PERGAMON

Computers and Structures 79 (2001) 1473–1486

Computers & Structures

www.elsevier.com/locate/compstruc

Stability and patch test performance of contact discretizations and a new solution algorithm

Nagi El-Abbasi, Klaus-Jürgen Bathe *

Department of Mechanical Engineering, Massachusetts Institute of Technology, Room 3-356, 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Received 17 November 2000; accepted 30 April 2001

Abstract

Stability and patch test performance are two important issues influencing the selection of discretization algorithms for flexible body contact problems. The patch test performance is dictated by the accuracy of evaluation of the contact integrals, not by the specific form of the gap, contact pressure, or contact surface definitions. Stability on the other hand, is related to the form of the assumed contact pressure distribution. In this paper, several commonly used contact algorithms are briefly summarized, and their stability in an inf–sup test and performance in a contact patch test are assessed. The existing algorithms do not satisfy both requirements. This fact has motivated the development of a new algorithm for flexible body contact problems that results in a symmetric coefficient matrix and satisfies both the stability and the contact patch conditions.

Keywords: Contact; Finite element solution; Stability; Patch test

1. Introduction

In order for any mixed finite element formulation to be attractive, it should satisfy the ellipticity and inf–sup conditions. With these conditions (and of course consistency) satisfied, stability and optimal convergence (within the interpolation assumptions used) are guaranteed [1–3]. In contact problems, the inf–sup condition relates the contact pressure to the no-penetration contact constraint. Obtaining a reliable contact discretization is contingent upon satisfying this inf–sup condition.

The patch test, on the other hand, originally proposed in Ref. [4], investigates whether an assemblage of displacement-based non-conforming elements is complete. The element passes the patch test if an arbitrary patch can represent a state of constant stress. This in turn ensures convergence for displacement-based elements as the mesh is refined. Similarly, patch tests can be used to test the ability of contact formulations to exactly transmit constant normal tractions between two contacting surfaces, regardless of their discretization. One way to pass the patch test is to use an approach employed in fluid-flow/structure interaction analysis. This procedure however, results in a non-symmetric coefficient matrix [5,6]. Algorithms that do not pass the patch test introduce solution errors at the contacting surfaces that do not necessarily decrease with mesh refinement. However, note that a patch test does not provide any information about the stability of an algorithm [1] and is not an issue when contact between a flexible and a rigid body is considered.

In this paper, we briefly summarize several contact algorithms for flexible bodies and discuss their stability as well as their performance in a contact patch test. Our results show that the two requirements – to satisfy the stability and the contact patch conditions – are not related. Finally, a new contact algorithm is presented, which is stable, optimal and passes the contact patch

* Corresponding author. Tel.: +1-617-253-6645; fax: +1-617-253-2275.

E-mail address: kjb@mit.edu (K.J. Bathe).

0045-7949/01/$ - see front matter
PII: S0045-7949(01)00048-7

condition/test. We classify the algorithm as a segment-to-segment procedure since it involves an accurate integration of the contact constraints over the surfaces of the contacting bodies, not just using values at the nodes.

2. Contact formulation

Consider a system consisting of two bodies in contact (Fig. 1). The contact kinematics dictate that for any admissible displacement $\mathbf{v}$, there is no inter-penetration between the bodies, and the contact pressure λ can only be zero or positive. These "normal contact conditions" can be represented as

$$g \geqslant 0, \quad \lambda \geqslant 0, \quad g\lambda = 0 \tag{1}$$

where g is the gap. For the "tangential contact conditions", we assume no resistance to sliding.

The gap g can be defined as

$$g := [\mathbf{x}_2 - \mathbf{x}_1] \cdot \mathbf{N} \tag{2}$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are appropriately selected points on the surfaces of bodies A and B, and $\mathbf{N}$ is a normal vector.

Let V be the Hilbert space of displacements $\mathbf{v}$ of the bodies. For our analysis we use

$$V := \{\mathbf{v} \in H^1; \mathbf{v} = \mathbf{0} \text{ on } \Gamma_D\} \tag{3}$$

where H^1 is the Sobolev space [1]

$$H^1(\text{Vol}) := \{\mathbf{v} \in L^2(\text{Vol}); \partial v_i/\partial x_j \in L^2(\text{Vol})\} \tag{4}$$

and the non-empty closed subset K satisfying the no-penetration contact constraint

$$K := \{\mathbf{v} \in V; g(\mathbf{v}) \geqslant 0 \text{ on } \Gamma_C\} \tag{5}$$

Under reasonable geometric assumptions K is convex.

The gap may be expressed as the sum of an initial gap g_0 (defined on Γ_C) and a normal trace operator γ mapping the functions $\mathbf{v}$ onto the boundary

$$g(\mathbf{v}) = g_0 + \gamma(\mathbf{v}) \tag{6}$$

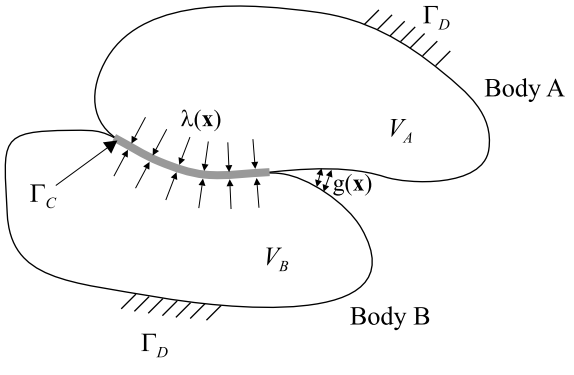

Fig. 1. Two bodies in contact.

where, in general

$$\gamma(\mathbf{v}) : H^1(V) \to H^{1/2}(\Gamma) \tag{7}$$

and $H^{1/2}$ is the usual fractional Sobolev space [7].

We then define the functional $J(\mathbf{v})$

$$J(\mathbf{v}) := \tfrac{1}{2}a(\mathbf{v}, \mathbf{v}) - (\mathbf{f}, \mathbf{v}) \tag{8}$$

where $a : V \times V \to \mathscr{R}$ is the bilinear form of the elasticity problem, and $\mathbf{f}$ is an element of V'. The contact problem can then be expressed as a constrained minimization problem

$$J(\mathbf{u}) = \min_{\mathbf{v} \in K} J(\mathbf{v}) \tag{9}$$

or equivalently as a variational inequality [8,9]

$$\left.\begin{array}{l} \text{find } \mathbf{u} \in K \quad \text{such that :} \\ a(\mathbf{u}, \mathbf{u} - \mathbf{v}) \leqslant (\mathbf{f}, \mathbf{u} - \mathbf{v}) \quad \forall \mathbf{v} \in K \end{array}\right\} \tag{10}$$

Several solution techniques can be used to solve this problem. Lagrange multiplier and penalty methods are the two most commonly used. In the Lagrange multiplier method, the contact traction μ is introduced as an additional unknown to the problem. For this purpose, we define the space

$$M := \{\mu \in H^{-1/2}(\Gamma_C)\} \tag{11}$$

and the convex cone

$$Q := \{\mu \in M; \mu \geqslant 0\} \tag{12}$$

We also define the continuous bilinear form b on $V \times M$

$$b(\mathbf{v}, \mu) := \langle g(\mathbf{v}), \mu \rangle \tag{13}$$

where $\langle ., . \rangle$ denotes the duality pairing between $H^{1/2}(\Gamma)$ and M.

The solution to the contact problem can now be expressed as a mixed variational inequality involving displacements and contact tractions (Lagrange multipliers)

$$\left.\begin{array}{l} \text{find } (\mathbf{u}, \lambda) \in V \times Q \quad \text{such that :} \\ a(\mathbf{u}, \mathbf{v}) - b(\mathbf{v}, \lambda) = (\mathbf{f}, \mathbf{v}) \quad \forall \mathbf{v} \in V \\ b(\mathbf{u}, \mu - \lambda) \geqslant 0 \quad \forall \mu \in Q \end{array}\right\} \tag{14}$$

Other solution techniques for the contact problem, such as the penalty method and the perturbed and augmented Lagrange methods, are closely related to the above Lagrangian formulation [1].

To obtain the discretized form of the Lagrange multiplier formulation of expression (14) we assume finite dimensional subspaces $V_h \subset V$ and $M_h \subset M$, and that Q_h is a closed convex cone in M_h. Then, we obtain

$$\left.\begin{array}{l} \text{find } (\mathbf{u}_h, \lambda_h) \in V_h \times Q_h \quad \text{such that :} \\ a(\mathbf{u}_h, \mathbf{v}_h) - b(\mathbf{v}_h, \lambda_h) = (\mathbf{f}, \mathbf{v}_h) \quad \forall \mathbf{v}_h \in V_h \\ b(\mathbf{u}_h, \mu_h - \lambda_h) \geqslant 0 \quad \forall \mu_h \in Q_h \end{array}\right\} \tag{15}$$

The specific form of V_h depends on the finite elements used, and the form of M_h depends on the contact dis-

cretization. Assuming that the contact surfaces are decomposed into segments Ψ_h (not necessarily matching the finite elements), we can consider the following space

$$\mathscr{M}^r_{h,s} := \{\mu_h \in H^r(\Gamma_h);\ \mu_h|_{\hat{K}} \in P_s(\hat{K})\ \forall \hat{K} \in \Psi_h\} \tag{16}$$

which involves polynomials of degree s, with C^{r-1}-continuity between contact segments.

One of the challenges in contact problems is the prediction of the actual contact surface (the active set of contact constraints). However, the objective of this paper is to assess the stability and patch test performance of different contact algorithms. Therefore, from hereon we assume that the contact surface Γ_C is known. In this case, Eq. (13) becomes

$$b(\mathbf{v},\mu) = \int_{\Gamma_C} \mu g(\mathbf{v})\,\mathrm{d}\Gamma_C \tag{17}$$

and expression (15) reduces to a standard mixed variational equality and the equations resulting from its variation can be assembled in the following matrix form

$$\begin{bmatrix} \mathbf{K}_{(m\times m)} & \mathbf{B}^{\mathrm{T}}_{(m\times n)} \\ \mathbf{B}_{(n\times m)} & \mathbf{0}_{(n\times n)} \end{bmatrix} \begin{Bmatrix} \mathbf{U}_{(m)} \\ \mathbf{\Lambda}_{(n)} \end{Bmatrix} = \begin{Bmatrix} \mathbf{F}_{(m)} \\ \mathbf{G}_{(n)} \end{Bmatrix} \tag{18}$$

where $\mathbf{U}_{(m)}$ is a vector of m displacement degrees of freedom, and $\mathbf{\Lambda}_{(n)}$ contains the n active Lagrange multipliers. The matrix $\mathbf{B}$, which corresponds to the active contact constraints, is used in assessing the solvability and stability of the contact discretization.

For mathematical analysis purposes, Eq. (18) represents the most severe case regarding the constraint condition (compared to penalty and mixed penalty-Lagrange methods). This solution is also most appealing because the contact conditions are exactly enforced. Note that the constraint function method can be used to solve the contact problem without the need for distinguishing between active and inactive contact constraints [1,10,11].

Considering the above discussion, we should mention that conditions similar to those of contact are encountered when domain decomposition techniques are used in parallel processing to ensure that domains are properly "glued" together [12]. Therefore, the mortar method and related techniques can potentially be used to model contact. However, in contact analysis the contact surfaces are generally unknown and possibly involve frictional effects. Hence, an algorithm with local behavior should be used. Also, some of the ideas described in the new contact algorithm presented below are directly applicable to mortar methods.

The contact integral of Eq. (17) is typically evaluated using numerical quadrature. We assume that this quadrature is performed with sufficient accuracy so that in the assessment of stability a numerical integration error can be neglected.

3. Criteria for assessing contact algorithms

In this section we discuss two essential requirements for contact algorithms. The first is stability, which can be represented by an ellipticity condition and an inf–sup condition. Satisfying the ellipticity condition depends on the selection of appropriate finite elements and boundary conditions, not on the contact formulation. Therefore, we devote our attention to the contact inf–sup condition and an inf–sup test. The second requirement pertains to passing a contact patch test. An algorithm passes the patch test if an arbitrary element patch can represent a state of constant contact pressure; hence we could refer to this requirement as the "patch condition". Note that, if any mesh configuration fails the inf–sup test, then the algorithm fails the condition, and the same holds for the patch test. We note that in Ref. [13], an inf–sup condition is used that can result in non-optimal estimates, and the patch condition is not considered.

3.1. Stability

It is necessary to use a contact algorithm that satisfies the stability requirements. This automatically guarantees solvability and convergence [1–3]. The first stability requirement is the ellipticity condition

$$a(\mathbf{v}_h,\mathbf{v}_h) \geqslant \alpha\|\mathbf{v}_h\|_1^2 \quad \forall \mathbf{v}_h \in V_h \tag{19}$$

where α is a constant strictly greater than zero. This condition is satisfied when appropriate finite elements and boundary conditions are used, and is not influenced by contact. The second requirement is the inf–sup condition, which follows the general form for constrained mixed formulations [2]

$$\sup_{\mathbf{v}_h\in V_h} \frac{b(\mathbf{v}_h,\mu_h)}{\|\mathbf{v}_h\|_1} \geqslant \beta \sup_{\mathbf{v}\in V} \frac{b(\mathbf{v},\mu_h)}{\|\mathbf{v}\|_1} \quad \forall \mu_h \in M_h \tag{20}$$

with β a constant strictly greater than zero.

Using the dual norm

$$\|\mu_h\|_{-1/2,\Gamma} := \sup_{\mathbf{v}\in V} \frac{b(\mathbf{v},\mu_h)}{\|g(\mathbf{v})\|_{1/2,\Gamma}} \quad \forall \mu_h \in M_h \tag{21}$$

the inf–sup condition can be expressed as

$$\inf_{\mu_h\in M_h} \sup_{\mathbf{v}_h\in V_h} \frac{b(\mathbf{v}_h,\mu_h)}{\|\mu_h\|_{-1/2,\Gamma}\|\mathbf{v}_h\|_1} = \beta_h \geqslant \beta > 0 \tag{22}$$

where the constant β is independent of the mesh size h. This inf–sup condition should be satisfied regardless of the technique used to enforce the contact constraints (penalty or Lagrange multipliers or both).

The fractional dual norm in Eq. (22) can be evaluated as given in Ref. [7]. Also, for properly supported structures the following inequality can be used

$$c\|\mathbf{v}_h\|_1 \leqslant \|\mathbf{v}_h\|_E \leqslant c\|\mathbf{v}_h\|_1 \tag{23}$$

where here and in the sequel c denotes any constant (taking on different values throughout the text and possibly even in a single equation), and $\|\mathbf{v}_h\|_E$ is the energy norm.

The above equations lead to a somewhat involved test, which however, strictly should be performed. To eliminate many discretization schemes that do not pass the above test, a somewhat simpler test can first be used. This test uses the inverse property

$$\|\mathbf{v}_h\|_{1/2,\Gamma} \leqslant ch^{-1/2}\|\mathbf{v}_h\|_{0,\Gamma} \tag{24}$$

With this relationship we obtain

$$\|\mu_h\|_{-1/2,\Gamma} \geqslant ch^{1/2}\|\mu_h\|_{0,\Gamma} \quad \forall \mu_h \in M_h \tag{25}$$

Hence, a "weaker" inf–sup condition is arrived at

$$\inf_{\mu_h\in M_h}\sup_{\mathbf{v}_h\in V_h}\frac{\int_{\Gamma_C}\mu_h g(\mathbf{v}_h)\,d\Gamma_C}{h^{1/2}\|\mu_h\|_{0,\Gamma}\|\mathbf{v}_h\|_1} = \beta_h^* \geqslant \beta^* > 0 \tag{26}$$

Namely, if this condition is not fulfilled then Eq. (22) is not satisfied. But, if Eq. (26) is fulfilled then we cannot make a conclusion as to whether Eq. (22) is also satisfied.

The matrix form of this inf–sup condition can be expressed as

$$\inf_{\mathbf{M}\in M_h}\sup_{\mathbf{V}\in V_h}\frac{\mathbf{M}^T\mathbf{B}\mathbf{V}}{h^{1/2}\sqrt{\mathbf{M}^T\mathbf{P}\mathbf{M}}\sqrt{\mathbf{V}^T\mathbf{S}\mathbf{V}}} = \beta_h^* > 0 \tag{27}$$

where the **S** and **P** matrices are defined as

$$\mathbf{V}^T\mathbf{S}\mathbf{V} = \|\mathbf{v}_h\|_1^2 \tag{28}$$

$$\mathbf{M}^T\mathbf{P}\mathbf{M} = \|\mu_h\|_{0,\Gamma_C}^2 \tag{29}$$

Note that matrix **S** could be replaced by **K** according to Eq. (23). The numerical evaluation of the inf–sup condition of Eq. (22) is more complicated and needs to be performed as described in Ref. [14].

The value of the numerical inf–sup condition of Eq. (27) is determined based on the following eigenvalue problem [1,15]

$$\mathbf{G}\phi = \lambda\mathbf{S}\phi \tag{30}$$

where

$$\mathbf{G} = \frac{1}{h}\mathbf{B}^T\mathbf{P}^{-1}\mathbf{B} \tag{31}$$

If Eq. (30) has $(k-1)$ zero eigenvalues, then the number of contact pressure modes, k_{pm}, in the system is

$$k_{pm} = k - (m - n + 1) \tag{32}$$

where m and n are the dimensions of the matrices in Eq. (18). If no contact pressure modes exist, then the inf–sup value is equal to $(\lambda_k)^{1/2}$, where λ_k is the smallest non-zero eigenvalue. Otherwise, it is equal to zero.

As stated earlier, satisfying the two stability conditions of Eqs. (19) and (22) guarantees solvability. However, it is valuable, as a first quick consideration, to analyze only the solvability conditions for the contact problem of Eq. (18) [1]

$$\mathbf{V}^T\mathbf{K}\mathbf{V} > 0 \quad \forall \mathbf{V} \in \text{kernel}[\mathbf{B}] \tag{33}$$

$$\mathbf{B}^T\mathbf{M} = \mathbf{0} \text{ implies that } \mathbf{M} = \mathbf{0} \tag{34}$$

The first solvability condition, Eq. (33), is satisfied for a properly supported structure. The solvability condition in Eq. (34) is satisfied if the **B** matrix is of full rank

$$\text{rank}(\mathbf{B}) = n \tag{35}$$

Algorithms that do not satisfy this condition can be solvable if a penalty-type formulation is used. However, the resulting formulation is not robust. Significant numerical errors can result, especially when the right-hand-side gap vector is non-zero [1].

In this study, the numerical evaluation of the solvability and stability of contact algorithms involving flexible bodies was performed using the model shown in Fig. 2 with a negative initial gap g_0. First solvability was checked, then the inf–sup test of Eq. (26) was performed and the algorithms that passed were further tested with the more elaborate inf–sup test of Eq. (22). In all inf–sup evaluations the number of elements on both surfaces was varied. Since the writing of this paper, an analytical study of the inf–sup condition in Eq. (22) has been performed and the results obtained are in agreement with the numerical predictions presented below [16].

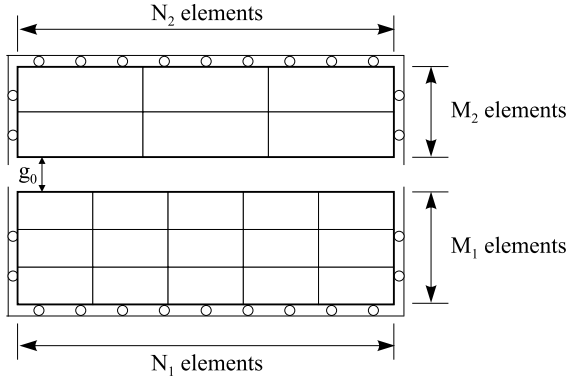

Fig. 2. Geometry used for inf–sup and patch tests.

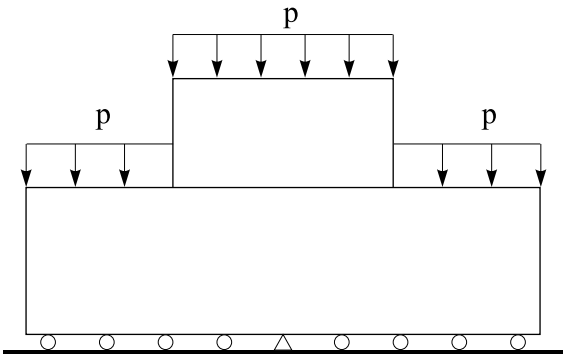

Fig. 3. Commonly used contact patch test.

3.2. Patch test

Fig. 3 shows the contact patch test used in Ref. [17]. In this work, we used the patch test shown in Fig. 4 (which is similar to the test used in Ref. [18]) and the model shown in Fig. 2. These simpler configurations provide all the required information. In the exact solution of the patch test of Fig. 4, the contact pressure field and the unit vector normal to the contact surface are constant throughout the contact region. Furthermore, both discretized contact areas are identical, hence $\Gamma_C = \Gamma_C^{Top} = \Gamma_C^{Bot}$. Based on these simplifications, Eq. (17) reduces to

$$b(\mathbf{u},\lambda) = \lambda_0 \int_{\Gamma_C^{Top}} u^{Top}\, d\Gamma_C^{Top} - \lambda_0 \int_{\Gamma_C^{Bot}} u^{Bot}\, d\Gamma_C^{Bot} \qquad (36)$$

where λ_0 is the constant contact traction, and u is the vertical component of displacement. Eq. (36) also represents the work done by two equal and opposite constant pressure loads acting on both bodies, which is the expected solution for this patch test. Hence, any contact algorithm can pass this patch test if its numerical integration of the contact term in Eq. (17), or the equivalent penalty or perturbed Lagrangian term, can be equal to the integral of a constant normal traction acting on both contacting surfaces.

The main obstacle hindering the exact evaluation of Eq. (17) in most contact algorithms is the form of the gap function $g(\mathbf{v})$. This function is piecewise continuous along Γ_C with possible discontinuities occurring at the nodes of either contact surfaces. This is illustrated in Fig. 5, which shows the variation of the gap function for a mesh of non-matching linear straight elements. The gap, in this case, is composed of four C^0-continuous segments with the derivative discontinuity occurring at the nodes of either surface. Accordingly, any integration scheme involving integration points that are dictated by only one of the two surfaces cannot exactly evaluate Eq. (17) regardless of the number of integration points used. If however, the integration intervals are based on "subsegments" corresponding to any two neighboring nodes regardless of their surface of origin, an exact evaluation is possible. A potential difficulty here is that the location of the integration points is not fixed, since the locations

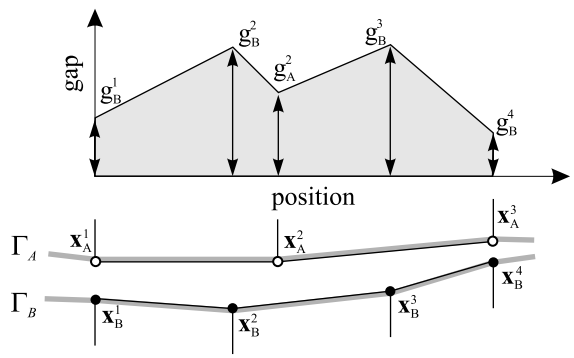

Fig. 5. Variation of gap between two surfaces (assuming a vertical normal vector).

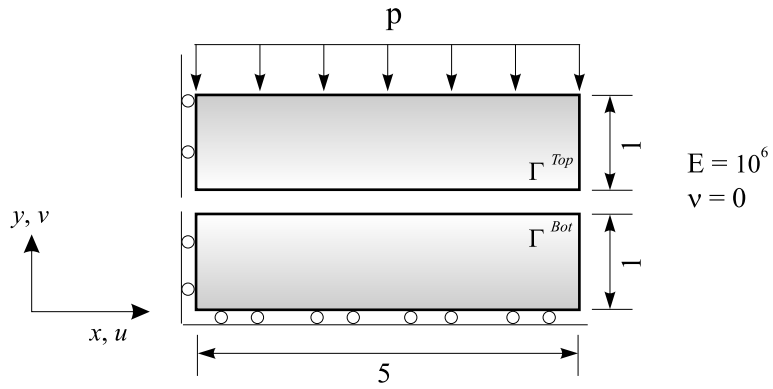

Fig. 4. Simple contact patch test problem.

of the segment boundaries change due to the relative motion between the two surfaces.

Note that another way to ensure that the patch test is passed is to use an approach adopted in fluid-flow/structure interaction analysis. In this case, the fluid tractions are applied to the structure using the consistent nodal force calculations and the fluid nodal velocities are constrained to be compatible with the structure nodal displacements [5]. However, this approach results in a non-symmetric coefficient matrix, which in fluid-flow calculations does not represent an inefficiency (because the coefficient matrices are already non-symmetric). However, in structural analysis, we aim to use symmetric matrices.

Different non-matching meshes were used for the patch test in Fig. 4. Fig. 6 shows three arrangements involving different combinations of linear and quadratic surfaces. The contact surface Γ_C was either defined on the top or the bottom body and several forms of the Lagrange multiplier field μ were investigated. Uniform meshes were used for both surfaces, since mesh distortion does not affect the contact patch test performance, as long as the chosen elements pass the regular patch test representing a state of constant stress.

4. Existing contact algorithms

In this section, we summarize various classes of contact algorithms and give an assessment of their stability and patch test performance.

4.1. Node-to-node contact

If both contact surfaces involve identical meshes with coincident nodes, a contact constraint can be applied for each node pair. The contact integral of Eq. (17) is then transformed to a summation of the form

$$b(\mathbf{v}_h, \mu_h) = \sum_i \mu^i [(\mathbf{v}_B^i - \mathbf{v}_A^i) \cdot \mathbf{N}^i + g_0^i] \tag{37}$$

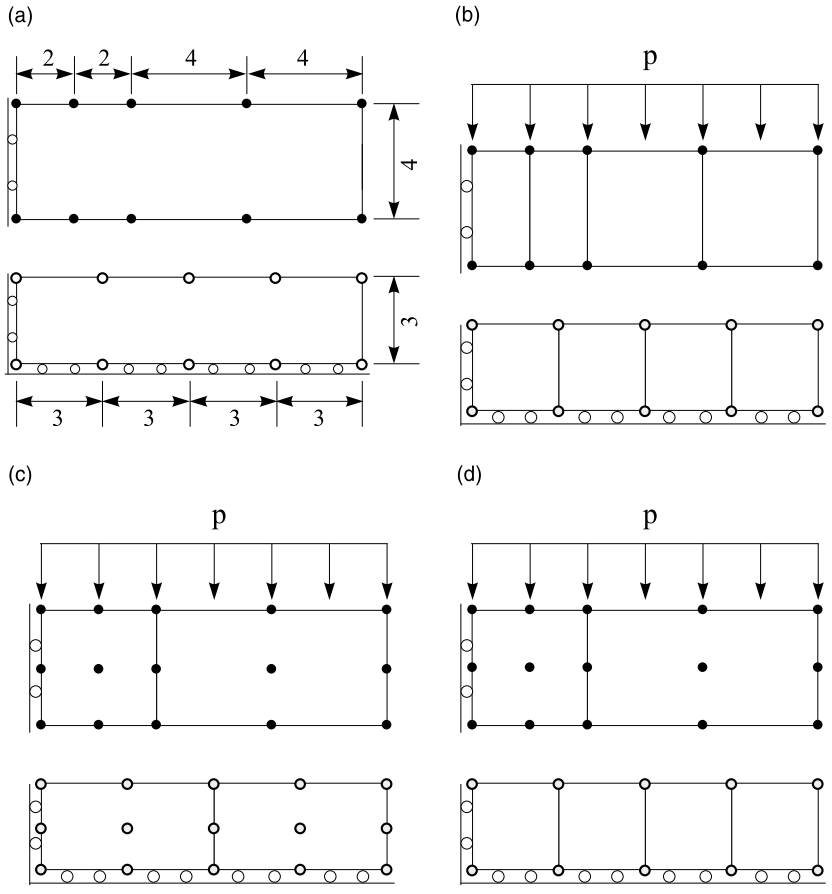

Fig. 6. Patch test: (a) dimensions, (b) linear on linear segments, (c) quadratic on quadratic segments, and (d) linear on quadratic segments.

where μ^i is the discretized contact force at node i. This algorithm is simple to implement, however, it is only applicable to a limited class of contact problems. The algorithm passes the patch test for linear elements. For quadratic elements, the patch test is passed only if the edge nodes on one surface are paired with the edge nodes on the other surface and mid-side nodes with mid-side nodes. This algorithm also satisfies the inf–sup stability condition.

4.2. Node-to-segment contact

The earliest published finite element formulations for contact problems enforce the no-penetration condition on a node-to-segment basis as shown in Fig. 7(a) (e.g., Refs. [19,20]). The nodes of one surface Γ_C commonly referred to as the slave or contactor are prohibited from penetrating the second surface Γ_T (referred to as the master or target surface). For each node on the contactor surface Γ_C with coordinates $\mathbf{x}_C^i$, the displacement of the target point on Γ_T is interpolated as follows (Fig. 7(b))

$$\mathbf{v}_T^{i^*} = \sum_j h_T^{ij} \mathbf{v}_T^j \tag{38}$$

where h_T^{ij} is an interpolation function relating the displacement of the target point with coordinates $\mathbf{x}_T^{i^*}$ to that of the target nodes with coordinates $\mathbf{x}_T^j$.

The no-penetration contact constraints are enforced only at the discrete contactor nodes on Γ_C and the integral of Eq. (17) is transformed to a summation over the target nodes

$$b(\mathbf{v}_h, \mu_h) = \sum_i \mu_C^i w^i [(\mathbf{v}_C^i - \mathbf{v}_T^{i^*}) \cdot \mathbf{N}^i + g_0^i] \tag{39}$$

where μ_C^i corresponds to the discretized contact traction at contactor node i, and w^i is a weight function representing the area attributed to node i, which is frequently set to 1. Then, μ_C^i corresponds to the discretized contact force. The resulting algorithm is computationally inexpensive and easily applicable to linear and higher order elements alike, as well as 2D and 3D problems. A modification of the node-to-segment approach involves enforcing the contact constraint at the integration points of the contactor surface (as opposed to at the nodes) [21,22]. Another modification involves applying the contact constraints between the nodes of the contactor surface and a higher order polynomial fit of the target surfaces [18,23].

This class of contact algorithms satisfy the inf–sup stability condition. Fig. 8 shows the value of the inf–sup test of Eq. (26) for five different meshes of the test problem of Fig. 2. The element size h is based on the lower surface mesh. Note that the inf–sup test is passed even when the less densely meshed surface is assumed to be the contactor. The algorithms also pass the inf–sup test of Eq. (22).

The node-to-segment algorithms do not pass the contact patch test of Fig. 4 due to the inaccurate transfer of forces from the contactor to the target surface. This

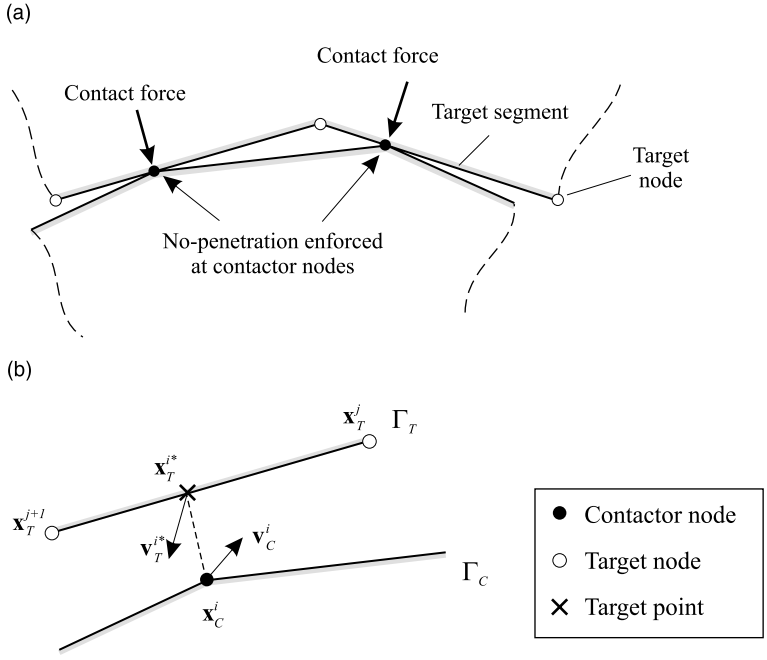

Fig. 7. Node-to-segment contact algorithm: (a) schematic, and (b) definitions.

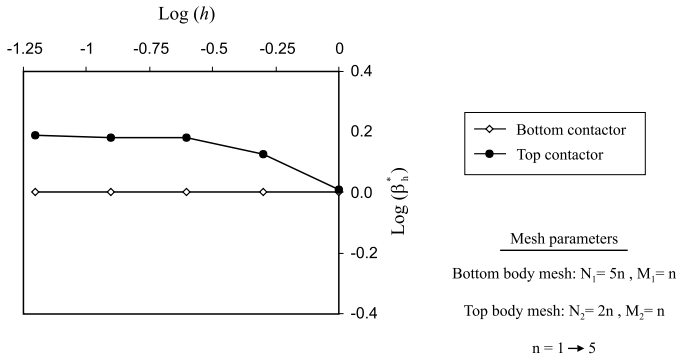

Fig. 8. Inf–sup value for two node-to-segment configurations using linear elements.

can be illustrated, for the linear element configuration of Fig. 6(b), by assuming the correct nodal forces at one of the two contact surfaces (the contactor) and examining the forces transferred to the other surface. The expected correct values are shown in Fig. 9(a). However, when the lower surface is selected as the contactor, the incorrect nodal forces transferred to the target are shown in Fig. 9(b). Similar inconsistent force transfers result when the roles of contactor and target are reversed, and when quadratic or higher order elements are used.

Another problem related to algorithms that do not pass the contact patch test is that despite convergence in the energy norm, the local errors at the contact region do not necessarily diminish with mesh refinement. This is illustrated in Fig. 10, which shows the maximum error in σ_x and σ_y for the meshes used in the inf–sup test of Fig. 8. The stresses are normalized by the exact value of σ_y.

Finally, two general observations are worth mentioning when using the node-to-segment approach: (a) selecting the surface with finer discretization as contactor usually produces better results, and (b) the error in the results is not necessarily smaller when using quadratic elements.

4.3. Two-pass node-to-segment contact

Two-pass node-to-segment contact algorithms eliminate the bias between master and slave surfaces by performing the nodal contact search twice [17,18]. In the first pass, one surface is treated as the contactor and the second one as the target and the roles are reversed in the second pass. The main advantage of this technique is the ease of evaluation of the contact term that is characteristic to the single-pass node-to-segment method.

However, the poor performance of two-pass node-to-segment algorithms can be illustrated by analyzing their solvability and stability. If any two nodes on both surfaces have identical locations, a duplicate constraint is created resulting in a rank deficient **B** matrix, which violates Eq. (34). Two approaches may be used to overcome this problem. The first involves using a penalty-based formulation, which is not reliable since Eq. (34) is still violated. The second approach involves

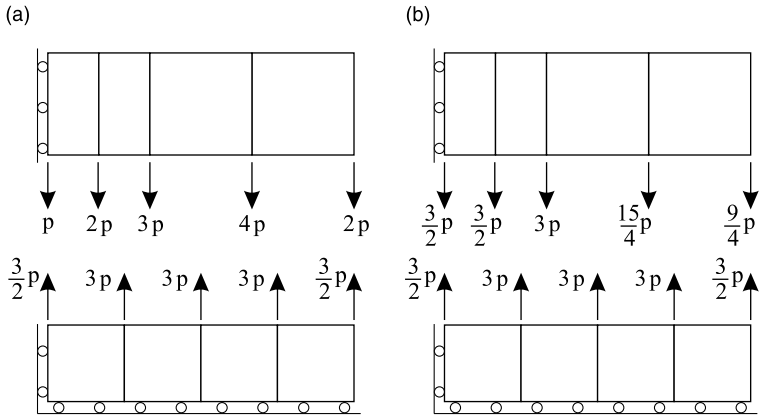

Fig. 9. Consistent force distribution: (a) exact, and (b) when lower surface is assumed to be contactor.

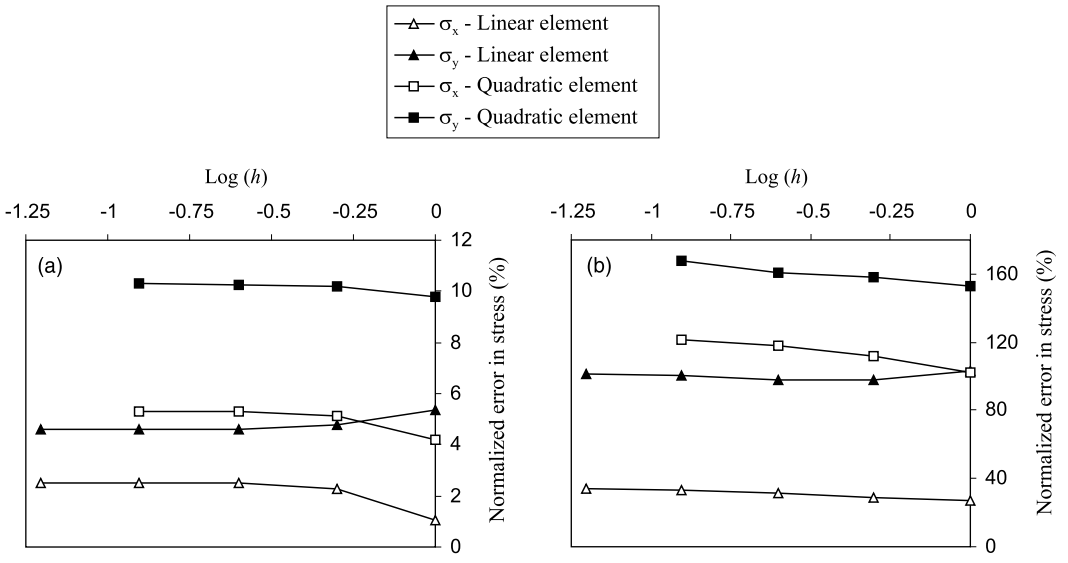

Fig. 10. Maximum error in stresses (normalized by exact σ_y) for linear and quadratic elements: (a) lower surface is contactor, (b) upper surface is contactor.

modifying the search algorithm to detect and remove duplications. The resulting algorithm satisfies Eq. (34), however, our results show that for certain element configurations the inf–sup test is not passed. This can be attributed to the overconstraining of the displacements on the contact surface. Fig. 11 shows the values of the inf–sup test of Eq. (26) for the different meshes of Fig. 2. The element configuration used in Fig. 11(a) leads to a constant value of β^*, while that of Fig. 11(b) leads to a decreasing β^*. However, the algorithm does not satisfy the inf–sup condition if any element configuration fails the test. Note the brief improvement in the inf–sup value for one of the meshes used in Fig. 11(b) (for $n = 2$). This is attributed to the matching of the meshes since at that configuration the lower surface has 10 elements and the upper one has five elements.

For quadratic and higher order elements the double pass node-to-segment algorithm results in a rank-deficient **B** matrix even if the multipliers corresponding to coincident nodes are removed. This is illustrated using the mesh shown in Fig. 12. Enforcing contact at the three nodes $\mathbf{x}_B^1$ through $\mathbf{x}_B^3$ dictates the form of the parabola connecting points $\mathbf{x}_A^1$ through $\mathbf{x}_A^3$. Hence, the gap at node $\mathbf{x}_A^2$ is guaranteed to be zero. Accordingly, the constraint at that point is redundant and the algorithm violates the solvability and stability requirements.

Although not stable, two-pass node-to-segment algorithms pass the contact patch test of Fig. 4 for linear elements. This can be explained by analyzing the kernel of the matrix **B**, which dictates the displacement modes that satisfy the contact constraints. Since the **B** matrix is overconstrained, its kernel includes only two displacement modes (in the vertical direction), regardless of the number of nodes (more modes result if any two nodes are coincident). These displacement modes correspond to a flat contact surface. The patch test is passed since the expected contact surface is horizontal (flat). However, this limited kernel means that any curved contact geometry cannot be achieved. In practice, some of the nodes must release from contact in order to maintain the curved geometry. This leads to a significant error in results, as reported in Ref. [18]. It is worth noting that quadratic and higher order elements with the double-pass node-to-segment algorithm do not pass the patch test [17,18].

4.4. Intermediate contact surface

In this approach, an intermediate contact surface is developed between the two contacting surfaces as shown in Fig. 13. The nodes on the two surfaces Γ_A and Γ_B are projected onto the intermediate surface Γ_C and a contact segment is defined between any two projected nodes (coordinates $\mathbf{x}_C^k$ and $\mathbf{x}_C^{k+1}$ in the figure). The main advantage is that the integrand of Eq. (17) is a continuous function over each contact segment. This algorithm was developed for 2D problems with linear elements in Ref. [24]. Similar algorithms based on an intermediate surface were developed for 2D quadratic elements [25] as well as 3D bilinear elements [26]. Both algorithms, however, are considerably more complex than their linear 2D counterpart.

The 2D version of this algorithm (for linear elements, see Ref. [24]) is similar to the two-pass node-to-segment algorithm (with duplicate constraints removed) and the same conclusions can be inferred regarding its stability

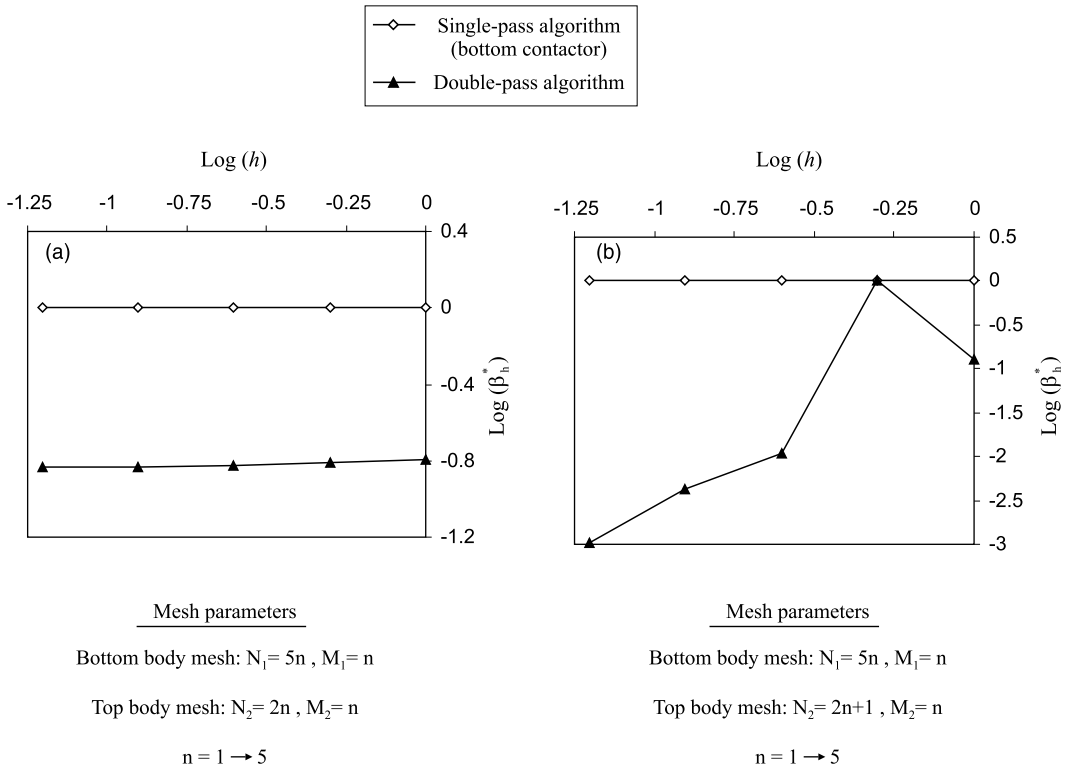

Fig. 11. Inf–sup value for single- and double-pass algorithms using two different mesh configurations, linear elements.

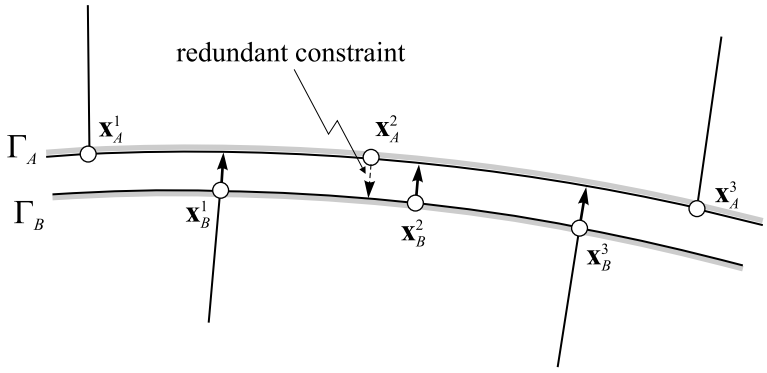

Fig. 12. Redundant constraint in two-pass node-to-segment algorithm for quadratic elements.

and its patch test performance: it does not satisfy the inf–sup stability condition (it is overconstrained) and it passes the patch test. The intermediate surface algorithm for 2D quadratic surfaces, developed in Ref. [25], avoids the redundant nodal constraints experienced in the two-pass quadratic node-to-segment algorithm. Accordingly, it is reported to pass the patch test. However, it still remains unstable. Note that the good patch test performance of this class of contact algorithms results from the manner in which the intermediate surface segments are defined. These segments identify the points of discontinuity in the gap function, which were illustrated in Fig. 5.

4.5. Segment-to-segment contact

We classify contact algorithms as segment-to-segment if they involve some form of integration of the

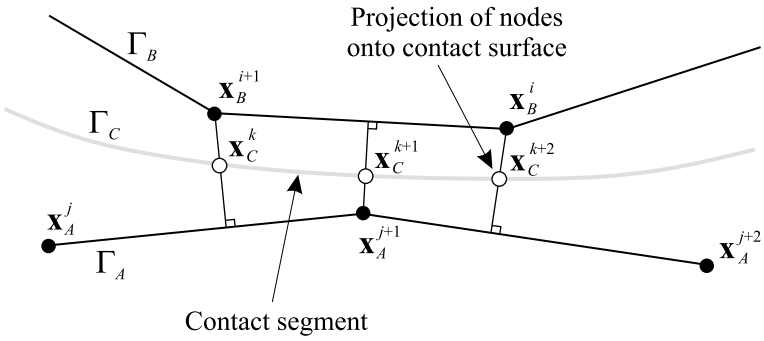

Fig. 13. Intermediate contact surface.

contact constraint equations over the surfaces of the contacting bodies, that is not just using values at the nodes (see for example, Ref. [27]). The intermediate contact surface algorithms can be considered a special case of this category of contact algorithms. In the following section we describe a new segment-to-segment algorithm that satisfies both the inf–sup stability condition and the contact patch test.

5. New contact algorithm

The algorithm involves a master-slave approach in which both the contactor and the target are treated as surfaces, not nodes. If the surface Γ_C is assumed to be the contactor, then the assumed contact pressure follows Eq. (16) with a polynomial order less than or equal to that of the element interpolation, and with segments Ψ_h defined as the element faces on Γ_C. Eq. (17) is then converted to a summation over its integration points (see Fig. 14)

$$b(\mathbf{v}_h, \mu_h) = \sum_i \mu_C^{iS} w^i [(\mathbf{v}_C^{iS} - \mathbf{v}_T^{i^*}) \cdot \mathbf{N}^i + g_0^{iS}] \tag{40}$$

where $\mathbf{v}_C^{iS}$, the displacement at integration point i, is interpolated from the nodal displacements on Γ_C as follows

$$\mathbf{v}_C^{iS} = \sum_k h_C^{ik} \mathbf{v}_C^k \tag{41}$$

The interpolation of the displacement of the target point $\mathbf{v}_T^{i^*}$ is identical to Eq. (38).

Unlike in the node-to-segment approach, the Lagrange multiplier μ_C^{iS} at integration point $\mathbf{x}_C^{iS}$ is not an independent variable. Instead, it is interpolated from the independent (usually nodal) multipliers μ_C^k on Γ_C

$$\mu_C^{iS} = \sum_k H_C^{ik} \mu_C^k \tag{42}$$

where the form of the H_C^{ik} interpolation functions depends on the polynomial order and inter-element continuity of the contact pressure field. Accordingly, the number of independent Lagrange multipliers is not related to the number of contact integration points. This is important in ensuring that adding integration points leads to a more accurate evaluation of the contact contribution without overconstraining the problem.

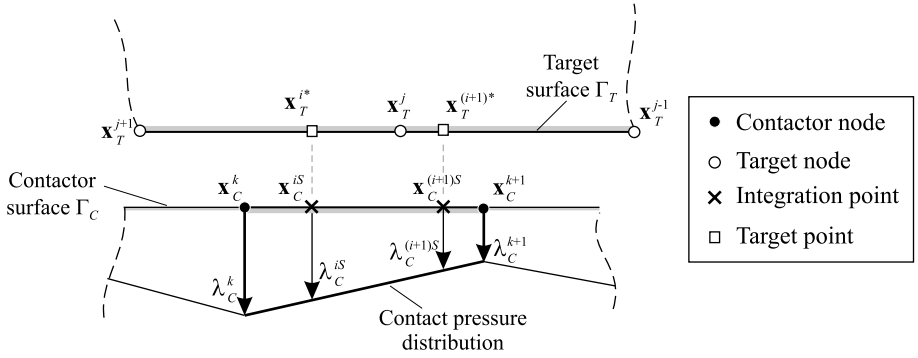

Fig. 14. Schematic of new contact algorithm.

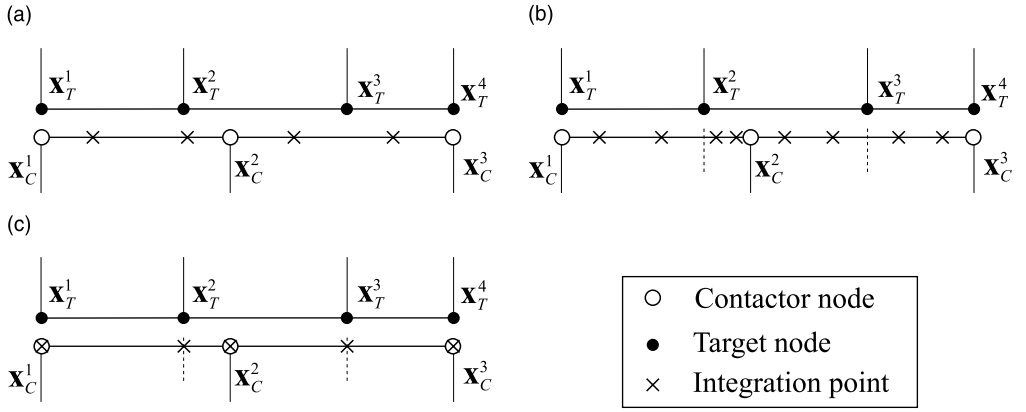

Fig. 15. Location of integration points based on: (a) second order Gaussian quadrature, (b) composite second order Gaussian quadrature, and (c) composite trapezoidal quadrature.

The integration point locations and their weight functions can be determined based on Gaussian or Newton–Cotes integration (Fig. 15(a)). However, this would not satisfy the patch test. Instead, a "composite" integration rule is used, which tracks the discontinuities in the nodes on both surfaces as shown in Fig. 15(b)–(c). The algorithm involves two steps. In the first, the sub-segment boundaries are determined by projecting the nodes of the target surface onto the contactor surface (only the edge nodes need to be projected for quadratic and higher order elements). In the second step, the contact expression on each sub-segment is integrated using Gaussian or Newton–Cotes integration rules. This accurate integration feature enables the algorithm to pass the patch test for both linear and higher-order elements.

The stability of the new algorithm governs the choice of the contact pressure field. Table 1 lists the possible interpolations for linear elements, their solvability and their stability. All cases were tested numerically using the contact inf–sup problem of Fig. 2. The results reveal that the constant pressure interpolation $\mathscr{M}^0_{h,0}$ leads to an unstable algorithm. The linear discontinuous interpolation $\mathscr{M}^1_{h,0}$ involves more pressure degrees of freedom than there are normal surface displacements. Hence the discretization is unsolvable and of course unstable. The only feasible option is the linear continuous contact pressure field $\mathscr{M}^1_{h,1}$ (see Theorem 3 in Ref. [16]). In all cases, stability was not affected when the roles of target and contactor were reversed. Also the choice of Gaussian or Newton–Cotes integration (of the same order) does not affect stability. Finally, applying higher orders of integration increases the accuracy (if needed) without adding new degrees of freedom that would compromise stability.

Table 2 shows the corresponding results for quadratic elements. The table shows that the $\mathscr{M}^1_{h,2}$, $\mathscr{M}^0_{h,0}$, $\mathscr{M}^1_{h,1}$ pressure interpolations result in stable algorithms (according to Theorems 3–5 respectively, in Ref. [16]). However, those with more pressure degrees of freedom have a higher convergence rate [1]. Accordingly, the quadratic continuous pressure distribution is recommended. Note that the patch test is passed for all stable contact pressure interpolations, even those that do not have optimal convergence in the solution.

Table 1
Stability and solvability of various pressure interpolations for linear elements

Contact pressure interpolation	Stable	Solvable	Optimal
Constant ($\mathscr{M}^0_{h,0}$)	No	Yes	No
Linear continuous ($\mathscr{M}^1_{h,1}$)	Yes	Yes	Yes
Linear discontinuous ($\mathscr{M}^0_{h,1}$)	No	No	No

The optimal convergence for this new contact algorithm results from using the highest order of contact pressure interpolation that satisfies stability (as well as the patch test). This is illustrated by applying a linearly varying displacement field to the top surface of the inf–sup problem of Fig. 2. Using linear elements results in linear convergence in the energy norm. This is true for both the new algorithm (with $\mathscr{M}^1_{h,1}$ contact pressure interpolation) as well as the classical node-to-segment algorithm. Using quadratic elements however, yields different results as shown in Fig. 16. The new algorithm (with $\mathscr{M}^1_{h,2}$ contact pressure interpolation) maintains quadratic convergence, while the node-to-segment algorithm results into only linear convergence.

This section has focused on the stability and patch test performance of the new segment-to-segment algorithm. While the theory given here is directly applicable

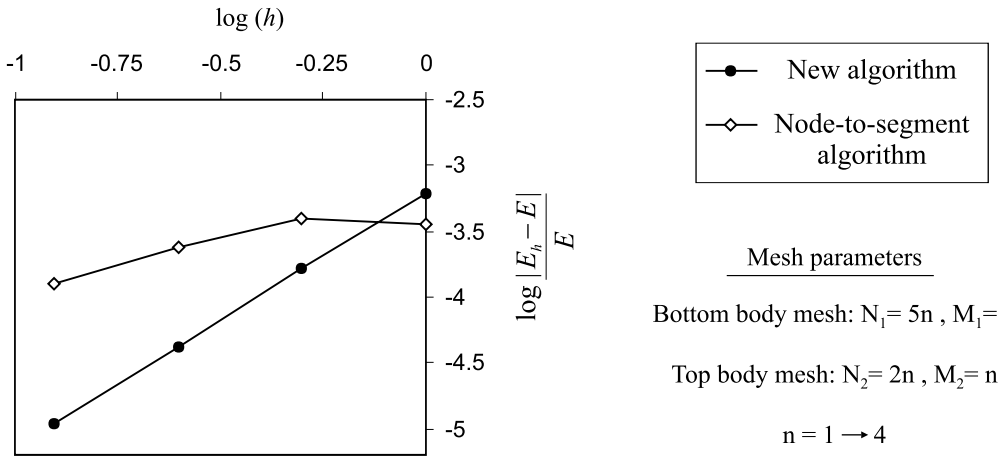

Fig. 16. Convergence in energy norm for quadratic elements; E_h is the strain energy for element size h, E is the "exact" strain energy.

Table 2
Stability and solvability of various pressure interpolations for quadratic elements

Contact pressure interpolation	Stable	Solvable	Optimal
Constant ($\mathscr{M}^0_{h,0}$)	Yes	Yes	No
Linear continuous ($\mathscr{M}^1_{h,1}$)	Yes	Yes	No
Linear discontinuous ($\mathscr{M}^0_{h,1}$)	No	Yes	No
Quadratic continuous ($\mathscr{M}^1_{h,2}$)	Yes	Yes	Yes
Quadratic discontinuous ($\mathscr{M}^0_{h,2}$)	No	No	No

to 3D contact problems, the actual detailed solution algorithm still needs to be developed.

6. Conclusions

The stability and patch conditions of finite element contact discretizations have been discussed, and different contact algorithms have been evaluated. The results revealed that these two requirements are not directly related. The performance in the contact patch test is predominantly dictated by the accuracy of evaluation of the contact integral, not by the specific assumption for the gap, contact pressure, or the contact surface. Stability, on the other hand, is governed by the interpolation of the contact pressure. Node-to-segment algorithms are stable, but do not pass the patch test. This leads to discretization errors that do not decrease with mesh refinement. On the other hand, some double-pass and intermediate surface algorithms satisfy the patch test, but are not stable. A new segment-to-segment algorithm was developed which satisfies both the stability and the contact patch conditions, using linear or quadratic element displacement interpolations.

References

[1] Bathe KJ. Finite element procedures. New Jersey: Prentice-Hall; 1996.
[2] Bathe K.J. The inf–sup condition and its evaluation for mixed finite element methods. Comput Struct 2001; 79:243–52, 971.
[3] Brezzi F, Bathe KJ. A discourse on the stability conditions for mixed finite element formulations. Comput Meth Appl Mech Engng 1990;82:27–57.
[4] Irons BM, Razzaque A. Experience with the patch test for convergence in finite elements. In: Aziz AK, editors. The mathematical foundations of the finite element method with applications to partial differential equations. Academic Press: New York; 1972. p. 557–87.
[5] Bathe KJ, Zhang H, Wang MH. Finite element analysis of incompressible and compressible fluid flows with free surfaces and structural interactions. Comput Struct 1995; 56:193–214.
[6] ADINA-F theory and modeling guide, ADINA R&D, Inc., 71 Elton Avenue, Watertown, MA 02472, 1994 (latest version August 2000).
[7] Brezzi F, Fortin M. Mixed and hybrid finite element methods. New York: Springer; 1991.
[8] Duvaut G, Lions JL. Inequalities in mechanics and physics. Berlin: Springer; 1976.
[9] Kinderlehrer D, Stampacchia G. An Introduction to variational inequalities and their applications. New York: Academic Press; 1980.
[10] Eterovic AL, Bathe KJ. On the treatment of inequality constraints arising from contact conditions in finite element analysis. Comput Struct 1991;40:203–9.
[11] Pantuso D, Bathe KJ. Finite element analysis of thermo-elasto-plastic solids in contact. In: Owen DRJ, Onate E, editors. Computational plasticity. Swansea: Pineridge Press; 1997. p. 72–87.
[12] Bathe KJ, editor. Computational fluid and solid mechanics, Elsevier, Amsterdam; 2001.
[13] Kikuchi N, Oden JN. Contact Problems in elasticity: a study of variational inequalities and finite element methods. Philadelphia: SIAM; 1988.

[14] Iosilevich A, Bathe KJ, Brezzi F. On evaluating the inf–sup condition for plate bending elements. Int J Numer Meth Engng 1997;40:3639–63.
[15] Chapelle D, Bathe KJ. The inf–sup test. Comput Struct 1993;47:537–45.
[16] Bathe KJ, Brezzi F. Stability of finite element mixed interpolations for contact problems. Proceedings della Accademia Nazionale dei Lincei, in press.
[17] Taylor RL, Papadopoulos P. On a patch test for contact problems in two dimensions. In: Wriggers P, Wagner W, editors. Computational methods in nonlinear mechanics. Springer: Berlin; 1991. p. 690–702.
[18] Crisfield MA. Re-visiting the contact patch test. Int J Numer Meth Engng 2000;48:435–49.
[19] Hughes TJR, Taylor RL, Sackman JL, Curnier A, Kanoknukulchai W. A finite element method for a class of contact–impact problems. Comput Meth Appl Mech Engng 1976;8:249–76.
[20] Bathe KJ, Chaudhary AB. A solution method for planar and axisymmetric contact problems. Int J Numer Meth Engng 1985;21:65–88.
[21] Qiu X, Plesha ME, Meyer DW. Stiffness matrix integration rules for contact-friction finite elements. Comput Meth Appl Mech Engng 1991;93:385–99.
[22] Kikuchi N. A smoothing technique for reduced integration penalty methods in contact problems. Int J Numer Meth Engng 1982;18:343–50.
[23] El-Abbasi N, Meguid SA, Czekanski A. On the modelling of smooth contact surfaces using cubic splines. Int J Numer Meth Engng 2001;50:953–67.
[24] Simo JC, Wriggers P, Taylor RL. A perturbed Lagrangian formulation for the finite element solution of contact problems. Comput Meth Appl Mech Engng 1985;50:163–80.
[25] Papadopoulos P, Taylor RL. A mixed formulation for the finite element solution of contact problems. Comput Meth Appl Mech Engng 1992;94:373–89.
[26] Papadopoulos P, Taylor RL. A simple algorithm for three-dimensional finite element analysis of contact problems. Comput Struct 1993;46:1107–18.
[27] Zavarise G, Wriggers P. A segment-to-segment contact strategy. Mathematical Comput Modell 1998;28:497–515.