comput Methods Appl Mech Eng 51:107–137. doi:10.1016/0045-7825(85)90030-1
12. Zienkiewicz OC (1995) Origins, milestones and directions of the finite element method—a personal view. Arch Comput Methods Eng 2:1–48. doi:10.1007/BF02736188
13. Wilson EA, Parsons B (1970) Finite element analysis of elastic contact problems using differential displacements. Int J Numer Methods Eng 2:387–395. doi:10.1002/nme.1620020307
14. Chan SK, Tuba IS (1971) A finite element method for contact problems of solid bodies—Part I. Theory and validation. Int J Mech Sci 13:615–625. doi:10.1016/0020-7403(71)90032-4
15. Wriggers P (1995) Finite element algorithms for contact problems. Arch Comput Methods Eng 2:1–49. doi:10.1007/BF02736195
16. Mijar AR, Arora JS (2000) Review of formulations for elastostatic frictional contact problems. Struct Multidiscipl Optim 20:167–189. doi:10.1007/s001580050147
17. Zavarise G, De Lorenzis L (2009) The node-to-segment algorithm for 2D frictionless contact: classical formulation and special cases. Comput Methods Appl Mech Eng 198:3428–3451. doi:10.1016/j.cma.2009.06.022
18. Taylor R, Papadopoulos P (1991) On a patch test for contact problems in two dimensions. In: Wriggers P, Wagner W (eds) Computer methods nonlinear mechanics. Springer, Berlin, pp 690–702
19. Laursen TA (2002) Computational contact and impact mechanics: fundamentals of modeling interfacial phenomena in nonlinear finite element analysis. Springer, Berlin
20. Hansson E, Klarbring A (1990) Rigid contact modelled by CAD surface. Eng Comput 7:344–348. doi:10.1108/eb023821
21. Heege A, Alart P (1996) A frictional contact element for strongly curved contact problems. Int J Numer Methods Eng 39:165–184. doi:10.1002/(SICI)1097-0207(19960115)39:1<165::AID-NME846>3.0.CO;2-Y
22. Heegaard JH, Curnier A (1996) Geometric properties of 2D and 3D unilateral large slip contact operators. Comput Methods Appl Mech Eng 131:263–286. doi:10.1016/0045-7825(95)00977-9
23. Pietrzak G, Curnier A (1999) Large deformation frictional contact mechanics: continuum formulation and augmented Lagrangian treatment. Comput Methods Appl Mech Eng 177:351–381. doi:10.1016/S0168-874X(00)00029-9
24. Padmanabhan V, Laursen TA (2001) A framework for development of surface smoothing procedures in large deformation frictional contact analysis. Finite Elem Anal Des 37:173–198
25. Wriggers P, Krstulovic-Opara L, Korelc J (2001) Smooth C1-interpolations for two-dimensional frictional contact problems. Int J Numer Methods Eng 51:1469–1495. doi:10.1002/nme.227

26. Al-Dojayli M, Meguid SA (2002) Accurate modeling of contact using cubic splines. Finite Elem Anal Des 38:337–352. doi:10.1016/S0168-874X(01)00088-9
27. Krstulovic-Opara L, Wriggers P, Korelc J (2002) A C1-continuous formulation for 3D finite deformation frictional contact. Comput Mech 29:27–42. doi:10.1007/s00466-002-0317-z
28. Stadler M, Holzapfel GA, Korelc J (2003) Cn continuous modelling of smooth contact surfaces using NURBS and application to 2D problems. Int J Numer Methods Eng 57:2177–2203. doi:10.1002/nme.776
29. Puso MA, Laursen TA (2002) A 3D contact smoothing method using Gregory patches. Int J Numer Methods Eng 54:1161–1194. doi:10.1002/nme.466
30. Lengiewicz J, Korelc J, Stupkiewicz S (2011) Automation of finite element formulations for large deformation contact problems. Int J Numer Methods Eng 85:1252–1279. doi:10.1002/nme.3009
31. Corbett CJ, Sauer RA (2014) NURBS-enriched contact finite elements. Comput Methods Appl Mech Eng 275:55–75. doi:10.1016/j.cma.2014.02.019
32. Parisch H, Lübbing C (1997) A formulation of arbitrarily shaped surface elements for three-dimensional large deformation contact with friction. Int J Numer Methods Eng 40:3359–3383. doi:10.1002/(SICI)1097-0207(19970930)40:18<3359::AID-NME217>3.0.CO;2-5
33. Mijar AR, Arora JS (2000) Study of variational inequality and equality formulations for elastostatic frictional contact problems. Arch Comput Methods Eng 7:387–449. doi:10.1007/BF02736213
34. Bonet J, Wood RD (1997) Nonlinear continuum mechanics for finite element analysis. Cambridge University Press, Cambridge
35. Laursen TA (1994) The convected description in large deformation frictional contact problems. Int J Solids Struct 31:669–681. doi:10.1016/0020-7683(94)90145-7
36. Klarbring A (1995) Large displacement frictional contact: a continuum framework for finite element discretization. Eur J Mech A Solids 14:237–253
37. Agelet de Saracibar C (1997) A new frictional time integration algorithm for large slip multi-body frictional contact problems. Comput Methods Appl Mech Eng 142:303–334. doi:10.1016/S0045-7825(96)01133-4
38. Alart P, Curnier A (1991) A mixed formulation for frictional contact problems prone to Newton like solution methods. Comput Methods Appl Mech Eng 92:353–375. doi:10.1016/0045-7825(91)90022-X
39. Heegaard J-H, Curnier A (1993) An augmented Lagrangian method for discrete large-slip contact problems. Int J Numer Methods Eng 36:569–593. doi:10.1002/nme.1620360403
40. Buczkowski R, Kleiber M (2009) Statistical models of rough surfaces for finite element 3D-contact analysis. Arch Comput Methods Eng 16:399–424. doi:10.1007/s11831-009-9037-2
41. Yastrebov VA, Anciaux G, Molinari J-F (2014) The contact of elastic regular wavy surfaces revisited. Tribol Lett 56:171–183. doi:10.1007/s11249-014-0395-z
42. Simo JC, Laursen TA (1992) An augmented lagrangian treatment of contact problems involving friction. Comput Struct 42:97–116. doi:10.1016/0045-7949(92)90540-G
43. Heege A, Alart P, Oñate E (1995) Numerical modelling and simulation of frictional contact using a generalised coulomb law. Eng Comput 12:641–656. doi:10.1108/02644409510799820
44. Areias P, Rabczuk T, Queirós de Melo FJM, César de Sá J (2014) Coulomb frictional contact by explicit projection in the cone for finite displacement quasi-static problems. Comput Mech 55:57–72. doi:10.1007/s00466-014-1082-5
45. Oden JT, Pires EB (1984) Algorithms and numerical results for finite element approximations of contact problems with non-classical friction laws. Comput Struct 19:137–147. doi:10.1016/0045-7949(84)90212-8
46. Hjiaj M, Feng Z-Q, de Saxcé G, Mróz Z (2004) On the modelling of complex anisotropic frictional contact laws. Int J Eng Sci 42:1013–1034. doi:10.1016/j.ijengsci.2003.10.004
47. Persson B (2000) Sliding friction: physical principles and applications. Springer, Berlin
48. Refaat MH, Meguid SA (1998) A new strategy for the solution of frictional contact problems. Int J Numer Methods Eng 43:1053–1068. doi:10.1002/(SICI)1097-0207(19981130)43:6<1053::AID-NME460>3.0.CO;2-L
49. Agelet de Saracibar C (1998) Numerical analysis of coupled thermomechanical frictional contact problems. Computational model and applications. Arch Comput Methods Eng 5:243–301. doi:10.1007/BF02897875
50. Luenberger DG, Ye Y (2008) Linear and nonlinear programming, 3rd edn. doi:10.1007/978-0-387-74503-9
51. Yastrebov VA (2013) Numerical methods in contact mechanics. Wiley, Hoboken
52. Courtney-Pratt JS, Eisner E (1957) The effect of a tangential force on the contact of metallic bodies. Proc R Soc A Math Phys Eng Sci 238:529–550. doi:10.1098/rspa.1957.0016
53. Hüeber S, Stadler G, Wohlmuth BI (2008) A primal-dual active set algorithm for three-dimensional contact problems with coulomb friction. SIAM J Sci Comput 30:572–596. doi:10.1137/060671061
54. Popp A, Gee MW, Wall WA (2009) A finite deformation mortar contact formulation using a primal-dual active set strategy. Int J Numer Methods Eng 79:1354–1391. doi:10.1002/nme.2614
55. Hestenes MR (1969) Multiplier and gradient methods. J Optim Theory Appl 4:303–320. doi:10.1007/BF00927673
56. Powell M (1969) A method for nonlinear constraints in minimization problems. In: Fletcher R (ed) Optimization. Academic Press, New York, pp 283–298
57. Cavalieri FJ, Cardona A (2015) Numerical solution of frictional contact problems based on a mortar algorithm with an augmented Lagrangian technique. Multibody Syst Dyn. doi:10.1007/s11044-015-9449-8
58. Alart PP (1997) Méthode de Newton généralisée en mécanique du contact. J Math Pures Appl 76:83–108. doi:10.1016/S0021-7824(97)89946-1
59. Mijar AR, Arora JS (2004) An augmented Lagrangian optimization method for contact analysis problems, 1: formulation and algorithm. Struct Multidiscipl Optim 28:99–112. doi:10.1007/s00158-004-0423-y
60. Mijar AR, Arora JS (2004) An augmented Lagrangian optimization method for contact analysis problems, 2: numerical evaluation. Struct Multidiscipl Optim 28:113–126. doi:10.1007/s00158-004-0424-x
61. Yoon J (1999) A general elasto-plastic finite element formulation based on incremental deformation theory for planar anisotropy and its application to sheet metal forming. Int J Plast 15:35–67. doi:10.1016/S0749-6419(98)00059-X
62. Cardoso RPR, Yoon J-W (2005) One point quadrature shell elements for sheet metal forming analysis. Arch Comput Methods Eng 12:3–66. doi:10.1007/BF02736172
63. Tekkaya AE, Martins PAF (2009) Accuracy, reliability and validity of finite element analysis in metal forming: a user's perspective. Eng Comput 26:1026–1055. doi:10.1108/02644400910996880
64. Alart P, Lebon F (1995) Solution of frictional contact problems using ILU and coarse/fine preconditioners. Comput Mech 16:98–105. doi:10.1007/BF00365863
65. Saad Y (2003) Iterative methods for sparse linear systems. Society for Industrial and Applied Mathematics, Philadelphia

66. Schenk O, Gärtner K (2004) Solving unsymmetric sparse systems of linear equations with PARDISO. Future Gener Comput Syst 20:475–487. doi:10.1016/j.future.2003.07.011
67. Gould NIM, Scott JA, Hu Y (2007) A numerical evaluation of sparse direct solvers for the solution of large sparse symmetric linear systems of equations. ACM Trans Math Softw 33:10-es. doi:10.1145/1236463.1236465
68. Chow E, Saad Y (1997) Experimental study of ILU preconditioners for indefinite matrices. J Comput Appl Math 86:387–414. doi:10.1016/S0377-0427(97)00171-4
69. Menezes LF, Neto DM, Oliveira MC, Alves JL (2011) Improving computational performance through HPC techniques: case study using DD3IMP in-house code. AIP Conf Proc 1353:1220–1225. doi:10.1063/1.3589683
70. Intel (2014) Intel math kernel library reference manual
71. Crisfield MA (2000) Re-visiting the contact patch test. Int J Numer Methods Eng 48:435–449. doi:10.1002/(SICI)1097-0207(20000530)48:3<435::AID-NME891>3.0.CO;2-V
72. Santos A, Makinouchi A (1995) Contact strategies to deal with different tool descriptions in static explicit FEM for 3-D sheet-metal forming simulation. J Mater Process Technol 50:277–291. doi:10.1016/0924-0136(94)01391-D
73. Hachani M, Fourment L (2013) A smoothing procedure based on quasi-C1 interpolation for 3D contact mechanics with applications to metal forming. Comput Struct 128:1–13. doi:10.1016/j.compstruc.2013.05.008
74. Hama T, Nagata T, Teodosiu C et al (2008) Finite-element simulation of springback in sheet metal forming using local interpolation for tool surfaces. Int J Mech Sci 50:175–192. doi:10.1016/j.ijmecsci.2007.07.005
75. Shim H, Suh E (2000) Contact treatment algorithm for the trimmed NURBS surface. J Mater Process Technol 104:200–206. doi:10.1016/S0924-0136(00)00555-0
76. Landon RL, Hast MW, Piazza SJ (2009) Robust contact modeling using trimmed NURBS surfaces for dynamic simulations of articular contact. Comput Methods Appl Mech Eng 198:2339–2346. doi:10.1016/j.cma.2009.02.022
77. Wang SP, Nakamachi E (1997) The inside-outside contact search algorithm for finite element analysis. Int J Numer Methods Eng 40:3665–3685. doi:10.1002/(SICI)1097-0207(19971015)40:19<3665::AID-NME234>3.0.CO;2-K
78. Farouki RT (1999) Closing the gap between cad model and downstream application. SIAM News 32:303–319
79. Zhu X-F, Hu P, Ma Z-D et al (2013) A new surface parameterization method based on one-step inverse forming for isogeometric analysis-suited geometry. Int J Adv Manuf Technol 65:1215–1227. doi:10.1007/s00170-012-4251-8
80. Chamoret D, Saillard P, Rassineux A, Bergheau J-M (2004) New smoothing procedures in contact mechanics. J Comput Appl Math 168:107–116. doi:10.1016/j.cam.2003.06.007
81. Belytschko T, Daniel WJT, Ventura G (2002) A monolithic smoothing-gap algorithm for contact-impact based on the signed distance function. Int J Numer Methods Eng 55:101–125. doi:10.1002/nme.568
82. Francavilla A, Zienkiewicz OC (1975) A note on numerical computation of elastic contact problems. Int J Numer Methods Eng 9:913–924. doi:10.1002/nme.1620090410
83. Jin S, Sohn D, Lim JH, Im S (2015) A node-to-node scheme with the aid of variable-node elements for elasto-plastic contact analysis. Int J Numer Methods Eng 102:1761–1783. doi:10.1002/nme.4862
84. Simo JC, Wriggers P, Taylor RL (1985) A perturbed Lagrangian formulation for the finite element solution of contact problems. Comput Methods Appl Mech Eng 50:163–180. doi:10.1016/0045-7825(85)90088-X
85. Sauer RA, De Lorenzis L (2015) An unbiased computational contact formulation for 3D friction. Int J Numer Methods Eng 101:251–280. doi:10.1002/nme.4794
86. Zavarise G, De Lorenzis L (2009) A modified node-to-segment algorithm passing the contact patch test. Int J Numer Methods Eng 79:379–416. doi:10.1002/nme.2559
87. Puso MA, Laursen TA (2004) A mortar segment-to-segment frictional contact method for large deformations. Comput Methods Appl Mech Eng 193:4891–4913. doi:10.1016/j.cma.2004.06.001
88. El-Abbasi N, Bathe K-J (2001) Stability and patch test performance of contact discretizations and a new solution algorithm. Comput Struct 79:1473–1486. doi:10.1016/S0045-7949(01)00048-7
89. Zavarise G, Wriggers P (1998) A segment-to-segment contact strategy. Math Comput Model 28:497–515. doi:10.1016/S0895-7177(98)00138-1
90. Bernardi C, Debit N, Maday Y (1990) Coupling finite element and spectral methods: first results. Math Comput 54:21–39. doi:10.1090/S0025-5718-1990-0995205-7
91. Wohlmuth BI (2001) Discretization methods and iterative solvers based on domain decomposition. doi:10.1007/978-3-642-56767-4
92. Belgacem FB, Hild P, Laborde P (1998) The mortar finite element method for contact problems. Math Comput Model 28:263–271. doi:10.1016/S0895-7177(98)00121-6
93. McDevitt TW, Laursen TA (2000) A mortar-finite element formulation for frictional contact problems. Int J Numer Methods Eng 48:1525–1547. doi:10.1002/1097-0207(20000810)48:10<1525::AID-NME953>3.0.CO;2-Y
94. Puso MA, Laursen TA (2004) A mortar segment-to-segment contact method for large deformation solid mechanics. Comput Methods Appl Mech Eng 193:601–629. doi:10.1016/j.cma.2003.10.010
95. Yang B, Laursen TA (2008) A large deformation mortar formulation of self contact with finite sliding. Comput Methods Appl Mech Eng 197:756–772. doi:10.1016/j.cma.2007.09.004
96. Puso MA, Laursen TA, Solberg J (2008) A segment-to-segment mortar contact method for quadratic elements and large deformations. Comput Methods Appl Mech Eng 197:555–566. doi:10.1016/j.cma.2007.08.009
97. Laursen TA, Puso MA, Sanders J (2012) Mortar contact formulations for deformable–deformable contact: past contributions and new extensions for enriched and embedded interface formulations. Comput Methods Appl Mech Eng 205–208:3–15. doi:10.1016/j.cma.2010.09.006
98. Farah P, Popp A, Wall WA (2014) Segment-based vs. element-based integration for mortar methods in computational contact mechanics. Comput Mech 55:209–228. doi:10.1007/s00466-014-1093-2
99. Wohlmuth BI (2000) A mortar finite element method using dual spaces for the Lagrange multiplier. SIAM J Numer Anal 38:989–1012. doi:10.1137/S0036142999350929
100. Christensen PW, Klarbring A, Pang JS, Strömberg N (1998) Formulation and comparison of algorithms for frictional contact problems. Int J Numer Methods Eng 42:145–173. doi:10.1002/(SICI)1097-0207(19980515)42::1<145:AID-NME358>3.0.CO;2-L
101. Batailly A, Magnain B, Chevaugeon N (2012) A comparative study between two smoothing strategies for the simulation of contact with large sliding. Comput Mech 51:581–601. doi:10.1007/s00466-012-0737-3
102. El-Abbasi N, Meguid SA, Czekanski A (2001) On the modelling of smooth contact surfaces using cubic splines. Int J Numer Methods Eng 50:953–967. doi:10.1002/1097-0207(20010210)50:4<953::AID-NME64>3.0.CO;2-P

103. Stadler M, Holzapfel GA (2004) Subdivision schemes for smooth contact surfaces of arbitrary mesh topology in 3D. Int J Numer Methods Eng 60:1161–1195. doi:10.1002/nme.1001
104. Qian X, Yuan H, Zhou M, Zhang B (2014) A general 3D contact smoothing method based on radial point interpolation. J Comput Appl Math 257:1–13. doi:10.1016/j.cam.2013.08.014
105. Farin G (2002) Curves and surfaces for CAGD. Curves Surf CAGD. doi:10.1016/B978-1-55860-737-8.50030-2
106. Piegl L, Tiller W (1997) The NURBS book. doi:10.1007/978-3-642-59223-2
107. Hughes TJR, Cottrell JA, Bazilevs Y (2005) Isogeometric analysis: CAD, finite elements, NURBS, exact geometry and mesh refinement. Comput Methods Appl Mech Eng 194: 4135–4195. doi:10.1016/j.cma.2004.10.008
108. De Lorenzis L, Wriggers P, Hughes TJR (2014) Isogeometric contact: a review. GAMM Mitt 37:85–123. doi:10.1002/gamm.201410005
109. Piegl L (1991) On NURBS: a survey. IEEE Comput Graph Appl 11:55–71. doi:10.1109/38.67702
110. Cox MG (1972) The numerical evaluation of B-splines. IMA J Appl Math 10:134–149. doi:10.1093/imamat/10.2.134
111. De Boor C (1972) On calculating with B-splines. J Approx Theory 6:50–62. doi:10.1016/0021-9045(72)90080-9
112. Nagata T (2005) Simple local interpolation of surfaces using normal vectors. Comput Aided Geom Des 22:327–347. doi:10.1016/j.cagd.2005.01.004
113. Neto DM, Oliveira MC, Menezes LF, Alves JL (2014) Applying Nagata patches to smooth discretized surfaces used in 3D frictional contact problems. Comput Methods Appl Mech Eng 271:296–320. doi:10.1016/j.cma.2013.12.008
114. Sekine T, Obikawa T (2010) Normal-unit-vector-based tool path generation using a modified local interpolation for ball-end milling. J Adv Mech Des Syst Manuf 4:1246–1260. doi:10.1299/jamdsm.4.1246
115. Boschiroli M, Fünfzig C, Romani L, Albrecht G (2011) A comparison of local parametric C0 Bézier interpolants for triangular meshes. Comput Graph 35:20–34. doi:10.1016/j.cag.2010.09.011
116. Neto DM, Oliveira MC, Menezes LF, Alves JL (2013) Improving Nagata patch interpolation applied for tool surface description in sheet metal forming simulation. Comput Aided Des 45:639–656. doi:10.1016/j.cad.2012.10.046
117. Neto DM, Oliveira MC, Menezes LF, Alves JL (2013) Nagata patch interpolation using surface normal vectors evaluated from the IGES file. Finite Elem Anal Des 72:35–46. doi:10.1016/j.finel.2013.03.004
118. IGES (1996) Initial graphics exchange specification, IGES 5.3. IGES/PDES Organization
119. Todd PH, McLeod RJY (1986) Numerical estimation of the curvature of surfaces. Comput Des 18:33–37. doi:10.1016/S0010-4485(86)80008-2
120. Meek DS, Walton DJ (2000) On surface normal and Gaussian curvature approximations given data sampled from a smooth surface. Comput Aided Geom Des 17:521–543. doi:10.1016/S0167-8396(00)00006-6
121. OuYang D, Feng H-Y (2005) On the normal vector estimation for point cloud data from smooth surfaces. Comput Des 37:1071–1079. doi:10.1016/j.cad.2004.11.005
122. Page DL, Sun Y, Koschan AF et al (2002) Normal vector voting: crease detection and curvature estimation on large, noisy meshes. Graph Models 64:199–229. doi:10.1006/gmod.2002.0574
123. Jin S, Lewis RR, West D (2005) A comparison of algorithms for vertex normal computation. Vis Comput 21:71–82. doi:10.1007/s00371-004-0271-1
124. Ubach P-A, Estruch C, Garcia-Espinosa J (2013) On the interpolation of normal vectors for triangle meshes. Int J Numer Methods Eng 96:247–268. doi:10.1002/nme.4567
125. Neto DM, Oliveira MC, Menezes LF, Alves JL (2016) A contact smoothing method for arbitrary surface meshes using Nagata patches. Comput Methods Appl Mech Eng. doi:10.1016/j.cma.2015.11.011
126. Lin J, Ball AA, Zheng JJ (2001) Approximating circular arcs by Bézier curves and its application to modelling tooling for FE forming simulations. Int J Mach Tools Manuf 41:703–717. doi:10.1016/S0890-6955(00)00100-0
127. Kamran K, Rossi R, Oñate E (2012) A contact algorithm for shell problems via Delaunay-based meshing of the contact domain. Comput Mech 52:1–16. doi:10.1007/s00466-012-0791-x
128. Yang B, Laursen TA (2008) A contact searching algorithm including bounding volume trees applied to finite sliding mortar formulations. Comput Mech 41:189–205. doi:10.1007/s00466-006-0116-z
129. Aragón AM, Yastrebov VA, Molinari J-F (2013) A constrained-optimization methodology for the detection phase in contact mechanics simulations. Int J Numer Methods Eng 96:323–338. doi:10.1002/nme.4561
130. Areias PMA, César de Sá JMA, Conceição António CA (2004) Algorithms for the analysis of 3D finite strain contact problems. Int J Numer Methods Eng 61:1107–1151. doi:10.1002/nme.1104
131. Zhi-Hua Z, Nilsson L (1989) A contact searching algorithm for general contact problems. Comput Struct 33:197–209. doi:10.1016/0045-7949(89)90141-7
132. Benson DJ, Hallquist JO (1990) A single surface contact algorithm for the post-buckling analysis of shell structures. Comput Methods Appl Mech Eng 78:141–163. doi:10.1016/0045-7825(90)90098-7
133. Oldenburg M, Nilsson L (1994) The position code algorithm for contact searching. Int J Numer Methods Eng 37:359–386. doi:10.1002/nme.1620370302
134. Fujun W, Jiangang C, Zhenhan Y (2000) A contact searching algorithm for contact-impact problems. Acta Mech Sin 16:374–382. doi:10.1007/BF02487690
135. Konyukhov A, Schweizerhof K (2008) On the solvability of closest point projection procedures in contact analysis: analysis and solution strategy for surfaces of arbitrary geometry. Comput Methods Appl Mech Eng 197:3045–3056. doi:10.1016/j.cma.2008.02.009
136. Belytschko T, Neal MO (1991) Contact-impact by the pinball algorithm with penalty and Lagrangian methods. Int J Numer Methods Eng 31:547–572. doi:10.1002/nme.1620310309
137. Konyukhov A, Schweizerhof K (2005) Covariant description for frictional contact problems. Comput Mech 35:190–213. doi:10.1007/s00466-004-0616-7
138. Parisch H (1989) A consistent tangent stiffness matrix for three-dimensional non-linear contact analysis. Int J Numer Methods Eng 28:1803–1812. doi:10.1002/nme.1620280807
139. Klarbring A, Björkourkman G (1992) Solution of large displacement contact problems with friction using Newton's method for generalized equations. Int J Numer Methods Eng 34:249–269. doi:10.1002/nme.1620340116
140. Laursen TA, Maker BN (1995) An augmented Lagrangian quasi-Newton solver for constrained nonlinear finite element applications. Int J Numer Methods Eng 38:3571–3590. doi:10.1002/nme.1620382103
141. Renard Y (2013) Generalized Newton's methods for the approximation and resolution of frictional contact problems in elasticity. Comput Methods Appl Mech Eng 256:38–55. doi:10.1016/j.cma.2012.12.008
142. Alart P, Heege A (1995) Consistent tangent matrices of curved contact operators involving anisotropic friction. Rev Eur des

Éléments Finis 4:183–207. doi:10.1080/12506559.1995.10511173
143. Menezes LF, Teodosiu C (2000) Three-dimensional numerical simulation of the deep-drawing process using solid finite elements. J Mater Process Technol 97:100–106. doi:10.1016/S0924-0136(99)00345-3
144. Oliveira MC, Alves JL, Menezes LF (2008) Algorithms and strategies for treatment of large deformation frictional contact in the numerical simulation of deep drawing process. Arch Comput Methods Eng 15:113–162. doi:10.1007/s11831-008-9018-x
145. Oliveira MC, Alves JL, Chaparro B, Menezes LF (2007) Study on the influence of work-hardening modeling in springback prediction. Int J Plast 23:516–543. doi:10.1016/j.ijplas.2006.07.003
146. Yamada Y, Yoshimura N, Sakurai T (1968) Plastic stress-strain matrix and its application for the solution of elastic-plastic problems by the finite element method. Int J Mech Sci 10:343–354. doi:10.1016/0020-7403(68)90001-5
147. Tur M, Fuenmayor FJ, Wriggers P (2009) A mortar-based frictional contact formulation for large deformations using Lagrange multipliers. Comput Methods Appl Mech Eng 198:2860–2873. doi:10.1016/j.cma.2009.04.007
148. NUMISHEET'93 (1993) Proceedings of the 2nd international conference numerical simulation of 3-D sheet metal forming processes
149. Neto DM, Oliveira MC, Alves JL, Menezes LF (2015) Comparing faceted and smoothed tool surface descriptions in sheet metal forming simulation. Int J Mater Form 8:549–565. doi:10.1007/s12289-014-1177-8
150. Oliveira MC, Menezes LF (2004) Automatic correction of the time step in implicit simulations of the stamping process. Finite Elem Anal Des 40:1995–2010. doi:10.1016/j.finel.2004.01.009
151. Klang M (1979) On interior contact under friction between cylindrical elastic bodies. Linköping University, Linköping
152. Hammer ME (2012) Frictional mortar contact for finite deformation problems with synthetic contact kinematics. Comput Mech 51:975–998. doi:10.1007/s00466-012-0780-0