

Algorithms for the analysis of 3D finite strain contact problems

P. M. A. Areias^{*,†,‡}, J. M. A. César de Sá and C. A. Conceição António

*Instituto de Engenharia Mecânica (IDMEC), Faculdade de Engenharia da Universidade do Porto,
Rua Dr Roberto Frias s/n 4200-465 Porto, Portugal*

SUMMARY

From the constraint imposition aspects in 3D to friction regularization, various ideas are exposed in this paper. A variation of the Rockafellar Lagrangian is proposed which results in continuous second-order derivatives if Lagrange multiplier estimates are greater or equal than one. This fact allows the adoption of a full second-order (i.e. Lagrange–Newton) method avoiding sequential unconstrained minimization techniques. An algorithm for global and local contact detection is presented which is developed for dealing with large step sizes typical of implicit methods. A modified constraint definition to deal with non-smooth situations is presented. Aspects of friction implementation, including a regularization scheme which ensures stepwise objectivity, are detailed. Finally, several illustrative examples are carried out with success. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: contact; friction; Rockafellar Lagrangian; finite elements; enhanced-strain elements; finite strains

1. GENERAL CONSIDERATIONS

It is clear that almost every mechanical deformation process involves contact forces, and also forces from distinct nature such as volume forces. Notwithstanding the existence of many papers presenting new advances and techniques for analysing contact problems with the finite element method, the finite deformation *implicit* contact analysis, including friction, of 3D deformable solids is not yet completely robust in large-scale applications.

Two key problems addressed in this work are the analysis of deformable–deformable contact and self-contact. In the context of metal forming processes, self-contact may emerge in a variety of situations, some of which are technically undesirable (e.g. Reference [1]). Additionally, it is

*Correspondence to: P. M. A. Areias, Instituto de Engenharia Mecânica (IDMEC), Faculdade de Engenharia da Universidade do Porto, Rua Dr Roberto Frias s/n 4200-465 Porto, Portugal.

†E-mail: p-areias@northwestern.edu, URL: www.fe.up.pt

‡Present address: Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Room B224, Evanston, IL 60208, U.S.A.

Contract/grant sponsor: Ministerio da Ciencia e da Tecnologia, FCT; contract/grant number: PRAXIS XXI/BD/18533/98

known that the consideration of tool deformation may also be important in certain circumstances. These two cases may be conceptually thought as contact between deformable bodies, as locally, there is no distinction between the two.

In the finite element analysis of contact between deformable solids, there are at least two physically observed phenomena that is imperative to represent: the non-penetration constraint and friction effects between the interacting surfaces. For an analytical introduction to the theme, Reference [2] is well known. For an introduction to the physical phenomena, References [3, 4] are very important.

The finite element treatment of the non-penetration is commonly based on the imposition of inequality constraints. These inequality constraints *can* be satisfied making use of modified equality constraint algorithms or, in alternative, employing active set strategies for the constraints. These are two paths discussed in non-linear mathematical programming textbooks [5, 6]. If second-order information is available, a class of effective (see some conclusions in Reference [7]) methods for solving *equality* constrained problems is the sequential quadratic programming (SQP) class of optimization methods, comprehensively exposed in References [6, 8]. The direct application of the classical version of this method to problems containing inequalities can become very expensive for large-scale situations, due to the fact that a inequality constrained quadratic subproblem must be solved in each SQP iteration. The solution of this subproblem is costly in the presence of both large number of inequality constraints and large number of variables (both active-set and Lemke's methods for solving this sub-problem are comprehensively described in Reference [9] and Lemke's method was employed in the finite element context in Reference [10]). Techniques to deal with the discontinuities that inequalities invariably imply were adopted in the finite element context. Reference [11] presents a strategy for 'smoothing' the Newton–Raphson method for dealing with contact constraints and Bathe and Bouzinov [12] propose regularizations that allow differentiability at the expenses of accuracy. Recently, Areias and César de Sá [13] proposed a new second-order algorithm which allows a smoother transition between active and non-active contact constraints without modifying the constraint satisfaction accuracy. Besides this innovation, new considerations of contact constraint definition are presented, which allow difficult non-smoothed 3D problems to be solved without smoothing procedures.

A review of some of the concepts involved in the study of contact mechanics (including some key-aspects to be represented in simulations) was carried out by Barber and Ciavarella [14]. It is interesting to refer that one of the first attempts to analyse classical contact problems with finite elements was carried out by Francavilla and Zienkiewicz [15] and a much more rigorous analysis was published in the following year by Hughes *et al.* [16] already including independent force variables.

2. INEQUALITY CONSTRAINTS ARISING FROM THE NON-PENETRATION CONDITION

If the discrete contact problem in study complies with the restrictive circumstance of being conservative (which among other consequences, eliminates friction forces from the model) then the discretized model can be classified as a non-linear programming problem.

Given two bodies denoted as \mathcal{B}_1 and \mathcal{B}_2 , an approach for the condition of non-penetration of one discretized body over another is achieved through the adoption of a node-to-face strategy

where every exterior node of body \mathcal{B}_1 is constrained not to penetrate in every face of body \mathcal{B}_2 , a reasonable discretization of contact in many situations. If body \mathcal{B}_1 contains $n_{\mathcal{B}_1}$ nodes and body \mathcal{B}_2 contains $m_{\mathcal{B}_2}$ faces, there should be $n_{\mathcal{B}_1} m_{\mathcal{B}_2}$ inequality constraints (usually not all active simultaneously). This key-aspect appears to be ignored in most bibliography concerned with the discretized treatment of contact problems:

- The maximum number of faces that each node is constrained not to penetrate in a given time step is pre-assigned.
- In most cases, this number is 1. Although during the analysis the faces from which each node is prevented to penetrate change, the faces are selected one at the time.

This universally accepted strategy allows the total number of constraints to be reduced to only $n_{\mathcal{B}_1}$, but is the root of some inadequacies in the application of optimization algorithms to the contact problems as the equidistance dilemma.

A general presentation of the inequality constrained problem is as follows:

$$\begin{aligned} \min \quad & \Pi(\mathbf{x}) \\ \text{s.t} \quad & \mathbf{h}(\mathbf{x}) \leq \mathbf{0} \end{aligned} \quad (1)$$

where \mathbf{x} is an n -dimensional vector, and $\mathbf{h}(\mathbf{x})$ is an m -dimensional vector of functions of \mathbf{x} , which, in the present work, is defined for all $\mathbf{x} \in \mathfrak{R}^n$. $\Pi(\mathbf{x})$ is the sum of the elastic potential energy and the potential of the conservative external forces. The variables x_i in \mathbf{x} represent nodal spatial position variables. The concept of signed distance function will be used for defining $\mathbf{h}(\mathbf{x})$ in a subsequent section.

A Lagrangian function corresponding to (1) is established as:

$$L(\mathbf{x}, \boldsymbol{\lambda}) = \Pi(\mathbf{x}) + \mathbf{h}(\mathbf{x}) \cdot \boldsymbol{\lambda} \quad (2)$$

where $\boldsymbol{\lambda}$ are additional variables, which will be related with the Lagrange multipliers at one solution of problem (1). Please note the difference between $\boldsymbol{\lambda}$, which are variables, and $\boldsymbol{\lambda}_*$ which are the values of these variables at the solution and are called Lagrange multipliers, as will be introduced (see Reference [6] for this nomenclature).

Functions $\Pi(\mathbf{x})$ and $\mathbf{h}(\mathbf{x})$ are, for now, assumed to be twice continuously differentiable.

At a given value of the variables \mathbf{x} , the set of indexes of *active* constraints can be written as:

$$A(\mathbf{x}) = \{j \mid h_j(\mathbf{x}) = 0\} \quad (3)$$

A comprehensive presentation of the *first-order* necessary conditions can be written as (see References [6, 17]):

Let \mathbf{x}_* be a local minimum of $\Pi(\mathbf{x})$ satisfying the constraints $\mathbf{h}(\mathbf{x}) \leq \mathbf{0}$ and assume that \mathbf{x}_* is regular. Then there exists a unique Lagrange multiplier vector $\boldsymbol{\lambda}_*$ such that:

$$\nabla_{\mathbf{x}} L(\mathbf{x}_*, \boldsymbol{\lambda}_*) = \mathbf{0} \quad (4)$$

$$\boldsymbol{\lambda}_* \geq \mathbf{0} \quad (5)$$

$$\lambda_{*j} = 0 \quad \forall j \notin A(\mathbf{x}_*) \quad (6)$$

A given point \mathbf{x}_* being regular, means, in the context of inequality constraints, that the gradients of all constraints whose indexes belong to the set $A(\mathbf{x}_*)$ are linearly independent.

From (4), it is possible to write the vectorial equality $\mathbf{h}(\mathbf{x}_*) \odot \boldsymbol{\lambda}_* = \mathbf{0}$ (such as $h_i \lambda_i = 0$ with no sum on i), usually designated as the complementary slackness condition [18]. If a Lagrange multiplier value is zero, the corresponding constraint *can* be non-active, but if the Lagrange multiplier is positive, then the constraint must be active. Situations exist where both the constraint is active and the correspondent Lagrange multiplier is null.

Let us discuss some methodologies for solving problem (4) for both \mathbf{x}_* and $\boldsymbol{\lambda}_*$ which hopefully can be adapted to the contact problem in hands.

A far from exhaustive list of suitable algorithms for the solution of contact problems (inequality constraints) which make use of available *second-order information* is:

1. Exterior penalty or interior barrier methods (usually approximated methods). Referred to generate ill-conditioned matrices for accurately satisfied constraints, these methods can be regularized [19] to avoid this effect. In fact, one motivation of the popular method of multipliers discussed for instance in Reference [7] is to shift the quadratic penalty (or a modified barrier [19–21]) to ensure that the Hessian matrix remains well conditioned.
2. The method of multipliers, where a penalized Lagrangian function (frequently denominated Augmented Lagrangian [6]) is adopted and a first-order update of the Lagrange multiplier estimates is carried out. The inequality constraints can be dealt with an active set strategy *or* with a special updating strategy of the multiplier estimates that complies with the modified Augmented Lagrangian for inequality constraints introduced by Rockafellar [22]. The variants with exponents [7, 20] overcome the absence of second derivatives (at a given point) of the Augmented Lagrangian by Rockafellar. The use of a modified barrier has also been adopted for contact problems in Reference [23].
3. Reduced-gradient methods, described in References [11, 24, 25], for example. Variants that use second-order information are based on a reduced Hessian matrix (e.g. References [8, 24]), which is formed by a projection of the original Hessian using a null space base of the constraint gradients. These methods are related with the so-called multi-point constraints (MPC), described for linear equality constraints in Reference [26], and demand carefully designed algorithms to determine which variables should be eliminated from the original unconstrained problem.
4. Direct applications of the Newton method extended to deal with equality constraints, subsequently modified for inequalities by a number of techniques:
 - (a) The use of active set strategies which are supported by the checking of the Lagrange multipliers signs after an equality constraint problem is solved, along with a strategy for incorporating activated constraints during the iteration process.
 - (b) The solution of the necessary conditions for the quadratic subproblem through the solution of a linear complementary problem (LCP), see Reference [9] for the mathematical details. The application of the Lemke's algorithm in the finite element context was carried out, among others, by Chertier and Chabrand [10].
 - (c) The use of slack variables on a slacked Lagrangian and a perturbation (which depends on a barrier parameter) of the complementary slackness condition, as presented in References [17, 25, 27].
 - (d) The use of a Rockafellar [22] Lagrangian or the exponential multiplier method [6, 7, 20] with a *second-order* updating of the Lagrange multiplier estimates, which

reduces the solution of the quadratic problem to the repeated solution of linear equations systems.

The method of multipliers is popular in contact finite element analysis. However, steepest ascent updating of multipliers, or any other first-order strategy that could be cogitated for the multipliers (see Reference [7] for an in-depth study about multiplier methods) damages the quadratic convergence properties of the Newton method adopted for the primary variables, \mathbf{x} . This fact is clearly emphasized in Reference [28], for instance in p. 232, ‘When a Newton-type method is applied to solve the unconstrained subproblem a second order multiplier estimate should be used in order to achieve the expected quadratic convergence...’.

This argument can be circumvented making use of the stepwise structure of the implicit finite element codes, freezing the Lagrange multiplier estimates between steps and artificially obtaining a quadratically convergent scheme for the variables \mathbf{x} in each step. This strategy is described in References [29–33] but leads to multiple sub-steps. Another difficulty with the method of multipliers is that its convergence to the correct solution depends on the penalty parameter value which should be greater than a (difficult to obtain) threshold, see Reference [6].

The use of a Rockafellar Lagrangian strategy (which emanates from a slacked Lagrangian subsequently liberated from the slack variables, see Reference [6]) can be adapted to second-order updates of the Lagrange multipliers. There is, however, a discontinuity of the Hessian matrix if at least *one* (j for example) constraint value equals the ratio between minus the Lagrange multiplier estimate and the penalty parameter, r :

$$\forall \mathbf{x} \mid h_j(\mathbf{x}) = -\frac{\lambda_j}{r} \Rightarrow \text{location of the Hessian matrix discontinuity} \quad (7)$$

A solution for this problem is discussed by Bertsekas and co-workers with the introduction of the exponential method of multipliers [6, 20] adopted within a first-order algorithm for the Lagrange multiplier estimates. A corresponding Lagrangian is not available for second-order algorithms. Therefore, a combination of the Rockafellar Lagrangian and the exponential method of multipliers is employed here. For a set of m inequality constraints, which are grouped in the vector $\mathbf{h}(\mathbf{x})$, it is possible to write, using a vector of m Lagrange multiplier estimates (MR indicates modified Rockafellar):

$$L_{\text{MR}}(\mathbf{x}, \boldsymbol{\lambda}) = \Pi(\mathbf{x}) + \frac{1}{2r} \sum_{i=1}^m \{\max[0, \lambda_i + e^{r h_i(\mathbf{x})} - 1]^2 - \lambda_i^2\} \quad (8)$$

from which first and second variations follow in a straightforward form. The following comments are applicable to definition (6):

- The effect in terms of gradient is similar to the modified barrier function presented in References [19, 21] but it is defined for all $\mathbf{x} \in \mathfrak{R}^n$, a fact which does not occur with classical barrier functions.
- The values of the Lagrangian L_{MR} , the Lagrange multiplier vector, $\boldsymbol{\lambda}_*$ and the variables \mathbf{x}_* are the same as the original Lagrangian (2) at a local constrained minimum of $\Pi(\mathbf{x})$.

Regarding the line-search algorithm, both a reduction of $\Pi(\mathbf{x})$ and satisfaction of constraints are aimed. The following merit function is employed:

$$\mathcal{M}_{\text{MR}} = \frac{1}{2} \|\nabla_{\mathbf{x}\boldsymbol{\lambda}} L_{\text{MR}}(\mathbf{x}, \boldsymbol{\lambda})\|^2 \quad (9)$$

Although other functions could be adopted (see References [6, 18]) this function (7) does *not* make direct use of the value of the function $\Pi(\mathbf{x})$ which is seldom available in the finite element codes, and it is particularly convenient for an inexact line search. Additionally this merit function can be used in non-conservative problems with a simple modification. The step size α is taken as the first α_i that satisfies the inequality:

$$\mathcal{M}_{\text{MR}}(\mathbf{x} + \alpha_i \Delta \mathbf{x}, \lambda + \alpha_i \Delta \lambda) \leq \chi \mathcal{M}_{\text{MR}}(\mathbf{x}, \lambda) \quad (10)$$

where χ is a parameter close to one and $\alpha_i = (\frac{1}{2})^i$ with $i = 0, \dots, m_i - 1$ with m_i being the maximum number of evaluations of the merit function at each Lagrange–Newton iteration. The values $m_i = 4$ and $\chi = 0.9$ are here adopted.

2.1. The necessary conditions and the slacked version of the proposed Lagrangian: derivations for one constraint

The common transformation of problem into an equality constraint problem with a side-constraint can be accomplished by re-writing problem (1) according to:

$$\begin{aligned} \min \quad & \Pi(\mathbf{x}) \\ \text{s.t} \quad & h(\mathbf{x}) + s = 0 \\ & \text{and to } s \geq 0 \end{aligned} \quad (11)$$

with the introduction of the slack variable s . Note that the use of only *one constraint* does not imply lack of generality.

The use of a penalty term and the modification of the constraint function h allows the reformatting of the previous problem as ($r > 0$):

$$\begin{aligned} \min \quad & \Pi(\mathbf{x}) + \frac{r}{2} \left[\frac{e^{rh(\mathbf{x})} - 1}{r} + s \right]^2 \\ \text{s.t} \quad & \frac{e^{rh(\mathbf{x})} - 1}{r} + s = 0 \\ & \text{and to } s \geq 0 \end{aligned} \quad (12)$$

whose solution is the same of the previous problem, as the penalty term disappears at the solution, because the first equation should be satisfied.

A Lagrangian corresponding to problem (10) consists of:

$$L_{\text{MRS}}(\mathbf{x}, s, \lambda, \lambda_2) = \begin{cases} \Pi(\mathbf{x}) + \lambda \left(\frac{e^{rh(\mathbf{x})} - 1}{r} + s \right) + \frac{r}{2} \left[\frac{e^{rh(\mathbf{x})} - 1}{r} + s \right]^2 + \lambda_2 s, & s < 0 \\ \Pi(\mathbf{x}) + \lambda \left(\frac{e^{rh(\mathbf{x})} - 1}{r} + s \right) + \frac{r}{2} \left[\frac{e^{rh(\mathbf{x})} - 1}{r} + s \right]^2, & s \geq 0 \end{cases} \quad (13)$$

where a second Lagrange multiplier was introduced, λ_2 , in the case $s < 0$. It is clear that a second penalty term could be introduced for the constraint $s \geq 0$. With the slacked Lagrangian

Table I. The Euler–Lagrange equations of the slacked Lagrangian (11), after simplification.

Variables	$s < 0$	$s \geq 0$
\mathbf{x}	$\nabla_{\mathbf{x}}\Pi(\mathbf{x}_*) + \lambda_*\nabla_{\mathbf{x}}h(\mathbf{x}_*) = \mathbf{0}$	$\nabla_{\mathbf{x}}\Pi(\mathbf{x}_*) = \mathbf{0}$
s	$\lambda_{2*} + \lambda_* = 0$	$\lambda_* = 0$
λ	$h_* = 0$	$\frac{e^{rh(\mathbf{x}_*)} - 1}{r} + s_* = 0$
λ_2	$s_* = 0$	*

(11) in hands, it remains to verify if the stationarity conditions of this Lagrangian (relative to $\{\mathbf{x}, s, \lambda, \lambda_2\}$) corresponds to the necessary conditions for a point \mathbf{x} to be a solution of problem (1). The stationarity conditions (after simplification) are represented in Table I for both branches. It is interesting to verify that the necessary conditions are included as Euler–Lagrange equations. The nullity of λ for the unconstrained case is satisfied and for the constrained case, the fact that a given point \mathbf{x}_* is a given minimum implies that if there is a given $\Delta\mathbf{x}$ such that $\nabla_{\mathbf{x}}\Pi(\mathbf{x}_*) \cdot \Delta\mathbf{x} < 0$ then $\nabla_{\mathbf{x}}h(\mathbf{x}_*) \cdot \Delta\mathbf{x} \geq 0$.

From these arguments it follows that $\lambda_* \geq 0$ and $\lambda_{2*} \leq 0$ for the constrained case and $\lambda_* = 0$ for the unconstrained case. If m constraints exist, the conclusions remain valid: $\mathbf{h}(\mathbf{x}_*) \leq \mathbf{0}$, $\mathbf{h}(\mathbf{x}_*) \odot \lambda_* = \mathbf{0}$ and $\nabla_{\mathbf{x}}L(\mathbf{x}_*, \lambda_*) = \mathbf{0}$ for $\lambda_* \in \mathfrak{R}^m$ and $\mathbf{h}(\mathbf{x}_*) \in \mathfrak{R}^m$.

Having illustrated the direct relation between the necessary conditions and the stationarity of L_{MRS} , it is possible to remove the slack variable s and obtain L_{MR} as (6).

With this conclusion in hands, it is straightforward to show that no discontinuity occurs for $\lambda > 1$.

3. THE ALGORITHM FOR CONTACT DETECTION

Two main steps are usually carried out for the detection: global detection (see the account on Reference [34]) and local search (see e.g. Reference [35]).

For the local search, whose purpose is to assign at most one face to each incident node, several strategies are possible: the direct projection onto each candidate face of each incident node [29, 36], the inside-outside algorithm [37] (based on the definition of an incident’s node normal vector) or a preliminary step before the projection over the face by the use of a box-like domain [38, 39] or a spherical domain [40, 41]. An alternative using simplex intersections was proposed in Reference [42].

To detail what is involved in contact detection with the proposed strategy, and what is the contact detection task, Figure 1 shows the connection between an incident node N_{kr} within slave group S_k being associated with a given target face F_{ks} belonging to the master group M_k (k indicates the group, r and s indicate the node and the face inside the group, respectively). It is assumed that the number of master and slave groups is equal to n_g . The grouping of slave nodes and master faces is carried out in a pre-processing stage.

Some incident nodes N_{kr} , represented in Figure 1, may not possess associated target face, therefore the actual domain of $F_{ks}(N_{kr})$ is a subset of S_k . Additionally $F_{ks}(N_{kr})$ is not bijective,

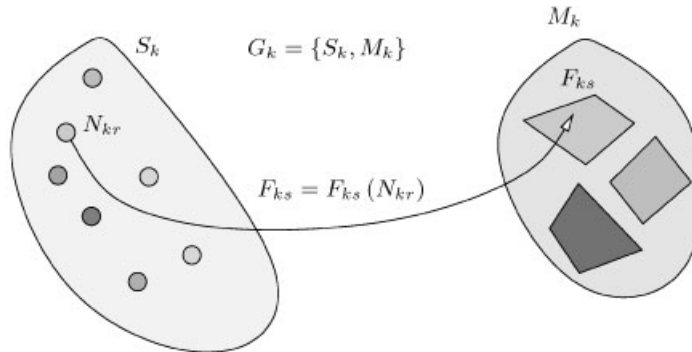


Figure 1. The task of the contact detection algorithm: for each incident node N_{kr} belonging to set S_k , to find a related target face F_{ks} belonging to set M_k . The set $G_k = \{S_k, M_k\}$ is denominated elements group G_k . This operation is repeated for all nodes N_{kr} in each elements group G_k with $k = 1, \dots, n_g$.

Table II. The contact detection algorithm.

1	Determine nodal average normal vectors
2	Define the bucket indexes of global detection
3	Do $k = 1, n_g$: number of groups
4	Do $r = 1, r_k$: number of incident nodes of group G_k
5	Determine the buckets to search for node r
6	Determine the slave node r radius of influence: R_r
7	Do $s = 1, s_k$: number of candidate faces of group G_k
8	Cycle if F_{ks} does not belong to one of the 27 neighbourhood buckets of N_{kr}
9	Cycle if F_{ks} contains any node coincident with N_{kr}
10	Determine the radius of influence of face s : R_s
11	Cycle if the distance between the centre of face F_{ks} and the node N_{kr} (denoted d_{rs}) is larger than the sum of the radius: $R_r + R_s < d_{rs}$
12	Check if node N_{kr} can be associated with face F_{ks} by verifying the violation of local co-ordinates (14), if not, cycle
13	Calculate, by a ranking strategy, the 'cost' of forming the element $N_{kr} \cup F_{ks}$: C_{krs} , which is a function of the gap estimate \bar{g} and of step 12
14	If the 'cost' C_{krs} is less than the previous minimum, C_{krs}^{\min} then $C_{krs}^{\min} \leftarrow C_{krs}$ and mark face F_{ks} for association with node N_{kr}
15	End do
16	If node N_{kr} has some face associated, then define the element $N_{kr} \cup F_{ks}$ forming the new connectivities
17	If node N_{kr} does not have some face associated, then delete the element from the connectivities if it existed previously
18	End do
19	End do
20	If there were changes in the connectivities, then recycle the code's pointers, and reorder degrees of freedom to minimize fill-in

each face F_{ks} can be the image of an arbitrary number of nodes N_{kr} . Furthermore, some faces F_{ks} can be associated with a null set of nodes.

The proposed contact detection algorithm is exposed in Table II.

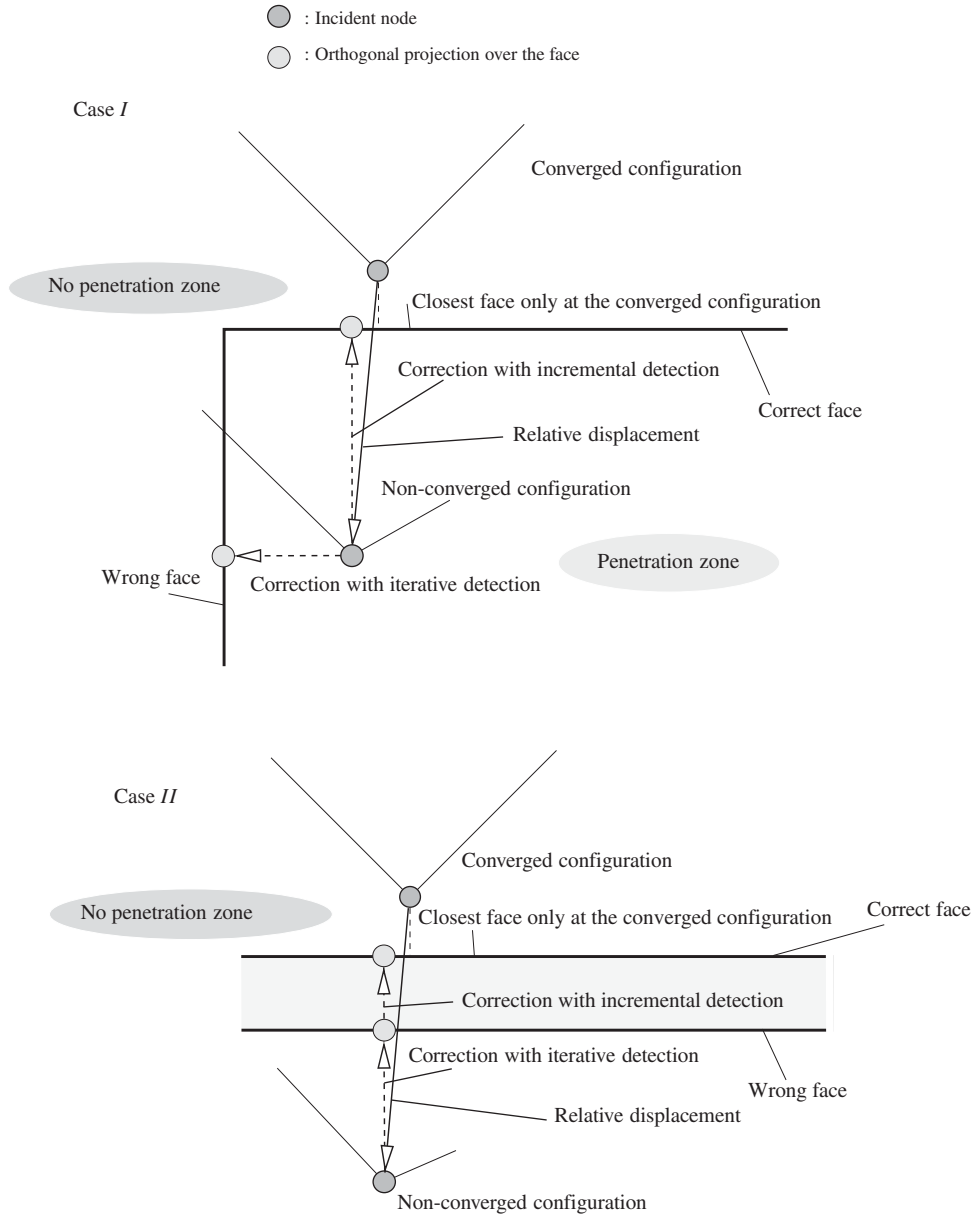


Figure 2. If the contact detection algorithm (see Table II) is applied during iterations, then the wrong (please see above definition) target face can be chosen and incorrect results can be obtained even in these common 2D cases.

The actual contact elements are established after the convergence of a given time step, so there is no possibility of the wrong face being chosen in the sense shown in Figure 2. In this figure, two strong arguments for the proposed algorithm are presented.

It is expected that, if iterative algorithms are used, in 3D cases complex situations can occur with wrong determination of faces. The nomenclature *wrong* is adopted here with the following interpretation:

- The position of the node over the candidate face implies penetration *or*.
- The resulting stepwise trajectory of the incident node implies penetration.

The paradigms represented in Figure 2 are seldom discussed in specialized literature. Both cases I and II can lead to incorrect results.

A closely related subject is the one of choosing the correct target face in the case of equidistance of the incident node relative to 2 faces. With the proposed strategy one of the three following situations can occur:

- The incident node is not violating any constraint and therefore the selection of either face is irrelevant, as no contact force exists after convergence.
- The incident node is violating one of the face's constraint, and therefore the related cost (step 13) is lower and this face will be selected.
- Both faces are violated, and either of each is selected, but information from both may be used in constraint enforcement, as will be described.

In terms of the equidistance dilemma, the arguments as those shown in Figure 2 are valid. Therefore, selecting either face indistinctly, as proposed in Reference [43] '...the (*orthogonal*) projection may fail to be unique... We circumvent these ambiguities by choosing either candidate projection' can lead to incorrect answers. In the same 2D analysis context, Reference [33], p. 154 states: 'In the unlikely event that both distances are equal, either projection is suitable for subsequent constraint calculation'. It is clear from the observation of the picture in this last reference that the incident point is represented *inside* the body, therefore detection is carried out during iterations, or, with the same consequences, convergence has occurred with violation.

Even in the case where the distances are distinct, it is far from obvious that the corresponding face is the correct candidate target face. The equidistance dilemma should therefore be analysed as a particular case of the problem illustrated in Figure 2.

If the orthogonal projection of the incident node onto the face is calculated, the association could be made if the local face co-ordinates satisfy the face constraints. However, the accurate procedure of calculating the orthogonal projection is too expensive for the contact detection stage. The approximate technique to ensure that at most one face is associated with N_{kr} is implemented using the average normal vectors calculated in step 1. The approximate projection of node N_{kr} onto face F_{ks} is carried out using the notation \mathbf{x}_1 for the co-ordinates of node N_{kr} , $\mathbf{x}_2 \dots \mathbf{x}_5$ for co-ordinates of nodes belonging to face F_{ks} . The following equation is solved for η_1 , η_2 and \bar{g} through the Newton–Raphson method:

$$\mathbf{x}_1 - \sum_{k=2}^5 [N_k(\eta_1, \eta_2)\mathbf{x}_k + N_k(\eta_1, \eta_2)\bar{g}\mathbf{n}_k] = \mathbf{0} \quad (14)$$

where N_k represent the quadrilateral shape functions, the terms η_1 and η_2 represent the local co-ordinates, and \bar{g} is an estimate of the gap.

The selection of a given face out of a number of candidates is carried out using a classification based on the gap estimate \bar{g} and the possible violation of the domain $[-1, 1] \times [-1, 1]$ by the

pair of local co-ordinates (η_1, η_2) . The cost C_{krs} in Table II is a function of these values and the previous history.

4. THE GAP FUNCTIONS AND RELATED VARIATIONS IN THE 3D CASE

A widely used constraint function is the so-called signed distance function, based on the definition of the closest projection point (or orthogonal projection in C^1 boundaries) of each point belonging to the *boundary* of body \mathcal{B}_1 into the *boundary* of body \mathcal{B}_2 .

The contact element for the 3D case is identified in Figure 3. In this figure, the vectors \mathbf{n}_k with $k = 2, \dots, 5$ identify average nodal vectors, which represent the contribution of each target face that contains node k . If a given node k is common to nel_k target faces, then:

$$\mathbf{n}_k = \frac{\sum_{i=1}^{nel_k} \mathbf{n}_i^{AV}}{\|\sum_{i=1}^{nel_k} \mathbf{n}_i^{AV}\|} \tag{15}$$

where \mathbf{n}_i^{AV} is a average normal vector to each face i .

If the set $1, \dots, 5$ is known (which is the task of the local detection phase), the orthogonal projection of point 1 in the target face containing nodes $2 \dots 5$ can be determined with the minimization of the function $0.5\|\mathbf{x} - \mathbf{x}_1\|^2$ in the parameterized face:

$$(\xi_1, \xi_2) = \operatorname{argmin}_{\eta_1, \eta_2} (0.5\|\mathbf{x}(\eta_1, \eta_2) - \mathbf{x}_1\|^2) \tag{16}$$

where η_i represent the curvilinear co-ordinates in face $2 \dots 5$.

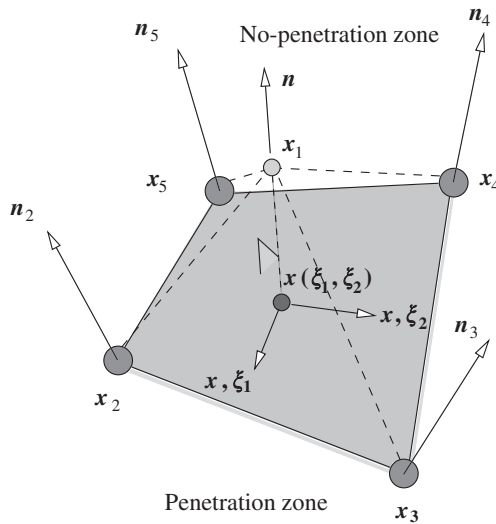


Figure 3. The 3D pentahedral element.

Therefore problem (16) reduces to the following system of non-linear equations in η_1 and η_2 :

$$\begin{aligned}\frac{\partial \mathbf{x}}{\partial \eta_1} \cdot [\mathbf{x}(\eta_1, \eta_2) - \mathbf{x}_1] &= r_1 = 0 \\ \frac{\partial \mathbf{x}}{\partial \eta_2} \cdot [\mathbf{x}(\eta_1, \eta_2) - \mathbf{x}_1] &= r_2 = 0\end{aligned}$$

this system converges to a local minimum, as $\|\mathbf{x}(\eta_1, \eta_2) - \mathbf{x}_1\|^2$ is unbounded above. Please note that \mathbf{x} is now an orthogonal projection of \mathbf{x}_1 over the target face.

In the present work, use is made of the Newton–Raphson method with a bisection line-search to solve this non-linear system. The derivatives of r_j with $j = 1, 2$ can be calculated as:

$$\frac{\partial r_j}{\partial \eta_k} = 0.25(1 - \delta_{jk})\eta_{m1}\eta_{m2}x_{mi}(x_i - x_{1i}) + \frac{\partial x_i}{\partial \eta_j} \frac{\partial x_i}{\partial \eta_k}$$

An important point in the check for nullity of r_1 and r_2 is the fact that the norm of the tangents $\partial \mathbf{x} / \partial \eta_i$ and the difference $\mathbf{x} - \mathbf{x}_1$ should be taken into account.

A tractable strategy can be used to deal in a simple way with projections outside each current target face. An important point in this strategy is the one of normality of the force. Strategies such as the PCE (planned contact element [44]) and the pinball algorithm (e.g. Reference [40]) or the smoothing of the original geometry [45,46] cause the non-penetration force to contain a tangential component. To detail some particular cases, let us consider the 2D case. In this case two situations can occur: the projection $\mathbf{x}(\xi_1)$ of the incident point with co-ordinates \mathbf{x}_1 is over a line segment $\xi_1 \in [-1, 1]$ and no special treatment is required, or the projection does not belong to the original segment, $\xi_1 \notin [-1, 1]$, and a special treatment is adopted.

The last case is represented in Figure 4, with a representation of what is found to be suitable in terms of point position correction.

Due to correspondence between constraint (the gap function) and its gradient, which is proportional to the corrective force, the gap must be defined according to orientation of the corrective force.

In the 3D case, the local geometry of the problem can be much more complex. A projection of a point with co-ordinates \mathbf{x}_1 can be found over a line segment with convex and concave parts or a node with a complex arrangement of neighbourhood faces. The Figure 5 illustrate two possible situations and the vector \mathbf{n} determined by the proposed algorithm.

Three vectors can be readily defined which are necessary to identify the situation. The face normal vector, \mathbf{n}_f , the average nodal vector \mathbf{n}_n and the difference \mathbf{n}_d between a corrected projection \mathbf{x}^* and the incident point \mathbf{x}_1 . These vectors can be written for the general 3D case as:

$$\mathbf{n}_f = \frac{\partial \mathbf{x} / \partial \xi_1 \times \partial \mathbf{x} / \partial \xi_2}{\|\partial \mathbf{x} / \partial \xi_1 \times \partial \mathbf{x} / \partial \xi_2\|} \quad (17)$$

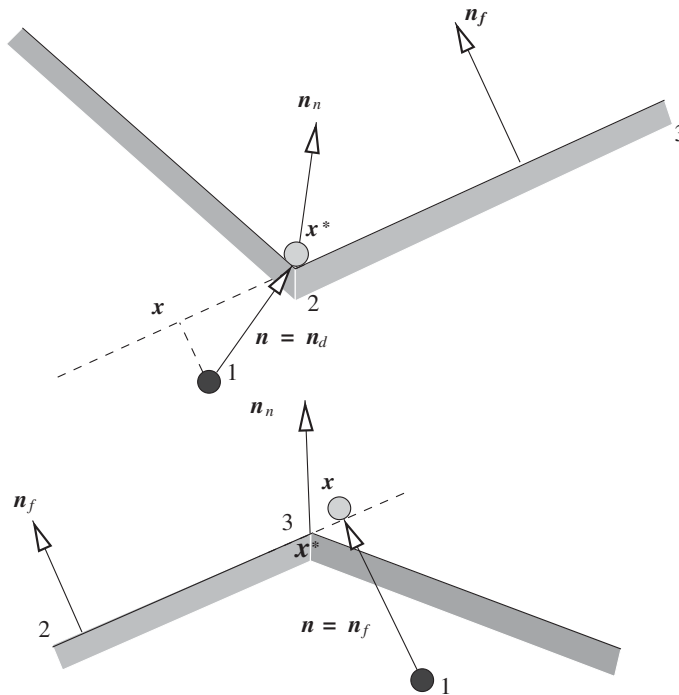


Figure 4. The 2D case: normals in the concave and convex cases. In the two cases the updated position of 1 should be the projection \mathbf{x} along vector \mathbf{n} .

$$\mathbf{n}_n = \sum_{i=2}^5 N_i \mathbf{n}_i \tag{18}$$

$$\mathbf{n}_d = \frac{(\mathbf{x}^* - \mathbf{x}_1)}{\|\mathbf{x}^* - \mathbf{x}_1\|} \tag{19}$$

The notion of corrected projection can be introduced as follows: If the orthogonal projection of \mathbf{x}_1 onto the face $2 \dots 5$ results in curvilinear co-ordinates $(\xi_1, \xi_2) \notin [-1, 1] \times [-1, 1]$ then it is clear that \mathbf{x} no longer belongs to face $2 \dots 5$ (i.e. the point 1 has a projection in other face or does not possess orthogonal projection in the considered set of target faces).

A set of corrected curvilinear co-ordinates can be defined as:

$$\xi_i^* = \max[-a, \min(a, \xi_i)] \tag{20}$$

where a is a number slightly larger than 1 to ensure that the correction of the curvilinear co-ordinates does not preclude sliding of points between faces in the concave case. The vector \mathbf{n}_d in (19) is evaluated using the corrected curvilinear co-ordinates (20).

$$\mathbf{x}^* = \sum_{i=2}^5 N_i(\xi_1^*, \xi_2^*) \mathbf{x}_i \tag{21}$$

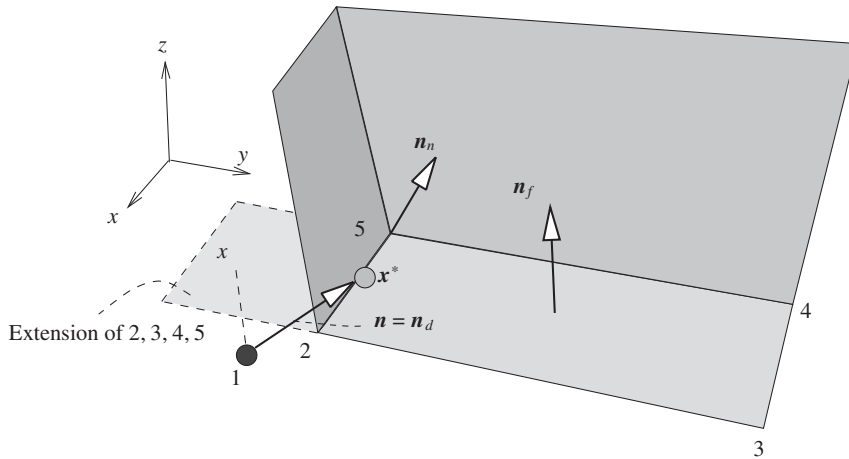


Figure 5. A 3D case of a box corner. In this case a point \mathbf{x}^* in the closest side is selected as projection of \mathbf{x}_1 . If point 1 moves in the direction of x in the negative sense, then \mathbf{x}^* can coincide with the co-ordinates of point 5.

Table III. The determination of contact force direction due to penetration and the gap function.

	$\mathbf{n}_d \cdot \mathbf{n}_f$	$[(\mathbf{I} - \mathbf{n}_f \otimes \mathbf{n}_f)\mathbf{n}_n] \cdot \mathbf{n}_d$	\mathbf{n}	g
Case 1	≤ 0	*	\mathbf{n}_f	$(\mathbf{x}_1 - \mathbf{x}) \cdot \mathbf{n}_f$
Case 2	> 0	≤ 0	\mathbf{n}_f	$(\mathbf{x}_1 - \mathbf{x}) \cdot \mathbf{n}_f$
Case 3	> 0	> 0	\mathbf{n}_d	$(\mathbf{x}_1 - \mathbf{x}^*) \cdot \mathbf{n}_d$

The candidate vector and therefore the direction of the contact force are calculated according to Table III.

The gap function obtained from Table III is always $(\mathbf{x}_1 - \mathbf{x}) \cdot \mathbf{n}_f$ in the case of $(\xi_1, \xi_2) \in [-1, 1] \times [-1, 1]$ and therefore reduces to the classical approach of signed distance gap function (in this situation g is twice differentiable).

However if point 1 does not project onto face 2...5 (which was selected after convergence of the previous time step) and there is penetration of 1 relative to the face 2...5 then two situations can occur, as a function of the classification of point \mathbf{x}^* .

- If the point is classified as a sticking point (Case 2) then \mathbf{x}_1 is constrained to be equal to \mathbf{x}^* . This situation precludes the slipping of the point during the current time step but not during the following time step, as due to the fact the parameter a in (16) is greater than 1, it is ensured that point 1 will belong to a corresponding adjacent face in the next time step.
- If the point \mathbf{x}^* is classified as slipping (Case 3) then the gap function is the signed distance to the current face and the point will belong to an adjacent face in the following time step.

The selection represented on Table III is based upon the assumption that $g \neq 0$. If a penalty strategy is used to model contact, this is assured. However, as a Lagrangian method is adopted, the condition $g = 0$ is approximately verified if convergence criteria are satisfied. To guarantee that the condition $g \neq 0$ is satisfied even if convergence is achieved, instead of satisfying the constraint $g = 0$, another constraint is satisfied:

$$g_2 = g + r_l t_g = 0 \tag{22}$$

where g represents the previously introduced gap, r_l represents a measure of the target face size and t_g is a user-defined tolerance. Alternative strategies for dealing with non-existent normal vectors consist in adopting an average normal vector in the *incident nodes* (see e.g. Reference [47]) or to use a steepest ascent method for the Lagrange multiplier estimates with a node-to-node contact enforcement with an additional strategy for ensuring the existence of the force direction for $g = 0$ (e.g. Reference [48]). Both these strategies contain shortcomings: the first one is inconsistent with the constraint and the second one is restricted to be adopted in a first-order strategy for the multiplier estimates. Due to the fact that each element's target face size r_l depends on the converged co-ordinates alone, then $\delta g_2 = \delta g$.

Finite element applications including a Rockafellar Lagrangian were presented in References [43, 49], however, and according to previous developments, the following constraint potential for $g_2 \geq 0$ is adopted:

$$\Pi_c[\mathbf{x}, \lambda] = \frac{1}{2r} \{ \max[0, \lambda - 1 + e^{-rg_2(\mathbf{x})}]^2 - \lambda^2 \} \tag{23}$$

Near $g_2 = 0$, the Lagrange multiplier λ starts to represent the normal contact force. The total Lagrangian L_{MR} for one constraint can be written as: $L_{MR}(\mathbf{x}, \lambda) = \Pi(\mathbf{x}) + \Pi_c[g_2(\mathbf{x}), \lambda]$. The first variation of Π_c can be written according to:

$$\delta \Pi_c = \begin{cases} (e^{-rg_2} - e^{-2rg_2} - \lambda e^{-rg_2}) \delta g + \left(\frac{e^{-rg_2} - 1}{r} \right) \delta \lambda, & \lambda - 1 + e^{-rg_2} \geq 0 \\ -\frac{\lambda}{r} \delta \lambda, & \lambda - 1 + e^{-rg_2} < 0 \end{cases} \tag{24}$$

It is noticeable that for the limiting case $\lambda = 1 - e^{-rg_2}$ it is possible to write:

$$\lim_{\lambda \rightarrow (1 - e^{-rg_2})^-} \delta \Pi_c = \lim_{\lambda \rightarrow (1 - e^{-rg_2})^+} \delta \Pi_c = \frac{e^{-rg_2} - 1}{r} \delta \lambda \tag{25}$$

which means that the contact force is defined for every $g_2 \in \mathfrak{R}$. This particular form (24) of the Lagrangian part of the total potential plays a role in the satisfaction of the non-penetration constraints.

The use of second variation of the potential $\Pi_c[g_2(\mathbf{x}), \lambda]$ can be written as:

$$d\delta \Pi_c = \begin{cases} r(\lambda - 1 + 2e^{-rg_2})e^{-rg_2} dg \delta g + (1 - \lambda - e^{-rg_2})e^{-rg_2} d\delta g - e^{-rg_2} \delta g d\lambda, & \lambda - 1 + e^{-rg_2} \geq 0 \\ -\frac{d\lambda \delta \lambda}{r}, & \lambda - 1 + e^{-rg_2} < 0 \end{cases} \tag{26}$$

5. FRICTION FORCE AND REGULARIZATION

Friction force can be expressed as (see Reference [3]) ‘...the force exerted by either of the contacting bodies tending to oppose relative tangential displacement of the other’. This denomination is adopted here. The definition is *not* unique, as according to Reference [50]: ‘...is the tangential force that must be overcome in order for one solid contacting body to slide over another’.

The limiting friction force f_a and the friction force magnitude $f = \|\mathbf{f}\|$ obey the following inequality:

$$f - f_a \leq 0 \quad (27)$$

with \mathbf{f} being the friction force. If large loads are present, plastic zones beneath the interacting asperities will grow and eventually the real area of contact will be close to the apparent area of contact (see also Reference [50]). In bulk forming high normal pressures are usually present, therefore the maximum friction force f_a should be limited by plastic shear.

The limit is such that, including all relevant effects, the resulting friction shear stress must be smaller than the yield shear stress, τ_y :

$$\tau_{\max} = m\tau_y \quad (28)$$

where m is denominated friction *factor*, due to the fact that the real area, A_r , would be always less than the apparent area, A_a [50, 51].

The maximum magnitude of friction force, f_a , can therefore be calculated as:

$$f_a = \min(\mu w, A_a m \tau_y) \quad (29)$$

with μ being the friction coefficient and w the normal load.

Wear effects (damage of surfaces with loss of material) conforming to Archard’s law are simply included (see References [3, 50, 52]) with the volume loss of a given surface, V , being calculated as:

$$V = K \frac{w u_t}{p} \quad (30)$$

with $u_t \geq 0$ being the distance of the relative displacement, p being the surface hardness and K being the wear coefficient. Both abrasive and adhesive wear mechanisms may be accounted with Equation (30) and incorporated in the value of the coefficient K .

In Equation (30) the normal load, w , is assumed to be constant. If w depends upon u_t , then the integration variable u is introduced:

$$V = \frac{K}{p} \int_0^{u_t} w(u) du \quad (31)$$

If the thickness decrease Δh is aimed, then the apparent area, $A_a(u)$ is introduced (this concept will be further inspected):

$$\Delta h = \frac{K}{p} \int_0^{u_t} \frac{w(u)}{A_a(u)} du \quad (32)$$

The difficulty of treatment of the exact Coulomb friction law, with a null tangential behaviour for $f < f_a$, lead several authors to adopt a regularized form of the Coulomb law which has

the side-effect of allowing some quantity of tangential displacement. Arguments supporting experimentally this tangential displacement are given in Reference [53].

Two approaches are commonly adopted for regularizing the Coulomb friction law:

- Direct regularization, with the use of a regularizing function for the relative tangential displacement (see References [54–57]).
- Return mapping with elastoplastic analogy of friction (analysed in Reference [58] and subsequently described in References [21, 33, 59, 60]).

The direct regularization procedure allows a simpler implementation, but the latter technique seems currently more popular, due to a stronger theoretical support [58], and also to the fact that the elastic sticking can be carried out during more than one time step (Reference [61] describes the technique).

In the present work, the elastic stick part of the tangential displacement is approximated by a *single* linear segment. To ensure frame-invariance of the frictional stress (or force), an objective rate can be used (Jaumann stress rate in References [44, 59], the Lie derivative of the tangential force in Reference [60]), or the definition of the tangential force scalar components in the moving surfaces [29, 61, 62].

In this work, a distinct strategy is adopted: the accurate determination of the elastic stick part of the tangential displacement is carried out from which the total tangential force is calculated. Therefore, there is no need of consistently integrating an objective stress rate, at least for isotropic friction laws.

The updating of the tangential displacement vector is carried out with its contravariant scalar components. The tangential stress (or force) is calculated directly as a function of the total tangential displacement.

If two instants are considered, identified as n and $n + 1$ then the tangential displacement vector at instant n can be written as:

$$\mathbf{u}_{t_n} = u^{1_n} \mathbf{t}_{1_n} + u^{2_n} \mathbf{t}_{2_n} \quad (33)$$

where \mathbf{t}_{k_n} represent the tangential vectors defined at the orthogonal projection of point 1 in instant n . The calculation of these vectors can be written as:

$$\mathbf{t}_{k_n} = \frac{\partial \mathbf{x}}{\partial \eta_k} = \left(\frac{\partial N_s}{\partial \eta_k} \right)_n \mathbf{x}_{s_n} \quad (34)$$

where \mathbf{x}_{s_n} are the co-ordinates of local node s at time step n . The N_s shape functions in (34) are calculated at instant n .

The contravariant scalar components, u^{1_n} and u^{2_n} can be therefore calculated as the following dot-product:

$$u^{k_n} = \mathbf{u}_{t_n} \cdot \mathbf{t}^{k_n} \quad (35)$$

where use was made of the definition of \mathbf{t}^{k_n} :

$$\mathbf{t}_{r_n} \cdot \mathbf{t}^{k_n} = \delta_r^k \quad (36)$$

with δ_r^k being the mixed-variant scalar components of the metric tensor.

Assuming that if there is no relative slip between the two instants n and $n+1$ the contravariant components are fixed, and that the local converged co-ordinates ξ_1 and ξ_2 remain fixed, then it follows that:

$$\mathbf{u}_{t_{n+1}}^* = (\mathbf{u}_{t_n} \cdot \mathbf{t}_{r_n}) \mathbf{t}_{k_{n+1}} m_n^{kr} \quad (37)$$

where m_n^{kr} represent the contravariant metric tensor components (in instant n), which can be written as:

$$m_n^{kr} = \mathbf{t}^{k_n} \cdot \mathbf{t}^{r_n} = m_{krn}^{-1} \quad (38)$$

The term $\mathbf{t}_{k_{n+1}}$ in Equation (37) is calculated as (34) but using the spatial co-ordinates of instant $n+1$:

$$\mathbf{t}_{k_{n+1}} = \left(\frac{\partial N_s}{\partial \eta_k} \right)_n \mathbf{x}_{s_{n+1}} \quad (39)$$

Terms m_{km}^{-1} in (38) represent the components of the inverse of the covariant metric tensor matrix. Please note that in Equation (39) the co-ordinates ξ_1 and ξ_2 remain the same as those adopted in Equation (34) which agrees with the assumption of null relative tangential displacement.

The updating formula presented in Equation (37) is purely of geometric nature. Additionally, any relative displacement that might occur due to modification of converged co-ordinates ξ_1 and ξ_2 is not included. This can be included using the following observation:

- The relative tangential displacement between instants n and $n+1$ is due to the modification of the converged co-ordinates ξ_1 and ξ_2 between instants n and $n+1$.

The relative tangential displacement vector must contain a null normal component. The relative tangential displacement vector can be expressed as:

$$\Delta \mathbf{u}_t = [\mathbf{I} - \mathbf{n} \otimes \mathbf{n}] (\mathbf{x}_{1_{n+1}} - \mathbf{x}_{1_n} - N_k \mathbf{x}_{k_{n+1}}) = \mathbf{x}_{k_{n+1}} \Delta N_k \quad (40)$$

with $\Delta N_k = N_{k_{n+1}} - N_{k_n}$ reflecting the change in shape functions values.

In Equation (40) the need for the tangential projection is due to the fact that even if n corresponds to a converged time step where $g \cong 0$, the current iteration of time step $n+1$ may result in $g \neq 0$ and therefore the vector $\mathbf{x}_{1_{n+1}} - \mathbf{x}_{1_n} - N_k \mathbf{x}_{k_{n+1}}$ is not guaranteed to be tangential. The total tangential displacement can therefore be calculated as:

$$\mathbf{u}_{t_{n+1}} = \begin{cases} \Delta \mathbf{u}_t + \mathbf{u}_{t_{n+1}}^* & \lambda - 1 + e^{-rg^2} \geq 0 \\ 0 & \lambda - 1 + e^{-rg^2} < 0 \end{cases} \quad (41)$$

In analogy with the small elastic strain theory of plasticity, the relative displacement (41) represents the sticking part of the tangential displacement, according to the notation:

$$\mathbf{u}_{t_{n+1}} = \mathbf{u}_{t_{n+1}}^T - \mathbf{u}_{t_{n+1}}^P \quad (42)$$

where $\mathbf{u}_{t_{n+1}}^T$ represents the total relative displacement (in practice never calculated) and $\mathbf{u}_{t_{n+1}}^P$ represents the slipping part of the tangential displacement.

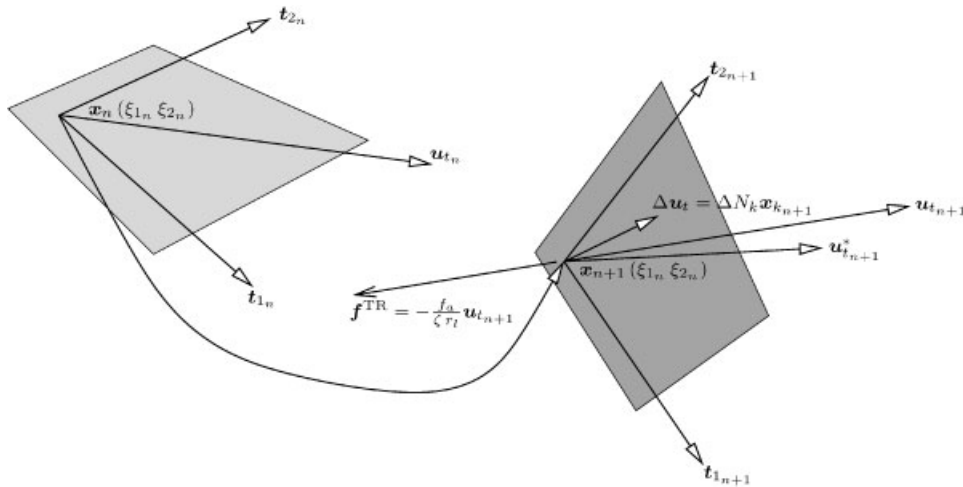


Figure 6. The updating of the elastic sticking displacement between two time steps, n and $n + 1$. If the relative tangential displacement is null ($\Delta \mathbf{u}_t = \mathbf{0}$), then the position of \mathbf{x}_{n+1} is given by the previous converged local co-ordinates ξ_{1n} and ξ_{2n} . Otherwise $\mathbf{x}_{n+1}(\xi_{1n+1}, \xi_{2n+1}) = \mathbf{x}_{n+1}(\xi_{1n}, \xi_{2n}) + \Delta \mathbf{u}_t$.

The friction force, \mathbf{f} , is caused by the sticking part of the relative tangential displacement, through the use of an elastic law:

$$\mathbf{f} = -\frac{f_a}{\max(\zeta r_l, \|\mathbf{u}_{t,n+1}\|)} \mathbf{u}_{t,n+1} \tag{43}$$

where ζ is the non-dimensional relative tangential admissible displacement.

The apparent area, A_a (see Equation (29)) is calculated using the action–reaction principle as:

$$\mathbf{n}^T \boldsymbol{\tau}_n \mathbf{n} A_a = \lambda_n \tag{44}$$

where $\boldsymbol{\tau}_n$ is the Kirchhoff stress tensor extrapolated from integration points to node 1.

Please note that in Equation (44) the Lagrange multiplier is the one representing the normal force at the last converged time step. In a similar way, the Kirchhoff stress tensor in Equation (44) represents the last converged value (Figure 6). This strategy ensures that f_a does not change during the iterations. Certainly, this induces a lack of precision that is subsequently corrected in a set of special friction evaluation iterations, as discussed. This partial decoupling between the normal force, represented by the Lagrange multiplier λ and the friction force f can be thought as a consequence of a well known fact in the friction analysis:

- The normal force, λ , is *explicitly* independent of the friction force, f .
- The maximum friction force magnitude, f_a , depends on the value of the normal force and it is limited by shear plastic flow.

Certain authors have proposed to use converged normal forces values in friction evaluation in the context of penalty method with no limiting friction stress (see Reference [63]) but without subsequent correction. The purpose there was to obtain a symmetric stiffness matrix (a result

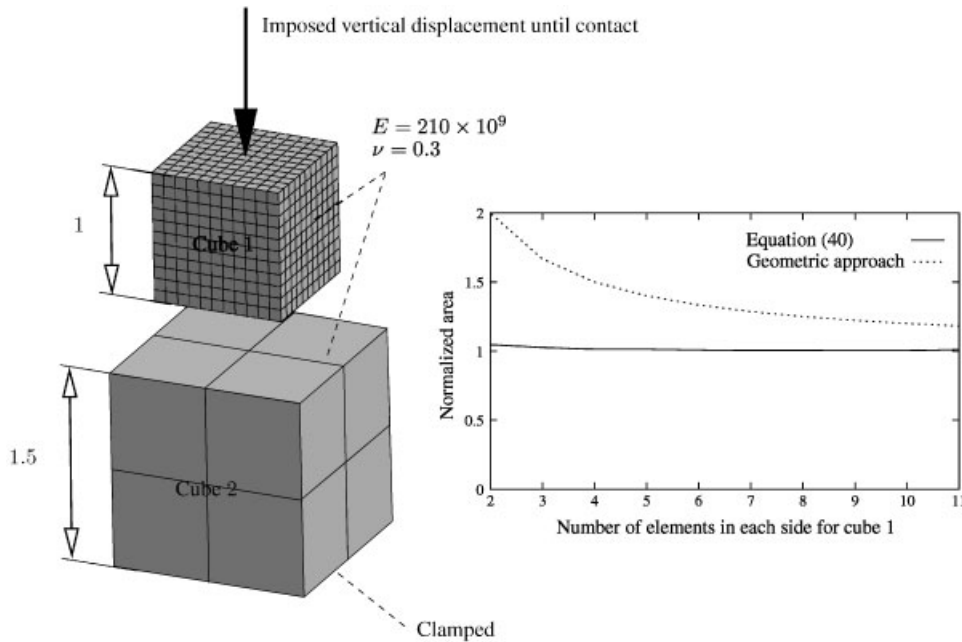


Figure 7. The contact between two deformable cubes along faces. The apparent contact area, A_a is unitary. Cube 1 has a varying number of elements in each side.

that could be obtained with the first order Augmented Lagrangian method [33] or a modification of the penalty contact law [64]).

The fundamental motivation for the use, in each time step, of a constant f_a , is related to the fact that with the particular strategy for the imposition of non-penetration the Lagrange multiplier estimate only has the meaning of normal force once the penetration is very close to zero, therefore at convergence. This is in contrast with the Rockafellar Lagrangian, for example.

In terms of apparent area calculation, the previous purely geometric proposal which is the underlying topic of Reference [65] can lead to poor results. As the apparent area (A_a) is a key aspect of many constitutive friction laws, this topic is particularly important. Additionally, the apparent area calculated as proposed in Reference [65] contradicts the action-reaction principle (40) if a Lagrangian method is adopted.

To inspect what occurs with the apparent area calculated with the present approach, a test is introduced. The test consists on the contact between two deformable cubes, with distinct size, as shown in Figure 7. The accuracy, in this particular case, of both the geometric approach described for the 2D case in Reference [65] and the proposed approach (see Equation (44)) is compared with the unitary analytical solution. If cube 1 contains $n1^3$ elements, then the geometric approach gives an area of $1 + 2/n1$ consistent units. The chart in Figure 7 shows that the present approach gives extremely accurate results for all mesh densities, in contrast with the geometric approach.

To ensure the satisfaction of condition (27), the algorithm in Table IV is adopted.

Table IV. Algorithm for calculating the friction force.

$$\mathbf{f}^{\text{TR}} = -\frac{f_a}{\zeta r_l} \mathbf{u}_{t_{n+1}}$$

$$f^{\text{TR}} = \|\mathbf{f}^{\text{TR}}\|$$

if $f^{\text{TR}} > f_a$ then

$$\mathbf{f} \leftarrow -\frac{f_a}{\|\mathbf{u}_{t_{n+1}}\|} \mathbf{u}_{t_{n+1}}$$

$$\mathbf{u}_{t_{n+1}} \leftarrow \frac{\zeta r_l}{\|\mathbf{u}_{t_{n+1}}\|} \mathbf{u}_{t_{n+1}}$$

else

$$\mathbf{f} \leftarrow \mathbf{f}^{\text{TR}}$$

endif

$$f = \|\mathbf{f}\|$$

The accounting for wear phenomena is based on Equation (32), with w replaced by λ :

$$\Delta h = K \int_0^{u_t} \frac{\lambda(u)}{pA_a(u)} du \tag{45}$$

where $\lambda(u)$ represents the normal load as a function of the tangential displacement u (in the present case the load is given by the Lagrange multiplier λ). Integrating (45) numerically using the backward Euler method, it results:

$$\Delta h_0 = 0$$

$$\Delta h_{n+1} = \Delta h_n + K \frac{\lambda_{n+1} \Delta u_t}{pA_{a_{n+1}}} \tag{46}$$

6. NODAL FORCE VECTOR, RESPECTIVE (APPROXIMATE) JACOBIAN MATRIX AND THE TWO-STEP ALGORITHM

The internal nodal force vector for the pentahedral element can be written as:

$$f_{ki} = \begin{cases} -N_k[n_i(1 - \lambda - e^{-rg_2})e^{-rg_2} + s_i f], & \lambda - 1 + e^{-rg_2} \geq 0 \\ 0 & \lambda - 1 + e^{-rg_2} < 0 \end{cases} \tag{47}$$

$$f_\lambda = \begin{cases} \frac{e^{-rg_2} - 1}{r}, & \lambda - 1 + e^{-rg_2} \geq 0 \\ -\frac{\lambda}{r} & \lambda - 1 + e^{-rg_2} < 0 \end{cases} \tag{48}$$

where $s_i = u_{t_i} / \|\mathbf{u}_t\|$ represent the *stick* direction components.

Table V. The approximate Jacobian matrix of the contact element force vector.

$$\text{if } \lambda - 1 + e^{-rg^2} \geq 0 \text{ then}$$

$$K_{kij} = rN_{k_{n+1}}N_{l_{n+1}}n_i n_j (\lambda - 1 + 2e^{-rg^2})$$

$$-N_{k_{n+1}}N_{l_n} \frac{f_a}{\mathbf{f} \cdot \mathbf{f} \max(\zeta r_l, \|\mathbf{u}_r\|)} f_i f_j H(\|\mathbf{u}_r\| - \zeta r_l)$$

$$+N_{k_{n+1}}N_{l_n} \frac{f_a}{\max(\zeta r_l, \|\mathbf{u}_r\|)} \delta_{ij}$$

$$K_{\lambda j} = N_{l_{n+1}}n_j e^{-rg^2}$$

$$K_{ki\lambda} = N_{k_{n+1}}n_i e^{-rg^2}$$

$$K_{\lambda\lambda} = 0$$

else

$$K_{kij} = K_{\lambda j} = K_{ki\lambda} = 0$$

$$K_{\lambda\lambda} = -\frac{1}{r}$$

endif

Table VI. Schematic algorithm for correction of the friction forces.

Step i —contact detection and pointer recycling
 Update stresses, internal variables, maximum friction forces
 Update load/displacement factor or arc-length
 Iterate to convergence for \mathbf{x} and λ with fixed maximum friction forces—abort and retry with smaller step if convergence does not occur
 Update maximum friction forces with the same load/displacement factor
 Iterate to convergence for \mathbf{x} and λ with fixed (but updated) maximum friction forces—re-start whole step if convergence does not occur
 $i \leftarrow i + 1$ and repeat procedure

An approximation to the Jacobian matrix of the element's forces can be determined as (neglecting the variation of \mathbf{n}) represented in Table V where $H(x)$ represents the Heaviside function of an argument x , and δ_{ij} represent the scalar components of the identity matrix.

This matrix is *exact* in the rigid-deformable case if the rigid body is represented by constant normal faces.

Two iteration loops are used. The first one iterates for both Lagrange multipliers and spatial position co-ordinates of the mesh with fixed maximum friction force, f_a with a load or displacement factor update. The second one iterates for both Lagrange multipliers (which now remain almost constant) and the spatial position co-ordinates with updated maximum friction forces and fixed load/displacement factor, see Table VI.

7. NUMERICAL EXAMPLES

Two goals are aimed with the following set of numerical examples:

- The inspection of the accuracy and robustness of the algorithms regarding both normal interaction and friction.
- The inspection of the accuracy of the 3D elements in problems which are usually analysed with shell elements.

Regarding the finite element technology employed, enhanced assumed strain elements are used, based on the 18 internal variable formulation proposed by Areias *et al.* [66]. Various material models are employed to represent the material behaviour of the bulk solids (for an accounting of the elastoplastic model employed, please consult References [67, 68]). Another point is that the applicability of the previous framework in the dissipative context is that plasticity algorithms do not allow, in a given point, both loading and unloading during the same time-step. This is equivalent, for each time-step, to use a point-wise deformation plasticity model.

All the examples were carried out in the finite element code SIMPLAS, a FORTRAN 90 code developed by the first author of this work.

7.1. Contact between two cylinders and comparison with Hertz solution

A 2D test of contact between two elastic cylinders is carried out, based on the geometry and properties presented in Reference [69]. The geometry, mesh, boundary conditions and material properties are identified in Figure 8.

The original mesh, adopted in Reference [69], consists of bi-quadratic elements. The element side size in the original mesh is here divided by 2 because bi-linear elements are adopted.

The normal stress contour lines over the deformed geometry are represented in Figure 9. It is clear that accurate results are obtained with the proposed formulation.

To further validate the model with this exercise, a comparison with the results obtained in Reference [69] is carried out. The results in that reference were obtained using the mortar finite element method for mesh compatibility of the contact lines. Figure 10 presents the comparison, where very good agreement with both the above reference and Hertz theoretical solution is observable.

7.2. A 2D friction test

This test was exposed in Reference [55] and represents an important verification test for the 2D analysis of friction. A more recent variant of this test (with a finer low order mesh) has been carried out by Armero and Petocz [70]. This last mesh is here adopted.

The geometry, mesh, boundary conditions and material properties are all represented in Figure 11, along with the final deformed mesh.

As indicated in Figure 11, the inspected zone corresponds to the lower edge of the block, which is in contact with the rough surface.

A comparison of the present results in terms of normal pressures and tangential stresses in the inspected zone is accomplished. Results obtained in References [55, 70] are also presented. Figure 12 shows the comparison.

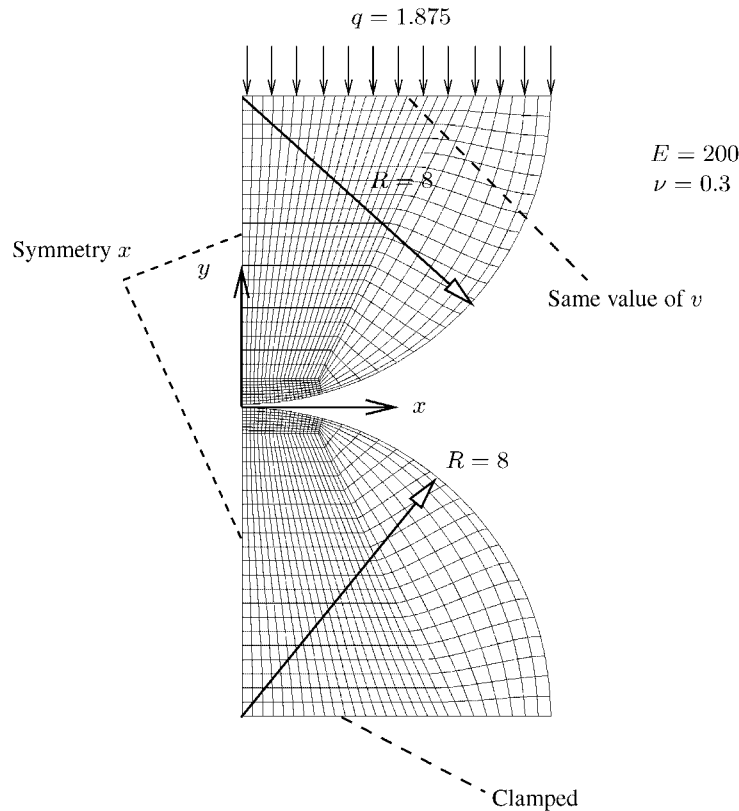


Figure 8. Contact between two elastic cylinders. Mesh of 2D plane strain elements.

7.3. Boot seal test: comparison of the Rockafellar Lagrangian and the proposed L_{MR} in terms of numerical efficiency

This 3D test consists in the analysis of the rotation motion of a shaft inside a boot seal and represents a benchmark for deformable-deformable 3D contact procedures in commercial finite element codes (see for example Reference [71]). An analogous problem, although with a distinct geometry, was analysed by Bathe and Bouzinov [12], with the purpose of illustrating the robustness of their algorithm.

Here, the test's purpose is to verify both the correctness of the detection algorithm and the efficiency of the present approach. For comparison, the penalty method and the Lagrange–Newton method with a Rockafellar Lagrangian are here also tested. The number of iterations and time-steps are monitored. Relative to Reference [71] two differences exist: the steel tube is hollow and it is deformable. This fact adds considerable difficulty to the solution as the problem becomes a truly deformable-deformable situation.

The problem geometry consists on two objects: a deformable circular tube made of steel: $E = 2.1 \times 10^5 \text{ N/mm}^2$ and $\nu = 0.3$ with an internal diameter of $\phi_i = 20 \text{ mm}$ and outer diameter of $\phi_o = 28 \text{ mm}$. The total length of the tube is $L_{TUBE} = 150 \text{ mm}$ and it is positioned

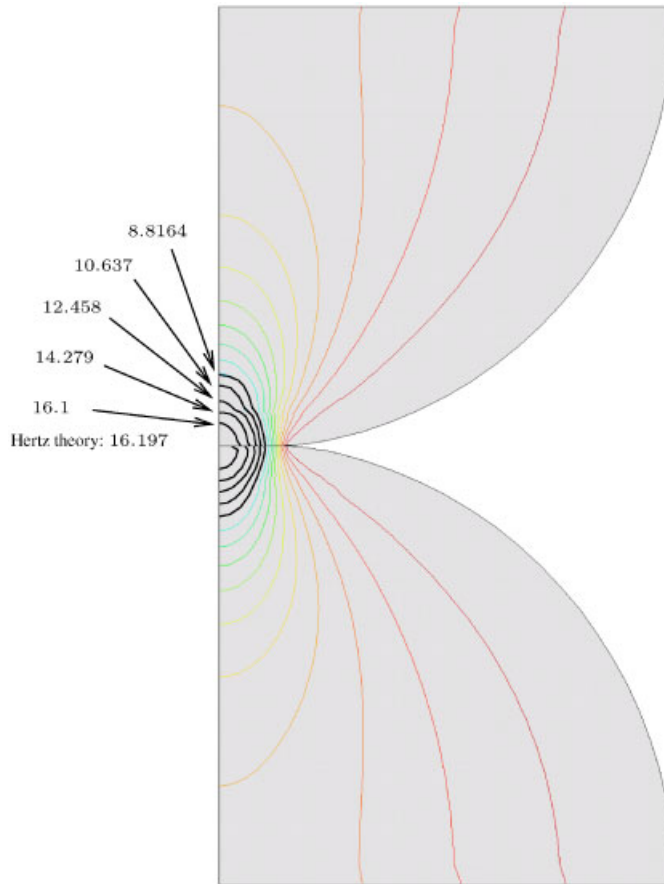


Figure 9. Contact between two elastic cylinders. Normal pressure contour lines.

33 mm below the boot seal. The second object is a boot seal whose geometry and dimensions correspond those described in Reference [71]. It is made of rubber, and a Mooney–Rivlin material model was adopted to represent the rubber mechanical behaviour. Its material properties are: $C_1 = 0.752 \text{ N/mm}^2$ and bulk modulus $\kappa = 7519.5 \text{ N/mm}^2$ (slightly compressible). The first parameter agrees with the value given in the above reference. As one symmetry plane exists, only half of the geometry is actually meshed. The tube contains 48 hexahedral elements and the boot seal contains 1680 hexahedral elements (12 elements circumferentially and 2 along the thickness). The original mesh and 3 steps of the analysis are presented in Figure 13.

In terms of boundary conditions, the boot seal is clamped in its lower faces and the tube rotates around a straight line perpendicular to the symmetry plane. Additionally, all external surfaces may contact each other: there are 3 incident and 3 target surfaces.

The analysis is actually carried out in 2 parts: the first one consists in a correction of pre-existent gap and is carried out in only one time step, and the second part with rotation of the tube (for convergence analysis a target angle of $\alpha_{\text{TUBE}} = 20^\circ$ is employed). The initial

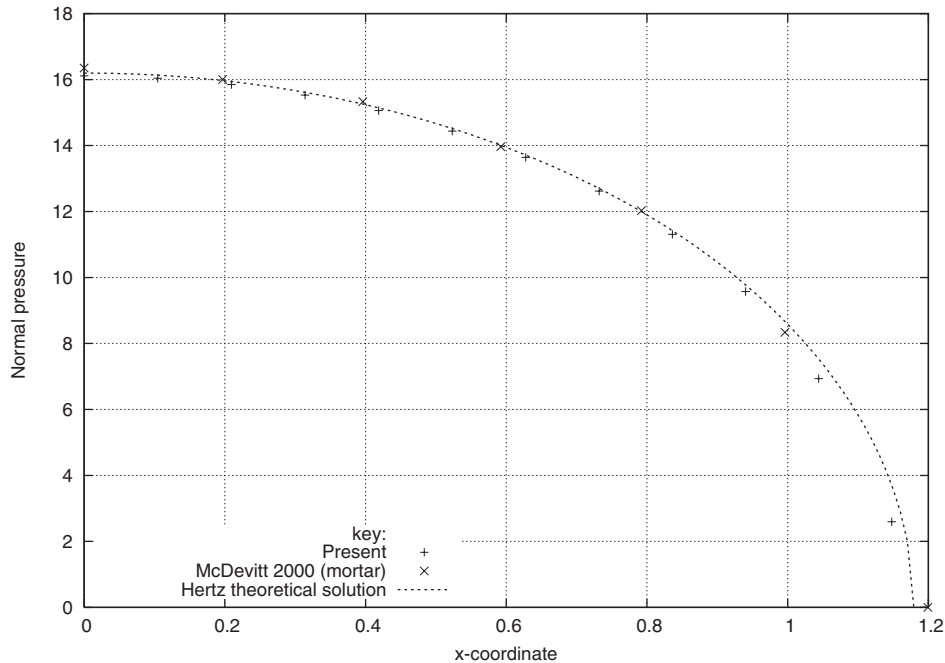


Figure 10. Contact between two elastic cylinders. Comparison with the results of Reference [69] and Hertz solution.

radial interference has a value of 2 mm. The problem is considered frictionless, the normal regularization parameter is $r = 1 \times 10^{-4}$ N/mm and the displacement relative displacement tolerance is fixed at $\text{tol}_D = 1 \times 10^{-6}$. The force *absolute* tolerance is $\text{tol}_F = 1 \times 10^{-3}$ N. For the penalty method, which is also employed for comparison, a value of $r_{\text{PENALTY}} = 1 \times 10^2$ N/mm is adopted.

As results presented in Table VII show, the new Lagrangian functional allows very large steps to be carried out without loss of convergence. It is noticeable that the original Rockafellar Lagrangian results in slower convergence. This example shows the improvement brought by the proposed solution in terms of efficiency.

The high rigor of the constraint satisfaction is depicted in Figure 14.

7.4. Verification of the Archard's wear law implementation

This is a simple verification test, where constant normal pressure is applied to a deformable block which is subsequently displaced along one of the diagonals of a larger deformable block. As normal pressure is imposed and constant, the actual wear that occurs due to contact and relative displacement of the blocks can be determined in closed-form. As the selected relative motion causes the upper block to cross diagonally the elements of the lower block, this test allows the accuracy verification of both the numerical integration of the wear law and the relative tangential displacement. In addition, the accuracy of the normal force is also verified.

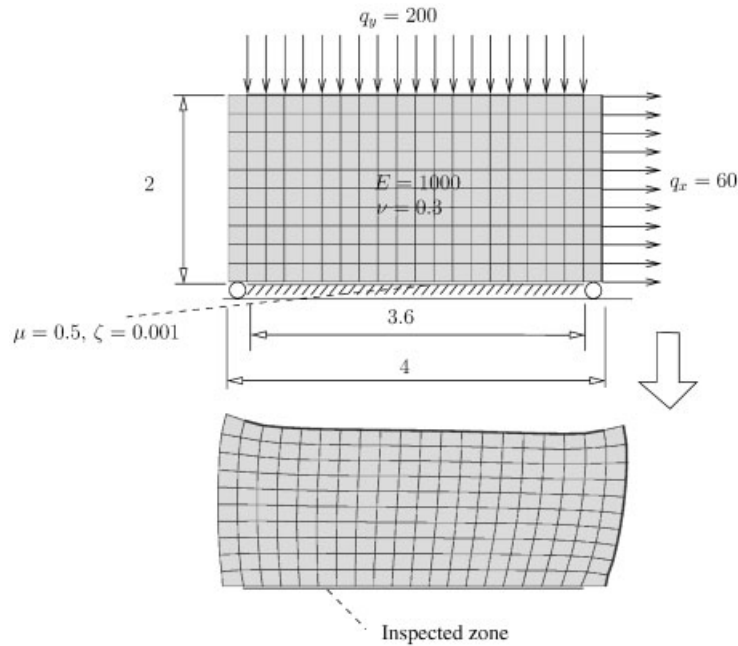


Figure 11. Elastic block pressed against a rough surface and pulled tangentially (see Reference [55] for additional details).

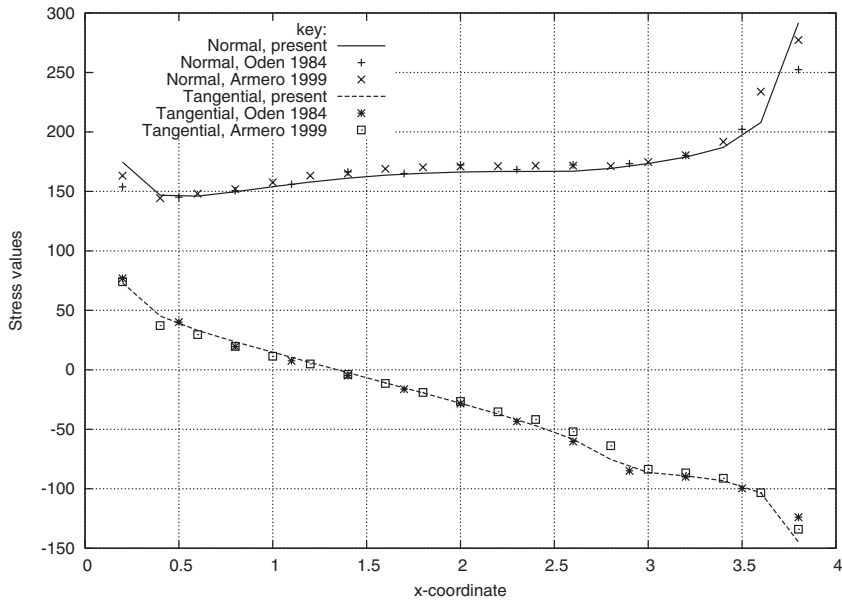


Figure 12. Comparison between the pressures and tangential stresses in the inspected zone. Results from References [55, 70] are also presented.

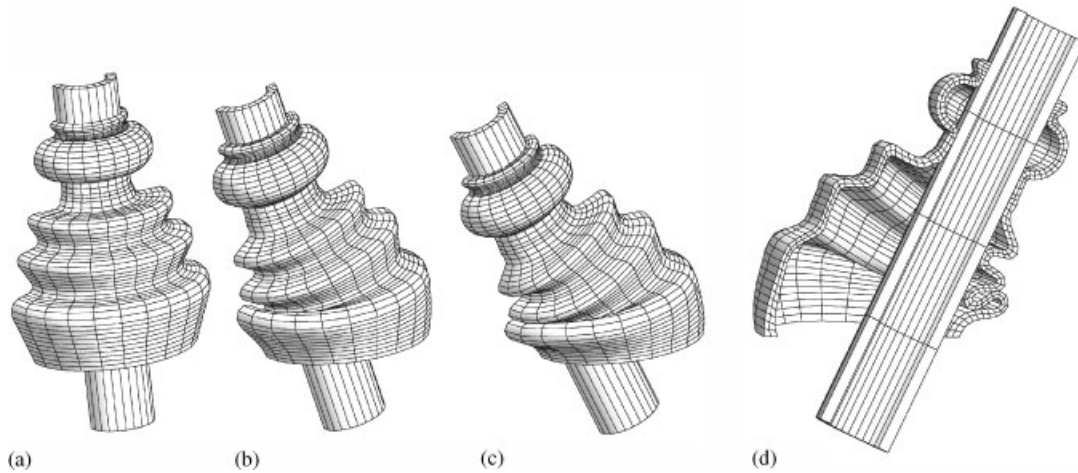


Figure 13. The mesh and steps in the boot seal deformation for 2 values of the rotation angle: (a) $\alpha_{\text{TUBE}} = 0^\circ$; (b) $\alpha_{\text{TUBE}} = 12.47^\circ$; (c) $\alpha_{\text{TUBE}} = 22.21^\circ$; and (d) $\alpha_{\text{TUBE}} = 22.21^\circ$.

Table VII. Boot seal problem: the number of iterations carried out using three methods: penalty method (P), Lagrange–Newton with Rockafellar Lagrangian (LNR) and Lagrange–Newton with proposed Lagrangian (LNMR). Please note that a target value of $\alpha_{\text{TUBE}} = 20^\circ$ is adopted.

	P	LNR	LNMR
Total number of iterations	484	161	70
Total number of time-steps	33	12	7

The geometry, boundary conditions and material properties are presented in Figure 15. The coefficient of wear is taken as $K = 7 \times 10^{-3}$, according to the tabulated value for mild steel in Reference [52]. Additionally, the hardness value is taken as $p = 1822.8$ MPa, also in agreement with the above reference. The problem is considered frictional, with a friction coefficient of $\mu = 0.62$. The smaller block is deformable in the normal direction only, as a measure to ensure that the normal stresses are reasonably constant in the contact area. Figure 15, besides the mesh and boundary conditions, also shows the distribution of the thickness reduction in the smaller block.

It is noticeable from Figure 15 that the accuracy in the calculation of thickness reduction is very high, and the thickness reduction in the contact area is practically homogeneous.

7.5. Axial crushing of a cylinder

This 2D axisymmetric example (see Reference [33] for the relevant data) involves self-contact consists on a cylinder which is forced into a conical die through the application of a downward displacement at its top. After some value of the downward displacement is attained, wrinkles start to form. The combination of axial and bending efforts results in a staircase-like force-displacement plot.

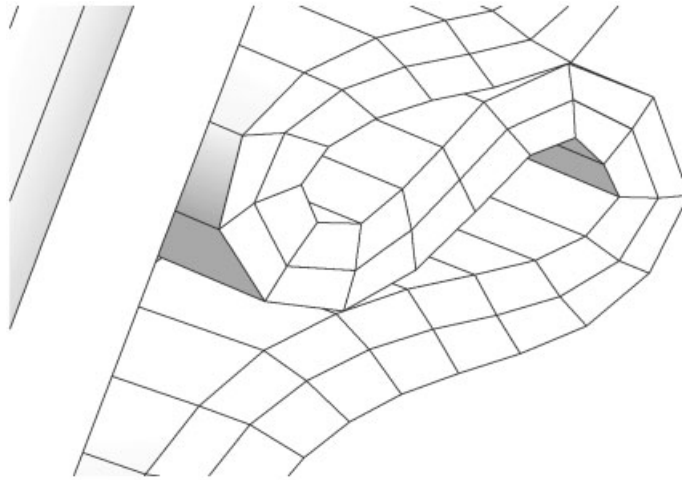


Figure 14. Boot seal test: detail illustrating the rigor in constraint satisfaction.

Two aspects are inspected in this example: the displacement/reaction curves and the number of wrinkles. The dimensions of the cylindrical shell are the following: the internal diameter is $\phi_i = 27$ mm, the external diameter: $\phi_o = 31.75$ mm and the total length of $L_{\text{CYLINDER}} = 180$ mm, the die is considered rigid and the pipe is made of steel with a Young's modulus of $E = 210$ GPa a Poisson coefficient of $\nu = 0.3$, an yield stress given by $\sigma_Y = 700 + 808\varepsilon_p$ MPa.

In terms of contact parameters, the normal regularizing penalty value is taken as $r = 1 \times 10^{-1}$ N/mm, and the non-dimensional regularization parameter for friction is taken as $\zeta = 1 \times 10^{-3}$ and no limit is imposed on the surface shear stress. For the sparser mesh, in the frictionless case, the deformed mesh at 5 distinct time steps, and the effective plastic strain contour lines are plotted in Figure 16. It is interesting to note that in Reference [33] only 3 wrinkles appear, as a much lower value of the total top displacement is imposed.

The comparison of the load-displacement curves with the results presented in Reference [33] is represented in Figure 17. A very close agreement is obtained with the results from the above reference, but one more wrinkle is formed.

7.6. Contact between two square plates

This example was introduced in Reference [72] in the context of testing contact between plates using shell elements. The plates measure 10×10 consistent units and are separated by a distance of 1 between their mid-surfaces. Furthermore, there is a separation of 2 units between a free side of the upper plate relative to the clamped side of the lower plate. The upper plate is subject to a downward pressure which follows the deformation. As the pressure intensity increases, the upper plate touches the lower plate and both move downward. The resistance to the pressure starts to decrease until a snap-through occurs. This fact has not been studied in the above reference and is here inspected. The plates are made of an elastic material with a Young's modulus $E = 4000$ consistent units and a Poisson coefficient of $\nu = 0.3$.

Two tests are carried out: one frictionless and another one with a friction coefficient of $\mu = 0.4$ and no limit on the shear stress. The regularizing penalty is chosen to be $r = 0.1$

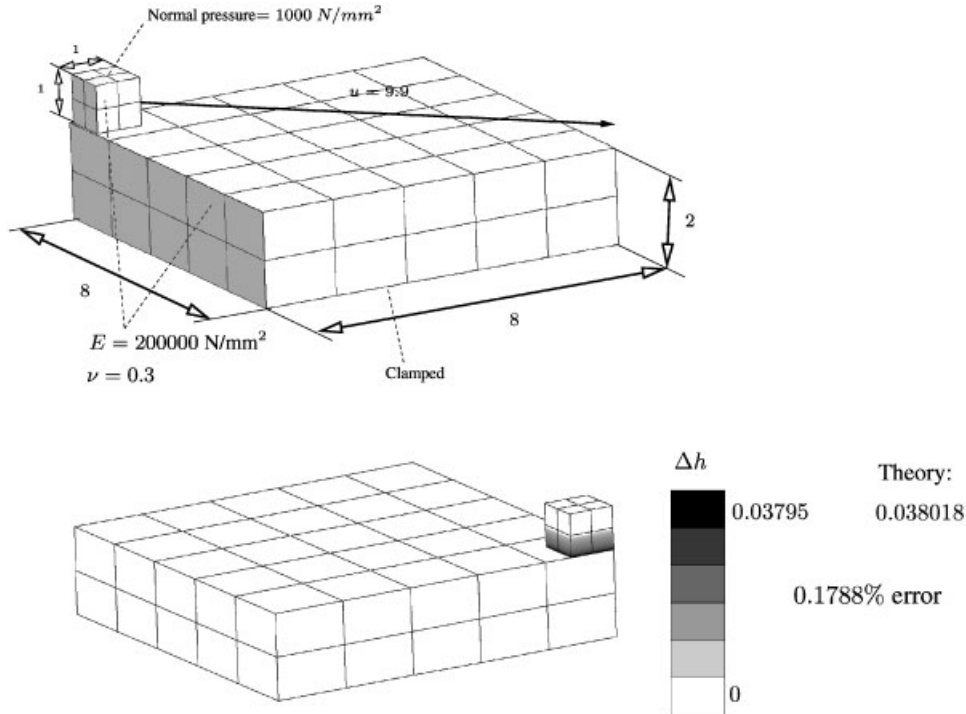


Figure 15. The geometry, mesh and boundary conditions of the wear verification test (all dimensions are in mm). The final distribution of the thickness reduction (Δh) over the deformed mesh is also represented.

and the regularizing parameter for friction is $\zeta = 1 \times 10^{-3}$. A spherical arc-length method is adopted with a linearized treatment if no real roots are obtained. Both the non-deformed and the deformed mesh of the frictional case are presented in Figure 18.

The pressure–displacement curve for the two studied cases is presented in Figure 19. It is clear that the inversion in the sense of pressure rate occurs and this has not been observed in the above reference.

7.7. A 3D variation of the two arches problem

This problem was presented in the recent book by Wriggers [73] and consists in the analysis of two circular arches including a snap-through behaviour which is analysed with an arc-length method. In the present work a variant of the arc-length procedure proposed by Ramm [74] is employed.

A mesh convergence test was carried out in Reference [73], with the conclusion that convergence only occurs with a 14000 element 2D mesh. Here, besides the comparison analysis with the results of the above reference, the inspection of higher load factors is accomplished. The selected mesh, boundary conditions and material properties is presented in Figure 20. Only one-half of the geometry is actually meshed, due to existence of the symmetry plane

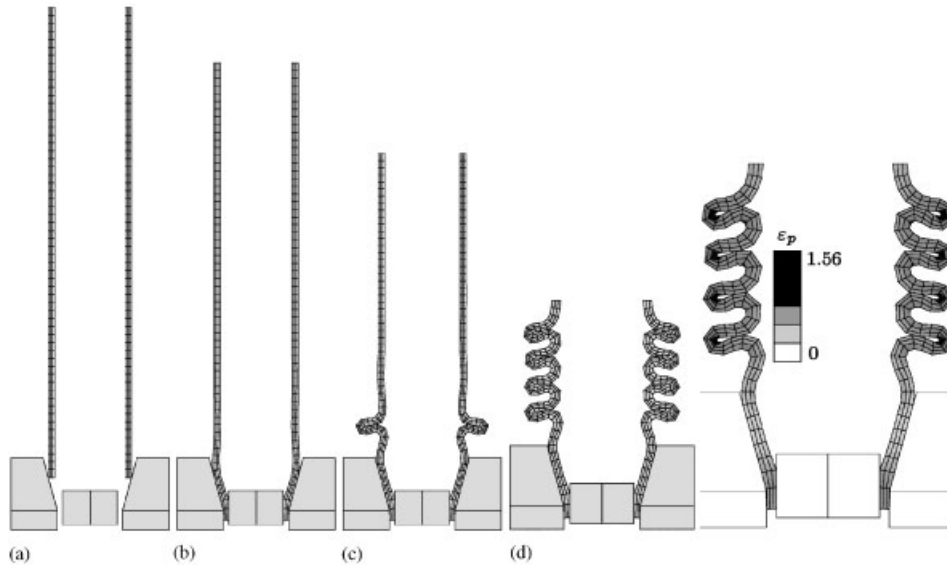


Figure 16. The deformed sparse mesh at 5 distinct time steps, and the effective plastic strain contour plot: (a) initial mesh; (b) displacement into the conical support; (c) first wrinkle; and (d) fourth wrinkle.

represented in the figure. The value of the normal regularization parameter adopted is $r = 0.1$ and no friction was considered.

The vertical displacement of the lower node of the larger arch (point *A* in 20) and the upper node of the smaller arch (point *B* in 20) are monitored. Three distinct time steps in the deformation process are represented in Figure 21. Please note the edge effect in the last 2 steps in the zone of force application, which is due to the plane stress condition near the free edge.

The load factor and the norm of the vertical displacement of point *A* are both represented in Figure 21 for three time steps.

The complete load factor versus displacement plots for both points *A* and *B* are represented in Figure 22. It is noticeable that when the two arches recover contact, the curves present a marked undefined tangent.

For comparison purposes, the previous curve for point *A* is reproduced and compared with the results of Reference [73] and exposed in Figure 23. It is an extraordinary fact that, although the present analysis was carried out with a very sparse mesh in the plane (only 480 elements in the plane) and 3 elements along the depth, the results coincide with a 14 000 plane elements mesh. Furthermore, the analysis is carried out to a much higher level of load (compare Figures 22 and 23). This analysis took only 58 variable size time steps to be concluded.

7.8. Pulling of a plug through a narrowing channel and the role of smoothing

This 3D deformable-deformable example has been proposed in the context of testing a surface smoothing procedure [75] for triangular target faces and tetrahedral elements. It can be described as a displacement of a parallelepiped along a channel with a narrowing open. The parallelepiped

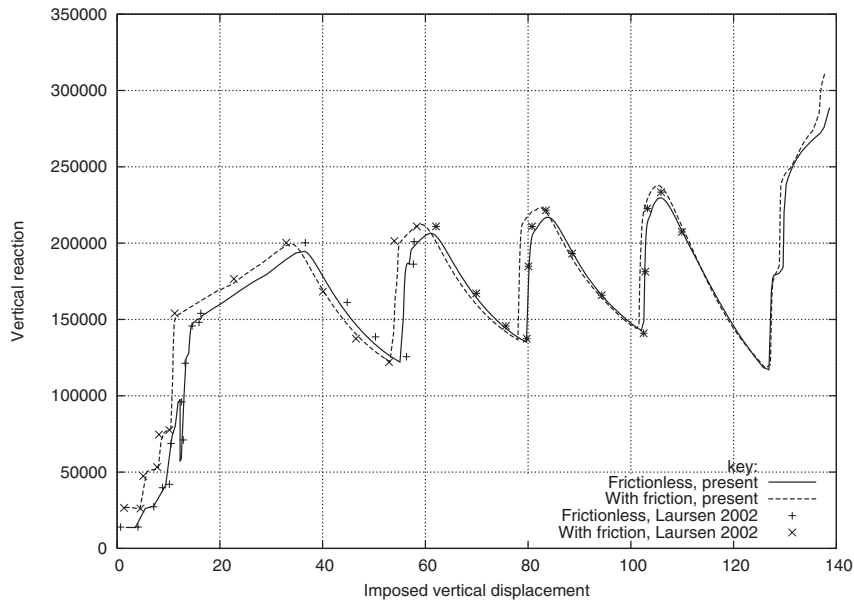


Figure 17. The load-displacement curves: both frictional and frictionless cases are represented and compared with the results of Reference [33].

is subject to imposed displacements in the longitudinal direction of the channel in a given edge. In its initial position, the plug overlaps the channel with the value $\delta = 0.1$. Therefore, initially, the overlapping must be eliminated in a single step, and after that, the plug slides inside the channel.

The purpose of this example in the present work is to show that, contrary to what occurs in the previous reference, the frictional case can be dealt without loss of convergence (and without smoothing strategies). Furthermore, as both the plug material and the channel material are considered hyper-elastic, in the frictionless case the final deformed shape of the plug should be a perfect parallelepiped. The mesh, boundary conditions and material properties are given in Figure 24.

Results are presented in Reference [75] for two situations: one where the coefficient of friction had a value of $\mu = 0.1$ and other without friction. It is an adequate example to test the present approach for *sliding* in both convex and concave arrangement of target faces. The values for the regularization parameters are $r = 1$ and $\zeta = 1 \times 10^{-2}$. The normal σ_{zz} stress component is represented through a contour plot over the deformed mesh in both cases: frictionless and frictional. Figure 25 presents the contour plot over the deformed meshes. In the frictionless case the stress field for that component (and also the others) is completely homogeneous, because the plug remains a perfect parallelepiped.

The difference between the normal stresses at the interface is consequence of a very sparse mesh.

The relation between the longitudinal displacement (\bar{u} in Figure 24) and the corresponding reaction is represented in Figure 26. The figure contains results from Reference [75] and also the theoretical friction force, which was calculated multiplying the friction coefficient by the

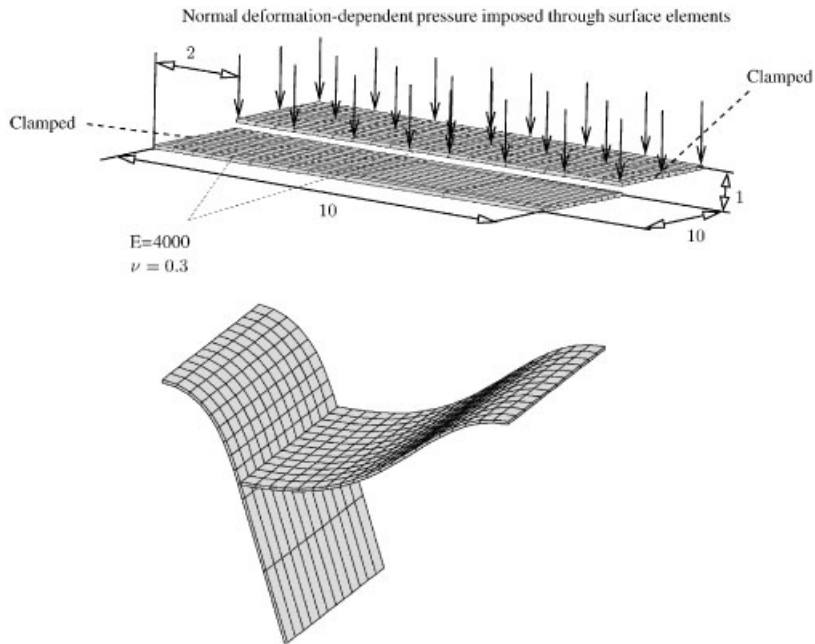


Figure 18. Contact between two square plates. The deformed mesh corresponds to the last step of the frictional case.

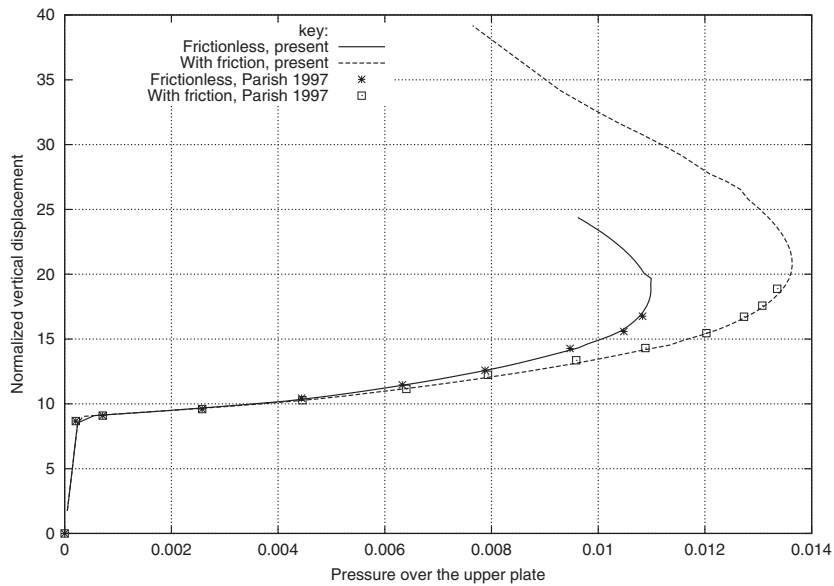


Figure 19. Pressure–displacement plot for the frictionless and the frictional cases, comparison with results obtained in Reference [72].

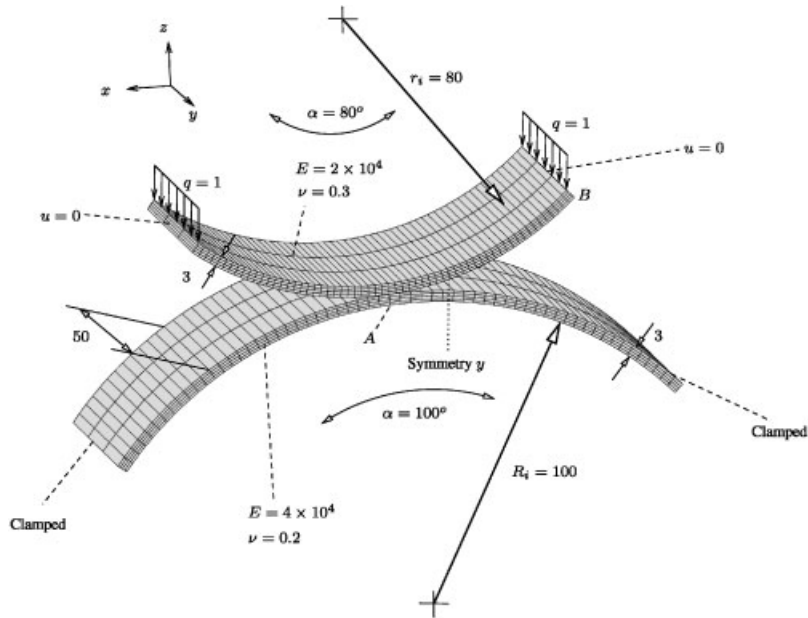


Figure 20. Two arches problem: geometry and mesh, boundary conditions and elastic material properties.

normal reactions. Two conclusions can be drawn:

- The present formulation allows convergence in the frictional case without smoothing. The use of averaged normal vectors induces the presence of tangential forces in the frictionless case, if the plug is sufficiently near the ramp.
- The exact value of the friction force is obtained with the present formulation.

7.9. Forming of an aluminium sheet into a S-rail

This benchmark was part of NUMISHEET 1996 benchmark test *B2* (see Reference [76]) and consists in carrying out the plastic forming of a metal sheet whose final shape should be a S-shaped rail. Tools are considered rigid. Only the aluminium material is tested with a blank holding force of 10 kN. This force is imposed through a bisection method applied to the displacement factor. The sheet thickness is 0.92 mm and the elastic material property values are: $E = 69$ GPa and $\nu = 0.3$. The average value for the friction coefficient is $\mu = 0.1$, in agreement with the above reference.

The yield stress adopted is given by: $\sigma_Y = 518.1(0.01041 + \epsilon_p)^{0.2366}$ MPa. All the data, including the sheet mesh containing elements is detailed in the above reference. Several pictures of the deformed mesh are presented, and wrinkles are found in the experimentally indicated zones in Figure 27.

Finally, the punch force curve is presented, which is found to be very close to average experimental results (please see Figure 28).

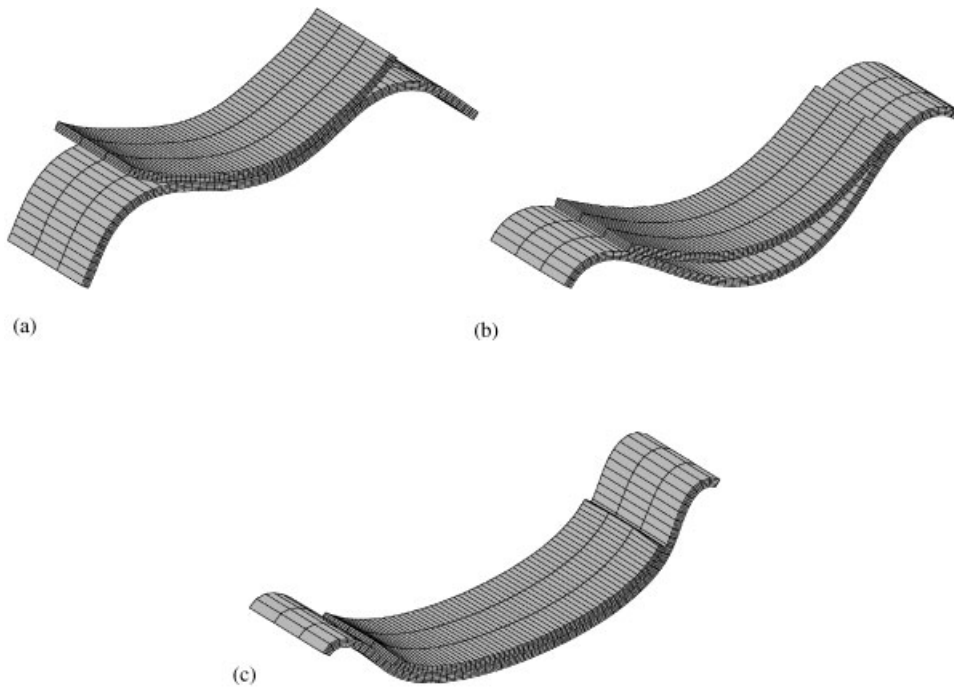


Figure 21. Two arches: three distinct time steps in the deformation process: (a) load factor = 91.23, Disp. A = 25.74; (b) load factor = 192.8, Disp. A = 63.64; and (c) load factor = 1153, Disp. A = 67.66.

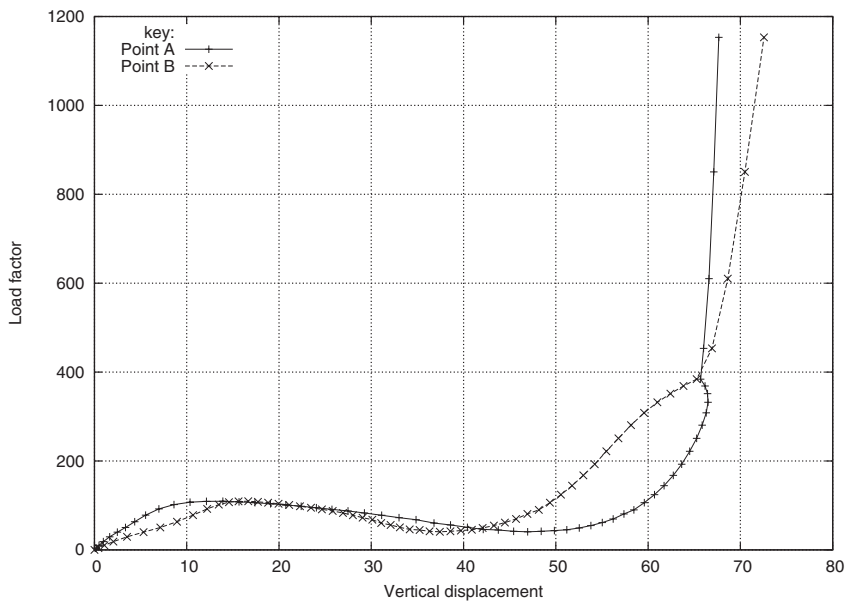


Figure 22. Load factor versus displacement for points A and B represented in Figure 20.

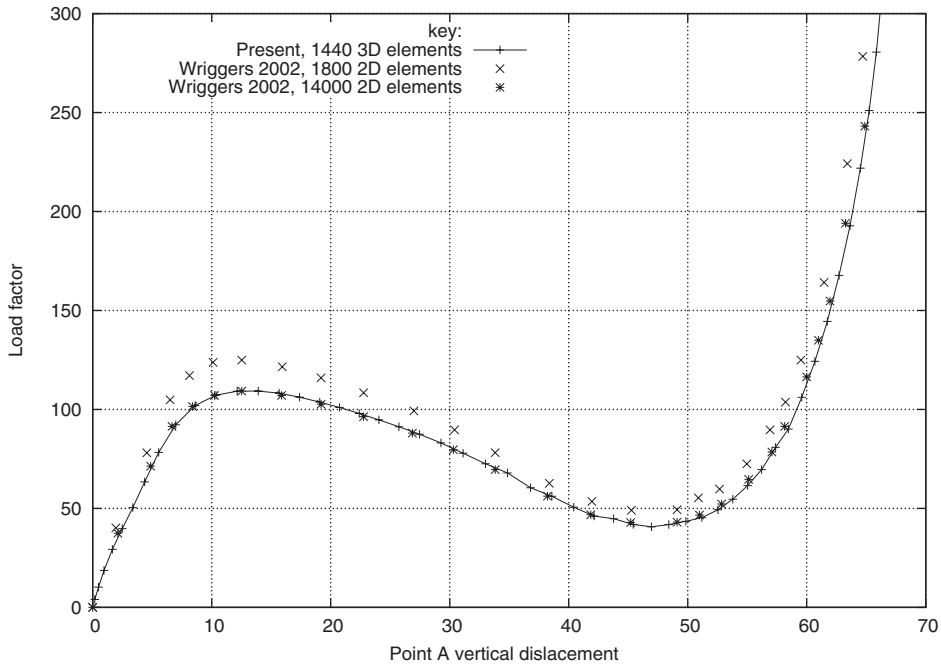


Figure 23. Comparison with the results obtained by Wriggers [73] for the arches problem.

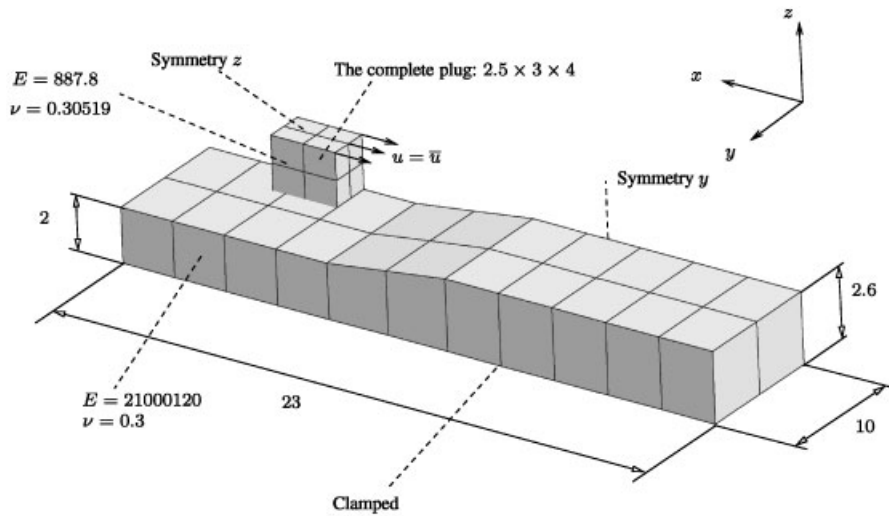


Figure 24. Pulling of a plug: mesh, boundary conditions and material properties. Only one-fourth of the geometry is actually meshed.

Figure 28 presents both the average of explicit and implicit analysis carried out by the participants of NUMISHEET 1996, and also the average, minimum and maximum values for the experiments carried out by the experimental participants in the benchmark. Please note that

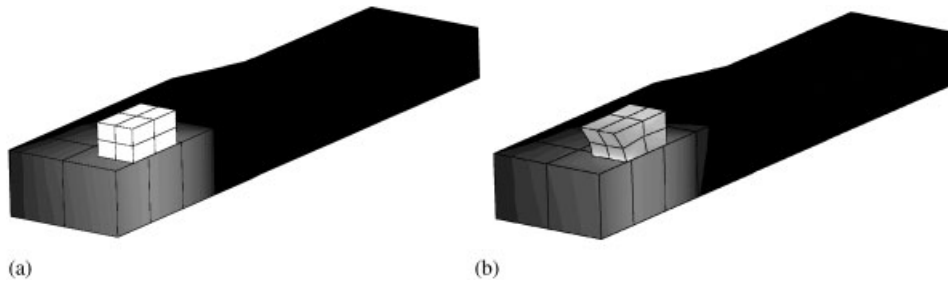


Figure 25. Pulling of a plug: final deformed mesh with the σ_{zz} stress component. Two cases are dealt: frictionless and frictional with $\mu = 0.1$: (a) frictionless case. Black: 0.8832; White: -334.5 ; and (b) frictional case. Black: 5.747; White -397.2 .

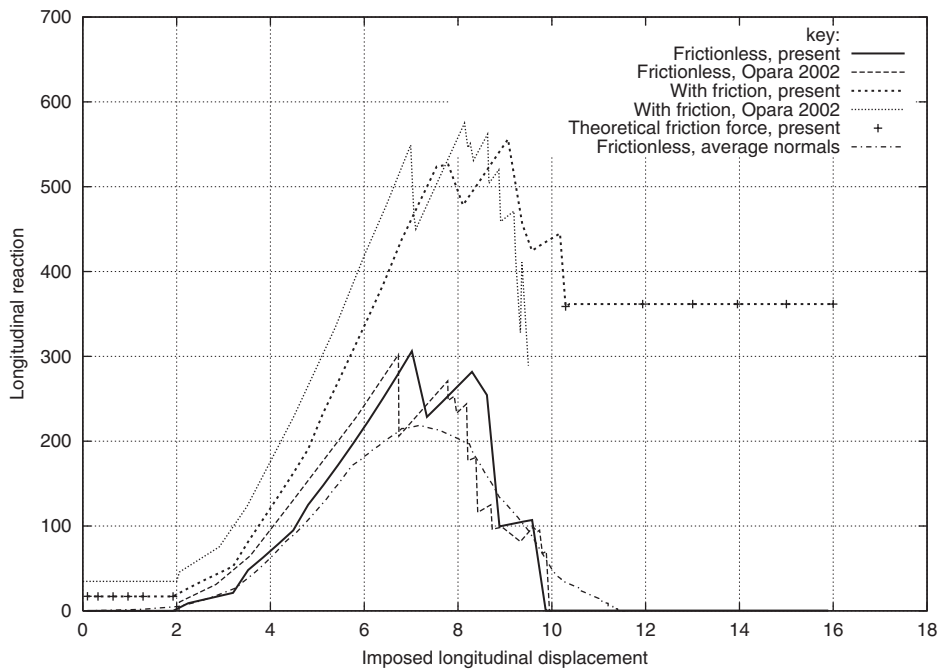


Figure 26. Pulling of a plug: reaction forces as a function of the imposed longitudinal displacement for the frictionless and the frictional cases. For comparison purposes, the results from Reference [75] are also reproduced.

the presented values correspond to all participants in NUMISHEET 1996, which induces a very wide amplitude between the minimum and maximum for the experimental values.

7.10. Spring-back in cylindrical bending

This test was a benchmark relative to the NUMISHEET 2002 conference [77] and was analysed, for example, in Reference [78]. The material and friction properties from this last source are adopted here. A 1 mm thick plate is subject to unconstrained bending through the application of

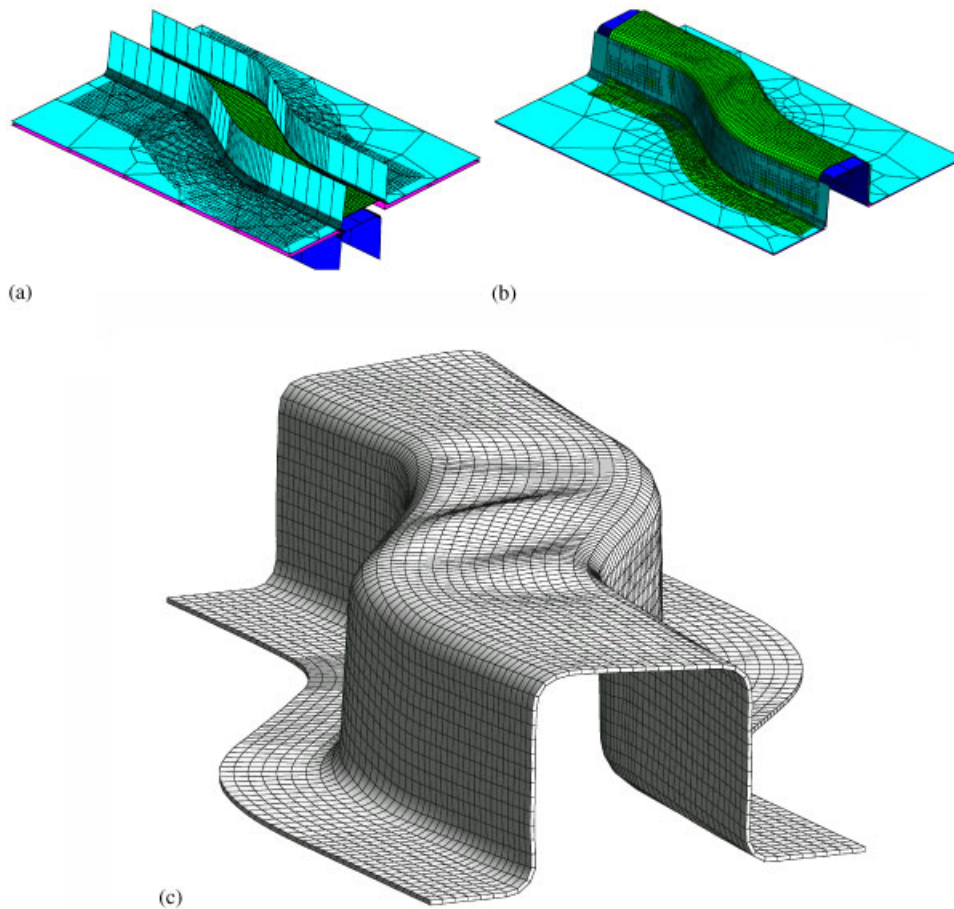


Figure 27. Various pictures of the S-rail forming process—the presence of wrinkles was verified experimentally (see Reference [76]): (a) initial configuration; (b) final configuration with all the tools; and (c) final configuration without the tools.

a cylindrical punch. Both the tool and the die are assumed rigid. The dimensions and motion of tools are the ones described in the above references. For comparison purposes, in this work 456 3D low-order elements are employed in the discretization of the plate. A steel plate is considered, with a Young's modulus of $E = 222.170$ GPa, Poisson coefficient of $\nu = 0.3$ and the following yield stress $\sigma_Y = 645.24(0.0102 + \varepsilon_p)^{0.25177}$ MPa. A friction coefficient of $\mu = 0.14817$ is adopted. The values of the regularizing parameters are $r = 2 \times 10^{-2}$ N/mm and $\zeta = 1 \times 10^{-3}$. Three instances of the loading process are represented, namely: unloaded initial mesh, fully loaded position and final position after spring-back in Figure 29.

The effective plastic strain contour plot is represented in Figure 30.

Several quantities are of concern in this problem: the final spring-back angle (the internal angle between the two arms), the angle prior to unloading, and the punch stroke when separation of the sheet occurs. Table VIII presents the results.

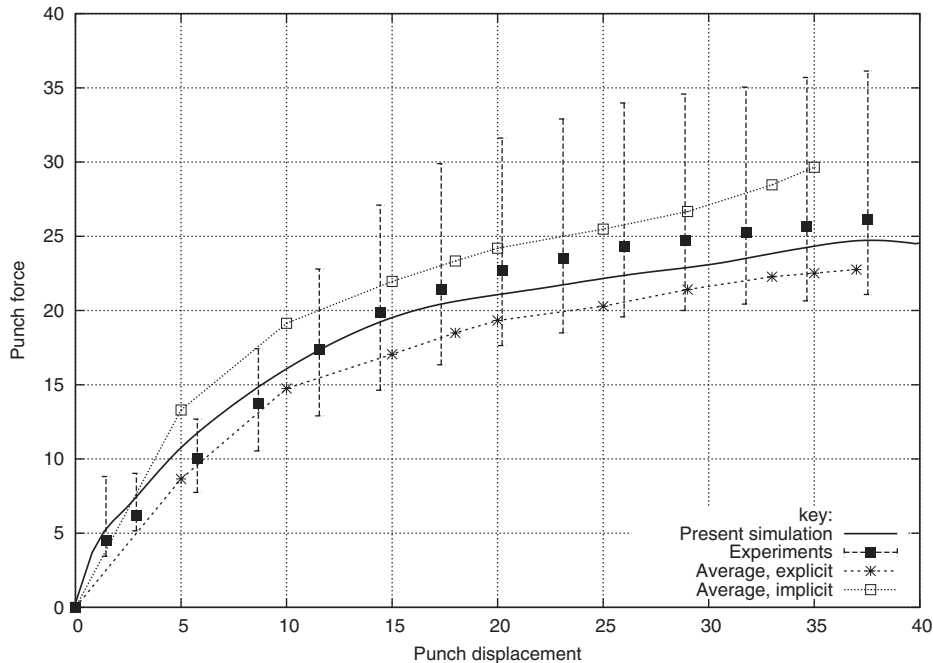


Figure 28. The punch force versus punch displacement for the S-rail problem. Comparison with experiments and results obtained from all the participants in NUMISHEET 1996.



Figure 29. Main time steps in the cylindrical bending.

7.11. V-die bending of a steel sheet

This example is taken from a detailed paper about V-die bending [79] of steel sheets. The present analysis is carried out using 3D elements, whereas in the above reference 2D plane strain elements were adopted. The main purpose of this test is to verify the ability of the proposed algorithms in predicting experimental results in terms of punch force–displacement curve. Information about geometry and boundary conditions can be consulted in the above reference for the problem with nomenclature V1. The material properties are the following: Young's modulus: $E = 205.8$ GPa, Poisson coefficient $\nu = 0.3$ and yield stress given by $\sigma_Y = 720.85(0.06 + \varepsilon_p)^{0.3282}$ MPa. A friction coefficient with value $\mu = 0.1$ is adopted, in agreement with the above reference. The plate has a thickness value of 2 mm. The sheet depth has a value of 100 mm and a width of 50 mm. The punch radius has a value of 6 mm and the die radius has a value of 15 mm. Further details about the geometry and boundary conditions are described in the above reference. A total punch stroke of 16.65 mm is imposed.

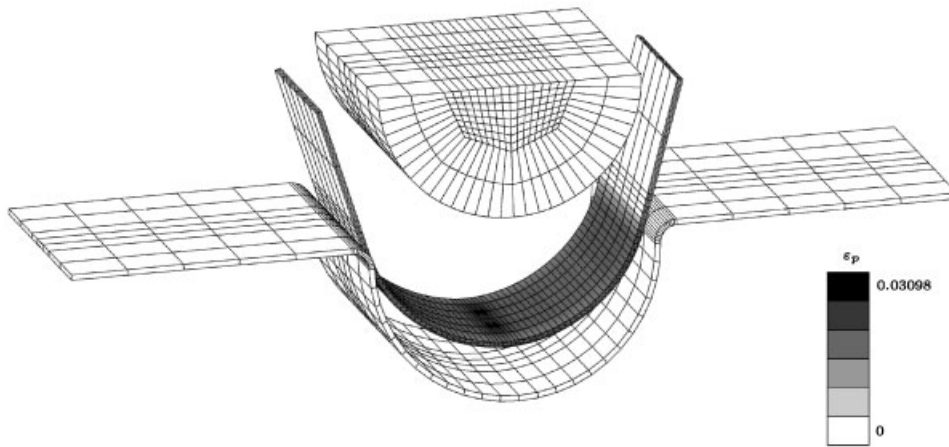


Figure 30. Effective plastic strain distribution *after* spring-back.

Table VIII. Cylindrical bending: comparison with experimental values.

Quantity	Present simulation	Min. experimental	Max. experimental
Angle at full stroke	21.347	20.75	2
Angle after spring-back	33.04	30.9	37.4
Stroke at separation	9.12	3	11.49

In terms of normal regularization a value of $r = 1 \text{ N/mm}$ and frictional regularization $\zeta = 1 \times 10^{-3}$. This problem contains two symmetry planes and therefore only one-fourth of the geometry is actually meshed. Both the non-deformed mesh and the deformed mesh with the effective plastic strain contour plot are presented in Figure 31 for the complete model.

The punch force–displacement curve is illustrated in Figure 32. It is clear, that although both experimental and numerical tests have been carried out by Huang and Leu [79], the present 3D results are more accurate in most parts of the force–displacement plot. The analysis took 88 time steps with automatic step size selection.

8. CONCLUSIONS

In this work, some innovations were introduced that allow a robust analysis of contact between deformable solids. A study and modification of established main steps in the algorithms of contact was carried out. Innermost details regarding both the mathematics of contact enforcement and the friction analysis were derived and the actual practical implementation was detailed.

The main contributions of this work were:

- The inspection of the existent suitable methods to carry out contact analysis with finite elements.

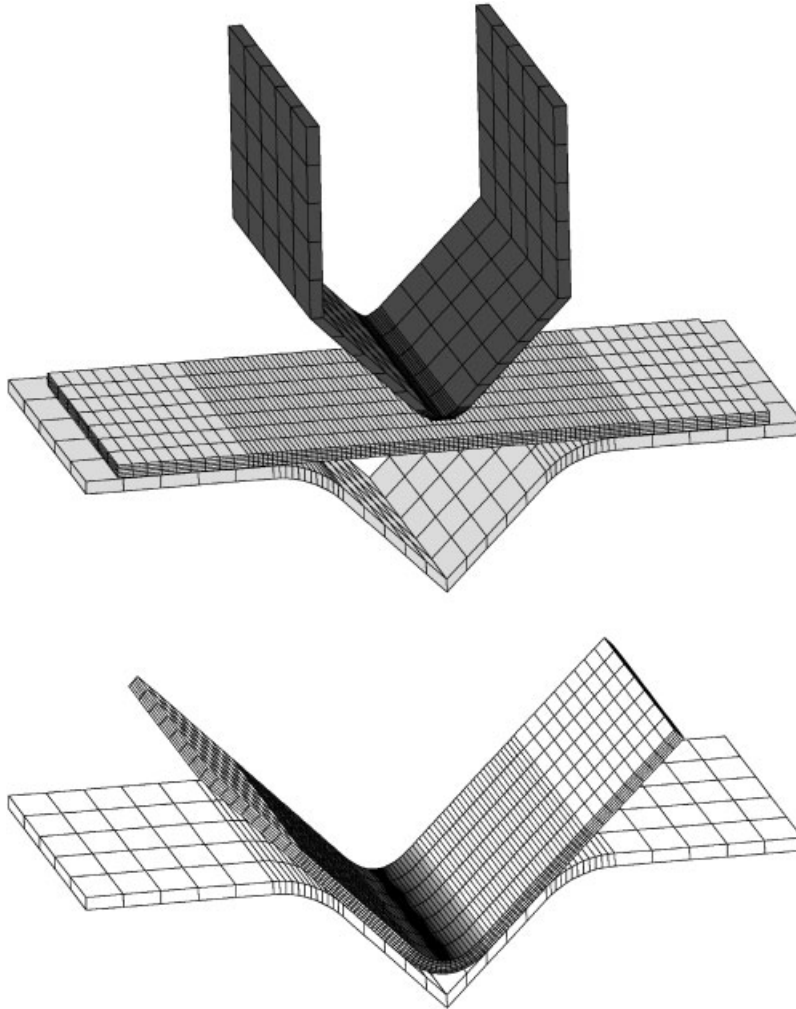


Figure 31. V-die bending of a steel sheet: the non-deformed mesh and the deformed mesh with the effective plastic strain contour plot. Black: $\varepsilon_p = 0.1839$; White: $\varepsilon_p = 0$.

- The investigation of new strategies in contact detection and the circumventing of the equidistance dilemma and face selection, including criteria for avoiding wrong selection of target faces.
- The proposition of a new Augmented Lagrangian function corresponding to a modification of the classical Rockafellar Lagrangian which allows the removal of the Hessian discontinuity for $\lambda_i \geq 1$.
- A new regularization approach for friction forces which does not rely on the selection of a particular stress rate but rather on the use of curvilinear co-ordinates.

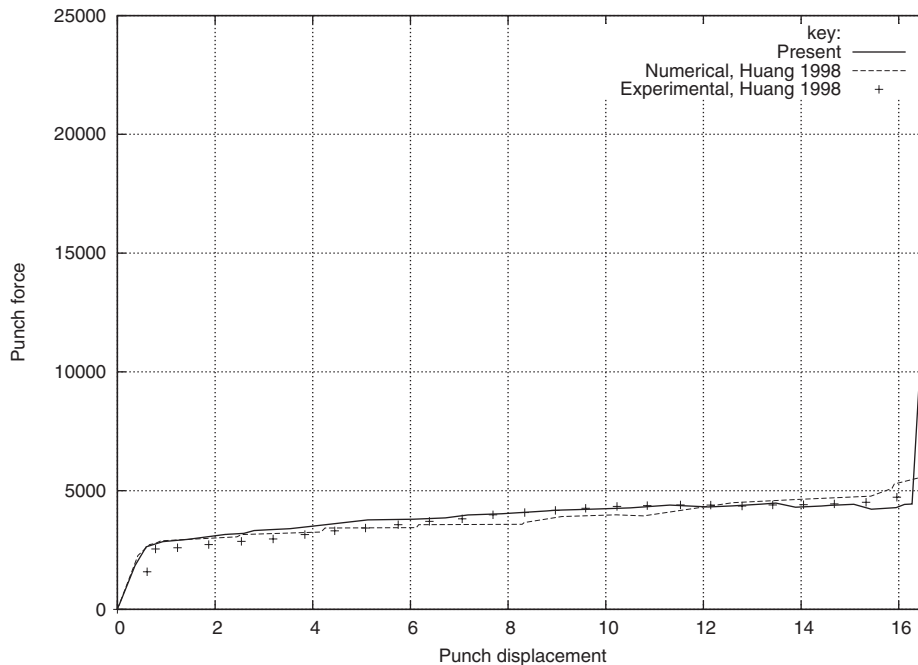


Figure 32. V-die bending of a steel sheet: punch force–displacement plot. For comparison purposes, the experimental and numerical results from Reference [79] are also reproduced.

ACKNOWLEDGEMENTS

The first author would like to acknowledge the financial support provided by ‘Ministério da Ciência e da Tecnologia, FCT’ under Grant PRAXIS XXI/BD/18533/98.

Some examples were carried out in a COMPAQ DS20 Workstation, which was made available for this work by Prof. Pedro P. Camanho at Faculdade de Engenharia da Universidade do Porto. This kindness is gratefully acknowledged.

REFERENCES

- Petersen SB. Radial extrusion of tubes—process modelling and development. *Ph.D. Thesis*, Instituto Superior Técnico, Universidade Técnica de Lisboa, Fevereiro de, 1996.
- Johnson KL. *Contact Mechanics*. Cambridge University Press: Cambridge, 1999.
- Rabinowicz E. *Friction and Wear of Materials*. Wiley: New York, 1965.
- Bowden FP, Tabor D. The friction and lubrication of solids. *The International Series of Monographs on Physics*. Oxford University Press: Walton Street, Oxford OX2 6 DP, 1986.
- Fiacco AV, McCormick GP. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley: New York, 1968.
- Bertsekas DP. *Nonlinear Programming*. Athena Scientific: Belmont, MA, 1995.
- Arora JS, Chahande AI, Paeng JK. Multiplier methods for engineering optimization. *International Journal for Numerical Methods in Engineering* 1991; **32**:1485–1525.
- Boggs BT, Tolle JW. Sequential quadratic programming. *Acta Numerica* 1995; **4**:1–52.
- Bazaraa MS, Sherali HD, Shetty CM. *Nonlinear Programming, Theory and Algorithms* (2nd edn). Wiley: New York, 1993.

10. Chertier O, Chabrand P. Etude du frottement pour des problèmes de contact en grandes déformations. *Revue européenne des éléments finis* 1998; **7**(1–2–3):163–176.
11. Leung AYT, Guiqing C, Wanji C. Smoothing Newton for solving two- and three-dimensional frictional contact problems. *International Journal for Numerical Methods in Engineering* 1998; **41**:1001–1027.
12. Bathe KJ, Bouzinov PA. On constraint function method for contact problems. *Computers and Structures* 1997; **64**(5/6):1069–1085.
13. Areias PMA, César de Sá JMA. Contact between deformable solids: a new second order approach. In Oñate E, Owen DRJ (eds) *VII International Conference on Computational Plasticity*, Barcelona, Spain, April 2003. CIMNE.
14. Barber JR, Ciavarella M. Contact mechanics. *International Journal of Solids and Structures* 2000; **37**: 29–43.
15. Francavilla A, Zienkiewicz OC. A note on numerical computation of elastic contact problems. *International Journal for Numerical Methods in Engineering* 1975; **9**:913–924.
16. Hughes TJR, Taylor RL, Sackman A, Curnier A, Kanoknukulchai W. A finite element method for a class of contact–impact problems. *Computer Methods in Applied Mechanics and Engineering* 1976; **8**:249–276.
17. Boggs PT, Tolle JW. Sequential quadratic programming for large-scale non-linear optimization. *Journal of Computational and Applied Mathematics* 2000; **124**:123–137.
18. Nash SG, Sofer A. *Linear and Nonlinear Programming*. McGraw-Hill International Editions: New York, 1996.
19. Nash SG, Polyak R, Sofer A. A numerical comparison of barrier and modified-barrier methods for large-scale bound-constrained optimization. *Large Scale Optimization: State of the Art*. Kluwer Academic Publishers: Dordrecht, 1994; 319–338.
20. Tseng P, Bertsekas DP. On the convergence of the exponential multiplier method for convex programming. *Mathematical Programming* 1993; **60**:1–19.
21. Kloosterman G. Contact methods in finite element simulations. *Ph.D. Thesis*, Netherlands Institute for Metals Research, The Netherlands, 2002.
22. Rockafellar RT. A dual approach to solving non-linear programming problems by unconstrained minimization. *Mathematical Programming* 1973; **5**:354–373.
23. Kloosterman G, van Damme RMJ, van den Boogaard AH, Huetink J. A geometrical-based contact algorithm using a barrier method. *International Journal for Numerical Methods in Engineering* 2001; **51**:865–882.
24. Biegler L, Schmid C, Terner D. Numerical experience with a reduced Hessian method for large-scale constrained optimization. *Computational Optimization and Applications* 2000; **15**(1):45–67.
25. Byrd R, Nocedal J, Waltz RA. Feasible interior methods using slacks for non-linear optimization. *Otc 2000/11*, Optimization Technology Center, Northwestern University, Evanston, IL, U.S.A., February 2000.
26. Abel JF, Shephard MS. An algorithm for multipoint constraints in finite element analysis. *International Journal for Numerical Methods in Engineering* 1979; **14**:464–467.
27. Byrd RH, Liu G, Nocedal J. On the local behaviour of an interior point method for non-linear programming. *Report Otc 98/02*, Optimization Technology Center, Northwestern University, Evanston, IL, U.S.A., January 1998.
28. Gill PE, Murray W, Wright MH. *Practical Optimization*. Academic Press: New York, 1981.
29. Laursen TA. Formulation and treatment of frictional contact problems using finite elements. *Ph.D. Thesis*, Stanford University, 1992.
30. Laursen TA, Simo JC. A continuum-based finite element formulation for the implicit solution of multibody, large deformation frictional contact problems. *International Journal for Numerical Methods in Engineering* 1993; **36**:3451–3485.
31. Wriggers P. Finite element algorithms for contact problems. *Archives of Computational Methods in Engineering* 1995; **2**(4):1–49.
32. Laursen TA, Oancea VG. On the constitutive modeling and finite element computation of rate-dependent frictional sliding in large deformations. *Computer Methods in Applied Mechanics and Engineering* 1997; **143**:197–227.
33. Laursen TA. *Computational Contact and Impact Mechanics*. Springer: Berlin, 2002.
34. Heinstein MW, Mello FJ, Attaway SW, Laursen TA. Contact-impact modeling in explicit transient dynamics. *Computer Methods in Applied Mechanics and Engineering* 2000; **187**:621–640.
35. Zhi-Hua Zhong. *Finite Element Procedures for Contact-Impact Problems*. Oxford University Press: Walton Street, Oxford, 1993.

36. Hallquist JO. *LS-DYNA3D Theoretical Manual* (2nd edn). Livermore Software Technology Corporation. July 1993.
37. Wang SP, Nakamachi E. The inside-outside contact search algorithm for finite element analysis. *International Journal for Numerical Methods in Engineering* 1997; **40**:3665–3685.
38. Zhong ZH, Nilsson L. A contact searching algorithm for general contact problems. *Computers and Structures* 1989; **33**:197–209.
39. Oldenburg M, Nilsson L. The position code algorithm for contact searching. *International Journal for Numerical Methods in Engineering* 1994; **37**:359–386.
40. Belytschko T, Neal MO. Contact-impact by the pinball algorithm with penalty and Lagrangian methods. *International Journal for Numerical Methods in Engineering* 1991; **31**:547–572.
41. Carleer BD, Atzema EH, Huetink J. Effective and efficient contact modelling in numerical simulations. In Huetink J, Baaijens FPT (eds). *Simulation of Materials Processing: Theory, Methods and Applications*, Balkema: Rotterdam, 1998; 187–193.
42. Kane C, Repetto EA, Ortiz M, Marsden JE. Finite element analysis of non-smooth contact. *International Journal for Numerical Methods in Engineering* 1999; **180**:1–26.
43. Heegard J-H, Curnier A. An augmented Lagrangian method for discrete large-slip contact problems. *International Journal for Numerical Methods in Engineering* 1993; **36**:569–593.
44. Bittencourt E, Creus GJ. Finite element analysis of three-dimensional contact and impact in large deformation problems. *Computers and Structures* 1998; **69**:219–234.
45. Wriggers P, Krstulovic-Opara L, Korelc J. Smooth C^1 -interpolations for two-dimensional frictional contact problems. *International Journal for Numerical Methods in Engineering* 2001; **51**:1469–1495.
46. Belytschko T, Daniel WJT, Ventura G. A monolithic smoothing-gap algorithm for contact-impact based on the signed distance function. *International Journal for Numerical Methods in Engineering* 2002; **55**:101–125.
47. Keum YT, Nakamachi E, Wagoner RH, Lee JK. Compatible description of tool surfaces and FEM meshes for analyzing sheet forming operations. *International Journal for Numerical Methods in Engineering* 1990; **30**:1471–1502.
48. Laursen TA. A note on the treatment of frictionless contact between non-smooth surfaces in fully non-linear problems. *Communications in Numerical Methods in Engineering* 1994; **10**:869–878.
49. Pietrzak G, Curnier A. Large deformation frictional contact mechanics: continuum formulation and augmented Lagrangian treatment. *Computer Methods in Applied Mechanics and Engineering* 1999; **177**:351–381.
50. ASM International. *Friction, Lubrication, and Wear Technology*, vol. 18. ASM International: Metals Park, 1992.
51. Wagoner RH, Chenot J-L. *Fundamentals of Metal Forming*. Wiley: New York, 1997.
52. Archard JF. Wear theory and mechanisms. *Wear Control Handbook*. ASME: New York, 1980; 35–80.
53. Oden JT, Martins JAC. Models and computational methods for dynamic friction phenomena. *Computer Methods in Applied Mechanics and Engineering* 1985; **52**:527–634.
54. Campos LT, Oden JT, Kikuchi N. A numerical analysis of a class of contact problems with friction in elastostatics. *Computer Methods in Applied Mechanics and Engineering* 1982; **34**:821–845.
55. Oden JT, Pires EB. Algorithms and numerical results for finite element approximations of contact problems with non-classical friction laws. *Computers and Structures* 1984; **19**(1–2):137–147.
56. Rebelo N, Nagtegaal JC, Hibbitt HD. Finite element analysis of sheet forming processes. *International Journal for Numerical Methods in Engineering* 1990; **30**:1739–1758.
57. Pantuso D, Bathe K-J, Bouzinov PA. A finite element procedure for the analysis of thermomechanical solids in contact. *Computers and Structures* 2000; **75**:551–573.
58. Curnier A. A theory of friction. *International Journal of Solids and Structures* 1984; **20**(7):637–647.
59. Cheng J-H, Kikuchi N. An incremental constitutive relation of unilateral contact friction for large deformation analysis. *Journal of Applied Mechanics* (ASME) 1985; **52**:639–648.
60. Kleiber M, Knabel J, Rojek J. Reliability assessment in metal forming operations. In Mang HA, Rammerstorfer FG, Eberhardsteiner J (eds), *Fifth World Congress on Computational Mechanics*, Vienna, Austria, 7–12 July, 2002.
61. Ibrahimbegovic A, Wilson EL. Unified computational model for static and dynamic frictional contact analysis. *International Journal for Numerical Methods in Engineering* 1992; **34**:233–247.
62. de Saracibar CA. A new frictional time integration algorithm for large slip multi-body frictional contact problems. *Computer Methods in Applied Mechanics and Engineering* 1997; **142**:303–334.

63. Garino CG, Ponthot J-P, Boman R, Oliver J. A symmetric tangent operator recovery via quasi-Coulomb model for friction contact between solids in large deformation analysis. In *European Conference on Computational Mechanics*, Munchen, Germany, 31 August–3 September 1999, ECCM99.
64. Ju SH, Rowlands RE. A three-dimensional frictional contact element whose stiffness matrix is symmetric. *Journal of Applied Mechanics* (ASME) 1999; **66**:460–467.
65. Stupkiewicz S. Extension of the node-to-segment contact element for surface-expansion-dependent contact laws. *International Journal for Numerical Methods in Engineering* 2001; **50**:739–759.
66. Areias PMA, César de Sá JMA, Conceição António CA, Fernandes AA. Analysis of 3D problems using a new enhanced strain hexahedral element. *International Journal for Numerical Methods in Engineering* 2003; **58**:1637–1682.
67. Areias PMA, César de Sá JMA, Conceição António CA. A gradient model for finite strain elastoplasticity coupled with damage. *Finite Elements in Analysis and Design* 2003; **39**:1191–1235.
68. Areias PMA, César de Sá JMA, Conceição António CA. Finite strain plasticity including softening: a gradient damage model. In Mang HA, Rammerstorfer FG, Eberhardsteiner J (eds). *Proceedings of the Fifth World Congress on Computational Mechanics (WCCM V)*, <http://wccm.tuwien.ac.at>, 7–12 July 2002, Vienna University of Technology, Austria (ISBN 3-9501554-0-6).
69. McDevitt TW, Laursen TA. A mortar-finite element formulation for frictional contact problems. *International Journal for Numerical Methods in Engineering* 2000; **48**:1525–1547.
70. Armero F, Petocz E. A new dissipative time-stepping algorithm for frictional contact problems: formulation and analysis. *Computer Methods in Applied Mechanics and Engineering* 1999; **179**:151–178.
71. Hibbitt, Karlsson, Sorensen. *Abaqus/Standard Example Problems Manual* (5.7 edn). vol. I, II, Pawtucket, RI, 1997.
72. Parish H, Lubbing CH. A formulation of arbitrarily shaped surface elements for three-dimensional large deformation contact with friction. *International Journal for Numerical Methods in Engineering* 1997; **40**:3359–3383.
73. Wriggers P. *Computational Contact Mechanics*. Wiley: New York, 2002.
74. Ramm E. *The Riks/Wempner Approach—An Extension of the Displacement Control Method in Nonlinear Analyses*. Recent advances in non-linear computational mechanics. Pineridge Press Limited: Swansea, U.K., 1982; 63–86.
75. Krstulovic-Opara L, Wriggers P, Korelc J. A c^1 -continuous formulation for 3d finite deformation frictional contact. *Computational Mechanics* 2002; **29**:27–42.
76. Lee JK, Kinzel GL, Wagoner RH (eds). *Proceedings of Numisheet 1996*, vol. 1. Dearborn, MI, U.S.A., 1996.
77. Yang D-Y, Oh SI, Huh H, Yong Hwan Kim (eds). *Proceedings of Numisheet 2002*, vol. 1, Jeju Island, Korea, October 2002.
78. Oliveira MC, Alves JL, Menezes LF. Springback evaluation using 3D finite elements. In Yang D-Y, Oh SI, Huh H, Kim YH (eds). *Numisheet 2002 Proceedings of 5th International Conference and Workshop on Numerical Simulation of 3D Sheet Forming Processes*, vol. 1, Jeju Island, Korea, October 2002; 189–194.
79. Huang Y-M, Leu DK. Effects of process variables on v-die bending process of steel sheet. *International Journal of Mechanical Sciences* 1998; **40**(7):631–650.