The University of Liverpool

Department of Mechanical Engineering

Contact Stress Analysis using the Finite Element Method

by

Steven Keith Pascoe

A thesis submitted in accordance with the requirements of the University of Liverpool for the Degree of Doctor in Philosophy.

December 1990

COPYRIGHT

Contact Stress Analysis using the Finite Element Method

by

Steven Keith Pascoe

Abstract

This thesis is concerned with the numerical solution of the stresses occurring between bodies in the regime of linear elastic contact. The computational modelling technique used is the finite element method.

Contact problems are generally non-linear in form with changing boundary conditions at the contact interface. To solve these problems several load incrementation solution algorithms have been developed. One change in boundary condition or loading in pre-defined load steps were the techniques used. The pre-defined load step method is similar in form to existing non-linear finite element analysis techniques. The method of identification of contact is undertaken from geometrically checking if 'potential' contact nodes of one surface, are touching or are inside the surface elements of the

other. This is achieved by the conversion of the contacting node's global coordinates into the equivalent isoparametric local coordinates of the other body's surface elements. Nodes are then defined as contacting if they are identified to be within the domain of an element on the other body. Constraint equations are formed between each contacting node and target element which are based on the exact location of contact as defined by the element's shape functions. This technique of defining constraint equations allows direct imposition of contact effects, as well as permitting mis-aligned meshes and the use of a wide range of different element types in the contact zone. The constraint equations applied are dependent upon the contact conditions. In the case of sticking contact, then either two or three constraint equations (depending upon the model's dimensionality) are applied, these being in the global x,y (and z if 3-D) directions. For sliding contact, a single constraint equation is applied in the normal direction relative to the contact surface.

Frictional effects as defined by Coulomb Friction has been implemented. These forces are included by an iterative technique, however, when the direction of sliding is known a priori then these forces have been included directly.

The constraint equations are imposed on the 'standard' stiffness matrix through the use of Lagrange multipliers, although it has been shown that the Penalty method can also directly impose identical constraints on the stiffness matrix in a single iteration. Additionally, if there is no overlap in the constraint equation then the Transformation matrix method can also apply identical constraints.

Substructuring is undertaken to improve solution times using frontal elimination, leaving just potential contact and externally loaded degrees of freedom in a reduced stiffness matrix. A hybrid equation solver is also developed which when coupled with Lagrange multipliers, permits just one decomposition of the structure's stiffness matrix. With this equation solver only the Lagrange multiplier constraint terms need to be decomposed in subsequent iterations.

Acknowledgements

I wish to express my sincere gratitude for the guidance given by my academic and industrial supervisors, Dr John E Mottershead and Dr Trevor K Hellen.

I wish to also thank Michael Rushton for his accurate typing and my sister Rosemary Rushton, for encouraging us whenever we tired. Thanks are also due to Russell English for his aid in preparing the equations, tables and flow diagrams.

I wish to acknowledge the services of Liverpool University Computer Department for the use of their facilities, and to Thackray Ltd and to Ford for providing applications for the contact analyses.

Finally I wish to thank the C.E.G.B. and S.E.R.C for sponsoring this work.

CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Contact in engineering

In engineering almost all devices are made up from an assemblage of different parts which are mechanically joined resulting in contact interfaces. At these interfaces contact stresses are likely to occur. Typical examples of this are the connection of flanged pipes, where contact stresses occur between the flange faces. A more complicated example is the mounting of turbine blades, where contact occurs between the blades' roots and the rotor. Other forms of contact are where components are initially separate, but due to relative displacements come together resulting in either sticking or sliding contact zones. Circumstances where this occur are in gear teeth contact and sliding contact in machine tools, for example as found in slideway and journal bearing contact. Applications of contact can also be seen in bio-engineering, with contacting surfaces in artificial joint replacements.

In all of the above examples, knowledge of the location and magnitude of the contact stresses is of particular interest in obtaining an estimate of the component's reliability and survivability.

There are essentially two distinctly different areas in engineering where the use of contact stress analysis is of benefit. The main area is in the early stages

of design, whereby initial component configurations and shapes can be analysed to provide indications of where the regions and magnitudes of high stress are likely to occur. This information can be used to modify either the component's shape or choice of material if initially unacceptably high stresses were occurring. Obvious time and economic savings can be obtained if the stress prediction can be calculated without recourse to actual manufacture during this design phase. Once the final design is decided however, it is wise for experimental confirmation of the structural performance to be undertaken.

The other area of engineering where contact stress analysis is of benefit occurs in situations where existing components require analysis. This may be needed for example in problems where an estimate of the fatigue life is desired or where failure has occurred and recommendations for future design improvements are required.

1.2 Scope of the present work

In circumstances where the applied loads and geometries of the contacting bodies are relatively simple, then 'closed form' analytical solutions can be used to predict the contact stresses. However, for a great many real structures this simplicity of shape and loading is not apparent. For these problems numerical modelling techniques are generally needed.

In this thesis several numerical computer based modelling techniques using the finite element method have been developed. The final computer algorithms produced by the author are general programs which allow the

modelling of any shaped, two or three dimensional structure with sticking or frictional sliding contact under elastic conditions. The methods compare well with known analytic results, and with work done by other researchers. Particular advances here are in the allowance of mis-aligned nodes in the contact zone and the use of curved elements. Additionally, a 'hybrid' matrix equation solver has been developed, which when combined with substructuring can produce substantial reductions in the solution times.

Although contact stress analysis using the finite element method has been in evidence for over fifteen years, little comparison has been made between the different techniques available. This is particularly the case with regard to highlighting their limitations. As a consequence, one of the aims of this thesis is to critically survey other techniques, as well as to develop and describe new methods.

1.3 Description of thesis content

The contents of the chapters in this thesis are now summarised.

A review of literature on the subject of computer modelling of contact problems is given in the remainder of this chapter for both the finite element method and other methods. The precise contribution by the author in the field is detailed at the end of this chapter in Section 1.5.

In Chapter two the standard finite element method is described and the basic theory recalled. Chapters three and four detail the complexities of modelling contact that occurs both in the real physical problem and the finite element

contact model. The essential stages within a finite element contact algorithm are identified and discussed in Chapter five and a unified constraint theory shown to exist. Chapter six contains the new theories developed by the author for including contact effects within the finite element method. In Chapter seven, detailed areas of the new algorithms including their structure are presented. The contact identification process is detailed in Chapter eight and a 'hybrid' equation solver explained in Chapter nine. Results for various test runs are shown and described in Chapter ten for both 'simple' and 'complex' models with Chapter eleven discussing the main findings. Finally the main conclusions of the work and further recommendations are given in Chapter twelve.

1.4 Review of Literature

The subject of contact stress analysis was probably first addressed by Hertz (1896) in his celebrated work on contact between elastic rollers. This work culminated in several fundamental theories which are still used today in contact stress analysis. These theories were developed to solve the contact conditions between parallel cylindrical frictionless rollers under the action of point loading. In the mathematical derivation, Hertz assumed the contact region to be a continuous quadratic profile which was small in size when compared to the contacting bodies. With these assumptions, equations were derived allowing the size of the contact zone and the pressure variation within it to be calculated. Unfortunately, the applicability of the method is restricted to simply shaped frictionless contact.

Other mathematical theories were developed for different shapes and loading. Hill et al (1947) allowed wedges to be analysed, with Mindlin (1949) producing a general theory for elastic bodies assuming circular or elliptical contact zones. Tangential forces were included by Johnson (1955) and slip analysed between cylinders by Bentall et al (1967) with Spence (1975) including friction into Hertz's contact problem. A full treatise of the different mathematical theories is given in Johnson's book (1985) on contact stress analysis. These mathematical methods however are restricted in their applicability due to the necessary simplifying assumptions made to facilitate a solution. That is they are restricted to simple 'easy to mathematically define' contact boundaries, which generally are flat or circular in form. Many real contact problems though do not fit into these categories. For these problems alternative techniques for evaluating the contact size and contact stresses were needed.

The finite element method developed in the late 1950's and early 1960's was a giant leap forward in the field of stress analysis. Amongst early contributors were Turner et al (1956) and Clough (1960). The method essentially involves discretisation of the structure to be analysed into a mesh of grid points (nodes) and small blocks (finite elements). A matrix representing the structure's stiffness can then be generated. Application of restraints and external loads then yields the nodal displacements from which stresses can be calculated. Use of the finite element method to solve contact problems was not successfully achieved until the late 1970's & 1980's. The reason for this late development was primarily due to inherent non-linearity in the contact problem, with the size of the contact region and the contact conditions unknown prior to solution.

For contact to be modelled effectively using the finite element method requires prevention of mesh overlap of the contacting bodies, with appropriate modification either to the stiffness matrix or the force vector. Over the years, distinctly different techniques for doing this have evolved. They can essentially be categorised as,

a, Penalty methods (gap elements)

b, Lagrange multiplier methods

c, Transformation matrix methods

d, Flexibility methods

e, Other specialist techniques

Developments using these techniques in contact stress analysis using the finite element method are now discussed.

1.4.1 Penalty methods

The penalty method is a general technique of constraint imposition where extra terms are added ***inside*** the original finite element stiffness matrix. These terms invoke the contact conditions. They are probably best understood in the context of gap elements connected across the potential contact interface.

The technique of modelling contact using finite elements in conjunction with the penalty method for the imposition of contact constraints was initiated by White and Enderby (1970),and Stadter and Weiss (1979). In both approaches, the gap element stiffnesses were initially defined by the user (low in a region where separation was thought may occur, and high where thought

to remain contacting). An iterative solution proceeded with checks made for mesh overlap after each stage, if overlap was detected then a re-solution was undertaken with the gap element's stiffness modified by the amount of overlap. The restriction with the methods were that in the first, nodal pairs were connected across the contact surface permitting only sticking contact to be modelled, whereas frictionless sliding could only be imposed in the second method. Additionally, node on node contact was necessary in the region of potential contact.

Oden (1981) and Kikuchi (1982) approached the contact problem from a more mathematical viewpoint. They showed how the penalty method was an addition to the standard finite element variational statement. The constraints were imposed by numerical integration at selected points, using reduced integration schemes such as Simpson's rule and the Trapezoid rule.

Friction occurring at the contact interface was included by Campos, Oden and Kikuchi (1982). In their method a frictionless run was initially undertaken to define the normal contact forces. Then, using Coulomb's friction relationship, frictional forces were calculated which were added to the force vector and a re-solve undertaken. Iteration continued in this manner until no noticeable change in the displacement vector occurred. The contact applications by these authors were however restricted to contact between a deformable body and a rigid body. This type of contact is considerably easier to solve than problems where both contacting bodies are deformable.

An alternative technique for including friction using gap elements (and with both contacting bodies deformable) was developed by Mazurkiewicz and Ostachowicz (1983), Zolti (1983), Ostachowicz (1984), and Hellen (1988). The methods involve including tangential gap elements whose stiffness is equal to the product of the coefficient of friction and the normal gap element stiffness. Although this technique intuitively seems to correctly add the frictional effects, it suffers severely depending upon the penalty value used. To impose the displacement constraints as accurately as possible necessitates the use of a high penalty value (equivalent to the gap element stiffness). In this circumstance the forces occurring in the system are in fact 'locking' the solution and hence applying sticking contact, and not the desired frictional sliding contact. If a low penalty value is used then the frictional condition can be imposed, however many iterations are necessary to eliminate the overlaps and satisfy the Coulomb Friction relationship.

Oden and Pires (1983a, 1983b) developed a variational statement that directly included friction. The friction relationship they produced was non-local and non-linear in form. The non-locality was induced as the decision on sticking or sliding was based not on the local contact forces alone, but on neighbourhood effects at a microscopic level. This essentially was defined by the roughness of the surface, selected by an 'asperity' indicator ρ. The non-linearity occurred due to elastic-plastic deformation, defined to be occurring at the junction of the contact interface and selected by a non-linear indicator ε . This method is probably more realistic in modelling the frictional effects than the Coulomb relationship. However, the need for user input of 'asperity' and 'non-linearity' variables complicate its use for the general 'non-expert' user. Another advanced friction relationship was

developed by Cheng and Kikuchi (1985), this also however required much experimental analysis to determine the appropriate extra friction variables.

The use of an incremental loading scheme with linear gap elements was used for the interesting problem of paper rolling by Wong (1984). Endo et al (1984) and Simo et al (1986) use a Riks load incrementation algorithm to allow contact and limit-point behaviour of snap-through problems to be analysed. Padovan et al (1985) highlighted some of the problems encountered in large displacement contact, where the area gap elements become so distorted that poor Jacobians are formed. This results in incorrect penalty stiffness terms being added to the stiffness matrix. They overcame this by pantographing and subsequent re-formation of the area gap elements.

Chang, Saleeb and Shyu (1987) use a special contact element which imposed constant pressure and had displacement continuity. However, their approach was limited to frictionless contact.

The use of constraints applied directly at Gauss points rather than nodes has been applied by Hitchings (1988). The advantage of this technique is that the decision on the contact state of sticking or sliding, is directly obtained from the Gauss point forces. However, as the gap elements are inserted at the beginning of the solution process, this method is best applied only when the contact zone is well defined in advance. With this condition, only one decomposition of the stiffness matrix is necessary, with changes only to the force vector in subsequent iterations for frictional effects.

To summarise, the penalty or gap element method involves the addition of extra terms *into* the original stiffness matrix. In its most common form in commercial finite element packages node on node contact is necessary to allow the gap element to be connected. Additionally, the rate of convergence of the solution is highly dependent upon the penalty value (gap stiffness) selected. However, its ease of implementation and the relatively minor architectural changes required, have made the gap element method the most widespread technique for contact stress analysis in current finite element packages.

1.4.2 Lagrange multiplier methods

The Lagrange multiplier method is a general technique of constraint imposition, where extra rows and columns are added ***around the outside*** of the original finite element stiffness matrix. These terms exactly apply the desired constraints. The extra variables in the solution vector represent the forces to enforce the constraint, and are the Lagrange multipliers. The size of the system of finite element equations is always increased by this approach.

The technique of Lagrange multipliers was first used in contact stress analysis by Hughes et al (1976). Their work was a particular advance as it allowed elastic and elastic-impact contact to be analysed. Restrictions were that node on node contact was necessary in the contact zone. Once contact was identified for a nodal pair, their displacements and forces were coupled by the addition of unity Lagrange terms in the relevant positions around the stiffness matrix. A Newmark time stepping algorithm was used for the

dynamic analysis. However, linear elements and sticking or frictionless contact could only be modelled.

Okamoto and Nakazawa (1979) permitted frictional effects to be directly included in the Lagrange terms. Again node on node contact was necessary, with the inclusion of friction causing the stiffness matrix to be unsymmetric. The loading was controlled to allow just one extra node to be included per load stage. This was achieved by fully loading the structure, then scaling such that the new node just contacted. Node on node contact was not necessary in the method of Guerra and Browning (1983). They used an incremental loading procedure and applied linear constraint equations. The concept of master and slave surfaces was also introduced (the current terminology now being contactor and target surfaces). They showed that penalty or Lagrange multiplier methods could be used to impose the constraints, although they highlighted that specific aspects of the implementation of the algorithm affected the solution obtained. In particular, the tolerance of contact identification was found to be of significance.

Bathe and Chaudhary (1985) developed a general two dimensional algorithm similar to Guerra and Browning. The decisions on contact state were made from the element forces with the subsequent constraints applied at the nodes. Chaudhary and Bathe (1986) further extended their method to include three dimensional and elastic dynamic contact. Only linear elements however were permitted in the contact zone.

In summary, the Lagrange multiplier method appears to have a distinct advantage over the penalty method in that constraints can be exactly applied. However, the size of the system of equations generally increases.

1.4.3 Transformation matrix methods

In the transformation matrix method, the contact constraints are applied by a pre- and post multiplication of the stiffness matrix by a 'transformation matrix'. This transformation matrix contains the constraints to be imposed on the system of equations. The resulting stiffness matrix is generally reduced in size, with the constraints exactly applied.

The first use of this method in contact analysis was by Wilson and Parsons (1970). Sticking contact was the imposed state resulting in applications to axisymmetric shrink-fit type contact problems. Fredriksson (1976) allowed both sticking and sliding contact with friction included as extra terms added to the force vector. The method was however restricted to node on node contact. A similar approach was developed by Gaertner (1977) although only one extra node was permitted per load stage. Mahmoud, Salamon and Marks (1982) proposed total removal of friction, giving a frictionless solution which admittedly was easier to install inside finite element coding. This method however was obviously restricted in its application. Particular advantage of the unity values occurring in the transformation matrix with node on node contact was realised by Torstenfelt (1983). This allowed Boolean operations to be undertaken on the pre- and post multiplication, reducing the computational time for this stage by a factor of about ten.

The allowance of non-aligned contacting nodes was permitted by Chen and Yeh (1988). Additionally, they directly included friction in a modified transformation matrix.

In summary, the transformation matrix method can apply constraints exactly on the system of finite element equations. However, a full decomposition of the 'reduced' stiffness matrix then needs to be carried out at each stage. This can become computationally quite time consuming.

1.4.4 Flexibility methods

There are several different variants under this heading, but essentially flexibility methods involve the inverse of stiffness in the solution technique.

Chan and Tuba (1971a, 1971b) calculated the local flexibility at the region of contact. This was then used to determine the corrective forces needed to eliminate mesh overlap. For its time the method was quite advanced as node on node contact was not necessary and friction could be included. A simpler flexibility technique was developed by Francavilla and Zienkiewicz (1975). Frictionless node on node contact was modelled, with re-inversion of the stiffness matrix for each change in boundary condition. Nodal pressures were the unknowns with gap displacements defined prior to solution. This method was restricted in its scope of applicability whilst Chan and Tuba's method was restricted to linear elements.

1.4.5 Other Techniques.

Many of the other techniques which have been developed are for specific applications and as such are restricted in their use.

Gangal (1972) developed a technique for interference fit contact problems requiring no modifications to standard finite element codes. This was achieved by inducing interference by applying a temperature field to the sleeve and shaft whilst setting the sleeve's coefficient of thermal expansion to zero. The corresponding expansion of the shaft subsequently inducing the interference stresses.

A method for modelling the interface between steel and concrete in reinforced structures was developed by Schafer (1975), and extended by Keuser et al (1983), Beer (1985) and Mehlhorn et al (1985). This method involved the use of a special 'bond' element, which had user defined properties for shear, obtained from experiment.

The contact problem of modelling tyre and road contact, where a travelling load and hence time enters the solution, was reduced to a purely spatial problem using a transformed coordinate system by Padovan and Zeid (1980), and Zeid and Padovan (1981).

Up to now, the contact methods described have all been based on the ***displacement*** derived finite element method. A discussion on some of the other techniques now follows.

The mixed or hybrid finite element method has been applied to contact stress analysis. This method involves the introduction of stress variables into the initial variational formulation. Tseng and Olsen (1981) used hybrid elements with two displacement and three stress degrees of freedom per node. Node on node contact was necessary with only one extra contacting node permitted per load stage. Heyliger and Reddy (1987a, 1987b) used Lagrange multipliers and mixed elements to allow non-aligned contacting nodes. A distinct advantage of these hybrid elements is that stresses are natural variables obtained in the solution vector. This allows the contact state of sticking or sliding to be easily identified. However, the number of unknowns is increased in the solution matrix because of these extra stress variables, resulting in larger matrices to be solved. Additionally, the hybrid elements are not familiar to many finite element users hence their use is probably restricted.

Another technique used in contact stress analysis is the boundary element method. In this method only the boundary of the structure is discretised into elements. Application of this technique to contact problems was made by Anderson et al (1980) and Becker and Plant (1987). Node on node contact is necessary in the contacting zone, hence large scale sliding problems require re-meshing of the elements. An advantage often quoted with the boundary element method over the finite element method is the reduced dimensionality of the model, resulting in smaller matrices to be solved, as the boundary and not the full domain is discretised. However, the use of substructuring techniques in the finite element method, whereby internal nodes are eliminated from the matrix, can result in comparable or smaller matrices being produced. Researchers who have used this technique of substructuring with the finite element method are Francavilla and Zienkiewicz (1975),

Fredriksson (1976), Mazurkiewicz et al (1983) and Nour-Omid and Wriggers (1986).

The non-linear nature of the contact problem and the subsequent large number of iterations required, coupled with the generally large structural matrices involved, stimulate the need for efficient equation solvers. Recent developments are the use of conjugate gradient methods by Nour-Omid and Wriggers (1986), May (1986) and Dilintas et al (1988). In this technique direct inversion of the stiffness matrix is avoided. Similarly simplex type algorithms were used by Fischer and Melosh (1987) and Vijayakar et al (1988) which allow minimisation without direct inversion.

1.5 Contribution by the author

In this thesis methods are developed for elastic, quasi-static finite element modelling of contact with aligned or mis-aligned meshes of linear and quadratic elements. The particular difficulties of identification of contact and definition of the constraints are overcome using the element's shape functions. Although only quadrilateral and brick elements have been implemented here, the technique of using shape functions is shown to hold for many types of finite element. The constraints are added as Lagrange multipliers, although a unified theory of constraint imposition is shown to exist which shows that the Penalty method could equally be used and if no overlaps exist, then so could the transformation matrix method.

Three new two-dimensional contact algorithms are developed, the first including Coulomb friction directly in the Lagrange constraint terms, which

results in an unsymmetric stiffness matrix. The other two methods are symmetrical variants of the first, although iteration is necessary to include friction. The load history is tracked by only allowing one change in boundary condition per load step.

A two and three-dimensional algorithm is also developed whereby nodal constraints are applied directly using Lagrange multipliers. The decision on contact state is selected at an element rather than nodal level. A load incrementation technique similar to that used in standard non-linear finite element analyses is developed, rather than allowing just one extra node per load stage. This method of loading was used because of the many nodes that may come into contact with 3-D problems. A modified version of the frontal solution technique (Hellen (1969) and Irons (1970)) was implemented, with just the potential contacting nodes remaining in the final 'reduced' stiffness matrix. Additionally, an active zone equation solver was used allowing the 'reduced' stiffness matrix to be decomposed just once. The Lagrange constraint terms due to contact are then the only terms requiring decomposition in each solution stage. This equation solver is shown to yield considerable computational savings.

CHAPTER 2

THE FINITE ELEMENT METHOD

2.1 Introduction

The finite element method is described in this chapter and its mathematical derivation presented.

In many structural analysis problems, direct closed form solutions can be obtained for the stresses or displacements. However, these solutions are generally restricted to 'simple geometries' and 'simple loading'. This approach though can give good results for real structural problems and some examples of which are given in Johnson (1985) and Timenshenko and Goodier (1951). The development and advancement of the computer in the early sixties, allowed structural analysis techniques which although simple in their concept, were phenomenally powerful in their applicability. A whole new range of structural problems could now be analysed in detail where before their geometry or loading were too difficult. A particular area of engineering which pioneered these techniques was the aerospace industry, where minimum weight and maximum strength priorities resulted in complex shapes which were difficult to analyse in any other way than by finite elements. In today's environment, the finite element method is used not just by aircraft designers, but also motor car manufacturers, the oil industry and construction companies to name but a few. It can be seen from its wide use,

that the method of finite elements has probably gained the reputation as being one of the most reliable and versatile methods for computational stress analysis.

The method of finite elements is fully described in many books on the subject, such as the reference text of Zienkiewicz (1977) and the texts of Bathe (1982) and Ahmad and Irons (1986). For new or unfamiliar readers in the subject, the book by Stasa (1985) is particularly recommended. A review of the fundamental method of finite elements now follows.

2.2 Description of the method

A component requiring analysis by the finite element technique must first be divided into an assemblage of blocks. Each block is of a finite size and is termed a finite element. At strategic points within each block are positions called nodes. Thus when using the finite element method the initial structure is converted into an equivalent geometrically formed shape consisting of nodes and elements. This assemblage is commonly referred to as the finite element 'mesh', a typical example of which is shown in Figure 2.1.

The discretisation of the structure is the key to the whole method and formation of a 'good' mesh is vital for obtaining reliable results. In fact this is quite a skillful process and mesh generation packages such as PATRAN have been developed to aid in this task. Once the mesh has been formed then a 'stiffness' matrix is calculated for each element. Summation of these

elemental stiffness matrices then gives a global structural stiffness matrix. The system of equations to then be solved is,

$$[K]\{U\} = \{R\} \qquad 2.1$$

where $[K]$ = global structural stiffness matrix

$\{U\}$ = global nodal displacement vector

$\{R\}$ = global external loads vector

Equation 2.1 represents a set of linear equations. With the external loads known, a solution for the unknown nodal displacements can be obtained, from which stresses can then be calculated.

To allow modification of this equation for the effects of contact requires the derivation and theory of its development to be understood.

2.3 Theoretical basis of finite elements

The total energy in any static structural system can be expressed in terms of the initial strain energy due to deformation and the work done due to external forces acting on the structure, i.e.

$$\Pi = S - W \qquad 2.2$$

For a discrete structural system under steady state Equation 2.2 can be rewritten as,

$$\Pi = \frac{1}{2}\{U\}^T[K]\{U\} - \{U\}^T\{R\} \qquad 2.3$$

In a real structure the deformation will be to a minimum energy condition, that is,

$$\frac{\partial \Pi}{\partial U_i} = 0 \qquad (\text{for all nodes , i}) \qquad 2.4$$

Thus differentiating Equation 2.3 with respect to the unknown nodal displacements and application of Equation 2.4 for minimisation, gives the familiar finite element equations of Equation 2.1.

The principles and techniques used to obtain the stiffness matrix now follow. Particular emphasis is placed on the element shape functions, which define the interpolation formulae inside each of the finite elements. It will be shown later that these shape functions play a key role in the contact algorithm developed by the author.

2.4 Element shape functions

Shape functions are a set of mathematical equations that define the interpolation rule within an element from its nodal values. They are defined

from the degrees of freedom per node and the number of nodes per element. The shape functions are well documented for each type of element. Zienkiewicz (1977) lists many of the commonly used element types and their appropriate shape functions, an example of which is the four noded quadrilateral element shown in Figure 2.2. The shape functions for this element are given in Equation 2.5.

$$\left.\begin{aligned} N_1 &= \tfrac{1}{4}(1-\xi)(1-\eta) \\ N_2 &= \tfrac{1}{4}(1-\xi)(1+\eta) \\ N_3 &= \tfrac{1}{4}(1+\xi)(1+\eta) \\ N_4 &= \tfrac{1}{4}(1+\xi)(1-\eta) \end{aligned}\right\} \qquad 2.5$$

Note that a local isoparametric coordinate system (ξ, η) has been used, with the maximum value of each local coordinate within the element being +1 or -1. These values represent the outer surfaces of the element. If the local coordinates of a point within an element (ξ_k, η_k) are known, then the global coordinates (x_k, y_k) of this point can simply be obtained from substitution of ξ_k and η_k into Equation 2.5 to give,

$$\left.\begin{aligned} x_k &= N_1x_1 + N_2x_2 + N_3x_3 + N_4x_4 \\ y_k &= N_1y_1 + N_2y_2 + N_3y_3 + N_4y_4 \end{aligned}\right\} \qquad 2.6$$

where $x_1 .. x_4$ = known x nodal coordinates

$y_1 .. y_4$ = known y nodal coordinates

Equation 2.6 can be rewritten in matrix notation as,

$$\begin{Bmatrix} x_k \\ y_k \end{Bmatrix} = \{X\}_k = [N_k]_i \{X\}_i^e \qquad 2.7$$

where

$[N_k]_i$ = shape function matrix for element i, evaluated at (ξ_k, η_k)

$\{X\}_i^e$ = nodal coordinates vector of element i

A similar expression for the displacements at any point within the element can also be obtained using the shape functions and the nodal displacements.

$$\begin{Bmatrix} u_{kx} \\ u_{ky} \end{Bmatrix} = \{U\}_k = [N_k]_i \{U\}_i^e \qquad 2.8$$

The strains within each element can be calculated from the nodal displacements,

$$\{\varepsilon\}_i^e = [B]\{U\}_i^e \qquad 2.9$$

The terms within [B] are global derivatives of the shape functions, and are obtained using the element's Jacobian matrix. The Jacobian matrix defines the transformation from global to local derivatives and for the two-dimensional element shown in Figure 2.2 can be written as,

$$[J] = \begin{bmatrix} \frac{\partial x}{\partial \xi} & \frac{\partial y}{\partial \xi} \\ \frac{\partial x}{\partial \eta} & \frac{\partial y}{\partial \eta} \end{bmatrix} \qquad 2.10$$

The terms in [B] therefore are,

$$[B] = \begin{bmatrix} \frac{\partial}{\partial x}[N_k]_i & 0 \\ 0 & \frac{\partial}{\partial y}[N_k]_i \\ \frac{\partial}{\partial y}[N_k]_i & \frac{\partial}{\partial x}[N_k]_i \end{bmatrix} \qquad 2.11$$

$$\text{with} \quad \begin{bmatrix} \frac{\partial}{\partial x}[N_k]_i \\ \frac{\partial}{\partial y}[N_k]_i \end{bmatrix} = [J]^{-1} \begin{bmatrix} \frac{\partial}{\partial \xi}[N_k]_i \\ \frac{\partial}{\partial \eta}[N_k]_i \end{bmatrix}$$

Assuming Hookean behaviour the element stresses can then be defined as,

$$\{\sigma\}_i^e = [D]\{\varepsilon\}_i^e \qquad 2.12$$

Where [D] in Equation 2.12 is dependent upon the Young's Modulus and Poissons ratio terms defined in the constitutive equations. For example in plane stress analysis it takes the form,

$$[D] = \frac{E}{(1-\nu^2)}\begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & (1-\nu)/2 \end{bmatrix} \qquad 2.13$$

A formulation for the strain energy within a system has been quoted in Equation 2.3 in terms of stiffness and displacement. It is also possible to express strain energy in terms of stresses and strain,

$$S = \frac{1}{2}\int_{\Omega} \{\sigma\}^T\{\varepsilon\} d\Omega \qquad 2.14$$

(where Ω = discretised domain i.e. 2-D or 3-D)

By the substitution of Equations 2.9 and 2.12 into 2.14, the following equation is formed,

$$S = \frac{1}{2}\int_{\Omega} \{U\}^T\{B\}^T[D]\{B\}\{U\} d\Omega \qquad 2.15$$

Inspection of the strain energy terms in Equation 2.3 and 2.15 shows the element stiffness matrix can be written as,

$$[K]^e = \int_{\Omega} \{B\}^T[D]\{B\} d\Omega \qquad 2.16$$

To summarise, the stiffness matrix for each element can be calculated knowing the element shape functions, their derivatives and the constitutive law applied.

The integration involved in Equation 2.16 can in some cases be carried out analytically, however it is impractical for complicated functions especially when ξ and η are curvilinear. In general, numerical quadrature is applied with weighting functions at specific positions within the element. The most commonly used method is Gaussian quadrature (see Zienkiewicz (1977) p196).

The global stiffness matrix representing the entire structure can be obtained by summation of the individual element stiffness matrices. This global stiffness is generally banded and symmetric in form.

CHAPTER 3

DISCUSSION OF PHYSICAL CONTACT PROBLEMS

3.1 Introduction

In contact analysis it is necessary to realise the wide range of different types of contact that can occur, for without proper understanding of the physical complexity of the real contact problem, then the computer modelling cannot be expected to predict the real situation reliably.

This thesis is concerned with the modelling of elastic contact. In the following sections, the physical complexities of this form of contact are identified and their effects discussed.

3.2 Physical complexity

The physical complexity for a contact problem can essentially be divided into two separate areas,

a, the shape and size of the contact zone

b, the prevalent contact state

Further discussion on these areas now follows.

3.3 Shape and size of the contact zone

The shape and size of the contact zone is dependent upon the shape of the potential contact surfaces, the material properties of the contacting bodies and the forces applied to them. There are two main types of sub-classification, this being problems where the shape and size of the contact region remains constant during all stages of loading, and problems where the shape and size varies. Obviously the latter type of contact problem is a much more difficult problem to solve than the former. Examples of non-changing contact area problems are those where generally large contact zones are present between conformingly shaped components, for example two slabs resting on top of each other under uniform vertical pressure loading as shown in Figure 3.1. Engineering examples of this form of contact can be seen in the contact of slideways in machine tools such as lathes and milling machines, and in interference fits between say a pin and a hole. In these problems, the following 'simplifying' features can be identified about the contact zone,

a, there is no change in contact size

b, the position of contact is known in advance

However, in the case of contact problems with varying contact areas, these two 'simplifying' features do not occur.

Varying contact areas generally occur when there is little conformity of the surfaces between the structures in the region of likely contact. For example this occurs in the contact of a cylindrical roller on a flat surface as shown in

Figure 3.2a, where the curvatures of the contact surfaces are quite different. In this case, contact occurs initially along a line and with the application of force the contact zone 'grows' to a finite size as shown in Figure 3.2b. This growth of contact zone occurs in many contact problems upon application of the loads. However, in some situations the increase in load can actually cause the contact zone to decrease in size. An example of this can be seen in Figure 3.3a, where a beam is initially resting on two wide supports. With application of load, the beam deforms into a curved profile resulting in small localised regions of contact at the inside edges of the supports as indicated in Figure 3.3b.

These contact examples, where the size of the contact zone changes during loading, highlight some of the difficulties involved in the contact modelling process. They show that a general reliable method must be capable of,

a, identifying whether contact is increasing or decreasing and

b, imposing these contact conditions on the contact model.

3.4 State of the contact conditions

Up to now, only geometric considerations have been discussed in obtaining the shape and size of the contact zone. There is however a further condition that is vitally important, and that is the contact state. The contact state identifies the type of contact that is occurring. There are three states that may exist, these are,

a, sticking contact

b, frictionless sliding contact

c, frictional sliding contact

These types of contact are fairly self-explanatory however a precise definition now follows, from which key features and differences can be identified.

3.4.1 Sticking contact

Sticking contact occurs when two bodies come together and no relative normal or tangential movement occurs at the contact interface. The displacements for the contacting zone of the two bodies are therefore identical, with equal and opposite contact forces also generated. Sticking contact can be described by the following equations for the contact zone.

$$\delta_n^A - \delta_n^B = 0 \qquad 3.1$$

$$\delta_t^A - \delta_t^B = 0 \qquad 3.2$$

$$F_n^A + F_n^B = 0 \qquad 3.3$$

$$F_t^A + F_t^B = 0 \qquad 3.4$$

where δ = surface displacement in the contact zone

F = contact force

n = the normal direction relative to the contact surface

t = the tangential direction relative to the contact surface

A = denotes body A involved in contact

B = denotes body B involved in contact

3.4.2 Frictionless sliding contact

Frictionless sliding contact occurs when no tangential force can be sustained on the contact interface. In this circumstance relative tangential sliding occurs between the two bodies. The amount of sliding is unknown prior to solution, although in certain problems it is governed by the normal interface forces. For example in Figure 3.4, a tapered wedge under end loading is being pushed into a similarly shaped cavity. With no load applied there are no interface forces occurring. However, with the application of the end load, sliding and deformation of the wedge occurs which results in normal interface forces being generated. The final position of the wedge is determined when the resolved normal contact forces acting in the direction of the applied loads are equal and opposite to the external loads. At this stage 'jamming' of the wedge has occurred.

The set of equations which define frictionless sliding contact are stated below,

$$\delta_n^A - \delta_n^B = 0 \qquad 3.5$$

$$\delta_t^A - \delta_t^B \neq 0 \qquad 3.6$$

$$F_n^A + F_n^B = 0 \qquad 3.7$$

$$F_t^A = F_t^B = 0 \qquad 3.8$$

From the above equations it can be seen that the equations in the normal directions (Equations 3.5 & 3.7) are the same as for sticking contact. The tangential direction equations are however quite different.

3.4.3 Frictional sliding contact

Frictional contact is similar to frictionless contact, with the same normal displacement and normal force equations. The difference with frictional contact though, is that although the sliding displacements are again unknown, frictional (tangential) forces are now present which are defined by some pre-determined relationship.

Researchers in the field of frictional contact have developed many different (and in general quite complex) relationships to define the friction forces present at a contact interface. These relationships are dependent upon a whole range of different parameters as can be seen in Bowden and Tabor's (1950,1964) definitive text on frictional contact. Some of the key variables which are defined as having effect are,

a, surface finish

b, contact size

c, material properties

d, temperature

e, lubrication etc.

No single relationship has been proven to be generally reliable due to the many different parameters that effect the problem. However, the friction relationship developed by Coulomb (1781) has been used extensively in solid mechanics due to its ease of understanding and acceptable results that it produces. As a consequence, Coulomb's friction relationship has been used to model frictional contact. The principles and equations used in this friction relationship are now described.

3.4.4 Coulomb's Friction Relationship

This relationship defines the maximum value of tangential force that can be sustained on a surface for sticking to be occurring, beyond which frictional sliding contact occurs. The magnitude of this limiting value is defined as the product of a 'coefficient of friction' and the normal forces present.

$$|F_t| = \mu |F_n| \qquad 3.9$$

Hence, if the magnitude of the tangential forces present in a contact problem are less than this limiting value, then sticking is occurring. Otherwise sliding results, with the tangential interface forces equal to this value. This tangential force acts in a direction opposite to the direction of sliding. In two-dimensional problems the direction of sliding can only be in one of two directions, these being along a tangent to the contact surface as shown in

Figure 3.5a. However, for three-dimensional problems the direction of sliding can be in any direction on the contact surface's tangential plane, as shown in Figure 3.5b.

The value for the friction coefficient is usually determined experimentally. Generally a single value of friction coefficient doesn't exist, this is because the tangential forces applied to a body to initiate sliding, are greater than those to maintain sliding. To include this effect properly, two coefficients of friction are needed, a static coefficient of friction μ_s to define the forces to initiate sliding, and a dynamic coefficient of friction μ_d defining the forces to maintain sliding . Normally μ_s is greater than μ_d. A list of typical coefficient of friction values for different materials is given in Table 3.1.

CHAPTER 4

DISCUSSION OF FINITE ELEMENT CONTACT MODELS

4.1 Introduction

The physical complexity encountered with real contact problems has been described in the previous chapter. There is however a range of complexity also apparent in the finite element modelling phase. The purpose of this section is to explain the different levels of complexity that may occur. This allows a fuller understanding of the contact solution methods developed here and by other researchers. Additionally, the applicability of a particular method and its relative advantages and disadvantages can be made more readily.

The main contact features that can occur in a whole range of different contact problems have been identified and listed in terms of complexity in Table 4.1. The headings 'easy' and 'difficult' have been used, these are general headings meant for relative comparison of the different features. A discussion on the complexity of each with regard to finite element modelling now follows.

4.2 Variation in contact area

If the region of contact is known in advance not to change in size, then this can greatly reduce the problems encountered as,

a, all nodes in this region can be set initially as contacting.
b, no checking routine is needed to identify separation.
c, a single linear solution may be acceptable.

However, if varying contact areas are likely then,

a, a method for identification of initial contact is necessary.
b, a method for eliminating any overlaps that may occur is required.
c, a method for identifying regions that were initially contacting but now may be separating is needed.
d, an incremental loading procedure is needed to allow for the changing contact area (and hence boundary conditions) during the loading phase.

4.3 Contact state

If sticking alone or frictionless sliding alone can be defined in advance as the prevalent contact state, then advantage of the following features can be taken,

a, this single state can be imposed upon identification of contact.

b, with sticking contact, once the position of contact has been identified, this remains the same for the whole solution.

c, with frictionless sliding no frictional forces need to be defined.

However, if mixed sticking and frictional sliding is possible, which may occur in real contact problems, or where the modeller is unsure of the actual contact state then,

a, a technique is needed to identify which contact state is prevalent and to change it if necessary as the solution progresses.

b, the magnitude of the frictional forces and the direction in which they act must be ascertained and included within the finite element solution.

These last two points are particularly difficult to incorporate within the finite element formulation and as could be seen in the literature survey, most early contact algorithms where sliding was incorporated, were for frictionless contact only.

4.4 Structural deformity

If the region of contact is between two structures of different materials with Young's Moduli values of E_1 and E_2 respectively, then if the ratio of $E_1/E_2 < .001$, then contact between a rigid body and a deformable body can be assumed to be occurring as very little deformation of the rigid body will

occur. This allows the stiffer of the bodies to be modelled as totally rigid with the other body deformable. The advantages of this strategy are,

a, no finite element stiffness matrix need be generated for the rigid body as it is wholly restrained, with the size of the corresponding matrices to be solved therefore reduced.

b, as one body is totally rigid, its surface does not move and hence checking and identifying of contact is simplified.

If the Young's Modulus values are similar for the contacting bodies, then selecting one body to be rigid is obviously an inaccurate assumption. In this type of contact (which accounts for many real contact problems) modelling techniques are necessary which allow both bodies to be deformable.

4.5 Dimensionality

All real structures are three dimensional in shape and form. Certain physical aspects though, allow a two dimensional computer model to give accurate results for some three dimensional problems. This can yield considerable advantages as models representing just the cross-sectional shape need be generated, allowing a reduction in model size and complexity. With regard to contact problems key advantages of two dimensional modelling are,

a, movements occur only in the plane of the cross section, making monitoring of contact geometrically easier.

b, line contact occurs (which is assumed to be constant in the depth

direction), whereas complex area contact may be present in some three dimensional problems. Hence, force and pressure monitoring at the interface is easier in 2-D.

c, generally an iterative solution technique is needed in contact modelling, thus the reduction in model size permitted with a two dimensional model yields significant computational savings.

It becomes apparent that advantage of two dimensional modelling should be taken whenever possible.

4.6 Mesh considerations

If the region in which contact will occur is known in advance, then special considerations can be made in the generation of the finite element meshes. Firstly, a fine mesh can be generated in this region to allow the high stress gradients which may occur to be detected. This is especially relevant in non-conforming contact, where small localised contact zones are likely.

Particular advantages in the solution process can be attained if the elements in the contact interface are arranged such that node on node contact occurs, as shown in Figure 4.1. This nodal arrangement allows that upon identification of contact, the separate meshes can become directly coupled at the nodal positions. Advantages of this feature are that,

a, the identification of contact becomes easier as only nodal pairs

(which may be defined in the initial data file or in some other manner) need to be checked for contact.

b, coupling in this manner allows the interface forces to be transmitted directly across the element boundaries at these coupled nodes.

However, due to intrinsic physical features in the real contact problem, node on node contact may not be possible in the contact zone. Circumstances where this may be so are,

a, where large scale sliding of one of the bodies over the other is occurring. In this case, mesh alignment may be initially achieved, however, due to sliding the meshes become mis-aligned as shown in Figure 4.2 (It is possible to reconstruct the meshes to ensure alignment as the sliding solution progresses, but this would be highly inefficient computationally as the stiffness matrix would need to be reformed).

b, where the bodies are initially separate and large scale movement is occurring, in which case the position of where initial contact commences may be difficult to predict.

c, where the geometric form of the surfaces of the contacting bodies are quite different, such that ensuring element alignment in the interface region in the mesh generation phase becomes excessively difficult. In this case a simpler, non-aligned mesh may have to be created. This problem is particularly likely in the three dimensional modelling of differently shaped contacting objects

(the knee prosthesis contact problem described in Appendix J being a typical example).

Further difficulties that can arise due to non-node on node contact are,

a, the checking routines to identify contact become more complicated as nodal pairs are not now occurring, with the location of contact being potentially anywhere on the contact surface. It therefore becomes necessary to identify with which element a node on one body is contacting and its precise position on the element.

b, interface forces cannot now be directly transmitted between adjacent nodes as they do not now occur, hence an alternative approach for transmission of these forces is necessary.

4.7 Contact curvature

In contact problems where the contact surface is curved, geometric difficulties are encountered when large amounts of sliding occur. This is because the curved sliding route necessary for a sliding node to remain in contact around a curved body, cannot be directly imposed in a single stage solution in the finite element method. This is due to the formulation of the finite element method yielding a linear system of equations. As a consequence, the nodal displacements are linear with respect to their initial positions. The sliding nodes therefore slide linearly in directions tangential to the contact surface. The result of this for large amounts of sliding is that

either geometric separation or penetration occurs as shown in Figure 4.3. These gaps or overlaps are unacceptable as they represent sliding around a shape different to the true profile, and can cause incorrect stress values to be obtained. Hence modelling of curved sliding contact requires techniques for elimination of possible induced gaps or overlaps.

4.8 Large scale sliding

Apart from the previously mentioned consequences of large scale sliding and contact curvature, there is also another problem encountered. This is that the position of one body on the other may change quite dramatically due to large scale sliding, when compared to the initial position of contact. In terms of finite element modelling, this may result in the element on which a node is finally contacting, being different to the element it was initially contacting as shown in Figure 4.4. It is therefore necessary in the solution procedure to identify when sliding from one element to another has occurred and with which element it is now contacting.

All of the above finite element difficulties need to be automatically overcome within a contact algorithm for its successful and reliable operation.

CHAPTER 5

INTRODUCTION TO CONTACT ALGORITHMS

5.1 Introduction

The extra stages that need to be included within a finite element package to allow modelling of contact are discussed in this chapter. It is shown that there are five essential stages, each of which must be automated and integrated to ensure convergence and satisfactory operation. Additionally, how these stages are solved is vital in defining the generality and applicability of a finite element contact algorithm.

The overall purpose of this chapter therefore, is to allow the 'best approaches' to be identified and subsequently for these to be selected. To achieve this selection requires a quite detailed analysis of each stage and this is now presented . The individual stages are,

a, the method of load application.

b, the method of identification of the size of the contact zone.

c, how the contact state is identified.

d, implementation of the contact constraints into the finite element equations.

e, the method of matrix solution.

A more detailed discussion of these stages follows.

5.2 Method of load application

There are primarily three different methods of applying the external loads which are,

a, fully loading the model in one stage.

b, loading in stages of one boundary condition change per step.

c, incrementally loading in defined load steps.

5.2.1 Fully loading

If this technique is used, then the loading history of the structure is obscured because complete re-solution is performed until force equilibrium and displacement compatibility are simultaneously satisfied. This is effectively linearising a non-linear problem into one load stage and as such creates errors when modelling contact with closing gaps and friction. However, modelling of non-varying contact areas using this approach is acceptable as implemented by Stadler and Weiss (1979) and Hitchings (1988). The disadvantage of this approach is its limited range of applicability.

5.2.2 One boundary condition change per load step

The other extreme in loading to fully loading in one stage is where only one extra boundary condition is permitted per load stage. This can be achieved by applying the full external loads and then scaling the resulting displacements such that no overlap occurs and just one new extra boundary condition is included in the next solution. The remainder of load is then applied in this next solution and the scaling process repeated. This procedure is continued until the structure is fully loaded. The overall solution is then obtained by summation of each of the scaled solutions. A diagram showing this scaling process is presented in Figure 5.1. An accurate build up of the contact zone is ensured with this technique. This method is probably the most obvious for tracking the solution whereby the non-linear problem has been converted into a series of linear solutions, with summation of the linear stages giving the overall solution. Some authors who have used this approach are Okamoto and Nakazawa (1979), Torstenfelt (1983) and Pascoe and Mottershead (1988,1989). However, the computational time with this approach may become excessive in problems where many contact nodes are present.

5.2.3 Incremental loading in defined stages

An alternative to the two approaches above is to apply load increments as in standard finite element non-linear analyses. This involves loading in prescribed load increments with iteration within each increment to eliminate any overlaps and force residuals to achieve convergence of nodal states,

displacements and contact forces. This technique has been used by Mazurkiewicz and Ostachowicz (1983) and Bathe and Chaudhary (1985) . It is probably the most attractive approach to problems involving many nodes in the contact region, as it allows reasonable tracking of the load history and is not restricted to one boundary condition change per load step.

To summarise, the method of loading depends to a great extent on the type of problem being considered. It should be noted that a variable load method, will due to its generality of applicability, be considerably more complex in its solution procedure for certain 'simple' problems than the 'straight forward' fully loading approach. However, a general method can always cope with a simple problem, whereas a 'straight forward' fully loading approach cannot always cope with complex models.

5.3 Identification of the size of the contact zone

From a finite element basis, this involves having a method for identifying which nodes and elements are in contact and precisely at which locations.

The gap element technique which is the most commonly used in finite element contact analysis has a particularly simple method of contact identification. This involves calculation of the forces occurring in the gap element. As an example, consider the line gap element inserted between adjacent nodes in Figure 5.2. The forces generated in the normal direction of the gap element indicate whether tension or compression is occurring, with compression indicating contact. They can be calculated from,

$$F_n = K_n(\delta_{1n} - \delta_{2n} - \delta_{gapn}) \qquad 5.1$$

In Equation 5.1, K_n is the element stiffness which is constant for each iteration and hence plays no role in identifying contact. It therefore becomes apparent that a direct comparison of the normal displacements of the end nodes is in fact being undertaken. Paradoxically, this simplicity of identification highlights the fundamental limitation of standard gap element theory. This being that node on node contact is necessary across the interface to allow this checking of relative nodal displacements through the calculation of gap element forces. Consequently, gap elements cannot be reliably used in problems involving large scale sliding, unknown contact zones or where nodal alignment cannot be achieved because of mesh generation difficulties. To solve the modelling of these problems, nodal identification of contact is necessary of not just *if* a node is contacting, but also *where* it is contacting. This involves finding the element against which a node is contacting and its exact position on the element's side or face. The contact checking routine to allow this must therefore be much more sophisticated than that used for gap elements.

The main technique used for this complex contact identification involves defining one surface as a master and the other as a slave, or as they are now more commonly phrased, 'contactor and target' respectively. Each node on the contactor surface is then checked for contact with each target element. This requires two checks to be carried out, one for the position of contact and the other for the overlap distance. Figure 5.3 shows a typical two-dimensional simple contact surface with one contacting node. Contact

can be identified by the intersection of a normal with respect to the target surface and the contactor node. From the position of intersection the values of β and δ_n can be evaluated for as,

$$\beta = \frac{x_c - x_1}{x_2 - x_1} = \frac{y_c - y_1}{y_2 - y_1} \qquad 5.2$$

$$\text{and } \delta_n = n_x(x_c - x_k) + n_y(y_c - y_k) \qquad 5.3$$

where n_x, n_y = normal direction cosines relative to the target surface

Contact is then defined to be occurring if,

$$\left.\begin{array}{l} 0 \leq \beta \leq 1 \\ \text{and} \quad \delta_n \leq 0 \end{array}\right\} \qquad 5.4$$

A major restriction with this approach is that only linear elements are permitted in the contact zone. The reason why this is so can be seen in Figure 5.4, where contact is occurring on a curved boundary. The values of β and δ_n from Equations 5.2 and 5.3 suggest in this case that no contact is occurring (as δ_n is calculated using the linear equation of 5.3, indicating a gap rather than overlap). The contact solutions for these quadratic and higher order elements (which would be encountered in curved boundary modelling) would therefore be seriously corrupted by this identification technique.

In three dimensional problems, contact identification is further complicated by the fact that contact is not occurring simply on an element edge, but anywhere on an element face. From the literature Chaudhary and Bathe (1986) have one of the few three- dimensional non-node on node contact algorithms. However, they only permit linear quadrilateral element faces in the contact zone, restricting the type of elements that can be used to eight noded brick elements. With these elements, the local position of contact on the target element is defined in terms of subtended area coordinates between the contacting node and each of the edges of the target element. The normal distance is again calculated from the intersection point on the target element face and the normal unit vector. This method of contact identification is also restricted to only linear elements for similar reasons to those previously described in two dimensional modelling.

Once the extent of the contact region has been identified, it is then necessary to determine whether the contact state of either sticking or sliding needs to be imposed.

5.4 Identification of the contact state

In many of the early contact algorithms, only wholly sticking or wholly sliding contact could be imposed. Examples of this are found in Francavilla and Zienkiewicz (1975), Stadter and Weiss (1979) and Mahmoud et al (1982). This is satisfactory for many types of contact problem if the contact conditions are known a priori and can considerably reduce the computational effort. However, the user may often be unaware of the contact conditions in

advance, in which case general purpose algorithms are required which can automatically calculate and apply the correct constraints.

The solution procedure most commonly used in these general purpose algorithms is to define newly contacting areas as sticking. From the forces generated after a solution the decision to change the contact state may be made. This change of state usually occurs when the tangential forces are greater than some limiting value, for example as defined from Coulomb's friction relationship.

The positions at which the forces are monitored and subsequently how the decisions on contact state are made is an important feature of the solution procedure. There are three different locations which are generally used, these being at nodes, elements or Gauss points. The techniques and advantages and disadvantages of each are now discussed.

5.4.1 Nodal decisions

Probably the easiest and most direct method of determining the contact forces and contact states is to use the forces occurring on the contacting nodes. These forces can be calculated from the element stiffnesses and displacements of the contacting nodes or by,

$$\{R_c\} = \{R\} - [K]\{U\} \qquad 5.5$$

where [K] is the overall stiffness matrix without any imposed constraints and {R} is the external applied loads vector. The resulting force vector $\{R_c\}$ contains zero and non-zero terms, the non-zero terms are the nodal global contact forces.

The decision for the nodal contact state can be made by transforming the global nodal forces into local tangential and normal forces with respect to the contact surface as shown in Figure 5.5. Note that when the master and slave technique is used for the contact surfaces, only the contact state on the master (contactor) surface is needed, as these forces will be subsequently distributed onto the slave (target) surface.

The contact state for each contactor node is then obtained from the following checks.

If,

$$F_n > 0 \qquad 5.6$$

then tension is occurring, indicating this node should be released from the contact conditions.

If,

$$|F_t| > |\mu_s F_n| \qquad 5.7$$

then the sticking condition previously applied needs to be changed to sliding, with the frictional (tangential) forces set as,

$$|F_t| = |\mu_d F_n| \qquad 5.8$$

If

$$|F_t| < |\mu_s F_n| \qquad 5.9$$

then the sticking condition should continue.

The nodal states need to be updated according to the above information and the appropriate constraints applied in the next iteration.

The nodal forces can only be directly used if they are truly representative of the contact conditions. To identify this, the nodal forces need to be examined from their derivation in the finite element method.

The discretisation of the domain into finite elements, with distinct interpolation functions (shape functions) per element, necessitates the forces acting on the finite elements to be consistent with these functions. Consequently, although uniformly distributed loading may be occurring on the real structure, the finite element nodal forces in this zone may not be similarly uniform. Examples of this can be seen in Figure 5.6, where the nodal forces to model uniform loading for different element types are shown. From this figure it becomes apparent why contact modelling using the finite element method has been almost entirely restricted to linear elements, as the nodal forces for these elements are directly indicative of the real forces.

In the case of the two-dimensional quadratic element, the sign of the contacting nodal forces does correctly identify whether separation or contact

is occurring. However, with the three-dimensional quadratic element there are tensile forces generated at the corner nodes when subjected to compressive loading. Thus using nodal forces alone to decide the contact condition is erroneous for this type of element.

5.4.2 Element decisions

An alternative decision making technique for identifying the contact state is to use forces acting on an element face, rather than nodal forces. This can be achieved by again initially calculating the nodal contact forces as in Equation 5.5. These discrete nodal forces can then either be averaged over an element by simply summing the nodal contributions as shown in Figure 5.7, or converted into an equivalent continuous pressure distribution over the surface, as shown in Figure 5.8. By numerical integration (usually Gaussian quadrature) the element forces can be calculated.

Once the element's total normal and tangential forces are known, then its state can be defined using the same conditions as for nodes (Equations 5.6 to 5.9). However, as only nodal constraints can be applied to the stiffness matrix, this requires that nodal states be defined from these element values. Bathe and Chaudhary (1985) use a system whereby if any of the adjacent elements to a node are sticking, then its state is also sticking. Similarly if all adjacent elements are sliding, or sliding and releasing, then this node is also sliding. Finally, release is defined for the node if all adjacent elements are releasing.

The technique of element decision making is more complex than for nodal checking, as 'conversion' matrices are required to convert the nodal forces into equivalent pressures and element force values.

5.4.3 Gauss point decisions

With the previously mentioned difficulties involved in three-dimensional quadratic elements in mind, Hitchings (1988) implemented the use of monitoring contact at surface Gauss point positions. This technique was used as the Gauss points are known to be positions at which the element stresses are 'best' determined (Barlow (1976)). The same contact checks for nodes and elements can also be applied at the Gauss points. One unfortunate effect with this method is that slight mesh overlap can occur. However, Hitchings results indicated that the stresses were still satisfactory.

The question arises as to which of these methods is the most reliable and hence which should be used. In Chapter ten, various test runs have been made using the different contact identification techniques. A summary of this analysis is that although slightly different results occur, they converge with increase of mesh refinement. The element contact state checking technique tends to smooth the results due to force averaging, whereas the discrete locations used in the other techniques can, in some applications, cause difficulties (see Appendix F). As previously mentioned, deciding the contact state directly at each node from its nodal forces is not suitable for three-dimensional elements of quadratic and higher order.

5.5 Implementation of the contact conditions

A key point that should be made here is that all finite element contact algorithms apply both displacement and force constraints on the structural stiffness matrix to impose the contact conditions. How these constraints are applied on a matrix level in terms of the changes and accuracy of the applied desired constraint, defines to a large extent the differences between the different contact algorithms.

In the field of finite elements the following three methods of constraint imposition have evolved.

a, Penalty method

b, Lagrange multiplier method

c, Transformation matrix method

The theoretical basis of each is now shown, from which the matrix modifications can be seen. The relative advantages and disadvantages of each method is also highlighted.

5.5.1 Penalty method

In this approach, each constraint is included by the addition of an extra term into the initial finite element variational statement (Equation 2.3), as shown in Equation 5.10.

$$\Pi^* = \Pi + \frac{\alpha}{2} C^T C \qquad 5.10$$

where α = penalty number

C = constraint equation

The 'penalty number' is analogous to stiffness with the constraint equations containing displacement relations. This extra term is physically representative of the strain energy to maintain the constraint.

As an example of this technique, consider the nodal overlap that would occur when nodes not previously contacting come into contact after an iteration, as shown in Figure 5.9. It is necessary to eliminate this 'physically unacceptable' nodal overlap in the subsequent iterations. This can be achieved if the following displacement constraint equation is imposed on the system of finite element equations.

$$u_{5n} - u_{2n} = \delta_n \qquad 5.11$$

Equation 5.11 can be re-written in matrix form, to be compatible with Equation 5.10 as,

$$\left.\begin{array}{l} [\,1 \quad -1\,] \begin{Bmatrix} u_{5n} \\ u_{2n} \end{Bmatrix} = \delta_n \\ \\ \text{or} \quad [L]\{U\} = \delta_n \end{array}\right\} \qquad 5.12$$

such that $C = [L]\{U\} - \delta_n$ 5.13

Inserting Equation 5.13 into 5.10 gives,

$$\Pi^* = \Pi + \frac{\alpha}{2}([L]\{U\} - \delta_n)^T([L]\{U\} - \delta_n) \qquad 5.14$$

Equation 5.14 can be expanded as,

$$\Pi^* = \frac{1}{2}\{U\}^T[K]\{U\} - \{U\}^T\{R\} +$$

$$\frac{\alpha}{2}(\{U\}^T[L]^T[L]\{U\} - 2\{U\}^T[L]^T\delta_n + \delta_n^2) \qquad 5.15$$

(note $[L]\{U\} = \{U\}^T[L]^T$ as this term is scalar)

Minimisation of Equation 5.15 by differentiation with respect to the displacements gives,

$$0 = [K]\{U\} - \{R\} + \alpha([L]^T[L]\{U\} - [L]^T\delta_n) \qquad 5.16$$

Rearrangement of Equation 5.16 with the unknown nodal displacements on the left hand side gives,

$$\left[[K] + \alpha[L]^T[L]\right]\{U\} = \left\{\{R\} + \alpha[L]^T\{\delta_n\}\right\} \qquad 5.17$$

The derivation of the finite element equations using the penalty method for constraint imposition has now been completed.

For the example constraint shown in Equation 5.11, the only extra terms to be added to the 'standard' finite element equations (Equation 2.1) are four terms in the stiffness matrix and two to the force vector. The rows and columns to which these terms apply is defined by the degrees of freedom in the initial constraint equation. Although only one constraint equation was applied here, the technique allows multiple constraints to be cumulatively applied. Hence global x,y,z constraints or local normal or tangential constraints can be imposed for multiple contacting nodes.

The modifications made here to the stiffness matrix have been derived entirely from a mathematical viewpoint. It is however also possible to intuitively obtain the same equations using gap elements. In this approach, a gap element is inserted between the potential contacting nodes with a very low stiffness initially defined. Once contact is identified, a very high stiffness value is inserted, effectively locking the nodes together. The contributions to the stiffness matrix for this approach are identical to that of the constraint approach (Equation 5.17). Some of the many authors who have used this gap element approach are Mazurkiewicz and Ostachowicz (1983), Zolti (1983) and Hellen (1988).

A major drawback of the penalty method is that the accuracy of the desired constraint on the system of equations is dependent upon the penalty number used. As a consequence, it is not possible to exactly apply the desired constraint. This can be realised by close inspection of Equation 5.17, where the greater the magnitude of the penalty number, the more likely these terms are to dominate the rows in which they occur. A typical row in which these terms occur is written out below.

$$\text{stiffness terms} \quad \ldots\ldots \quad -\alpha u_2 \ldots\ldots +\alpha u_5 = \alpha\delta_n \qquad 5.18$$

The dominance of the penalty terms in the rows in which they appear effectively explain why this method of constraint imposition works. The force equilibrium equation initially represented by a row in the stiffness matrix has been converted to what amounts to a displacement constraint equation, with the stiffness terms made to have little effect by the selection of a very high penalty number. However, if too high a penalty value is selected, then an ill-conditioned matrix can result. Felippa (1977) in his paper gives advice on recommended values. From results conducted in this thesis, a penalty value in the order of 10^{+6} times greater than the largest term in the initial stiffness matrix, gave convergence of results to within five significant figures of the exact solution.

A distinct advantage of the penalty method is its ease of implementation into existing finite element codes through the use of gap elements, although in this form the contact is restricted to being node on node. Further advantages are that the size and positive definiteness of the matrix remain unchanged.

The allowance of free movement when not contacting and constrained movement when contacting, requires two different stiffness values to be defined for each gap element. When switching from one to the other, the structural stiffness matrix is changed, necessitating it to be re-decomposed at the solution stage. This can be a computationally expensive procedure.

Element numbering to minimise this effect can be introduced at the mesh generation phase where the elements of potential contact are the last to be numbered. Then in the full stiffness matrix only the bottom right hand corner of the matrix requires re-decomposition.

To avoid this re-decomposition, some authors (Ishinabe (1987) and Hitchings (1988)) define only one 'medium' stiffness value for the gap elements. This allows the stiffness matrix to be decomposed just once. However, the use of medium stiffnesses results in the constraints being only 'weakly' applied with overlaps still existing after a solution. This overlap is eliminated by further iteration using new force vectors calculated after each solution. This can result though in many iterations in certain problems, especially those with varying contact areas.

In summary, the penalty method using gap elements is relatively 'easy' to install and monitor in existing finite element codes. However, matrix decomposition and re-solution can become a major phase in the solution procedure due to swapping of gap element stiffnesses or 'weakly' imposed constraints. Additionally, the technique in gap element form is restricted to node on node contact.

5.5.2 Lagrange Multipliers

With the Lagrange multiplier approach the contact constraints are included by the addition of a different term than that of the penalty method to the initial variational statement (Equation 2.3). The variational formulation for Lagrange multipliers is shown in Equation 5.19,

$$\Pi^{*} = \Pi + \lambda^{T} C \qquad 5.19$$

where λ = Lagrange multiplier

C = Constraint equation

The Lagrange multiplier is a scalar value and represents the force to maintain the constraint. The 'energy product' of the Lagrange multiplier and the constraint equation represents the 'work done' in applying the constraint.

The finite element formulation using Lagrange multipliers to impose contact will now be developed. Inserting the constraint equation of 5.13 into 5.19 gives,

$$\Pi^{*} = \Pi + \lambda^{T}([L]\{U\} - \delta_{n}) \qquad 5.20$$

Although λ^{T} is scalar, it now becomes an additional variable in the variational formulation. Hence differentiation with respect to both U and λ must be undertaken,

$$0 = [K]\{U\} - \{R\} + [L]^T\lambda \qquad 5.21$$

$$0 = [L]\{U\} - \delta_n \qquad 5.22$$

(note $\lambda^T[L]\{U\} = \{U\}^T[L]^T\lambda$ as this is a scalar term)

Equations 5.21 and 5.22 can be combined to form their more familiar Lagrange multiplier matrix form of,

$$\begin{bmatrix} K & L^T \\ L & 0 \end{bmatrix} \begin{Bmatrix} U \\ \lambda \end{Bmatrix} = \begin{Bmatrix} R \\ \delta_n \end{Bmatrix} \qquad 5.23$$

For the example constraint shown in Equation 5.11, there would be no changes inside the initial stiffness matrix, with four new terms added in extra rows and columns around the outside of the stiffness matrix and one to the augmented force vector. The rows and columns to which these terms are added are defined by the degrees of freedom in the initial constraint equation. Multiple constraints can be applied, with the dimension of the stiffness matrix and the augmented force vector increasing accordingly.

One of the main advantages of Lagrange multipliers is that the constraint equations are exactly imposed on the system of equations, although this is at the expense of increasing the size of the initial matrix.

Difficulties can occur for some equation solvers for this type of matrix due to the zero terms on the main diagonal. For example, the Cholesky decomposition method (explained in Wilkinson and Reinsch (1971)) results

in the square root of negative terms occurring during the decomposition process, which prevents solution by this technique. However, solutions can be obtained using Gaussian elimination or Crout's method.

Authors who have used the Lagrange multiplier method of constraint imposition include Hughes et al (1976), Okamoto et al (1979) , Bathe and Chaudhary (1985) and Pascoe and Mottershead (1988,1989).

5.5.3 Transformation matrix method

This technique of constraint imposition is derived not from a variational viewpoint, but from matrix operations of coupling degrees of freedom in the stiffness matrix. The constraint equation, by its very definition, connects one variable with respect to several others. This variable therefore is a dependent ratio of the other variables. This permits the initial variable to be directly eliminated from the system of equations with its contribution being added to the other terms.

The matrix operations to carry out this elimination for a single constraint are shown in Equation 5.24.

$$\left[[T]^T[K][T]\right]\{U\}^* = [T]^T\{R\} \qquad 5.24$$

where $[T]$ = Transformation matrix

$\{U\}^*$ = reduced system of unknown nodal displacements

The transformation matrix defines the matrix operations to combine the coupled degrees of freedom and is defined by the desired constraint to be imposed on the system. Its general form can be written as,

$$[T] = [I] - [L]^* \qquad 5.25$$

$$\text{where} \quad [L]^* = \begin{bmatrix} 0 \\ L \\ 0 \end{bmatrix}$$

and $[I]$ = unit matrix

Note that in Equation 5.25, the [L] constraint terms only occur in the row of the dependent degree of freedom and the columnar positions of this row as defined by the other degrees of freedom in the constraint equation. The diagonal term in the transformation matrix for the dependent degree of freedom should be zero. This can be guaranteed by ensuring the constraint coefficient term in [L] for the dependent degree of freedom is +1 (This can be achieved by dividing through the equation by the coefficient of the dependent degree of freedom). Most authors using the transformation matrix method for contact analysis use node on node contact in which case the elements of [T] can be intuitively obtained (Fredriksson (1976) & Torstenfelt (1983)). The general method of calculation of the transformation matrix for any type of constraint, as shown in Equation 5.25 has not previously been encountered by the author. Confirmation of the validity of

the technique is contained in the results chapter, where a range of different constraint problems are solved and compared with the penalty and Lagrange multiplier methods.

Advantages of the transformation matrix method are that the constraints can be exactly applied and that a reduction in dimension of the matrix occurs. Disadvantages are that extra matrix operations are required for each change in contact conditions and that re-decomposition of the reduced structural stiffness matrix is necessary. Additionally, it is necessary to have no overlaps hence one boundary condition change per load step is the only permitted loading technique.

To summarise, the main techniques of constraint imposition have now been discussed and described. From the generalised concept of defining an [L] constraint array, it has been shown that any of the methods of constraint imposition described here could be used. In the examples shown here, only node on node contact was considered. In the following chapter, it will be shown that the [L] matrix does not necessarily have to define node on node contact, in which case completely general and unified theories have been developed. The extra complications in allowing non-node on node contact however, requires much more sophisticated geometric identification routines to allow the general [L] matrix to be constructed.

The overriding question that arises from this section is which method of constraint imposition is 'best'. The definition of 'best' is a somewhat awkward quantity to define. Probably a satisfactory definition would be,

'the 'best' technique is one which provides
sufficiently accurate results in a minimum of time'.

It can be seen from this definition that the method of constraint alone is not sufficient to define the 'best' technique as time considerations become important. This is particularly relevant in varying boundary condition type contact problems, as their inherent non-linearity results in generally many tens, if not hundreds, of iterations being necessary to obtain 'sufficiently accurate' results. Thus the type of matrix solution solver used and any reduction in matrix size can have dramatic effects on the computational effort required.

5.6 Matrix solution solver techniques

The technique of substructuring where internal degrees of freedom are condensed out of the initial stiffness matrix, yielding a 'reduced' matrix is ideally suited to contact problems. The method of substructuring is well documented in the literature (for example, Stasa pp337-339). The method has in the past been used in several contact papers (Francavilla and Zienkiewicz (1975), Fredriksson (1976) and Mazurkiewicz et al (1983)). Its use however, is imperative for efficient iterative solution.

A major part in most finite element contact algorithms is the computational cost incurred with re-decomposition of the stiffness matrix, which is necessary due to the different contact constraints that become operative as the solution progresses. Hence a reduction in matrix size, achievable by using

a frontal elimination technique (Hellen (1969) and Irons (1970)), can yield considerable computational savings as smaller matrices require decomposition (the frontal elimination procedure in effect automatically imposes substructuring techniques).

Additionally, how and where the constraint terms due to contact are added to the stiffness matrix, can also have a dramatic effect on the computation required in the re-decomposition. For example, a full re-decomposition may not be necessary with certain equation solvers.

In Chapter nine, a 'hybrid' equation solver is developed which only requires one decomposition of the reduced stiffness matrix. Furthermore, the re-decomposition at each iteration stage is restricted to only the Lagrange constraint terms, and even these may not need to be re-decomposed for subsequent iterations. The overall computational efficiency with these features, can in certain applications, make the hybrid equation solver tens or hundreds of times faster in decomposition, when compared to full stiffness matrices and traditional equation solvers.

CHAPTER 6

CONTACT CONSTRAINTS

6.1 Introduction

In the preceding chapters many of the general problems encountered when modelling contact have been discussed. These problems can essentially be summarised as to how a contact is identified, and what are the 'appropriate' constraints that need to be imposed on the system of finite element equations to enforce the prevailing contact conditions.

In this chapter, a general theory is developed which allows these problems to be resolved. Towards the end of the chapter, several specific contact techniques for directly including frictional effects are described.

6.2 Contactor and target definition

Consider two separate structures, each divided into a finite element mesh as shown in Figure 6.1 (for reasons of simplicity, only one interface region and two dimensional contact will be considered in the foregoing constraint analysis, however the principles developed here can be directly applied to multi-interface contact and three dimensional problems).

Let one of the bodies be labelled the contactor and the other the target. The labelling of these bodies is important as they define the role of master and slave in the analysis. Nodes on the contactor surface are the masters, whilst the target nodes which define the element surfaces are the slaves. Contactor nodes are prevented from penetrating the target surface, however no similar constraint applies to the target nodes. This system is satisfactory for the prevention of mesh overlap provided care is taken in the selection of which body is the contactor. General recommendations are that fine meshes should be used in the region of potential contact (which the 'aware' finite element user should use for reliable stress calculations anyway) and that the body with the most curved or sharp contact zone be selected as the contactor. This final point avoids a problem that can occur of overlapped meshes not being detected as shown in Figure 6.2, where in the second diagram none of the contactor nodes have overlapped, yet considerable penetration of the target has occurred.

6.3 Identification of contact

The method of identification of contact developed here is based on geometric considerations. It uses the displacements from the last iteration to update the surface nodal coordinates. Geometric routines are then entered to identify contactor overlap.

The geometric routines used need to be accurate, as usually the displacements are small compared to the initial coordinates and hence only small differences

in values are occurring. Additionally, they need to be general to allow for the different types of element that may be used in the contact zone.

6.3.1 Local isoparametric coordinate identification

The technique for identification of contact developed here takes advantage of key characteristics of the element's local isoparametric coordinate system. For example consider the elements shown in Figure 6.3. With these elements, their surfaces are exactly defined by one of the local coordinates ($\xi, \eta,$ or ζ if 3D) taking the value of +1 or -1. With triangular elements local area coordinates are usually defined within the the elements as shown in Figure 6.4. In these cases, their outer surfaces are exactly defined by one of the area coordinates (L_1, L_2 or L_3) taking the value of 0.

The geometric identification routine relies upon the target elements on which contact is likely to occur, having their potential contact surfaces being defined at the beginning as 'active surfaces'. The associated normal local coordinates to these surfaces are also labelled as 'active' (the other local coordinates being labelled as 'non-active'). The information defining this can either be input by the user, possibly in a similar labelling system as already used in many finite element packages to define edge or facial element pressures from the element nodal topology. Alternatively, this process can be automated whereby only the target elements are defined. Then, from calculation of nodal outward normals from each element and summation of normals for shared nodes, permits the 'free' surface to be defined from non-zero normal vectors as shown in Figure 6.5.

With the active surface and the active isoparametric direction known, then identification can be undertaken of whether a node is touching, and if it is, with which target element and exactly where on its surface. This is achieved by conversion of the global coordinates of the contactor node into local coordinates of the target element in question.

This coordinate conversion is attained by the solution of Equation 6.1 for the unknown local coordinates used in the shape function matrix.

$$\{X\}_k = [N_k]_T \{X\}^e_T \qquad 6.1$$

For linear target elements of the type shown in Figure 6.3, the ξ_k and η_k terms in $[N_k]$ are linear, which allows the matrix to be directly solved for these local coordinates. However, for elements of quadratic and higher order, the shape function expressions for these elements contain similarly quadratic or higher order local coordinate terms. Consequently the system of equations to be solved is of a similar order.

The direct solution of these equations can be difficult, hence a Newton Raphson iterative technique is used here. The method is an important part of the solution algorithm and is fully detailed in Chapter eight. A key feature in the implementation is that the Newton Raphson equations are derived from the element's shape functions and its local derivatives. Hence, as these are standard routines in finite element codes, then this method of contact identification is general for any element type and independent of the complexity of shape of the actual surfaces.

Once the local coordinates have been obtained, then whether contact is occurring or not, is checked from the type of target element and the local coordinates. If the element is of a type where local coordinates of +1 or -1 indicate a surface, then contact is occurring if,

$$|\Phi_k| \leq 1.0 + TOL \qquad 6.2$$

$$\text{and} \quad |\Psi_k| \leq 1.0 \qquad 6.3$$

where Φ_k = active local coordinate

Ψ_k = the non-active local coordinates

TOL = tolerance of acceptable deviation from the exact target surface for contact.

Values of Φ_k and Ψ_k outside the limits defined in Equations 6.2 and 6.3 indicate that the node is outside the domain of the target element.

For elements in which local area coordinates are used, such as triangular elements, contact is occurring if,

$$-TOL \leq \Phi_k \leq 1 + TOL \qquad 6.4$$

$$\text{and} \quad 0 \leq \Psi_k \leq 1 \qquad 6.5$$

The above techniques can be used to identify contact irrespective of the curvature or dimensionality of the target elements. In the next section, the local coordinates of contact defined above are used in the derivation of the contact constraint equations.

6.4 Contact constraints

There are two types of contact condition that may be prevalent in a contact zone, these are sticking contact and sliding contact. In the following sections, constraints for these conditions are derived which can be generally applied for any type of element. Additionally, the contacting nodes between the two meshes can either be aligned (node on node contact) or mis-aligned (non-node on node contact).

6.4.1 Sticking contact

Consider the individual contactor node k coming into contact with the target body shown in Figure 6.6. For sticking contact to be defined between the contactor node and the target surface, requires this node to be effectively adhering to the exact position of initial contact on the target. For this to occur, the subsequent displacements of the contactor node must be identical to the subsequent displacements of the initial point of contact on the target body. From the finite element formulation, the displacements at any point within an element can be defined from the element's nodal displacements and

its shape functions. Therefore, for sticking contact the following displacement constraint can be defined,

$$\{U\}_k = [N_k]_T \{U\}_T^e \qquad 6.6$$

The displacement constraint defined by Equation 6.6 assumes the contactor node was initially just touching the target surface. However, in the general case node k may well be overlapping the target element mesh. This is likely to occur when a contactor node was previously not contacting, and due to displacements in the last solution now overlaps.

The removal of overlaps (which is necessary, as physically structural overlap cannot be accepted) is achieved by the addition of a global overlap vector to Equation 6.6. i.e.

$$\{U\}_k = [N_k]_T \{U\}_T^e + \{\Delta\} \qquad 6.7$$

where in 2-D $\{\Delta\} = \begin{Bmatrix} \delta_x \\ \delta_y \end{Bmatrix}$

In circumstances where overlap has occurred, the position of initial contact (ξ_k and η_k for the target element in Equation 6.6) on the target surface is again required. However, this is not as readily available as now the coordinates of the overlapped node do not represent this 'just touching' position. The local coordinates of initial surface contact in this circumstance can be calculated from the intercept of the updated target surface, and a line joining the

current and previous coordinates of the contactor node in question, as shown in Figure 6.7. A particularly efficient technique is to take advantage of inherent features of the target element's local coordinate system. For example in quadrilateral and brick elements, the outer surfaces are as previously stated, exactly defined by one of the local coordinates taking the value of +1 or -1 as indicated in Figure 6.3.

The ratio to initial contact γ for all the elements in Figure 6.3, can be calculated using the 'active' local coordinate,

$$\gamma = \frac{A_s - \Phi^{i-1}}{\Phi^i - \Phi^{i-1}} \qquad 6.8$$

where A_s = +1 or -1, dependent upon the active surface

Φ^{i-1} = previous active local coordinate

Φ^i = current active local coordinate

The local coordinates of the position of initial contact can now be found,

$$\xi_k = \xi^{i-1} + \gamma(\xi^i - \xi^{i-1}) \qquad 6.9$$

$$\eta_k = \eta^{i-1} + \gamma(\eta^i - \eta^{i-1}) \qquad 6.10$$

$$\zeta_k = \zeta^{i-1} + \gamma(\zeta^i - \zeta^{i-1}) \quad \text{..... (if 3-D)} \qquad 6.11$$

Obviously, the active surface local coordinate in the above will take the value of +1 or -1. The global overlap vector for sticking contact can be evaluated from,

$$\{\Delta\} = \left[[N_k]_T - [N_i]_T\right]\{U\}_T^e \qquad 6.12$$

The constraint coefficients and the overlap vector in Equation 6.7 are now fully defined. Rearrangement of this equation with separation of the known and unknown quantities to the right and left hand sides respectively gives,

$$\{U\}_k - [N_k]_T\{U\}_T^e = \{\Delta\} \qquad 6.13$$

For the two dimensional contact shown here, there are two separate constraint equations defined in Equation 6.13, one representing an 'x' direction constraint and the other a 'y' direction constraint. Writing out the first row of Equation 6.13 gives,

$$U_{kx} - [N_{kx}]_T\{U\}_{Tx}^e = \delta_x \qquad 6.14$$

which can be rearranged into the constraint form of Equation 5.12 to form,

$$\left.\begin{aligned} &\left[1 \quad -[N_{kx}]_T\right]\begin{Bmatrix} U_{kx} \\ U_{Tx}^e \end{Bmatrix} = \delta_x \\ &\text{or} \qquad [L]\{U\} = \delta_x \end{aligned}\right\} \qquad 6.15$$

Equation 6.15 is identical in form to the general constraint of Equation 5.13. Hence, either Lagrange multipliers or the Penalty method of constraint

imposition discussed in Section 5.5 could be used for applying this constraint on the stiffness matrix.

The constraint method used here is that of Lagrange multipliers. This was used because it permits exact imposition of the contact conditions. Additionally, the use of Lagrange multipliers when coupled with the 'hybrid' equation solver discussed in Chapter nine, permits 'fast' iterative decomposition which can yield considerable computational savings.

The sticking type of contact described here is probably the easiest to define as there is no relative sliding occurring. In sliding problems it is necessary to define different contact constraints.

6.4.2 Sliding contact

There are essentially two types of sliding contact which can be defined to occur, these being frictionless and frictional contact. In the following analysis, frictionless contact is initially addressed. The effects of friction are then included either iteratively or directly in the initial stiffness matrix if the direction of sliding is known.

Consider again Figure 6.6, where an individual contactor node is touching the target body, however, now the node is desired to undergo frictionless sliding. In the case of small scale sliding, the node will slide from its initial point of contact along a tangent to the target surface. For this to occur in the finite element model requires the subsequent displacements of the contactor

node in the normal direction, to be identical to the normal displacement of the initial point of contact on the target body. In terms of the nodal displacements, the following individual normal displacement constraint can be defined.

$$\{n\}^T\{U\}_k = \{n\}^T[N_k]_T\{U\}_T^e \qquad 6.16$$

where $\{n\}^T$ = normal direction cosine vector relative to the point of initial contact

In the definition of frictionless contact, there is no tangential direction constraint occurring at the contact interface. Therefore, imposition of Equation 6.16 alone (in the rearranged form of Equation 6.17) on the stiffness matrix, will cause frictionless sliding of the contactor node until 'jamming' occurs.

$$\left[\{n\}^T - \{n\}^T[N_k]_T\right]\begin{Bmatrix} U_k \\ U_T^e \end{Bmatrix} = 0 \qquad 6.17$$

6.4.3 Large scale sliding

If the contact surface is curved and large scale sliding occurs, then nodal overlap or separation as defined in Section 4.7 may result.

The method to correct this effect is dependent upon the loading procedure used. If one boundary condition change per load step is used, then the contact zone must be accurately tracked through the loading stages to permit correct identification of newly contacting nodes. This necessitates that the sliding solution be re-solved with a modified normal direction cosine vector in Equation 6.17, such that the sliding node's final position is on the target surface (or to within some pre-defined tolerance of contact). This new normal direction cosine vector can be defined from the average of the previous and new normal direction cosines.

When using the loading procedure of pre-defined load increments (Section 5.2.3), the displacements obtained from each iteration contribute to the final solution. Hence this method necessitates the overlap (or separation) distance due to large scale sliding, to be eliminated directly in the next iteration. This can be achieved by the addition of this normal overlap (or separation) distance to Equation 6.17, producing the constraint of Equation 6.18.

$$\left[\{n\}^T - \{n\}^T[N_k]_T\right]\begin{Bmatrix} U_k \\ U_T^e \end{Bmatrix} = \delta_n \qquad 6.18$$

This displacement constraint can, as for sticking contact, be directly imposed on the stiffness matrix using Lagrange multipliers.

Up to now, the contact constraints have only been derived in terms of displacements. There is however the important consideration of how the contact forces are defined and how they are distributed across the interface.

6.4.4 Distribution of contact forces

The contact forces have not been discussed up to now, this is because they are automatically defined from the minimisation of the variational statement with the applied displacement constraint. For example, consider the normal reaction force between the contactor node and the target element shown in Figure 6.8. Correct distribution of this force onto the target element is required. From standard finite element theory, point loads which are not applied directly at a node, should from virtual work considerations be distributed in the ratio of the shape functions onto the element's surface nodes, i.e.

$$\{F\}_T^e = F_n[N_k]_T^T\{n\} \qquad 6.19$$

In Lagrange multiplier theory, the Lagrange multiplier represents the force to maintain the constraint. Hence from rewriting Equation 5.23, the physical significance of the $[L]^T$ term becomes apparent,

$$\begin{bmatrix} K & L^T \\ L & 0 \end{bmatrix} \begin{Bmatrix} U \\ \lambda \end{Bmatrix} = \begin{Bmatrix} R \\ \delta \end{Bmatrix} \qquad 6.20$$

The constituents in the $[L]^T$ term and the Lagrange multiplier λ in the above equation precisely imposes the force constraint of Equation 6.19 onto the relevant contacting nodes. This is confirmed by examination of the affected rows of the stiffness matrix, where the force equation that each column of $[L]^T$ represents, imposing the contributions due to the contact forces of that specific constraint.

The physical significance of the $[L]^T$ term above is an important realisation, as in Section 6.4.6 it will be shown that it can be modified for the direct inclusion of friction.

6.4.5 Frictional contact

A major difficulty in modelling frictional contact is that it is dependent upon the load history. In particular the loads exerted on a system will define in which direction sliding will occur, and as the friction forces by their nature oppose the direction of sliding, then they cannot be accurately included unless this direction is known.

To overcome this difficulty, many authors initially carry out a frictionless solution when sliding is defined. Then from analysis of the resulting normal contact forces and the direction of sliding, Coulomb's friction relation can be used to define the 'missing' frictional forces as,

$$|F_t| = \mu_d |F_n| \qquad 6.21$$

This frictional force can be included into the system of finite element equations by resolution of the tangential friction forces occurring at the contactor nodes into global forces. Using these forces and Equation 6.19 permits the target element nodal friction forces to be determined, as shown in Equation 6.22.

$$\{F\}_T^e = -F_t[N_k]_\Gamma^T\{t\} \qquad 6.22$$

where $\{t\}$ = tangential direction cosine of the sliding direction relative to the target surface

(note, -ve force in Equation 6.22 as the target friction forces act in an opposite direction to those of the contactor's)

These forces are added to the right-hand side force vector of Equation 6.20 and re-solution undertaken. Generally, because of the effect of the new friction forces, the new normal forces generated are slightly different from those obtained with the frictionless run. As a consequence, the tangential friction forces also need to be modified to maintain the Coulomb relation. An iterative process involving re-calculation of the friction forces and re-solution, generally results in converged solutions being obtained.

However, in certain situations divergence can occur and hence no solution can be formed using this technique of 'adding the friction forces post-solution'. Campos, Oden and Kikuchi (1982) identified that there could

possibly be problems in applying friction in this manner, although they did not show in what type of problem it may occur.

In Appendix H, a class of wedge type contact problems (similar to Figure 3.4) are shown to suffer from this problem. It is indicated that provided the difference in angle between the applied forces and the taper surface is greater than the friction angle, then the problem of divergence does not occur. However,when this is not the case solutions can be obtained, provided the total contact force in the direction of the applied force is scaled to equal the applied force in this direction.

It should be pointed out that this problem of divergence rarely occurs and on the whole, 'post-solution' addition of the friction forces is satisfactory.

The number of iterations to attain frictional convergence is problem dependent. However, in general from the author's experience at least four iterations are needed assuming no change in boundary conditions. In the next section, a method is developed for the direct inclusion of the friction forces in one iteration provided the direction of sliding is known.

6.4.6 Direct inclusion of friction

The normal and tangential force relationships present under frictional sliding can be summed to define the global nodal contact forces acting on the contactor and target respectively as,

$$\{F\}_k = F_n\{n\} - F_t\{t\} \qquad 6.23$$

$$\{F\}_T^e = F_n[N_k]_T^T\{n\} - F_t[N_k]_T^T\{t\} \qquad 6.24$$

As F_t is defined from Coulomb's friction relation in terms of F_n, then these equations can be rewritten as,

$$\{F\}_k = F_n[\{n\} - \mu_d\{t\}] \qquad 6.25$$

$$\{F\}_T^e = F_n\left[[N_k]_T^T(\{n\} - \mu_d\{t\})\right] \qquad 6.26$$

Equations 6.25 and 6.26 represent the interface contact forces between a contactor node and target element. For force equilibrium these relationships need to be coupled as equal and opposite, i.e.

$$\left.\begin{aligned} F_n\left[(\{n\} - \mu_d\{t\}) - [N_k]_T^T(\{n\} - \mu_d\{t\})\right] &= \begin{Bmatrix} F_k \\ F_T^e \end{Bmatrix} \\ \text{or} \quad \lambda_n[J]^T &= \{F\} \end{aligned}\right\} \qquad 6.27$$

Equation 6.27 is a force constraint for frictional sliding and defines the distribution of F_n (and its induced F_t) on the affected contacting nodes. From Section 6.4.4 it became apparent that the $[L]^T$ terms in the Lagrange multiplier equation physically represents the force constraint on the system.

Hence this term can be replaced by $[J]^T$ to directly impose frictional sliding, i.e.

$$\begin{bmatrix} K & J^T \\ L & 0 \end{bmatrix} \begin{Bmatrix} U \\ \lambda \end{Bmatrix} = \begin{Bmatrix} R \\ \delta \end{Bmatrix} \qquad 6.28$$

It should be emphasised that Equation 6.28 can only be used for the direct inclusion of sliding, when the sliding direction defined by {t} is known. Probably the simplest procedure for obtaining this direction is to initially apply frictionless conditions for when sliding has been identified. The displacements from this solution can then be used to obtain the vector defining the direction of sliding.

An unfortunate characteristic that occurs with the direct inclusion of friction, is that it causes the 'total' stiffness matrix to become unsymmetric. This can cause difficulties in some matrix equation solvers due to increased storage and loss of symmetry in the decomposition process. In the following section, a symmetric method of friction inclusion is developed, which although not direct, has about half the iterations of the 'post-solution' friction technique discussed in Section 6.4.5.

6.4.7 Symmetric friction method

This method is essentially a hybrid version of Equation 6.28. It originated from examination of the terms in the 'stiffness' matrix and identifying what

happens when the matrix is made symmetric. It was noted that the 'post-solution' method of including friction was a symmetric variation of this equation with $[L]$ and $[L]^T$ contributions. Therefore, what would be the effect of $[J]$ and $[J]^T$ contributions ?

With the Lagrange multiplier method, it has become clearly evident that the extra rows added to the stiffness matrix define the displacement constraints on the system, and that the extra columns define the force constraints. Thus application of $[J]$ as the displacement constraint, will actually apply incorrect displacement constraints as the coefficients in the Lagrange terms now include friction (see Equation 6.27). The effect of this is that the normal which was previously used to define the sliding surface becomes modified by the coefficient of friction. This causes sliding to occur along a different slope as shown in Figure 6.9.

If a contact problem has sliding occurring in one direction, then all the contactor nodes on that surface will because of application of the $[J]$ displacement constraint, slide along the 'wrong' slope creating a gap as indicated in Figure 6.10. However, because the frictional force constraints are correctly applied (by $[J]^T$), this causes the induced contact forces to be close to the forces obtained from the 'exact' direct inclusion of friction method. As a consequence of this, the deformation of the target body, which is controlled by the generated contact forces acting on it, is close to the frictional solution obtained in the direct method.

The gap though that has resulted between the contactor and target is unacceptable, as these contactor displacements do not represent sliding

around the true contact surface. However, with the calculation and inclusion of 'corrective gap terms' into the Lagrange overlap vector, subsequent re-solution produces convergence to the same results as obtained by the direct friction method.

The 'corrective gap terms' were defined by a rigid body translation of the contactor nodes in the direction of the applied loads as shown in Figure 6.11, such that the sum of the normal gaps above the target surface were equal to the sum of the normal gaps beneath it, that is,

$$\Sigma(+ \delta_n) = |\Sigma(- \delta_n)| \qquad 6.29$$

Confirmation of the improved convergence with this method over the 'post-solution' technique of including friction is highlighted in the paper 'Two new finite element contact algorithms' contained in Appendix K, where several example problems are compared.

Although there are advantages in symmetrification of the matrix in this manner, the method relies on calculation of translation distances and normal corrective gaps. In three dimensional problems, this process could become quite complex, hence this method has only been implemented for two-dimensional contact. The main motivation for developing this method was to allow existing symmetric equation solvers to be used. The 'hybrid' equation solver developed in Chapter nine allows unsymmetric matrices to be solved and hence friction can be directly included with this equation solver.

This chapter has described and defined contact constraint equations between contactor nodes and target element surfaces for sticking, frictionless sliding and frictional sliding contact. A key feature of these constraints was that the element shape functions were used. As a result, the equations developed here can be modified by selection of the appropriate shape functions for the particular type of target element (evaluated at the point of contact), which allows the technique to be applied to almost all area (2D) and volume (3D) elements.

The constraints on the contactor body defined above were applied at the contactor's nodes. It is possible however to apply constraints at different positions on the contactor surface, for example at 'surface' Gauss points. This technique and the resulting constraints are now described.

6.5 Gauss point constraints

It was explained in Section 5.4.1 that nodal forces could only be used for deciding the contact state for certain element types, where their naturally occurring nodal contact forces were representative of the real contact conditions.

The motivation for using Gauss point constraints is derived from the fact that stresses calculated at these points are known to be the most accurate (Barlow (1976)). Thus sampling the contact forces at positions as close to the Gauss points as possible, is likely to give better indication of the contact

condition than at the nodes. These surface points can be found from extrapolation of the Gauss points to the surface as shown in Figure 6.12.

The principles defined in the previous section of this chapter for constraints at the contactor nodes also apply for the contactor surface Gauss points. The only difference is in the definition of the position of point of contact on the contactor. Hence the nodal contactor term becomes replaced by the Gauss point position defined from the shape function of the contactor element, i.e.

$$\{U\}_k \equiv [N_G]_C \{U\}^e_C \qquad 6.30$$

where $[N_G]_C$ = Contactor shape functions evaluated at the surface Gauss point

$\{U\}^e_C$ = Contactor global nodal displacement vector %%

Hence the general constraints would for surface Gauss points become,

For sticking contact,

$$\left[[N_G]_C - [N_k]_T\right]\begin{Bmatrix} U^e_C \\ U^e_T \end{Bmatrix} = \{\Delta\} \qquad 6.31$$

For frictionless sliding contact,

$$\left[\{n\}^T[N_G]_C - \{n\}^T[N_k]_T\right]\begin{Bmatrix} U^e_C \\ U^e_T \end{Bmatrix} = \delta_n \qquad 6.32$$

For frictional sliding contact,

$$[J]^T = \left[[N_G]_C^T(\{n\} - \mu_d\{t\}) - [N_k]_T^T(\{n\} - \mu_d\{t\}) \right] \qquad 6.33$$

This chapter has described and discussed how contact is identified and developed many different types of contact constraint and shown how they can be included in the finite element equations. In the next chapter, four different contact algorithms are derived which incorporate these features.

CHAPTER 7

NEW CONTACT ALGORITHMS

7.1 Introduction

Four complete contact algorithms are described in this chapter, including the architecture of the algorithms for automated operation. The first three algorithms have been coded for two dimensional elements only. The final algorithm allows both two and three dimensional modelling.

7.2 Linearised contact algorithms

Three algorithms are presented here which use the one boundary condition change per load step approach. By using this technique, the non-linear nature of the contact problem is converted into a series of linear solutions. Summation of the displacements from each solution provides the overall deformation profile. Note that as these algorithms have only been implemented in 2-D and involve linearised steps, that the Lagrange multipliers λ (which represent the forces to maintain the constraint) are in fact the nodal contactor contact forces. Therefore in these specific algorithms these contactor forces have been used to define the contact state. The solution procedure when using the linearised approach is now summarised.

Initial contact conditions are identified and the appropriate constraints and full external loads imposed. The resulting displacements and contact forces are then scaled such that the next new boundary condition just becomes operative. A new boundary condition could include any of the following,

a, change of contact state (e.g. sticking to sliding)

b, new nodal contact

c, nodal sliding into a new target element

d, nodal release

To just begin to introduce a new boundary condition a scale factor β is calculated. The displacements and contact forces from this iteration are then scaled by this value and stored. The remainder of load is then calculated. The next iteration includes the new boundary condition amongst its imposed constraint equations along with application of the remainder of load. The scaling procedure is then re-entered for the results from this iteration. This process is repeated until the structure is fully loaded and is shown below.

Solve,

$$[K^*]^i \begin{Bmatrix} U \\ \lambda \end{Bmatrix}^i = \begin{Bmatrix} R \\ 0 \end{Bmatrix}^i \qquad 7.1$$

where $[K^*]^i$ = the current stiffness matrix including the Lagrange constraint terms

and i = the current iteration (if $i=1$ then apply the full external loads).

Calculate the scale factor β to introduce the next boundary condition and then update the displacements and contact forces,

$$\sum_{j=1}^{i} \{U\}^j = \sum_{j=1}^{i-1} \{U\}^j + \beta\{U\}^i \qquad 7.2$$

$$\sum_{j=1}^{i} \{\lambda\}^j = \sum_{j=1}^{i-1} \{\lambda\}^j + \beta\{\lambda\}^i \qquad 7.3$$

Calculate the remainder of load to be applied,

$$\{R\}^{i+1} = (1-\beta)\{R\}^i \qquad 7.4$$

Increment i and re-enter Equations 7.1-7.4 until the loads have been fully applied.

More details on aspects of the individual algorithms now follow.

7.3 Direct inclusion of friction algorithm

A flow diagram showing the solution procedure when using this method is contained in Figure 7.1. Key features are that unsymmetric constraints are applied for frictional sliding, with the transition from sticking to sliding involving a frictionless sliding run to be initially imposed. This allows the direction of sliding to be correctly defined in the next frictional iteration.

7.4 Post-inclusion of frictional forces algorithm

A flow diagram outlining the solution procedure for this algorithm is shown in Figure 7.2. In particular with this method, the 'missing' nodal frictional forces are calculated from the normal contact forces and included in the force vector of the next iteration. Re-solution is continued until the Coulomb friction relation is correctly satisfied.

7.5 Post-inclusion of normal gaps algorithm

The flow diagram for this algorithm is shown in Figure 7.3. With this method, the translation distance of the contactor to equalise the normal gaps above and beneath the target surface after a solution is desired. From this translated position, the normal gap terms can be calculated and inserted in the constraints of the next iteration.

7.6 Algorithmic similarities

The flow diagrams in Figures 7.1-7.3 show many similarities between the methods. For example, the scaling and load application procedures are identical. They also highlight the main disadvantages of the one boundary condition change technique. This being that scaling can only be reliably undertaken if the displacement results in this iteration are accurate. To enable this requires several re-solves to be undertaken, to ensure the correct friction forces are acting and the elimination of any induced overlaps or gaps

due to large scale sliding. The number of re-solves coupled with the allowance of only one extra boundary condition change per load step, can cause these methods to be computationally expensive.

The following method is derived from an incremental loading approach and does not have a re-solve stage, as every solution contributes to the overall solution, and as such permits multi-changes in contact conditions per iteration.

7.7 Incremental loading algorithm

The energy functionals developed in Chapter five related to individual linear stages. For non-linear problems the external loads are applied in increments, with iteration within each increment for convergence. The general non-linear form of these equations for the i^{th} iteration within the p^{th} increment can be written as,

$$\Pi^p = \frac{1}{2}\left\{\{\Delta U_p\}^i + \sum_{j=1}^{i-1}\{\Delta U_p\}^j\right\}^T [K]\left\{\{\Delta U_p\}^i + \sum_{j=1}^{i-1}\{\Delta U_p\}^j\right\} -$$

$$\left\{\{\Delta U_p\}^i + \sum_{j=1}^{i-1}\{\Delta U_p\}^j\right\}^T \left\{\sum_{s=1}^{p}\{R_s\}\right\} +$$

$$\left\{ \{\Delta U_p\}^i + \sum_{j=1}^{i-1} \{\Delta U_p\}^j \right\}^T \left\{ \sum_{s=1}^{p-1} [K]\{U_s\} \right\} \quad 7.5$$

where

$\{\Delta U_p\}^i$ = incremental displacements from the i^{th} iteration

$\sum_{j=1}^{i-1} \{\Delta U_p\}^j$ = incremental displacements within the p^{th} increment up to the $(i-1)^{th}$ iteration

$\sum_{s=1}^{p} \{R_s\}$ = sum of external loads applied up to the p^{th} increment

$\{U_s\}$ = total displacements from each previous increment, s

Equation 7.5 is a non-contact energy formulation. The general form of constraint to be imposed on Equation 7.5 for contact conditions is shown in Equation 7.6.

$$[L]^{i-1} \{\Delta U_p\}^i - \{\Delta\}^{i-1} = 0 \quad 7.6$$

$[L]^{i-1}$ = an m x n constraint coefficient array derived after the $(i-1)^{th}$ iteration,

$\{\Delta\}^{i-1}$ = vector of m overlaps

where m = number of constraint equations

and n = dimension of stiffness matrix

Including Equation 7.6 into Equation 7.5 and allowing for previously calculated contact forces gives,

$$\Pi = \Pi^p + (\{\Delta\lambda\}^i)^T([L]^{i-1}\{\Delta U_p\}^i - \{\Delta\}^{i-1})$$

$$- \left\{\{\Delta U_p\}^i + \sum_{j=1}^{i-1} \{\Delta U_p\}^j\right\}^T \{R_c\}^{i-1} \qquad 7.7$$

Where

$\{\Delta\lambda\}^i$ = incremental contact forces generated in this iteration

$\{R_c\}$ = previously evaluated contact forces

Taking the first variation of Equation 7.7 with respect to the incremental displacements and incremental Lagrange multipliers gives,

$$0 = [K]\{\Delta U_p\}^i - \sum_{s=1}^{p} \{R_s\} + [K]\sum_{j=1}^{i-1} \{\Delta U_p\}^j$$

$$+ [K]\sum_{s=1}^{p-1} \{U_s\} - \{R_c\}^{i-1} + ([L]^{i-1})^T\{\Delta\lambda\}^i \qquad 7.8$$

and

$$0 = [L]^{i-1}\{\Delta U_p\}^i - \{\Delta\}^{i-1} \qquad 7.9$$

These two equations can be combined to form,

$$\begin{bmatrix} [K] & ([L]^{i-1})^T \\ [L]^{i-1} & 0 \end{bmatrix} \begin{Bmatrix} \Delta U_p^i \\ \Delta\lambda^i \end{Bmatrix} = \begin{Bmatrix} \sum_{s=1}^{p} R_s \\ 0 \end{Bmatrix} -$$

$$\begin{Bmatrix} [K]\sum_{s=1}^{p-1}\{U_s\} + [K]\sum_{j=1}^{i-1}\{\Delta U_p\}^j \\ 0 \end{Bmatrix} + \begin{Bmatrix} R_c^{i-1} \\ \Delta^{i-1} \end{Bmatrix} \qquad 7.10$$

which can be written as,

$$[K^*]\begin{Bmatrix} \Delta U_p^i \\ \Delta\lambda^i \end{Bmatrix} = \{R\} - \{F\}^{i-1} + \begin{Bmatrix} R_c \\ \Delta \end{Bmatrix}^{i-1} \qquad 7.11$$

or more simply,

$$[K^*]\begin{Bmatrix} \Delta U_p^i \\ \Delta\lambda^i \end{Bmatrix} = \begin{Bmatrix} \Delta R \\ \Delta \end{Bmatrix}^i \qquad 7.12$$

where $[K^*]$ = global stiffness matrix including the current Lagrange constraints

$\{\Delta U_p^i\}$ = incremental displacements from this iteration (should obviously decrease in each iteration as convergence is approached for each load increment)

$\{\Delta\lambda^i\}$ = incremental contactor contact forces from this iteration

$\{R\}$ = sum of the incremental load so far

$\{F\}^{i-1}$ = sum of the total element stresses upto the $(i-1)^{th}$ iteration of the present load increment

$\{R_c\}^{i-1}$ = sum of the total nodal contact forces upto the $(i-1)^{th}$ iteration of the present load increment

The total displacements are found by summation of the incremental displacements from every iteration. The incremental form of Equation 7.10 is not new, as other authors (e.g. Bathe and Chaudhary (1985)) have developed similar equations. However, the formation of a general constraint equation [L] makes these equations universally applicable. Similarly, provided the total contactor forces are known then they can be distributed in the ratio of the constraint equations to define the forces on the target nodes. An important feature of the algorithm is how these contact forces are obtained.

The form of Equation 7.10 is similar to standard non-linear finite element equations. The (R-F) term on the r.h.s. usually represents 'out-of-balance' force terms which are generally removed by further iteration. However, in the contact formulation, these 'out-of-balance' terms represent the nodal contact forces. Hence for convergence, the nodal contact forces (defined in the $\{R_c\}$ vector) should be equal and opposite to the (R-F) terms at these nodes. In fact after each iteration, the (R-F) and $\{R_c\}$ terms do exactly satisfy this condition (in elastic analyses). However, because of changes in contact subsequently identified after this iteration, the $\{R_c\}$ contact force vector or $\{\Delta\}$ overlap vector may become modified. For example, if a change of state from sticking to frictionless sliding is defined, then the $\{R_c\}$ tangential contact force components calculated due to the sticking constraint need to be set to zero, in which case $\{R_c\}$ is modified and a nett force residual is produced from summation of the r.h.s. force terms of Equation 7.11. Additionally, new contact constraints may be defined, in which case overlap terms in the r.h.s. of Equation 7.11 also need to be eliminated. Thus residuals in the r.h.s. of Equation 7.11 (force or displacement) occur as a consequence

of changes in contact conditions when compared to the last iteration. Convergence can therefore be defined within a load increment when the load residual and the overlap vector to be applied in the next iteration are both below pre-determined tolerance values.

7.7.1 Convergence criteria

The Euclidean norms of the residual displacements and the residual forces were calculated after each iteration and convergence defined to have occurred for that load increment, when these were beneath pre-determined tolerance values. The norms were calculated as,

$$E_U = \sqrt{\sum_{k=1}^{n} U_k^2} \qquad 7.13$$

$$E_R = \sqrt{\sum_{k=1}^{n} \Delta R_k^2} \qquad 7.14$$

where n = total displacement degrees of freedom in the stiffness matrix

The magnitude of the tolerance values selected are quite important. If too low a value is selected, then a high precision is being defined, requiring extra iterations that probably do not significantly contribute to the accuracy of the final solution. Conversely, if too high a value is selected, then convergence may be accepted for quite a poor (inaccurate) solution. Comparison of the results after many iterations for different models have indicated that a convergence criteria of E_R to 0.1% of the load increment, and E_U to 0.1% of

the total displacements in that increment, to be satisfactory convergence tolerances.

7.7.2 Incremental solution procedure

A flow diagram showing the solution stages for the incremental loading approach is shown in Figure 7.4. Key features of this algorithm are that no re-solutions are involved as every iteration contributes towards the converged incremental solution. Additionally, multi-changes in boundary condition can be facilitated within each iteration. The contact force checking routine and geometric contact checking routines are now discussed in more detail.

7.7.3 Contact force checking routine

In the first contact iteration only {R} terms exist in the {ΔR} force vector. The nodal forces equivalent to element stresses $\{F\}^{i-1}$ can be evaluated as,

$$\{F\}^{i-1} = \{F\}^{i-2} + [K]\{\Delta U\}^{i-1}$$

The nodal contact forces are then obtained from,

$$\{R_c\}^{i-1} = \{R\} - \{F\}^{i-1}$$

Specifically, only the contactor nodal forces are needed which are converted from global to local forces relative to the contact surface. The total normal and tangential force acting on each element is then obtained by dividing the nodal forces by the number of elements each node shares. These divided nodal forces are then summed for each element, to give the total normal and tangential forces per element face. The contact state of the element is then defined by application of the Coulomb friction check, i.e.

a, if the tangential forces exceed the limiting value defined by the product of the normal force and the static coefficient of friction, then the element is defined as sliding.

b, if the tangential forces are less than this limiting value, then the element is defined as sticking.

c, if the normal forces are positive, then release of the element is defined.

The decision on contact state was made on an element rather than nodal level because of the fluctuations in nodal force values that can occur in the initial iterations. The use of element forces tends to smooth out these effects. However, the constraints imposed on the stiffness matrix must be at a nodal level, hence the contact state at each node must be decided from the element states. The technique used here is to decide the nodal state from the state of adjacent elements.

a, if any of the adjacent elements to a node are sticking, then the

node's state is also sticking.

b, if all of the adjacent elements are sliding or sliding and releasing, then the nodal state is sliding.

c, if all of the adjacent elements are releasing, then the nodal state is also releasing.

The nodal states, which define the type of constraint equation to be applied, have now been identified. It is also necessary to modify the contactor contact nodal forces $\{R_c\}$ to be consistent with these nodal states.

a, if the nodal state is sticking, then no change is made to the nodal forces.

b, if the nodal state is sliding, then the tangential nodal force is scaled to the frictional value defined by the normal force and the dynamic coefficient of friction. This tangential force is applied in the direction opposing motion.

c, if the nodal state is releasing, then its contact forces are set to zero.

The $\{R_c\}$ vector for the contacting contactor nodes has now been defined. To obtain the forces on the target nodes these forces are distributed in the ratio of the latest constraint equations (as defined by the shape functions)

onto the relevant target nodes. The $\{R_c\}$ vector is now fully set-up allowing the residual force vector for the next iteration to be calculated.

7.7.4 Geometric contact checking routine

The procedure for identifying contact is fully explained in the next chapter and essentially involves identifying whether or not a contactor node is contacting the target body. The process involves checking each contactor node with each potential target element. Two checks are undertaken, the first is a rough check which quickly identifies if a node is in the possible domain of the target element. If it is, a more accurate check is then undertaken to precisely identify the local coordinates of the node's final position. If the node is found to be inside the target element, then the target element number, the position of initial contact, the amount of overlap, and the contactor node number are stored (which are then used to define the constraint equation in the next iteration).

CHAPTER 8

Contact Identification Procedure

8.1 Introduction

The identification of contact is a vital stage in a general contact algorithm. The process used here essentially involves checking each contactor node for contact with the potential target surface elements of the other body. Different contact checking procedures were used, depending upon the contactor node's previous state.

8.2 Previously non-contacting nodes

For nodes not previously in contact it is necessary to identify for each potential contactor surface node whether it has contacted the target body, and if it has, with which target element. The contact checking technique employed here involves checking each contactor node against each target element. This comprises of two stages, the first stage is a rough geometric check which identifies quickly whether a node is in the possible domain of a target element. This is undertaken by obtaining the maximum and minimum global x,y (and z if 3-D) coordinates of the target element in question. The updated global coordinates of the contactor node are then checked for whether they lie in the target element's coordinate range. If any of the contactor node's coordinates lie outside the x,y or z range, then contact with this target element is not occurring.

It is possible however, for a contactor node to pass through the element against which it should actually be touching as shown in Figure 8.1a. To overcome this difficulty, the maximum displacement occurring in the previous iteration is added to either side of the target element's coordinate range. This maximum displacement is generally small and results in an envelope extending around the true element's shape as shown in Figure 8.1b, allowing rough contact to be detected. Once the rough contact checking routine identifies contact, it is necessary to enter the accurate contact checking routine.

The purpose of the accurate contact checking routine is to identify the precise location of the previous and current global coordinates of the contactor node within the isoparametric local coordinate system of the relevant target element. Then by using linear interpolation, the position of initial contact on the target element's surface can be identified using Equations 6.8 to 6.11.

8.3 Previously sticking nodes

For nodes that were previously sticking, then the element with which it is contacting is known, in which case it is necessary only to check whether there is any overlap as stated in Equation 6.12.

8.4 Previously sliding nodes

For these nodes the element against which it was previously contacting is known, the first check therefore is to identify whether it is still contacting this element. This is done by converting its new position into local isoparametric

coordinates of this element and checking that the non-active local coordinates lie in the range +1 to -1. If they do not, then this node must be checked with the other potential target elements until the new target element against which it is touching is identified.

8.5 Global to isoparametric coordinate conversion

The conversion of global x,y (and z if 3-D) coordinates to local $\xi, \eta,$(and ζ if 3-D) coordinates for quadratic and higher order elements requires the solution of non-linear equations. The technique developed here is iterative and makes use of the target element shape functions. The method had to be capable of accurately locating contact as generally only small changes in displacement occur. Additionally, reasonable convergence characteristics were desired.

As previously stated the global coordinates of the contactor node need to be converted into the equivalent isoparametric coordinates of the target element, the general equations therefore which need to be solved are,

$$\left.\begin{aligned} x_k &= [N(\xi_k, \eta_k, \zeta_k)]_T \{x\}_T^e \\ y_k &= [N(\xi_k, \eta_k, \zeta_k)]_T \{y\}_T^e \\ z_k &= [N(\xi_k, \eta_k, \zeta_k)]_T \{z\}_T^e \end{aligned}\right\} \qquad 8.1$$

Where ξ_k, η_k and ζ_k are the unknown isoparametric coordinate values of the target element exactly equivalent to the global coordinates of the contactor node (x_k, y_k and z_k) . The shape function relationships containing ξ_k, η_k and ζ_k in Equation 8.1 are of the same order as the target element (i.e.

linear if linear elements and quadratic if quadratic elements etc.). Hence these equations for quadratic and higher order elements are non-linear.

One method for solving the system of equations in Equation 8.1 is to apply the Newton Raphson iterative solution procedure (e.g. see Jeffrey p651). This procedure essentially involves,

1. calculation of the shape functions using initial estimates for ξ, η and ζ
2. calculation of an improved estimate of the local coordinates from Newton Raphson theory using an equation of the form,

$$X^{i+1} = X^i - \frac{f(X^i)}{f^1(X^i)} \qquad 8.2$$

Here three separate equations in x, y and z need to be solved. In the following analysis it is necessary to allocate each local coordinate to its 'nearest in orientation' global coordinate. Assuming that $\xi, \eta \wedge \zeta$ are nearest to x,y, and z respctively then rearrangement of Equation 8.1 and application of the Raphson theory gives,

$$\xi^{i+1} = \xi^i - \frac{[N(\xi^i, \eta^i, \zeta^i)]_T \{x\}^e_T - x_k}{\frac{\partial}{\partial \xi}[N(\xi^i, \eta^i, \zeta^i)]_T \{x\}^e_T} \qquad 8.3$$

$$\eta^{i+1} = \eta^{i} - \frac{[N(\xi^{i}, \eta^{i}, \zeta^{i})]_{T}\{y\}^{e}_{T} - y_{k}}{\frac{\partial}{\partial \eta}[N(\xi^{i}, \eta^{i}, \zeta^{i})]_{T}\{y\}^{e}_{T}} \qquad 8.4$$

$$\zeta^{i+1} = \zeta^{i} - \frac{[N(\xi^{i}, \eta^{i}, \zeta^{i})]_{T}\{z\}^{e}_{T} - z_{k}}{\frac{\partial}{\partial \zeta}[N(\xi^{i}, \eta^{i}, \zeta^{i})]_{T}\{z\}^{e}_{T}} \qquad 8.5$$

The updated values for ξ, η and ζ obtained after the first solution from Equations 8.3-8.5 are used to calculate updated shape function terms. These equations are then re-entered and this process repeated until convergence in the local coordinate values occurs. The use of shape functions and their local derivatives in the above equations takes advantage of standard shape function routines which generally return both of these quantities.

If the element shapes are relatively 'uncurved', then Equations 8.3-8.5 converge in relatively few iterations (less than ten). However, because these equations are coupled, then for curved elements very poor convergence (and in some cases divergence) can occur. To overcome this problem the following technique has been developed, which is based on the Newton Raphson theory and has accelerated and reliable convergence characteristics.

8.6 Accelerated accurate identification routine

Consider the node and target element shown in Figure 8.2. Two fictitious positions are calculated from this node. The first (x_k, y_k, z_l) is obtained assuming constant 'non-active' global coordinates (x and y in this case) and

a constant active local coordinate ($\zeta = +1$, representing the top surface in this case). The second position (x_2, y_2, z_k) is obtained assuming a constant 'active' global coordinate (z in this case) and constant 'non-active' local coordinates Ψ_k as calculated from position 1.

The values of the active local coordinate Φ at positions 1 and 2 bound the exact solution of the active local coordinate Φ_k for node k. However, a close approximation can be obtained for Φ_k , by construction of a normal from k to the line formed by positions 1 and 2. The close approximation for Φ_k can then be obtained from,

$$\Phi_k^{i+1} = \Phi_1 + \beta(\Phi_2 - \Phi_1) \qquad 8.6$$

where β = ratio along the line 1-2 to the intersection point of the normal with the element surface

The above process (of forming new positions 1 and 2 and improved estimates of Ψ_k and Φ_k) can be repeated until suitably accurate results are obtained.

The equations to execute this technique are now presented for 3-D (a somewhat similar approach also applies to 2-D involving the definition of two fictitious positions, although ζ and z obviously do not enter the solution).

Position 1

Using an initial (constant) active local coordinate, calculate the non-active local coordinates to achieve the same non-active global coordinates of k, i.e.

1. Use constant x_k, $\zeta^0 = 1.0$, $\eta^0 = 0$ variable ξ ($\xi^0 = 0$)

$$\text{solve} \qquad \xi^1 = \xi^0 - \frac{[N(\xi^0, \eta^0, \zeta^0)]_T \{x\}_T^e - x_k}{\frac{\partial}{\partial \xi}[N(\xi^0, \eta^0, \zeta^0)]_T \{x\}_T^e} \qquad 8.7$$

Let $\xi^0 = \xi^1$ and re-solve Equation 8.7 until convergence is attained to give ξ^1.

2. Use constant y_k, $\zeta^0 = +1$, $\xi = \xi^1$, variable η ($\eta^0 = 0$)

$$\text{solve} \qquad \eta^1 = \eta^0 - \frac{[N(\xi^1, \eta^0, \zeta^0)]_T \{y\}_T^e - y_k}{\frac{\partial}{\partial \eta}[N(\xi^1, \eta^0, \zeta^0)]_T \{y\}_T^e} \qquad 8.8$$

Let $\eta^0 = \eta^1$ and re-solve Equation 8.8 until convergence is attained to give η^1.

Position 2

Using the non-active local coordinates calculated for position 1, calculate the active local coordinate to achieve the same active global coordinate of k, i.e.

3. Use constant z_k, $\xi = \xi^1$, $\eta = \eta^1$, variable ζ ($\zeta^0 = +1$)

$$\text{solve} \qquad \zeta^1 = \zeta^0 - \frac{[N(\xi^1, \eta^1, \zeta^0)]_T \{z\}_T^e - z_k}{\frac{\partial}{\partial \zeta}[N(\xi^1, \eta^1, \zeta^0)]_T \{z\}_T^e} \qquad 8.9$$

Let $\zeta^0 = \zeta^1$ and re-solve Equation 8.9 until convergence is attained for ζ^1.

4. Use ξ^1, η^1 andζ^0 to calculate z_1

5. Use ξ^1, η^1 andζ^1 to calculate x_2 and y_2

From similar triangles in Figure 8.2 the value of β can be obtained as

$$\beta = \frac{b^2}{b^2 + c^2} \qquad 8.10$$

where $b = z_k - z_1$

and $c = \sqrt{(x_k - x_2)^2 + (y_k - y_2)^2}$

Using Equation 8.6, an improved estimate of the active local coordinate is obtained. This value and the latest non-active local coordinates can be used as new initial guesses in Equation 8.7 etc., and the process repeated. This procedure then continued until the local coordinates obtained for these fictitious positions are to within some pre-defined tolerance of each other, indicating converged values.

The derivation shown here has assumed that contact is occurring on a surface where ζ is the active coordinate. For contact on other surfaces, i.e. ξ or η active coordinate systems, then the same basic principles hold although different local coordinates would represent the non-active directions. Similarly the terms for b and c in Equation 8.10 would also be correspondingly changed.

CHAPTER 9

HYBRID MATRIX EQUATION SOLVER

9.1 Introduction

A new 'hybrid' equation solver facilitating 'fast' solution of iterative contact problems with varying boundary conditions has been developed. It takes advantage of the following features,

a, Lagrange multipliers

b, the frontal elimination method

c, substructuring and

d, an active zone equation solver

The particular advantages in using these features are now discussed.

9.2 Features and operation of the 'hybrid' technique

Lagrange multipliers have been used to apply the contact constraints, the reasons why they were used are two fold, firstly they allow the constraints to be exactly imposed and secondly they do not effect the 'inside' of the initial stiffness matrix, i.e. no terms are added as in the penalty method and no row or column multiplications are needed as is necessary with the transformation

matrix method. The advantage of this will become clear when the active zone equation solver technique is described.

The frontal elimination method permits only the elements and nodes of the wave 'front' to be retained as the solution progresses with Gaussian elimination of the fully summed terms. The advantages of this are that the full stiffness matrix need never be formed, permitting considerable space savings compared to storage of the full matrix. The technique is well explained in Irons and Ahmed pp203-213. A slightly modified version is used here where the elimination process is aborted when all the non-contact degrees of freedom have been eliminated. More details on the method now follows.

The technique of substructuring, is to a large degree automatically incorporated in the frontal elimination process. How the 'front' identifies which degrees of freedom are either externally loaded nodes or potential contacting nodes, and hence are not to be eliminated, is defined here by the use of a fictitious element of highest element number containing this degree of freedom (d.o.f.) list. An element versus d.o.f. lookup table is set up with this fictitious element included, as shown for a typical example in Figure 9.1. The frontal elimination procedure involves the calculation of one element stiffness matrix at a time. This is added to the overall current stiffness matrix. Each degree of freedom in this is then checked to see if it is fully summed. This can easily be identified from scanning along the row in which this degree of freedom occurs in the table shown in Figure 9.1. If no further reference in the remaining elements is made, then this d.o.f. can be eliminated. The frontal elimination process is aborted when all the elements have been added

into the 'front' and only the last fictitious element's degrees of freedom remain in the reduced stiffness matrix.

The active zone equation solver decomposes the reduced stiffness matrix into lower and upper triangular form as shown in Equation 9.1,

$$[K] = [L][U] \qquad 9.1$$

A review of the technique is shown here along with its particular advantages. Full documentation of the method can be found in Stasa pp270-275.

A key advantage of the active zone equation solver is that the terms for [L] and [U] are calculated using only previously evaluated [L] and [U] values, and from stiffness terms from the current 'active zone' of the original stiffness matrix. These active zones can be seen in Equation 9.2, where a simple 3x3 stiffness matrix has been partitioned into its three active zones.

$$\begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \left[\begin{array}{c|c|c} K_{11} & K_{12} & K_{13} \\ \hline K_{21} & K_{22} & K_{23} \\ \hline K_{31} & K_{32} & K_{33} \end{array}\right] \end{array} = \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \left[\begin{array}{c|c|c} 1 & 0 & 0 \\ \hline L_{21} & 1 & 0 \\ \hline L_{31} & L_{32} & 1 \end{array}\right] \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \left[\begin{array}{c|c|c} U_{11} & U_{12} & U_{13} \\ \hline 0 & U_{22} & U_{23} \\ \hline 0 & 0 & U_{33} \end{array}\right] \end{array} \qquad 9.2$$

The calculation of the [L] and [U] terms is undertaken from the use of each active zone in turn as,

zone 1 $\qquad U_{11} = K_{11}$

zone 2 $U_{12} = K_{12}$

$$L_{21} = K_{21} / U_{11}$$

$$U_{22} = K_{22} - L_{21} / U_{12}$$

zone 3 $U_{13} = K_{13}$

$$L_{31} = K_{31} / U_{11}$$

$$L_{32} = (K_{32} - L_{31}U_{12}) / U_{22}$$

$$U_{23} = K_{23} - L_{21}U_{13}$$

$$U_{33} = K_{33} - L_{31}U_{13} - L_{32}U_{23}$$

The above equations are shown in their general form for any size of matrix in Appendix E.

A vital feature of this decomposition is that the [L] and [U] terms for the current active zone can be calculated entirely from the previously decomposed active zones of [L] and [U], and the current [L], [U] and [K] active zone. The essence of why this is such a good method can be realised from the general trend of the propagation of the active zones, these being from the top left of the matrix, with adjacent positions to the right and below being used in the next and subsequent zones. This permits changes to be made to the outside of the stiffness matrix, i.e. the addition of extra active zones, which are completely isolated from the initial matrix. The advantage

of this being that changes can be made to the outside of the reduced stiffness matrix, with these extra terms having no effect whatsoever on the previously calculated [L] and [U] terms of the initial reduced stiffness matrix. There are obvious advantages when this technique is allied with Lagrange multipliers, as only one decomposition of the reduced stiffness matrix is necessary.

With regard to contact modelling, any changes in the contact conditions identified after each iteration doesn't need a full decomposition, as only the decomposition of the Lagrange constraint terms is required. In the case of relatively few contact constraints as is often encountered in the early stages of 'growing contact', then this extra computation is quite low. Further advantage can be gained if the region of contact does not involve gross sliding. In these circumstances, the constraint equations for previously contacting nodes may not change, and hence their previously obtained [L] and [U] decompositions would remain valid.

Solutions are obtained for each iteration using two backward substitution phases as shown below,

$$\left.\begin{aligned} [L]\{z\} &= \{R\} \\ [U]\{u\} &= \{z\} \end{aligned}\right\} \qquad 9.3$$

Due to the relatively small coefficients of the Lagrange constraint terms when compared to the stiffness values, round-off errors can be encountered causing inaccurate displacements to be obtained. However, using a process of

iterative refinement, accurate displacements result where the error residuals are calculated as shown in Equation 9.4.

$$\{r\} = [K]\{u\} - \{R\} \qquad 9.4$$

By back-substitution of this error residual, replacing {R} in Equation 9.3, corrective displacement terms can be calculated.

The active zone equation solver is a fast technique of constraint imposition and equation solver when coupled with Lagrange multipliers and the frontal elimination method. Other methods of constraint imposition such as the Penalty method or the Transformation matrix method involve changes to the actual stiffness matrix. These changes being either the addition of extra terms, or row and column multiplications. As a consequence of this, the advantage of the active zone technique as shown above is not realised with these methods.

One of the commonest equation solver methods in finite element codes is Gaussian elimination. There are not, however, any 'active zone characteristics' in this decomposition technique. Hence, if Lagrange multipliers (or the Penalty method) are used, then re-decomposition from the row in which the first columnar contact term occurs becomes necessary with Gaussian elimination. This is because the technique involves direct elimination of variables using row subtraction, thus the first row in which a Lagrange term occurs, has effect for the remainder of the decomposition.

CHAPTER 10

RESULTS

10.1 Introduction

One of the purposes of this work was to develop a successful and reliable contact solution method. However, from previous chapters, it becomes apparent that there are many different features that need to be 'optimised' to attain a desirable algorithm.

In this chapter, some results for the two-dimensional algorithms presented earlier are discussed. The majority of the results though, relate to the final general purpose algorithm CONTACT FORTRAN, which can solve both two and three-dimensional contact problems.

10.2 CONTACT FORTRAN

The finite element contact code developed here has evolved from the standard NAGFE finite element program SEG1P1DP FORTRAN on the I.B.M. 3081 mainframe computer. The program (SEG1P1DP FORTRAN) in its virgin form only allowed elastic static analysis of two dimensional plane strain linear quadrilateral elements. The advantage of using this code was that it was fully documented and simply structured to allow its manipulation

by finite element researchers. The final finite element program developed by the author, CONTACT FORTRAN, bears little resemblance to the initial SEG1P1DP code, as substructuring using the frontal elimination technique, load incrementation or prescribed displacements and a new equation solver have been included, along with the addition of three extra element types. These elements being the eight noded quadratic two dimensional quadrilateral element, the eight noded linear three dimensional brick element and the twenty noded quadratic three dimensional brick element. To facilitate these extra element types, the three dimensional constitutive law was added (Zienkiewicz (1977)), along with the appropriate shape function and numerical integration routines for the varying number of Gauss points needed for the different elements. To confirm the correct implementation of these routines, standard (non-contact) finite element runs were undertaken using CONTACT FORTRAN, the results of which were validated using BERSAFE.

Commercial finite element packages such as BERSAFE make use of highly efficient data storage techniques. These techniques are quite advanced and could not be implemented inside CONTACT FORTRAN within the scope of this work. As a consequence of this, the models analysed here cannot hope to represent the truly complex models that can occur in real problems. The principles and techniques, however, for modelling contact can be identified and implemented for relatively simple meshes. Comparison of these results with different techniques and known analytic results give confidence in the validity and accuracy of the approaches developed.

All the finite element meshes were generated using PATRAN. A translator program converted the PATRAN 'neutral' files to CONTACT FORTRAN input data format with manual modification to include the extra contact information. Appendix D details the main changes needed to permit contact analysis. The final output from CONTACT FORTRAN contains the interface nodal contact forces, the final nodal displacements and a summary of the nodal contact states as the solution progressed, as well as convergence criteria regarding displacements and forces after each iteration.

Stress results were obtained using BERSAFE which allowed advantage to be taken of their existing stress calculation routines. This was achieved by translation of the CONTACT finite element input data files into a BERSAFE compatible input file, with prescribed nodal displacements at every node as defined by the CONTACT displacements results file. BERSAFE results were then subsequently obtained. Stress and displacement plots were then produced by conversion of the BERSAFE results into a PATRAN compatible results file.

10.3 Verification

The results presented in this chapter can be separated into two different classifications, these being 'initial verification' problems and 'complex' problems. The initial verification involved results for very simple models with the purpose of confirming the basic validity of the constraint theory and the correct implementation of the finite element code. Only a sample of these are shown here, as the results from these analyses run into several hundred

pages. The complex models were run when the code was known to be satisfactory with the purpose of checking the accuracy and convergence criteria for different techniques and features.

10.4 Simple verification runs

As previously mentioned, the purpose of running simple verification runs was to confirm the constraint theories and ensure the correct implementation of the finite element code. Many test runs were undertaken and a selection of typical models used and their results are given in Appendix F. The type of checks that were carried out during these runs were,

a, confirmation that the displacement constraint was correctly imposed by comparison and plotting of the displacements of the constrained nodes and the target surface.

b, in certain cases, symmetric models and symmetric loading were used. In these circumstances, the displacements that result should obviously also be symmetric. This symmetry check is particularly useful for confirmation of the correct positional insertion of the constraint terms into the stiffness matrix and correct equation solution.

c, the sum of the contact forces at the interface in the direction of the applied load, should (provided no restraints exist on the contactor body in that direction) be exactly equal and opposite to the applied external loads in that direction.

d, the sum of the contactor interface normal and tangential forces

should be exactly equal and opposite to the sum of the target interface forces in those same directions.

e, in the case of sticking contact with node alignment across the interface, then standard (non-contact) results for an equivalent single mesh should yield identical results.

The results in Appendix F validate the nodal constraint imposition theory of using Lagrange multipliers and constraint equations derived from shape functions. The specific features of sticking, frictionless sliding, direct and iterative inclusion of friction and the use of different element types has also been confirmed. Presented in the following section are results for more complicated finite element models.

10.5 Advanced models

In this section, more advanced contact models are run and the results compared with known analytical solutions.

The first set of results included here show the accuracy of results that can be obtained. Additionally, important features of nodal contact state and convergence criteria as the solution progresses are highlighted and discussed. The second phase of results are comparative runs to highlight differences of loading, constraint type, constraint method and mesh effects. In particular, what is their influence on the accuracy and computational efficiency of the solutions obtained ?

Very few analytical solutions are available for frictional contact conditions. The commonest technique of comparison of finite element contact results, therefore, is with analytical frictionless solutions. Probably the most well known of these are the Hertzian contact equations, listed in Appendix C. These equations quite reliably predict the contact stresses and the contact area for cylindrical contact provided the contact area is small. The confirmation of frictionless finite element results with these and other frictionless analytical results, gives confidence in the general finite element contact method. The finite element results with friction are then subsequently compared with the frictionless results and particular trends are looked for. For example, in sliding contact, the amount of sliding would be expected to decrease with the inclusion of friction, due to the induced frictional tangential forces opposing the motion. Additionally, tangential forces should be introduced which comply with the imposed friction relationship.

10.5.1 The Hertz contact problem

The Hertz contact problem consists of a cylinder subjected to normal loading on a frictionless rigid flat surface as shown in Figure 10.1. Two different finite element meshes were used to allow comparison of the results between four noded linear elements and eight noded quadrilateral elements. The shape of the mesh was kept constant for both models, whilst advantage of symmetry allowed just a quarter of the cylinder to be modelled. The finite element mesh for the linear elements is shown in Figure 10.2, with mid-side nodes present in the quadratic element mesh. The load incrementation solution procedure was used with a prescribed vertical displacement on the

top surface of 0.2mm. Converged solutions were obtained in five iterations for both models with the contact width being two contactor elements wide. The displacement plot obtained from PATRAN for the linear mesh is shown in Figure 10.3. This plot is misleading as it shows that the contactor mesh has overlapped the target body. This is not actually the case, as examination of the nodal displacement values shows no overlap to be present. The mesh 'overlap effect' is caused by magnification in the plotting routine of displacements for nodes which were not initially contacting. For this reason in the remainder of this thesis displacement plots have mainly been avoided with the displacement values tabulated. The normal contact stress results obtained using CONTACT FORTRAN are shown in Figures 10.4 and 10.5. As would be expected, the maximum σ_y stress (equivalent to normal stress in this case) is at the point of contact and a high stress concentration is observed in this zone. The linear elements yield a maximum contact stress of 212.55 MN/m^2 whereas the quadratic elements give 197.95 MN/m^2. From the Hertzian equations (Appendix C), the maximum normal contact pressure can be calculated for this problem, provided the equivalent load to cause 0.2mm deformation is known. This load was obtained by summing the normal contact forces on the contactor body, as these values are equal and opposite to the applied force. The maximum contact pressure using the Hertzian equations with this force is 196.8 MN/m^2. Thus finite element results to within an accuracy of 1% of the Hertzian results have been attained for the quadratic mesh. Additionally, the size of the contact zone from the finite element run (two elements wide) is 0.868mm compared to the predicted Hertzian width of 0.869mm. Furthermore, the plots of maximum shear stress

τ_1 contained in Figures 10.6 and 10.7, confirm the sub-surface location of the maximum value.

The iterations involved to obtain these final converged results for the quadratic mesh are now discussed. The contactor nodal states after each iteration are shown in Table 10.1. It can be seen that after the first iteration where just a single node was initially contacting, that deformation of the contactor has caused five new nodes to overlap. Sticking constraints to eliminate these overlaps were applied in the second iteration. After this stage, the end contactor node was defined as releasing, as tensile forces were encountered on this element face. The sticking nodes were also changed to sliding in this iteration as the tangential forces exceeded Coulomb's friction relation. In the next three iterations no change in contact state occurred, however, these iterations were necessary to remove residual forces and to eliminate overlaps. Convergence occurred when the Euclidean norm of the force vector to be applied in the next iteration, and the Euclidean norm of the displacements in the last iteration were beneath a pre- defined tolerance value. For all the incremental solution runs undertaken in this thesis, a force tolerance of 0.1% of the load per increment and a displacement tolerance of 0.1% of the total displacements, were defined as the convergence criteria.

In the next set of results, the effect of different loads on the quadratic mesh was investigated in terms of the number of contacting nodes and the contact stresses that result, with comparison to Hertz's solution. Figure 10.8 shows the theoretical Hertzian normal contact stress variation and the finite element maximum contact stress results for four different prescribed displacements. The number of contacting nodes increased with increase of prescribed

displacement as shown in Figure 10.9. The difference in the contact area size between the Hertz and finite element results is predominantly due to the coarseness of the mesh.

A more difficult Hertzian problem rarely attempted by finite element contact researchers, is where the flat surface is also deformable. This, as mentioned in previous chapters, is more complicated to model as the shape of the target surface varies with application of load. The mesh used for this problem is shown in Figure 10.10 and consists of 77 quadratic eight noded elements. Different material properties were defined for the cylinder and base, with the cylinder having a Young's Modulus ten times less than the base. A prescribed displacement of 0.2mm was applied to the top surface.

The Hertzian solution for this problem gives a maximum contact stress of 181.7 MN/m^2. The maximum σ_y contact stress from the finite element model was 185.3 MN/m^2. This represents an error of 3% which, considering the relative coarseness of the mesh, is an acceptable result. The contact stresses should be equal in magnitude at the interface between the contactor and target, and this is confirmed in the σ_y plot shown in Figure 10.11. The number of iterations for convergence in this problem was five.

10.5.2 Pin in a hole contact model

The contact problem to be analysed consists of a frictionless pin in a conforming hole with small radial clearance as shown in Figure 10.12. The Hertzian theory of contact between cylinders can give satisfactory results for

this type of problem by definition of a negative radius in the equivalent radius equation (Equation C.2 in Appendix C). However, when very small radial clearances are present (typically $\Delta R/R$ less than 0.01) the Hertzian equation becomes ill- conditioned and poor results for contact stress and contact area subsequently occurs. Johnson (1985) in his book on contact mechanics lists two different analytic solutions by Steuermann (1939) and Persson (1964), which more accurately predict the contact conditions in these circumstances. Results are presented here for the one boundary condition change per load step approach and the incremental loading approach which are both compared with the analytic solutions.

The finite element meshes initially used for this problem are shown in Figure 10.13. A radial clearance of 0.01mm was used with a pin of 5mm radius.

The nodal contact states occurring in each iteration under a load of 1030 N are shown in Tables 10.2 and 10.3 for the one boundary condition change and load incrementation methods. It can be seen from these results that the one boundary condition method took eighteen iterations to converge whereas the incremental method converged in seven iterations. The larger number of iterations for the one boundary condition method were necessary as scaling was undertaken to allow just one extra node to contact in each iteration. Also, initially defining newly contacting nodes as sticking and then a change of state to sliding, required a further two iterations per newly contacting node.

The size of the final contact zone for the one boundary condition change per load step approach was 60^0 (equivalent to nine contacting nodes) and 52.5^0

(equivalent to eight contacting nodes) for the incremental loading approach. This difference in contact area is probably because the applied force of 1030N was that required with the one boundary condition change approach to just bring the ninth node into contact. However, with the incremental method, residuals of load remain (although small) which may cause this ninth node to not come into contact.

A plot of the average normal pressure per element for both models is compared with Persson's theoretical solution in Figure 10.14.

The next series of runs investigated the effect of changing the applied load on the size of contact area between the pin and hole. These plots are shown in Figure 10.15 and are also compared with the Steuermann, Persson and Hertz solutions, the latter known to be unsatisfactory. It can be seen that both the one boundary condition and the load incrementation approach results lie in the range defined by the former analytic solutions. The solution by Persson is accepted as the more accurate (Johnson (1985)) and it can be seen that the one boundary condition change results are closer than the load incrementation approach to this solution. This is probably because of the gradual application of load with the one boundary condition change per load step method.

The effect of whether a closer solution to Persson's results could be attained for the incremental load method by increase of mesh refinement was next investigated. The refined mesh used is shown in Figure 10.16. Results are plotted in Figure 10.15 which confirm closer values to Persson's with increase of mesh refinement.

Up to now, only frictionless results have been obtained. The following section compares results from the one boundary condition change loading approach, using the three different techniques for including friction as discussed in Chapters six and seven.

10.5.3 Comparison of friction techniques

Two simple models covering flat and curved frictional contact of mis-aligned meshes were analysed using the one boundary condition change loading techniques described in Chapter seven. The meshes and results for these runs are described in the paper by Pascoe and Mottershead (1989) (section 5.1 and 5.2 of Appendix K).

In summary, with all three methods the models analysed converged to the same frictional solution. The method of directly including friction using unsymmetric constraints converged the fastest, whilst the 'post-inclusion' of friction method took approximately twice as many iterations. The corrective gap method was approximately half way between these results in the number of iterations for convergence.

10.5.4 Sliding frictional cylinder contact problem

A cylinder sliding over a frictional flat surface, with both bodies having the same elastic properties, has been analytically solved by several researchers including Sackfield and Hills (1983). Finite element results for this contact

problem have been obtained here using the one boundary condition change with direct inclusion of friction method, and the load incrementation method with friction included in subsequent iterations. The finite element mesh used is shown in Figure 10.17 with a coefficient of friction of 0.2. The number of iterations for convergence with the two methods was 80 iterations for the one boundary condition change loading method and 10 iterations for the load incrementation method with the load applied in one increment. The analytic contours defined by Sackman and Hills of principal shear stress are compared with the finite element one boundary condition results in Figure 10.18, and maximum shear stresses plotted alone for the load incrementation method in Figure 10.19. In Figure 10.19 the contactor stresses have been set to zero, this was necessary because high stresses occurred at the point of load application which dominated the stress pattern, masking out stresses in the target body. The location of the maximum shear stress values by the finite element method in both cases is sub-surface and offset from the centre as predicted by the analytical results. The values of this shear stress are over-estimated by 15% for the one boundary condition change loading method and under-estimated by 10% for the load incrementation method.

10.5.5 Large scale frictional sliding problem

Conforming contact problems are particularly difficult to analyse by closed form analytical solutions because of problems in defining suitable displacement profile and pressure distribution functions. A feature of conforming type contact problems is that small changes in the contact

conditions (such as the coefficient of friction) can have a significant effect on the solutions obtained.

Consider the contact problem of a wedge sliding into a similarly shaped cavity as shown in Figure 3.4. This problem was analysed using the load incrementation method with the mesh shown in Figure 10.20. Scaling of the contact forces after each iteration was implemented, to overcome convergence difficulties due to 'excessive' tangential forces being defined with the post-inclusion of friction (see Appendix H for further discussion on this effect). The finite element model consists of 48 quadratic elements and a mis-aligned mesh in the contact zone. Several runs were undertaken to investigate the effect of different coefficients of friction. A summary of these results is now presented. In Figure 10.21, the displacement profiles of the loaded contactor 'front' surface is shown with increasing coefficient of friction. As would be expected, the sliding displacements of the contactor gradually decrease with increase of coefficient of dynamic friction. Additionally, the resisting 'x' direction forces at the contact interface cause a shear type displacement profile of the contactor. The average normal and tangential contact pressure per element is plotted in Figure 10.22. The tangential pressures exactly satisfy the Coulomb friction relation with a noticeable decrease in normal contact pressure with increase of coefficient of dynamic friction. This occurred because the tangential frictional forces act to reduce the amount of sliding displacement and hence reduce the subsequent normal pressures induced. The number of iterations to attain convergence with increase of coefficient of friction is shown in Figure 10.23. This plots shows that the introduction of friction has little effect on the number of iterations for convergence.

10.5.6 Variation of slope models

The effect of varying the slope in the wedge model above, using the same load and a constant dynamic coefficient of friction of 0.2 was investigated. The number of elements in the meshes were maintained as in Figure 10.20, the only difference was the angle of the slope defined for the contact surface.

Results for two slope values either side of the 0.1 tangent slope used above are now presented. From Figure 10.24 and Figure 10.25, it can be seen that the amount of sliding and the normal contact pressures both decrease with increase of slope. These are trends that would be expected in the real contact problem.

10.5.7 Comparison of aligned and mis-aligned mesh results

The effect of variation of mesh characteristics between node on node contact and mis-aligned nodal contact are now compared. The discretisation process of the contacting bodies should ideally (provided a fine enough mesh is used) produce similar contact results irrespective of whether the mesh is aligned or mis-aligned.

The previously obtained results for the 0.1 tangent slope wedge contact problem with a mis-aligned mesh, are now compared with results obtained using the aligned mesh shown in Figure 10.26. From the displacement plot in Figures 10.27, it can be seen that almost identical results are obtained.

10.6 Comparison of constraint techniques

In Chapter five, it was shown that provided a constraint equation could be formed, that the Penalty method, Lagrange multiplier method or the Transformation matrix method (if no overlap) could be used to impose the constraint. In this section, results are obtained for a single 'simple' model which allow comparisons to be made between the accuracy of the technique applied. A second 'advanced model' is then used which allows the differences in computational efficiency to then be compared.

10.6.1 Simple model analysis

The mesh shown in Figure 10.28 was used for the simple model. A single frictionless sliding constraint was imposed.

The final global displacement of typical nodes for this mesh are tabulated for the different methods in Table 10.4. There are four sets of results for the penalty method, where gradually higher values of penalty number have been used.

From Table 10.4, it becomes apparent that identical results were obtained by the Lagrange multiplier and Transformation matrix methods. The Penalty method results are quite severely in error for low penalty values, although a penalty value 10^{+6} times greater than the highest stiffness term, resulted in identical results to those of the Lagrange multiplier and Transformation matrix methods being obtained. It should be noted that the Transformation

matrix method can only solve problems where no overlap occurs, this limits its use to 1 b.c. change loading.

10.6.2 Advanced model analysis

The 'advanced model' used in this case was the wedge contact problem shown in Figure 10.20. In this model slight overlaps are induced as the solution progresses. The Transformation matrix method is therefore not able to solve this problem by the load incrementation method. The Penalty and Lagrange multiplier methods are compared with a penalty value 10^{+6} greater than the highest term in the stiffness matrix.

The displacement and force results obtained by the different techniques were identical. However, the computational solution times to obtain these results, varied depending upon the constraint imposition technique. The Lagrange multiplier method total c.p.u. time was 17.8 secs, whilst the Penalty method's c.p.u. was 21.1 secs. Examination of the contributions to these times showed 10.7 secs was due to initial element formation and frontal elimination. The contact iterations therefore took 7.1 secs and 10.4 secs respectively. This c.p.u. benefit by Lagrange multipliers is entirely due it being necessary to decompose only the Lagrange constraint terms in each iteration, whereas full elimination was necessary with the Penalty method.

10.7 Variation of the number of load increments

All the models and results using the incremental loading approach up to now have had the loads fully applied in just one load increment. The effect of increasing the number of load increments on several of the previous frictionless and frictional models is now examined.

For the pin in a hole contact problem, the contact pressures for one and ten load increments were within 0.1%. Similarily, for the sliding frictional cylinder problem, results to within 0.1% with one and ten load increments were also obtained.

10.8 Further constraint investigations

In this section, results for some further constraint techniques have been coded and verified for some simple models. The first method shows how constraints using Lagrange multipliers can be applied at Gauss points. In the second method, a technique for imposing slope continuity constraints on elements not containing slope degrees of freedom is presented. Finally in the third method, a variation of the Penalty method is described which allows frictional effects to be directly included in the stiffness matrix.

10.8.1 Lagrange multipliers and Gauss point constraints

The constraint theory for this technique is explained in Chapter six. Displacements have been obtained for the simple two block contact problem described in Section F.1 of Appendix F, using constraints imposed at the surface Gauss points. These results were found to be identical to those obtained by imposition of nodal constraints, as undertaken in Section F.1.

10.8.2 Lagrange multipliers and slope constraints

In the course of this work, the development of displacement nodal constraint techniques was extended to slope constraints. The slope constraint theory is detailed in Appendix G. It essentially involved either equating local derivatives of the shape functions of one element to local derivatives of the adjacent element, if slope continuity is desired, or setting the local derivatives to zero if prescribed displacements are desired.

As a consequence of applying these slope constraints, novel constraints can be applied which previously were difficult to define. For example, consider the situation where a prescribed uniform displacement needs to be imposed such that the total equivalent force due to this displacement equals some pre-determined value. By applying slope constraints, the exact desired force and the corresponding uniform displacements can be automatically imposed.

This constraint technique has been applied to the finite element mesh shown in Figure 10.29. Seven local derivative constraints were imposed, two at three

of the corner nodes and one at the other. (Note, application of eight constraints over defines the problem, causing a singular matrix to be formed). A single vertical load of 100N was applied in separate runs at different corners on the top surface and was found to have no effect on the final solution. The displacement results for this mesh are also shown in Figure 10.29. It can be seen that a uniform prescribed displacement has been imposed on the top surface.

10.8.3 Penalty method and the direct inclusion of friction

This method hasn't been developed from a strict mathematical approach, but from inspection and speculative modification of the contributions to the stiffness matrix with the Penalty method. The technique is described in Appendix I, and involves replacing the 'normal' $\alpha[L][L]^T$ penalty terms added to the stiffness matrix by $\alpha[L][J]^T$ terms. Initial investigations using this technique have shown that frictional effects can be directly included. For example the sliding contact problem shown in Figure 10.28 produced identical frictional results with this method in one iteration, when compared to the post-inclusion of friction results and the Lagrange multiplier direct inclusion of friction results.

10.9 Special application contact problems

In this section, two special applications of the contact methods developed here are presented which were undertaken under contracts with Thackray

Ltd and Ford U.K.. The contact problem for Thackray involved analysis of the contact stresses occurring in an artificial knee joint prosthesis. The contact problem for Ford involved the dynamic analysis of a disc brake assembly.

10.9.1 Knee joint prosthesis analysis

The knee joint prosthesis that was analysed is shown in Figure 10.30. The purpose of the analysis was to identify and locate the position of maximum stress on the two separate contact interfaces, between the 'femoral' and 'meniscal' component and the 'meniscal' and 'tibial' components. The procedure and results are fully described in Mottershead and Pascoe (1988) of which the main features are described in Appendix J. A summary of the analysis is discussed here.

The contact modelling for the knee prosthesis involved three-dimensional curved frictional contact between three separate components, resulting in two separate contact interfaces. The finite element mesh used is shown in Figure 10.31.

The mesh, although being quite coarse, gave results that indicated the positions of maximum stress to be consistently at the outer regions of the contact surfaces, as shown in Figure 10.32. As a consequence of this work, further analysis is to be carried out with more refined meshes and possible modifications are to be made to the prosthesis design. It is hoped to obtain a more uniform distribution of the contact stresses, thereby improving the design life of this component.

10.9.2 Dynamic brake assembly analysis.

The finite element mesh of the brake assembly to be dynamically analysed is shown in Figure 10.33 and consists of five separate components. The purpose of the analysis was to identify if there were any natural frequencies around a specific frequency of 2500 Hz, which was known to be a frequency at which 'brake squeal' occurred in some brake assembly designs.

The key role of the contact analysis in this case was not to obtain a contact stress solution, but to use the contact algorithm to form the constraint equations allowing the individual finite element meshes to be coupled. In total over 500 constraint equations were defined to allow complete coupling of the five individual meshes. These equations were formed by entering the geometric contact identification routines within CONTACT FORTRAN. Each interface was analysed in turn to identify the location of the potential contactor nodes on one surface, with the potential target elements on the other. The constraint equations were then applied as 'multi-point constraints' in MSC-NASTRAN.

A dynamic solution of the assembled mesh was undertaken using modal analysis, giving the natural frequencies and mode shapes of the brake assembly. The results from this work are commercially confidential, although it was shown that the method of joining meshes in this manner was successful.

CHAPTER 11

DISCUSSION

11.1 Introduction

The main findings of the work in this thesis are discussed in this chapter, including discussion on,

a, the method of load application
b, the method of identifying contact
c, the method of defining the contact state
d, the method of imposing the contact constraints
e, the method of equation solution

Additionally, features such as mesh effects (e.g. refined or mis-aligned meshes) and the use of sub-structuring are also discussed.

11.2 Method of load application

From the results obtained in Chapter ten, contact results have been obtained which compare well with the analytical solutions for the different methods of load application, i.e. allowing one boundary condition change or using pre-defined load steps. However, some of the variation in results that occurred, were found to be caused by the coarseness of the meshes. For

example, in the pin in a hole contact problem, the initial differences in calculated contact angle between the different loading approaches was 7.5° (equivalent to one node), however with mesh refinement, closer results were obtained by the incremental method. The contact interface pressures were also found to improve with mesh refinement.

Results for the sliding cylinder problem, confirmed that similar results between the incremental and one boundary condition loading approach could be obtained, which were close to the analytic solution. The load incrementation technique however, allowed solution in significantly less iterations, as re-solves to correct induced overlaps (or gaps) occurring due to the relatively large amount of sliding, were not necessary.

A significant feature with the incremental method was its rate of convergence and number of total iterations to obtain the final solution, this being quicker than the one boundary condition change technique. One of the main reasons for this was because multi-changes in contact were permitted per iteration. The amount of reduction in number of iterations is problem dependent, but for the problems shown here was of the order of a factor of three or four. In the larger 3-D models which can be analysed in commercial packages, even greater gains would be realised, with many more nodes permitted in the likely contact zone.

The effect of varying the number of load increments on the displacement results was found to be marginal, with the variation in displacements between one load step and ten load steps producing results that varied by less than 0.1%. This 'closeness' of results was unexpected, as it was felt that

allowing gradual build up of the contact zone (as is achieved when using 10 load increments) would yield different results. A possible reason for the similarity in results may be due to the method of contact force evaluation and distribution. To calculate the contact forces requires the total contactor surface forces to be evaluated after each iteration, these are then distributed in the ratio of the 'latest' constraint equation onto the target nodes. This procedure is valid when the displacements are small compared to the element sizes, in which case the constraint equations for a each contactor node do not appreciably change with sliding. However, if the displacements are of the same order as the elements, then the constraint equations will change significantly during loading in which case the target forces may need to be stored and 'built-up' as the solution progresses.

11.3 Method of identifying contact

The geometric identification of where contact is occurring (especially with mis-aligned meshes) represents one of the main difficulties in developing general contact algorithms. The method described in Chapter eight has proven to be reliable and accurate for the quadrilateral and brick elements analysed in this work. Slight modifications would need to be made to this process for triangular shaped elements, although the concept of evaluating accurate local coordinates (area coordinates in this case) remains the same.

It is possible that contactor nodes, which are undergoing sliding, may ***just*** slide off the edge of the target element into free space. In this circumstance the geometric identification routines would identify the node as not touching any element, and hence it would be removed from the contacting constraints.

Geometrically this is correct, but from a modelling viewpoint this can cause poor results to occur. This is because the region adjacent to this node and the next constrained node on the contactor body now has no constraints applied to it and hence may result in overlap of the target body in the next iteration. In this type of problem it is best to include this sliding node in the contact constraints, since it has only just slid off the element. This can be achieved by defining an extended tolerance around the target element when checking for contact of sliding nodes in this special circumstance.

11.4 Method of identifying the contact state

As mentioned in Chapter five, deciding the contact state from the nodal forces is only reliable for elements whose generated nodal contact forces are representative of the real forces, and in particular for compression and tension, to allow identification of nodal release. For the two dimensional algorithms the use of nodal or element contact state decision making processes generally results in the same solutions. However, in certain circumstances the end regions of the contact zone may have less contacting nodes when the contact state is decided by nodal rather than element forces. This effect can occur for example on a 2-D linear element face, where the end zone of a region of contact has a large compressive force on the pen-ultimate node and a small tensile force on the last contacting node. If the nodal force decision making process is used, then release of this end node would be defined. With the element force decision making process though, this end node would be defined as contacting because its state is defined from the state of adjacent element faces. It can therefore be seen that a different number of contacting nodes may be defined depending upon the method

used. The reason why this occurs is because the end of the 'real' contact zone lies somewhere between this pen-ultimate and end node.

The question arises as to which of the contact state decision making processes is nearer to the true solution and what is their influence on the solution? In the case of the nodal approach, the contact stresses will be higher because the interface forces are transmitted over fewer nodes and hence a smaller area.

If the nett contact forces on an element face indicate compression (even though the end node is in tension) then this generally indicates that the 'real' contact zone extends to nearer the tensile node than the compressive node. In which case it is best to include the tensile node in the contact constraints, as would be achieved when using the element approach.

The influence of this effect on the maximum stresses depends on the problem considered. If the contact region tends to grow in size with the application of load which is one of the most common forms of contact, then the maximum stresses occur in the 'central' region of the ensuing contact zone and the effect of these 'remote nodes' becomes of less importance on the contact stresses. Additionally, the use of many elements in the contact zone would reduce this effect.

11.5 Method of imposing the contact constraints

From the results in Chapter ten, it has been shown that the Penalty, Lagrange multiplier and Transformation matrix method can all be used to impose the contact constraints (although no overlaps are permitted with the

Transformation matrix approach). Furthermore, the exact imposition of constraints for non-aligned meshes in just one iteration was achieved with all these methods. To achieve this the penalty value was about 10^{+6} times greater than the highest term in the stiffness matrix. The technique of applying high penalty numbers with normal and tangential gap elements though, causes 'locking' of the results to the sticking solution. Because of this effect, medium stiffness values (penalty numbers) are a necessity when tangential gap elements are used, which as a consequence requires extra iterations to remove overlaps due to the weakly imposed constraints. It has been shown here, that by application of a normal direction constraint alone and using the Penalty method with a very high stiffness value, allows frictionless sliding to be accurately imposed. Then using the induced normal contact forces calculated from the (R-F) force term, the frictional forces can be included in the force vector of the next iteration.

The technique of including the friction forces 'post-solution' was implemented within the incremental contact algorithm and is not restricted to the Penalty method of constraint imposition. However, as stated in Appendix H, for models where shallow contact slopes and horizontal loading occur, scaling may be necessary to ensure 'excessive' frictional forces are not imposed which can cause slow convergence and even possibly divergence.

11.6 Elimination of internal degrees of freedom

All the results obtained using the incremental algorithm in Chapter ten incorporated frontal elimination, which left just the potential contact degrees of freedom (d.o.f.) and the d.o.f.'s at which external loads were applied in a

reduced stiffness matrix. At each iteration stage therefore, solution of a reduced matrix rather than the total structural stiffness matrix was undertaken. Compared to the solution of the initial global stiffness matrix this yields considerable computational savings.

In the case of the Penalty and Transformation matrix methods, a complete re-decomposition of the reduced matrix, including the new constraints was needed for each iteration. The savings per iteration for these methods with the reduced matrix, compared to the full system are generally quite significant.

The solution time for a symmetric banded matrix is dependent upon the decomposition method used, for Gaussian elimination is approximately equal to,

$$\text{time} = N(2H - 1)^2 \qquad 11.1$$

The reduced stiffness matrix produced after frontal elimination is generally not banded but fully populated. In Equation 11.1 the reduced matrix size (N) can be substituted for the bandwidth (H). The time to solve this reduced 'fully populated' matrix can still be dramatically less than for the initial banded matrix. For example, with a model of 1000 d.o.f.'s and a bandwidth of 100, of which say 50 d.o.f.'s are on the potential contact region, then inserting these values into Equation 11.1 gives an 80 fold reduction in solution time for the reduced matrix. For larger models the savings can be even greater.

With the use of the active zone equation solver and Lagrange multipliers, still further computational savings can be made. Only one decomposition of the reduced stiffness matrix is necessary and decomposition in subsequent iterations is only required for the outer constraint terms. The computational savings by this process are dependent upon the number of constraint equations added. For increasing area type contact problems this method is particularly advantageous, as the initial iterations where relatively few nodes are contacting, can be solved with minimal extra decomposition.

11.7 Large deformation and plasticity

Although contact coupled with plasticity has not been addressed here, the technique of applying nodal constraints between the potential contact surfaces is equally valid in the plastic regime. With regard to plasticity the elimination of all internal degrees of freedom now needs to be modified, such that the nodes connected to elements that may become plastic, are now prevented from elimination. This then allows the effects of variation of element stiffness associated with plasticity to be included on these nodes. Another important point is that the incremental contact algorithm developed here, is similar in its form to other non-linear approaches, and as such is well suited to the merging with current plasticity finite element theories.

With large deformation problems, the elimination of all internal degrees of freedom can cause problems with the contact identification routine. This difficulty occurs due to the current method of updating only the degrees of freedom retained in the reduced stiffness matrix. This results in distorted shape functions because the sub-surface nodal coordinates aren't updated.

Provided the deformations are small (which is generally the case in linear elastic analysis) then the shape function relationships are adequate to identify contact. However with large deformations, the shape function relationships can become quite distorted.

This problem can be overcome by either including all the nodes of the surface elements to be retained in the reduced stiffness matrix, or by entering a back-substitution phase after each iteration, permitting the sub-surface nodal coordinates to be updated.

11.8 Mesh difficulties

The refined mesh that was initially used in the pin and hole contact problem, involved just refinement of the contactor (pin) mesh alone and is shown in Figure 11.1. For this mesh extremely poor contact pressure results were obtained, with the maximum normal contact pressure not under the pin's centre, but at its edge. This was found to be occurring due to the coarse nature of the target mesh and the geometric inaccuracy in modelling a 15° curve with a single quadratic element. Magnification of the mesh geometry between the refined contactor element mesh and the coarse target element mesh is also shown in Figure 11.1 and highlights the cause. It can be clearly seen that the contactor mesh will not progressively increase in contact from the pin's centre, but touch initially at the edges of the target element. This will obviously cause poor stress results to be obtained as contact is occurring not on the desired target profile.

This effect is apparent in problems where quite coarse mis-aligned meshes are used to model curved contact regions. Provided either refined meshes are used or if this is not possible then aligned meshes (which ensures the geometric inaccuracies are identical in both contacting bodies) then this problem can be avoided.

In linear elastic contact problems generally only small amounts of displacement occur, it is therefore necessary that the accuracy of the contacting surfaces be maintained as high as possible during the mesh generation phase for reliable finite element contact stress calculation. Appendix B details procedures that allow accurate mesh generation.

CHAPTER 12

CONCLUSIONS

Constraint imposition theories have been developed and applied based on the element shape functions. Both sticking and sliding constraint equations have been defined. These allow the contact conditions to be imposed between any set of meshes, and are not restricted to node on node contact. Frictional effects have been included iteratively in a general load incrementation algorithm, although direct techniques have been implemented in two-dimensional analysis.

The method of constraint imposition used has been the Lagrange multiplier method, although it has been shown that either the Penalty or Transformation matrix method could also be used, even for mis-aligned meshes. However, no overlaps are permitted with the Transformation matrix method. The advantage gained with the Lagrange multiplier method coupled with an active zone equation solver, yield this technique as the most computationally efficient. Several one boundary condition change per load step algorithms were developed, although the results indicated that the incremental loading algorithm was considerably quicker in obtaining the overall solution. Additionally, the use of this algorithm and in particular its solution procedure, make it compatible for extension into other non-linear finite element areas such as plasticity and creep.

The process of substructuring using the frontal elimination method has been implemented. This allows just the degrees of freedom relating to the potential contact nodes and the loaded nodes to remain in a 'reduced' stiffness matrix. Subsequently, the size of the matrices to be solved at each iteration is considerably less than that of the initial structure's 'full' stiffness matrix.

Further investigation on larger models, permitted by coding this work into commercial finite element packages and in particular for 3-D problems is recommended, as this would allow further comparison of the different contact state decision making processes such as using Gauss points or element forces.

REFERENCES

Andersson, T., Fredriksson, B. and Allan-Persson, G. 'The boundary element method applied to two dimensional contact problems', *New developments in Boundary Element Methods* ,(Ed. Brebbia) CML publications, Southampton (1980)

Barlow, J., 'Optimal stress locations in finite element models', *Int. J. Numer. Meth. Engng.* V10, pp 243-251 (1976)

Bathe, K-J. and Chaudhary, A., 'A solution method for planar and axisymmetric contact problems', *Int. J. Numer. Meth. Engng.* ,V21 pp65-88 (1985).

Becker, A.A. and Plant, R.C.A., 'Contact mechanics using the boundary element method', *Proc. Inst. Mech. Eng.* V201, pp975-980 (1987)

Beer, G., 'An isoparametric joint/interface element for finite element analysis', *Int. J. Numer. Meth. Engng.* , V21, pp585-600 (1985)

Bentall, R.H. and Johnson, K.L., 'Slip in the rolling contact of two dissimilar elastic rollers', *Int. J. Mech. Sci.* , V9, pp389-404 (1967)

Belytschko, T. and Lin, J.I., 'A three-dimensional impact penetration algorithm with erosion', *Comput. Struct.* , V25, pp95-104 (1987)

Campos, L.T., Oden, J.T. and Kikuchi, N., 'A numerical analysis of a class of contact problems with friction in elastostatics', *Comp. Meth. Appl. Mech. Engng.* , V34, pp821-845 (1982)

Chan, S.K. and Tuba, I.S., 'A finite element method for contact problems of solid bodies, Part I - Theory and Validation', *Int. J. Mech. Sci.* , V13, pp615-625 (1971)

Chan, S.K. and Tuba, I.S., 'A finite element method for contact problems of solid bodies, Part II - Application to turbine blade fastenings', *Int. J. Mech. Sci.* , V13, pp627-639 (1971)

Chang, T.Y., Saleeb, A.F. and Shyu, S.C., "Finite element solutions of two-dimensional contact problems based on a consistent mixed formulation', *Comput. Struct.* , V27, pp455-466 (1987)

Chaudhary, A. and Bathe, K-J, 'A solution method for static and dynamic analysis of three-dimensional contact problems with friction', *Comput. Struct.* , V24, pp855-873 (1986)

Chen, W.H. and Tsai, P., 'Finite element analysis of elastodynamic sliding contact problems with friction', *Comput. Struct.* , V22, pp925-938 (1986)

Chen, W.H. and Yeh, J.T., 'Finite element analysis of finite deformation contact problems with friction', *Comput. Struct.* , V29, pp423-436 (1988)

Cheng, J.H. and Kikuchi, N., 'An incremental constitutive relation of unilateral contact friction for large deformation analysis', *J. Appl. Mech.* , V52, pp639-648 (1985)

Clough, R.W., 'The finite element in plane stress analysis', *Proc. 2nd ASCE Conf. on Electronic Computation* , Pittsburgh, Sept 1960

Coulomb, C.A., 'Theorie des machines simples' (1781)

Dilintas, G., Laurent-Gengoux, P. and Trystram, D., 'A conjugate projected gradient method with preconditioning for unilateral contact problems', *Comput. Struct.* , V29, pp675-680 (1988)

Endo, T., Oden, J.T., Becker, E.B. and Miller, T., ' A numerical analysis of contact and limit-point behaviour in a class of problems of finite elastic deformation', *Comput. Struct.* , V18, pp899-910 (1984)

Felippa, C.A., 'Error analysis of penalty function techniques for constraint definition in linear algebraic systems', *Int. J. Numer. Meth. Engng.* , V11, pp709-728 (1977)

Fischer, U. and Melosh, R.J., 'Solving discretized contact problems using linear programming', *Comput. Struct.* , V25, pp661-664 (1987)

Francavilla, A. and Zienkiewicz, O.C., 'A note on numerical computation of elastic contact problems', *Int. J. Numer. Meth. Engng.* , V9, pp913-924 (1975)

Fredriksson, B., 'Finite element solution of surface non-linearities in structural mechanics with special emphasis to contact and fracture mechanics problems', *Comput. Struct.* , V6, pp281-290 (1976)

Gaertner, R., 'Investigation of plane elastic contact allowing for friction', *Comput. Struct.* , V7, pp59-63 (1977)

Gangal, M.D., 'Direct finite element analysis of elastic contact problems', *Int. J. Numer. Meth. Engng.* , V5, pp145-147 (1972)

Guerra, F.M. and Browning, R.V., 'Comparison of two slideline methods using ADINA', *Comput. Struct.* , V17, pp819-834 (1983)

Hellen, T.K., 'A frontal solution for finite element techniques', *C.E.G.B. Report No RD/B/N1459.* (1969)

Hellen, T.K., 'A gap element facility in BERSAFE', *C.E.G.B. Report No. TPRD/B/1043/R88* (1988)

Hertz, H., 'Uber die beruhrung fester elastischer korper (on the contact of elastic solids). *J.reine und angewandte mathematik* ,V92, pp156-171 (1882) (For English translation see Miscellaneous Papers by H. Hertz. Eds. Jones and Schott, London, Macmillan press. 1896)

Heyliger, P.R. and Reddy, J.N., 'A mixed computational algorithm for plane elastic contact problems - I: Formulation.', *Comput. Struct.* , V26, pp621-634 (1987)

Heyliger, P.R. and Reddy, J.N., 'A mixed computational algorithm for plane elastic contact problems - II. Numerical examples', *Comput. Struct.* , V26, pp635-653 (1987)

Hill, R., Lee, E.H. and Tupper, S.T., 'Theory of wedge indentation of ductile materials', *Proc. Royal. Soc.* , A188, pp273-283 (1947)

Hitchings, D., 'A finite element algorithm for solving two and three dimensional contact problems', *Imperial College, report for Rolls Royce.* (1988)

Horrigmoe, G. and Bergan, P.G., 'Incremental variational principles and finite element models for non-linear problems', *Comp. Meth. Appl. Mech. Engng.* , V7, pp201-217 (1976)

Hughes, T.J.R., Taylor, R.L., Sackman, J.L., Curnier, A. and Kanoknukulchai, W., 'A finite element method for a class of contact-impact problems', *Comput. Meth. Appl. Mech. Engng.* , V8, pp249-276 (1976)

Hung, N.D. and deSaxce, G., 'Frictionless contact of elastic bodies by finite element method and mathematical programming technique', *Comput. Struct.* , V11, pp55-67 (1980)

Irons, B.M., 'A frontal solution program for finite element analysis', *Int. J. Numer. Meth. Engng.* , V2, pp5-32 (1970)

Ishinabe, M., 'An algorithm of contact problems for forming process of thin plates - a numerical analysis for the seaming process of a can', *Comput. Struct.* , V27, pp23-26 (1987)

Johnson, K.L., 'Surface interaction between elastically loaded bodies under tangential forces', *Proc. Royal. Soc.* , A230, pp531-546 (1955)

Kalker, J.J., 'The computation of three dimensional rolling contact with dry friction', *Int. J. Numer. Meth. Engng.* , V14, pp1293-1307 (1979)

Keuser, M., Mehlhorn, G. and Cornelius, V., 'Bond between prestressed steel and concrete - computer analysis using ADINA', *Comput. Struct.* , V17, pp669-676 (1983)

Kikuchi, N., 'A smoothing technique for reduced integration penalty methods in contact problems', *Int. J. Numer. Meth. Engng.* , V18, pp343-350 (1982)

Kwak, B.M. and Lee, S.S., 'A complementary problem formulation for two dimensional frictional contact problems', *Comput. Struct.* , V28, pp469-480 (1988)

Lazaridis, P.P. and Panagiotopoulos, P.D., 'Boundary variational principles for inequality structural analysis problems and numerical applications', *Comput. Struct.* , V25, pp35-49 (1987)

Madsen, N., 'Numerically efficient procedures for dynamic contact problems', *Int. J. Numer. Meth. Engng.* , V20, pp1-14 (1984)

Mahmoud, F.F., Salaman, N.J. and Marks, W.R., 'A direct automated procedure for frictionless contact problems', *Int. J. Numer. Meth. Engng.* , V18, pp245-257 (1982)

May, H.O., 'The conjugate gradient method for unilateral problems', *Comput. Struct.* , V12, pp595-598 (1986)

Mazurkiewicz, M. and Ostachowicz, W., 'Theory of finite element method for elastic contact problems of solid bodies', *Comput. Struct.* , V17, pp51-59 (1983)

Mehlhorn, G., Kollegger, J., Keuser, M. and Kolmar, W., 'Nonlinear contact problems - a finite element approach implemented in ADINA', *Comput. Struct.* , V21, pp69-80 (1985)

Mindlin, R.D., 'Compliance of elastic bodies in contact', *J. Appl. Mech.* , V71, pp259-268 (1949)

Mottershead, J.E., *Technical report for Thackray Ltd* , (1988)

Mottershead, J.E., *Technical report for Ford U.K.* , (1989)

Nayak, L. and Johnson, K.L., 'Pressure between elastic bodies having a slender area of contact and arbitrary profiles', *Int. J. Mech. Sci.* , V21, pp237-247 (1979)

Nour-Omid, B. and Wriggers, P., 'A two-level iteration method for solution of contact problems', *Comput. Meth. Appl. Mech. Engng.* , V54, pp131-144 (1986)

Oden, J.T., 'Exterior penalty methods for contact problems in elasticity', *Non-linear Finite Element Analysis* , Ed. Wunderlich, Stein and Bathe. Pub. Springer-Verlay, Berlin (1981)

Oden, J.T. and Pires, E.B.,'Nonlocal and nonlinear friction laws and variational principles for contact problems in elasticity', *J. Appl. Mech.* , V50, pp67-75 (1983)

Oden, J.T. and Pires, E.B., 'Numerical analysis of certain contact problems in elasticity with non-classical friction laws', *Comput. Struct.* , V16, pp481-485 (1983)

Okamoto, N. and Nakazawa, M., 'Finite element incremental contact analysis with various frictional conditions', *Int. J. Numer. Meth. Engng.* , V14, pp337-357 (1979)

Ostachowicz, W., 'Mixed finite element method for contact problems', *Comput. Struct.* , V18, pp937-945 (1984)

Padovan, J., Moscarello, R., Stafford, J. and Tabaddor, F., 'Pantographing self adaptive gap elements', *Comput. Struct.* , V20, pp745-758 (1985)

Padovan, J. and Zeid, I., 'On the development of travelling load finite elements', *Comput. Struct.* , V12, pp77-83 (1980)

Padovan, J., Tovichakchaikul, S. and Zeid, I., 'Finite element analysis of steadily moving contact fields', *Comput. Struct.* , V18, pp191-200 (1984)

Pascoe, S.K. and Mottershead, J.E., 'Linear elastic contact problems using curved elements and including dynamic friction', *Int. J. Numer. Meth. Engng.* , V26, pp1631-1643 (1988)

Pascoe, S.K. and Mottershead, J.E., 'Two new finite element contact algorithms', *Comput. Struct.* , V32, pp137-144 (1989)

Pascoe, S.K., Mottershead, J.E. and Hellen, T.K., 'A comparison of finite element techniques for contact stress analysis', *Modern Practice in Stress and Vibration Analysis* , (Ed. Mottershead) Pergamon Press, pp 299-312 (1989)

Pian, T.H.H., Chen, D.P. and Kang, D., 'A new formulation of hybrid/mixed finite element', *Comput. Struct.* , V16, pp81-87 (1983)

Pian, T.H.H. and Kubomura, K., 'Formulation of contact problems by assumed stress hybrid elements', *Non-linear Finite Element Analysis'* , Ed. Wunderlich, Stein and Bathe., Springer-Verlag, Berlin (1981)

Post, D., 'Developments in moire interferomotry', *Opt. Engng.* , V21, pp458-467 (1982)

Rahman, M.U., Rowlands, R.E., Cook, R.D. and Wilkinson, T.L., 'An iterative procedure for finite element stress analysis of frictional contact problems', *Comput. Struct.* , V18, pp947-954 (1984)

Sackfield, A. and Hills, D.A., 'A note on the Hertz contact problem:Correlation of standard formulae', *J. Strain. Anal.*, V18, pp195-217 (1983)

Schafer, H., 'A contribution to the solution of contact problems with the aid of bond elements', *Comput. Meth. Appl. Mech. Engng.*, V6, pp335-354 (1975)

Simo, J.C., Wriggers, P., Schweizerhof, K.H. and Taylor, R.L., 'Finite deformation post-buckling analysis involving inelasticity and contact constraints', *Int. J. Numer. Meth. Engng.*, V23, pp779-800 (1986)

Spence, D.A., 'The Hertz contact problem with finite friction', *J. Elast.*, V5, pp297-307 (1975)

Stadter, J.T. and Weiss, R.O., 'Analysis of contact through finite element gaps', *Comput. Struct.*, V10, pp867-873 (1979)

Telles, J.C.F. and Brebbia, C.A., 'Boundary element solution for half-plane problems', *Int. J. Solid. Struct.*, V17, pp1149-1158 (1981)

Torstenfelt, B., 'Contact problems with friction in general purpose finite element computer programs', *Comput. Struct.*, V16, pp487-493 (1983)

Tsai, P. and Chen, W.H., 'Finite element analysis of elastoplastic contact problems with friction', *AIAA Jnl*, V24, pp344-346 (1986)

Tseng, J. and Olson, M.D., 'The mixed finite element method applied to two-dimensional elastic contact problems', *Int. J. Numer. Meth. Engng.* , V17, pp991-1014 (1981)

Turner, M.J., Clough, R.W., Martin, H.C. and Topp, L.J., 'Stiffness and deflection analysis of computer structures', *J. Aero. Sci.* , V23, pp805-823 (1956)

Vijayakar, S.M., Busby, H.R. and Houser, D.R., 'Linearization of multibody frictional contact problems', *Comput. Struct.* , V29, pp569-576 (1988)

White, D.J. and Enderby L.R., 'Finite element stress analysis of a non-linear problem: A connecting rod eye loaded by means of a pin', *J. Strain. Anal.* , V5, pp41-48 (1970)

Wilson, E.A. and Parsons, B., 'Finite element analysis of elastic contact problems using differential displacements', *Int. J. Numer. Meth. Engng.* , V2, pp387-395 (1970)

Wong, C.J., 'Applications of non-linear finite element method to contact problems and paper handling problems', *Comput. Struct.* , V19, pp315-320 (1984)

Zeid, I. and Padovan, J., 'Finite element modelling of rolling contact', *Comput. Struct.* , V14, pp163-170 (1981)

Zolti, E., 'A finite element procedure to time dependent contact analysis', *Comput. Struct.* , V17, pp555-561 (1983)

BIBLIOGRAPHY

Bathe, K.J., 'Finite element procedures in engineering analysis', Prentice-Hall, New Jersey (1982)

Bowden, F.P. and Tabor, D., 'The friction and lubrication of solids - Part I', Clarendon Press, Oxford (1958)

Bowden, F.P. and Tabor, D., 'The friction and lubrication of solids - Part II', Clarendon Press, Oxford (1964)

Irons, B. and Ahmad, S., 'Techniques of finite elements', John Wiley and Sons Ltd, New York (1986)

Jeffrey, A., 'Mathematics for engineers and scientists', Thomas Nelson and Sons, London (1979)

Johnson, K.L., 'Contact mechanics', Cambridge University Press (1985)

Rabinowicz, E., 'Friction and wear of materials', J. Wiley, New York (1965)

Stasa, F.L., 'Applied finite element analysis for engineers', Holt, Rinehart & Winston (1985)

Timoshenko, S. and Goodier, J.N., 'Theory of Elasticity' 2nd Edition, McGraw-Hill (1951)

Zienkiewicz, O.C., 'The finite element method', 3rd Edition, McGraw-Hill, London (1977)

APPENDIX A

Initial Contact Checks

The standard finite element discretisation procedure results in meshes being formed for the structure(s) to be analysed. However, whether or not contact constraints should be initially imposed depends on the type of contact problem. For example, if both of the potential contacting bodies have rigidly fixed nodes (such that non-singular stiffness matrices are naturally formed for each body after application of restraints) then no contact conditions need be imposed unless geometric contact or overlap is initially identified. An example of this form of contact problem can be seen in the case of a cantilevered beam and deformable base as shown in Figure A.1.

Other forms of contact problem are where one of the bodies is insufficiently restrained, such that its stiffness matrix is singular. This is quite a common situation in contact problems and occurs frequently when restraints are only applied in one direction (generally seen in sliding problems). If the meshes are initially generated such that they are just touching in their assembled form, then a geometric contact checking routine can identify the appropriate contact constraint to 'couple' the meshes.

In some circumstances however, the complexity of the contact surface profiles make it extremely difficult to identify just where initial contact will occur. In this case it is difficult to generate the initial meshes in a 'just touching' form. This problem can be seen in the rack and pinion type contact shown in Figure A.2, where the exact position of contact cannot easily be identified.

The difficulty is overcome here for this sort of problem by constructing vectors in the direction of likely movement for the 'under-restrained' body (usually the contactor) from its surface nodes. The length of each vector to intersect with the target surface is calculated. The minimum length is identified and used as a translation distance which is added onto the global coordinates of the 'under-restrained' body. Entrance into a geometric contact checking routine then defines the appropriate contact constraints to couple the meshes.

APPENDIX B

Accurate Surface Modelling

The precision of the finite element mesh representing the structural shape, as defined by the nodal coordinates in the input data file, is extremely important for accurate contact solutions. This is because generally only small displacements are occurring in the solution, hence the reliable detection of contact is dependent upon the fact that the initial surface profiles are accurately defined.

At first glance, it would appear that this can be ensured by increasing the precision of the input geometry. For example, most finite element packages have the facility of allowing coordinates of higher precision than normal to be defined (e.g. in BERSAFE 8 d.p. instead of 3 d.p.). This feature is in fact necessary in contact analysis, but does not guarantee the accuracy of the actual coordinates. This is defined by the process of mesh generation.

Many of the mesh generators, such as PATRAN, generate a geometric model from which the finite element model is derived. Hence, it is the accuracy of this geometric model that must be maintained for the definition of accurate contact surfaces. Straight or flat surfaces generally do not present any problem, as the lines or surfaces can be precisely defined. However, curved surfaces can cause difficulties because they are generally defined by an isoparametric function, which for PATRAN is cubic in form. For example, with circular arcs, deviation of the cubic function from the true circular profile increases with increase of subtended arc. General recommendations

to ensure accurate circular surface definition are to restrict individual isoparametrics to subtend an angle of less than 30°. Thus, if a potential contact zone was defined by a curved surface of 90° as shown in Figure B.1, then at least three (rather than one) isoparametrics should be used in the generation of this surface. For example, if PATRAN was being used, then this would involve at least three separate 'patches' being defined around this surface.

Additionally, the type of element used affects the accuracy of the contact surfaces. Obviously linear elements only allow straight line segments to define the contact surface. Thus to use these to model curved contact is unwise, as even with the use of many elements, the true surface profile cannot be accurately produced. Quadratic or higher order elements are recommended for curved contact although again problems of inaccurate surface modelling can occur if the meshes are too coarse.

APPENDIX C

Hertz Contact Equations

The following equations apply to linear elastic frictionless contact between two parallel cylinders subjected to a load of F per unit length. An 'equivalent' Young's modulus E^* and equivalent radius R^*, for the contacting cylinders are calculated as shown in Equations C.1 and C.2. These values in conjunction with Equations C.3 and C.4 allow the maximum contact pressure p_{max} and semi-contact width 'a' to be calculated.

$$E^* = \left\{ \frac{(1-\nu_1^2)}{E_1} + \frac{(1-\nu_2^2)}{E_2} \right\}^{-1} \qquad \text{C.1}$$

$$R^* = \left\{ \frac{1}{R_1} + \frac{1}{R_2} \right\}^{-1} \qquad \text{C.2}$$

$$p_{max} = \sqrt{\frac{FE^*}{\pi R^*}} \qquad \text{C.3}$$

$$a = \sqrt{\frac{4FR^*}{\pi E^*}} \qquad \text{C.4}$$

APPENDIX D

Input Contact Data

Described here is the extra input data required, compared to a 'standard' finite element input data file, to allow contact to be defined using CONTACT FORTRAN.

There are essentially three extra features which need to be included, these are,

1, the static and dynamic coefficients of friction
2, a list of 'potential' contactor nodes
3, a list of 'potential' target elements and their respective 'active' surfaces

To minimise the effort of the modeller and avoid the possibly tedious task of identifying the potential contactor nodes, (which for a complex analysis could involve many hundreds of nodes) just the potential contactor elements are required and their 'active' surfaces, rather than a list of contactor nodes. For quadrilateral and brick elements the 'active' surface labelling scheme is shown Figure 6.3. This labelling scheme is linked to the element's node numbering topology. With the NAGFE finite element topology known (e.g. for a 2-D quadrilateral element start the local node numbering from the bottom left hand corner of the element and number sequentially clockwise) and the particular element and its active surface known, allows routines to automatically identify and list the potential contactor surface nodes and the

potential target surface nodes. This information is then used internally within CONTACT FORTRAN in the frontal elimination procedure to prevent their elimination, as well as for the contact geometric identification process.

The contact input data for a typical model is shown in Table D.1. This is in fact the contact input data used for the sliding wedge problem (Section 9.5.5), where eight quadratic contactor elements and eight quadratic target elements were present. The node numbering topology for each element was such that the active surface for the contactor elements was defined by -2 , and +2 for the target elements.

APPENDIX E

General Active Zone Equations

The active zone equation solver equations described in Chapter 9, can be generalised for an n x n matrix [K] as follows,

For the first active zone,

$$U_{11} = K_{11} \qquad \text{E.1}$$

$$L_{11} = 1 \qquad \text{E.2}$$

and for each subsequent active zone j, from 2 to n

$$L_{j1} = \frac{K_{j1}}{U_{11}} \qquad \text{E.3}$$

$$U_{1j} = K_{1j} \qquad \text{E.4}$$

and

$$L_{ji} = \frac{K_{ij} - \sum_{m=1}^{i-1} L_{jm}U_{mi}}{U_{ii}} \qquad \text{for } i = 2,3...,j\text{-}1 \qquad \text{E.5}$$

$$U_{ij} = K_{ij} - \sum_{m=1}^{i-1} L_{im}U_{mj} \qquad \text{for } i = 2,3,...,j\text{-}1 \qquad \text{E.6}$$

and finally,

$$L_{jj} = 1 \qquad \text{E.7}$$

$$U_{jj} = K_{jj} - \sum_{m=1}^{j-1} L_{jm}U_{mj} \qquad \text{E.8}$$

If [K] is symmetric, then the L terms can be obtained more simply as

$$L_{ji} = \frac{U_{ij}}{U_{ii}} \qquad \text{for } j \geq i \qquad \text{E.9}$$

APPENDIX F

Simple Verification

The models used here were simple two and three element models. The purpose of running these problems was to confirm the constraint theories and ensure the correct implementation of the finite element code. All the runs listed are single iterations unless stated otherwise using the Lagrange multiplier method of constraint imposition.

F.1 Two element sticking contact

The model in Figure F.1a contains eight nodes and consists of two separate elements. Sticking contact is occurring with equations of the type shown in Equation 6.6 applied. To verify the sticking contact results, a standard (non-contact) solution was undertaken with a different mesh of only six nodes and two coupled elements as shown in Figure F.1b. The displacement results obtained by both methods were identical confirming the sticking constraint theory and code. The displacements are plotted in Figure F.1c and tabulated in Table F.1. Within the contact algorithm specific features that were confirmed by this run were,

a, correct identification of position of contact

b, correct definition of direction cosines defining the normal and tangent directions

c, correct sticking or sliding constraints being formed

d, correct insertion of Lagrange terms into the stiffness matrix

e, correct equation solution

F.2a Two element 2-D frictionless sliding contact

The model shown in Figure F.2 is more difficult to solve than that of Figure F.1a in that contact is not occurring between adjacent nodes. The shallow slope in the region of contact and the horizontal contactor force cause sliding to be the prevalent state. With this model, an initially frictionless run was performed by application of a single constraint of the form shown in Equation 6.17. In this circumstance large scale sliding of the single contacting node occurs. From force equilibrium considerations, the normal force generated due to contact is dependent upon the applied force and the contact slope. The 'x' component of this normal force should be equal and opposite to the applied force, i.e.

$$F_x = F_n \mathrm{Sin}\theta \qquad \text{F.1}$$

The finite element contact results for this model are summarised in Table F.2. The contact force generated exactly satisfies the force considerations of Equation F.1. Additionally, the displacement of the sliding node is along the tangent from the initial point of contact as desired. The magnitude of displacement however was quite large, causing the final position of the sliding node to be above the deformed target surface. The overlap elimination procedure in the subsequent iteration removed this positional error.

F.2b Two element frictional sliding contact

An identical model to Figure F.2 was used here with the only difference being that friction was present at the interface. In this case, from force equilibrium conditions the total resolved normal and tangential (friction) force in the 'x' direction should equal the applied horizontal force, i.e.

$$F_x = |F_n \mathrm{Sin}\theta| + |F_t \mathrm{Cos}\theta| \qquad \text{F.2}$$

Constraints allowing the direct inclusion of friction were applied (Equation 6.27) The results are summarised in Table F.2. From these results it can be seen that the constraints exactly impose the force equilibrium of Equation F.2 and that the Coulomb friction relation between the normal and tangential forces is also exactly imposed. Additionally, the displacement of the sliding node has decreased with inclusion of the coefficient of friction. This is as desired as the frictional forces will cause less sliding displacement. It should be noted that in these runs two stages were required for the final solution. In the first stage the direction of sliding was known, allowing the direct inclusion of friction in the constraint equations. From this solution the normal positional error was corrected in the next iteration. In a general contact algorithm there would need to be two earlier stages, the first applying sticking contact from which a Coulomb friction check would in this case indicate sliding, the second stage applying frictionless contact permitting the direction of sliding to be identified.

F.3 Three element overlap contact

A contactor body consisting of two elements contacting a single target element is shown in Figure F.3. Initially only one node is contacting, however after the first solution nodal overlap of the initially 'unconstrained' end nodes has occurred. Application of the sticking overlap constraints in the next iteration causes these overlaps to be eliminated. The results are summarised in Table F.3. It should be noted that one of the main arguments for using element decision making for deciding the contact state is highlighted here. The forces generated at the end nodes exceed the limiting value of static friction and hence indicate sliding, however the physical contact of normal contacting faces and normal forces suggests sticking should be the prevalent state. Averaging the forces over the face of each contacting element generally gives a better decision on the true contact state. In this numerical example 'element force averaging' does in fact produce an element contact state of sticking.

F.4 Three dimensional contact

Two blocks touching over a face are shown in Figure F.4. Frictionless conditions were imposed, which resulted in eight normal constraints being applied. The finite element model consists of twenty nodes per element. The displacements from this run are shown in Figure F.4 and show the desired symmetry and prevention of mesh overlap.

F.5 Mis-aligned contact

A single element undergoing contact at a single node with the middle of a face of another element is shown in Figure F.5. In this case sticking contact

was imposed. The displacement of the contacting node obtained from the results, was precisely such that it remained in contact with the deforming target surface. Also, no tangential forces were generated with the normal contactor nodal force exactly equal to the external applied loads.

F.6 Three body contact

The finite element mesh shown in Figure F.6 was used to model three bodies in contact. This is a particularly interesting problem as no restraints apart from contact conditions were imposed on the middle block. Frictionless conditions were imposed with applied loads causing the deformation shown in Figure F.6.

APPENDIX G

Slope Constraint Theory

The following theory allows slope constraints to be applied on a finite element model which has no slope degrees of freedom in its initial stiffness matrix.

Generally, to allow slope constraints to be imposed, necessitates that slope degrees of freedom exist in the formulation of the finite element and its shape functions. The equations developed here, permit slope constraints (such as slope continuity between adjacent elements) to be imposed when there are no such slope degrees of freedom in the element (e.g. as is the case with the commonly used twenty noded brick element).

Consider the two adjacent elements at the free surface at node k in Figure G.1. To impose slope continuity across this node, requires that the slope approaching node k in element 1, is equal to the slope approaching the same node in element 2, i.e.

$$\frac{\partial y}{\partial x}\Big|_{1k} = \frac{\partial y}{\partial x}\Big|_{2k} \qquad \text{G.1}$$

In the region of node k, a local slope constraint equivalent to Equation G.1 can also be defined,

$$\frac{\partial y}{\partial \xi}\Big|_{1k} = \frac{\partial y}{\partial \xi}\Big|_{2k} \qquad \text{G.2}$$

The value of 'y' in the above equation can be defined for each element from its nodal values {Y}, and its shape functions. Hence, differentiating the shape functions with respect to ξ , allows Equation G.2 to be written as,

$$\frac{\partial}{\partial \xi}[N(\xi_1, \eta_1, \zeta_1)]_1\{y\}_1 = \frac{\partial}{\partial \xi}[N(\xi_2, \eta_2, \zeta_2)]_2\{y\}_2 \qquad G.3$$

This equation can be rearranged into the general form of

$$[L]\begin{Bmatrix} y_1 \\ y_2 \end{Bmatrix} = 0 \qquad G.4$$

Equation G.4 simply defines a constraint equation which needs to be imposed on the system of equations, and is identical in form to the displacement constraints already discussed in the main body of this thesis. Hence using Lagrange multipliers this constraint can be directly imposed.

The equations developed above are specifically for two-dimensional elements, with slope continuity required on a constant global 'y' surface. It is not essential that global 'y' constraints be imposed, although if global 'x' constraints are imposed ($\partial x/\partial \xi$) , then these slopes would approach infinity (in this particular problem) creating ill-conditioning in the applied constraint. The general rule therefore, is that provided the local derivative of the global variable takes a finite value (this can in fact always be guaranteed by selection of the appropriate global variable) then slope constraints can be satisfactorily applied. Additionally, the local derivatives with respect to

which differentiation is occurring, must be the 'non-active' local coordinates. In 3-D analysis therefore, slope constraints need to be applied in the two 'non-active' local coordinate directions.

Up to now slope continuity between adjacent elements has been discussed. It is also possible to apply individual slope constraints at particular locations. For example if the top surface of a structure is desired to have a prescribed displacement corresponding to a pre-determined load, then this can be applied by imposing a set of slope constraints on the model. With the simple mesh shown in Figure 10.29, application of the following seven slope constraints, imposed a uniform prescribed displacement equal to the applied force.

$$\frac{\partial z}{\partial \xi}\Big|_1 = 0$$

$$\frac{\partial z}{\partial \xi}\Big|_3 = 0$$

$$\frac{\partial z}{\partial \xi}\Big|_5 = 0$$

$$\frac{\partial z}{\partial \xi}\Big|_7 = 0$$

$$\frac{\partial z}{\partial \eta}\Big|_1 = 0$$

$$\frac{\partial z}{\partial \eta}\Big|_3 = 0$$

$$\frac{\partial z}{\partial \eta}\Big|_5 = 0$$

APPENDIX H

Divergence with Post-Inclusion of Friction

H.1 Introduction

A divergence effect is discussed here which applies to a class of contact problem where sliding over a shallow slope occurs. Furthermore, the divergence effect is directly due to the post-inclusion of friction in the solution method, and hence is only encountered with this method of frictional modelling. The cause is now described and a technique is suggested for overcoming the divergence problem. Finally a numerical example highlighting the effect is presented.

H.2 Cause

Consider the sliding wedge problem shown in Figure 3.4. When frictionless sliding is imposed, the normal contact forces generated, are from force equilibrium conditions equal to,

$$F_n = \frac{F_x}{\text{Sin } \phi} \qquad \text{H.1}$$

Using this value of F_n to obtain the first estimate of the frictional forces gives,

$$F_t = \mu F_n \qquad \text{H.2}$$

Equation H.2 can result in the magnitude of the resolved friction forces in the 'x' direction, being greater than the applied external loads in this same direction. Thus including these friction forces in the right-hand side force vector of the next iteration, is to effectively be solving a model where the nett external forces in the 'x' direction on the contactor are negative rather than positive. The consequence of this being that sliding in the next iteration occurs in the opposite direction to that as would be expected with the initial external forces. This is obviously undesirable and leads to a divergent solution, as the friction forces which are defined to occur in the opposite direction to sliding, now act ***with*** the applied external loads. This therefore creates even greater normal contact forces from which greater tangential forces are produced, which now act in the opposite direction, and hence the divergence effect is accentuated.

H.3 Results

The divergence effect is solely occurring because the initial frictionless normal contact forces are not 'close' to the final frictional values. Therefore, the estimated tangential forces are also in error. Examining the resultants of the initial normal and tangential contact forces in the 'x' direction, shows them to exceed the applied external 'x' direction load. From force equilibrium conditions though, these resolved normal and tangential 'x' direction forces should exactly equal the applied external loads. Therefore, a valid correction would be to scale the initial normal and tangential forces, such that the resolved 'x' component exactly equals the applied external 'x' loads. This has been implemented and found to solve the divergence problem.

H.4 Occurrence

When divergence is likely to occur and when scaling is necessary is of importance in the force calculation routines. A possible option is to ***always*** have force scaling. However, results undertaken with this option, have shown that this can introduce poor convergence in problems where there is not a large amount of sliding occurring. Examination of Equations H.1 and H.2 shows that the effect of estimated tangential forces being greater than the applied external forces, is a function of the difference in angle between the direction of applied loads and the contact surface (ϕ), and the coefficient of friction. In fact, provided the difference in angle between the forces and contact slope is greater than the friction angle (as stated in Equation H.3), then the tangential forces will not exceed the applied external loads and convergence will result.

$$\mu < \text{Tan}\ \phi \qquad \text{H.3}$$

H.5 Examples

Consider the sliding wedge model discussed in Chapter ten and shown in Figure 9.20. With frictionless conditions a converged solution is obtained in eight iterations. However, inclusion of a dynamic coefficient of friction of 0.2, causes divergence to occur. Introducing scaling, such that the resolved normal and tangential forces in the 'x' direction equal that of the applied load results in convergence in seven iterations.

APPENDIX I

Penalty Method for the Direct Inclusion of Friction

I.1 Introduction

In the following analysis it is shown that the direct inclusion of friction technique, developed and installed in the one boundary condition change per load step algorithms using Lagrange multipliers, can also be imposed using the Penalty method.

I.2 Theory

In penalty techniques, the additional term added to the variational statement is representative of the strain energy distribution associated with the constraint. The extra terms that become added to the stiffness matrix, are defined by the $\alpha[L][L]^T$ product. This imposes frictionless constraints. From Chapter six, it was shown how the use of a $[J]^T$ term (which essentially is a $[L]^T$ term with frictional effects included) could be used with Lagrange multipliers to impose friction directly. The following theory hasn't been strictly derived, but derived from comparison of the $[L]^T$ and $[J]^T$ terms. When using Lagrange multipliers, the $[J]^T$ term imposed the frictional force conditions, what therefore would be the effect of replacing the $[L]^T$ term by $[J]^T$ in the $\alpha[L][L]^T$ product added to the stiffness matrix?

I.3 Results

Imposition of $\alpha[L][J]^T$ terms into the stiffness matrix was found to exactly impose the frictional conditions, with identical results to those obtained by the $[L]$ and $[J]^T$ Lagrange multiplier technique. This was confirmed by analysis of the sliding block problem shown in Figure F.2, whereby identical frictional results to those using the Lagrange multiplier frictional technique were obtained in a single iteration, using the $\alpha[L][J]^T$ contributions to the stiffness matrix.

APPENDIX J

Knee Prosthesis Analysis

The stress analysis of a three-dimensional knee prosthesis using the direct inclusion of friction algorithm is summarised here.

The knee prosthesis analysed is shown in Figure 10.30 and consisted of three parts, the femoral and tibial components manufactured from a high grade stainless steel and the meniscal component made from high density polyethylene. When assembled, the three parts allow universal motion of the joint. Cylindrical surfaces on the femoral component and on the top of the meniscal component allow the knee to bend. Conforming conical surfaces on the underside of the meniscal component and on the tibial component allow rotation of the tibia about its vertical axis.

The finite element mesh for the prosthesis is shown in Figure 10.31. With this particular contact problem there are two separate contact interfaces, one between the femoral and meniscal components, and the other between the meniscal and tibial components. To allow this to be defined required two sets of contact information to be input for the interfaces.

The geometric contact checking routines were then entered for the contactor nodes of each interface with their respective target elements. Once the positions of contact had been identified then frictionless sliding constraints were imposed (sliding was known to occur because of the different material properties of the components). The force applied to the mesh was 28 stones

(this value was obtained from estimating the shock load due to sudden jarring on the prosthesis, as two times the body weight of a 14 stone person). This force was applied as a uniformly distributed load on the top flat surface of the femoral component. Restraints preventing nodal movement were applied on the underside of the tibial component.

The solution procedure involved an initial solution with frictionless sliding imposed. From this result direct inclusion of friction constraints were added in the next iteration (μ_d = 0.15). Several different runs were undertaken with this mesh and other more locally refined meshes. Some of the key results are now presented.

The σ_z direct stresses on the contact surfaces were of particular interest. These stresses although not truely representative of the normal contact forces, could be used as a guide to the normal stresses, as their values greatly exceeded the σ_x and σ_y stresses. From these stress results it was found that approximately equal in magnitude compressive stresses were occurring between adjacent contact surfaces. This is obviously as would be expected and desired. A typical σ_z stress plot for the underside of the meniscal component is shown in Figure 10.32. From this plot it can be seen that the maximum stresses occur at the outer region of the meniscal component's surface. They are probably occurring here because the femoral component, to which the external loads are applied, rests on the outer region of the top of the cylindrical surface of the meniscal component. Hence the induced contact stresses are concentrated primarily in this outer zone and transmitted through to the meniscal's underside surface.

The mesh used here is admittedly quite coarse (this was because of computational storage limitations) however, further results on modified meshes have all endorsed these locations of maximum stress. The maximum σ_z stress obtained here is in the order of 3 MN/m^2. This is within the yield stress of the meniscal component (this component suffering the most wear, due to its relatively low material properties compared to the other two components).

This work has confirmed the technique and principles of using constraint equations and Lagrange multipliers for the complex analysis of a 3-D multi-body conforming sliding contact problem.

APPENDIX K

Publications

INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING, VOL. 26, 1631–1643 (1988)

LINEAR ELASTIC CONTACT PROBLEMS USING CURVED ELEMENTS AND INCLUDING DYNAMIC FRICTION

S. K. PASCOE AND J. E. MOTTERSHEAD

Department of Mechanical Engineering, The University of Liverpool, Liverpool, U.K.

SUMMARY

A finite element solution for two-dimensional contact between elastic bodies is presented. Equations of constraint and equilibrium governing 'sticking' and 'sliding' contact are imposed at mesh contact boundaries and incorporated in the usual displacement solution routine. The method allows sliding over curved bodies and the dynamic coefficient of friction between the bodies is included in a direct (non-iterative) solution at each loading stage.

A progressive loading approach is implemented whereby full loads are initially applied. If a change in the contact boundary condition occurs, then scaling (reduction) of loads and displacements is undertaken until the first new boundary constraint becomes just operative again. The remainder of the load is then applied with a modified stiffness matrix and the process is repeated. Finally the mesh is fully loaded and the contact boundary conditions are satisfied.

Sample problems include comparison of the finite element results with classical solutions from the literature.

INTRODUCTION

Contact stress problems occur in the design of cams, valves, pistons, rolling element bearings and a wide variety of other engineering components and structures. Not only is the range of engineering products which require an analysis of contact stresses very wide but also there is great variety in the treatment which is most appropriate to these products. Johnson[1] in his book *Contact Mechanics* deals with elastic and non-elastic contact, dry friction and lubricated surfaces, rolling and sliding contact, thermal effects and surface roughness. In the present paper the authors restrict their attention to linear elastic contact problems with Coulomb friction. This 'limited' range of problems still represents a huge area of vital interest to the design of many engineering components. Industrial interest has been such that in-house developments for contact stress analysis have been widely used. For example, the work of Hartnett,[2] who provided a numerical solution to the Boussinesq equations of a half space, has been widely applied in the rolling element bearing industry.

Contact problems have been particularly difficult to treat using the finite element method because of the problems of determining the region of contact between two (or more) meshes representing separate components and ensuring overall equilibrium of forces including applied external forces, contact forces and reactions at nodal constraints. One of the most widely applied numerical algorithms to be installed in a finite element program is that of Francavilla and Zienkiewicz.[3] The main disadvantage with the method was that it required a flexibility (rather than stiffness) approach which was applied locally in the contact zone.

0029–5981/88/071631–13$06.50

Received 3 August 1987
Revised 11 November 1987

A significant advance in the finite element modelling of contact stress problems was made by Hughes *et al.*[4] who developed a method which could be installed in the usual finite element displacement solution routines, but node on node contact was necessary in the contact region. Bathe and Chaudhary[5] removed this restraint for two-dimensional finite element meshes with sticking and sliding contact using four-noded quadrilateral elements. In a further paper[6] the work was extended to allow the solution of three-dimensional problems and dynamic contact using the Newmark direct integration scheme.[7] Displacement constraints, designed to prevent overlapping of meshes in contact regions, and contact forces, were determined using linear interpolation formulae between the nodes. The dynamic coefficient of friction was not included explicitly in the formulation and this omission made it necessary to perform an iterative solution of sliding contact problems.

A rather different approach using a non-local and non-linear model was proposed by Oden and Pires.[8] Oden and Martins[9] conducted a thorough survey of experimental results appertaining to frictional contact problems before implementing in a simplified form the results of Reference 8 in a finite element algorithm.

In the present paper the authors extend the work of Bathe and Chaudhary to include eight-node isoparametric elements using the shape functions to determine displacement constraints and forces in the contact zone. For sliding contact problems the dynamic coefficient of friction is included explicitly, thereby allowing a direct (non-iterative) solution. The formulation and the results presented apply specifically to linear elastic contact but the method can be adapted to deal with material and geometric non-linearity in the contacting bodies.

FORMULATION OF THE FINITE ELEMENT CONTACT APPROACH

We begin by recalling the familiar finite element displacement equation,

$$\mathbf{Ku} = \mathbf{f} \tag{1}$$

where $\mathbf{K}$ is the global finite element stiffness matrix, $\mathbf{u}$ is the vector of unknown element nodal displacements and $\mathbf{f}$ is a vector of externally applied loads. If two or more separate finite element meshes are considered then equation (1) cannot prevent overlapping unless some further constraint is applied. Bathe and Chaudhary have introduced the notion of contactor and target bodies. Following this approach we allow target nodes to overlap into the contactor but prevent the entry of contactor nodes into the target. Thus in the contact region a set of displacement constraint equations coupling the motions of the contactor and target nodes is defined as follows:

$$\mathbf{Lu} = \mathbf{0} \tag{2}$$

where $\mathbf{L}$ is an $m \times n$ constraint coefficient array, n is the number of degrees of freedom in the complete finite element model and m is the number of nodal contact variables. Since the contactor nodes are not allowed to penetrate the target then the element nodal deflections in the contact region are constrained according to the shape functions as

$$\mathbf{u}_C = \mathbf{N}_T(\xi, \eta)\mathbf{u}_T \tag{3}$$

Here subscripts C and T refer to the contactor and the target. Vector $\mathbf{u}_C$ contains the deflections of all contactor nodes incident upon a target element. $\mathbf{u}_T$ contains the full set of nodal displacements on a target element. Thus, in the case illustrated in Figure 1, $\mathbf{u}_C$ contains the deflections of node k and the deflections of nodes a, b and c (which are contained in $\mathbf{u}_T$) are constrained such that a parabola passing through a, b and c also passes through k. Equation (2) is formed by re-arranging equation (3).

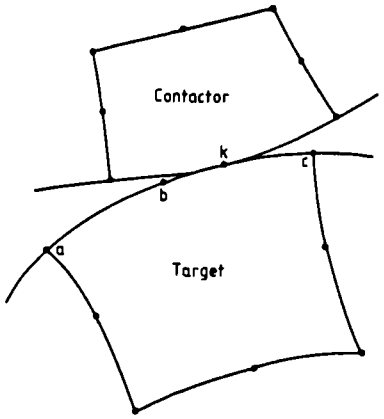


Figure 1. Displacement constraints in the contact region

Having implemented a constraint linking the deflections of the contactor and target bodies it is necessary to ensure equilibrium of forces over the complete structure. This will also provide a mechanism whereby separation of previously connected meshes can be achieved. We return to equation (1) but now include the interface forces acting between the contactor and the target. Thus,

$$\mathbf{K}\mathbf{u}+\mathbf{J}\mathbf{c}=\mathbf{f} \tag{4}$$

where $\mathbf{c}$ is a vector containing the contact forces governing sticking/sliding conditions in the contact zone. Since node-on-node contact is unlikely to occur the contact forces must be distributed along the target element nodes. The contact forces can be distributed according to the target element shape functions

$$\mathbf{c}_C = \mathbf{N}_T(\xi, \eta)\mathbf{c}_T \tag{5}$$

The elements of the force coefficient array $\mathbf{J}$ in equation (4) are then given according to the shape function distribution of forces defined by (5). Alternative schemes for the distribution of contact forces on target nodes have been implemented but the authors have observed that the shape function distribution provides reliable results. The shape function distribution also yields a symmetric system of equations for computation of displacements in the case of sticking friction.

Combining equations (2) and (4) we obtain a set of algebraic equations which can be solved using the conventional finite element displacement routines,

$$\left[\begin{array}{c|c} \mathbf{K} & \mathbf{J}' \\ \hline \mathbf{L} & \mathbf{0} \end{array}\right] \left\{\begin{array}{c} \mathbf{u} \\ \hline \mathbf{c}' \end{array}\right\} = \left\{\begin{array}{c} \mathbf{f} \\ \hline \mathbf{0} \end{array}\right\} \tag{6}$$

$\mathbf{c}'$ and $\mathbf{J}'$ are condensed versions of $\mathbf{c}$ and $\mathbf{J}$. When $\mathbf{c}'$ is of order m it contains all the non-zero terms contained in $\mathbf{c}$. $\mathbf{J}'$ is an $n \times m$ force coefficient array. The contact forces $\mathbf{c}'$ are identical to the Lagrange multipliers which appear in References 5 and 6.

The approach taken in this work with respect to forces and displacements in the contact zone has been to separate them into normal and tangential components. This involves a co-ordinate transformation, affecting the elements of $\mathbf{L}$ and $\mathbf{J}'$ but has significant advantages in the treatment of contact problems because tangential sliding then becomes easy to implement.

Further discussion of the transformation to normal and tangential components can be found in Appendix I.

In the sticking contact regime the following condition applies to the normal and tangential contact forces at the contactor nodes,

$$|c_t| < |\mu_s c_n| \tag{7}$$

where μ_s represents the static coefficient of friction and the subscripts n and t indicate normal and tangential directions. When the threshold for the onset of sliding has been overcome by large tangential contact forces, then

$$c_t = \pm \mu_d c_n \tag{8}$$

depending upon the direction of sliding. μ_d is the dynamic coefficient of friction.

The relationship (8) can be implemented in equation (6), thereby eliminating one column in $\mathbf{J}'$ for each contactor node in sliding contact. Consequently for each column eliminated in $\mathbf{J}'$ a corresponding row in $\mathbf{L}$ must be deleted if the 'stiffness' matrix is to remain square. The rows to be deleted in $\mathbf{L}$ are those defining the tangential displacement constraints which must be released if sliding is to occur between the contactor and target bodies. The explicit inclusion of (8) in (6) enables a direct (non-iterative) solution of sliding, frictional contact problems but it has the disadvantages that the 'stiffness' matrix becomes non-symmetric. A scheme for improving the bandedness of the 'stiffness' array $\begin{bmatrix} \mathbf{K} & \mathbf{J}' \\ \mathbf{L} & \mathbf{0} \end{bmatrix}$ is presented in Appendix II.

SOLUTION ALGORITHM

The solution algorithm uses a progressive application of the external loads. Initially the full external load is applied and changes in the state of interference of the contactor and target meshes are noted. The load is then scaled back to the point where the first contactor node just touches the surface of a target element. Contact constraints are then applied at this contactor node and the remaining load is reapplied. The loading is again scaled back to allow new contact constraints to be imposed or released. Then the remaining load is reapplied, and so on until the contactor and target meshes are fully loaded and the contact region is fully defined. For each node on the contactor surface there are four changes of boundary condition which may occur during this process:

(a) nodal state change, e.g. sticking to sliding;
(b) penetration of a new node into the target body;
(c) separation of a node; and
(d) sliding of a contactor node from one target element to another.

If the state of contactor node is found to change from sticking to sliding then a complete re-solution of the previous loading stage is undertaken with new contact constraints. The state of a contactor node cannot change within a load stage because the friction angle remains constant in each stage.

If a contactor node penetrates the target mesh then a scaling factor, γ, is applied to the computed displacements and forces. This situation is illustrated in Figure 2. γ is given as follows:

$$\gamma = \frac{u_\gamma}{u_k} \tag{9}$$

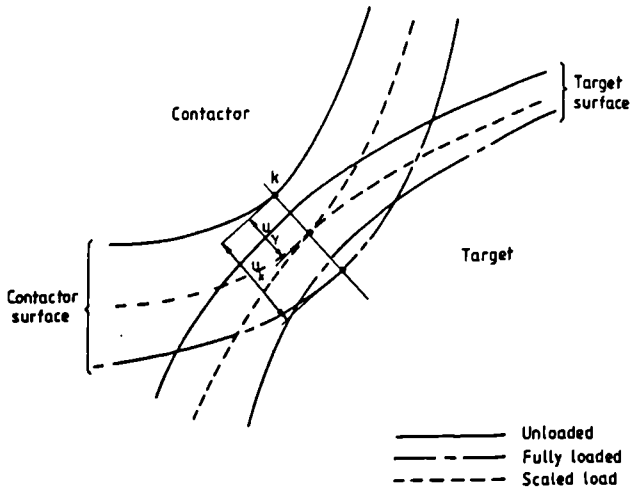


Figure 2. Scaled penetration model

Separation of a contactor node is identified when the normal force at that node becomes negative (i.e. tensile). γ is then calculated such that separation of the node just occurs. In this case γ is given by the ratio of the sum of the compressive (positive) normal contact forces computed at previous load stages to the absolute value of the tensile (negative) normal force computed in the current load stage, i.e.

$$\gamma = \frac{\sum \text{compressive normal forces from previous stages}}{|\text{tensile normal force in current stage}|} \tag{10}$$

Values of γ which are greater than unity indicate that separation is not occurring.

When a contactor node slides from one target element to another then it is necessary to allow for a change of equation of the target surfaces (defined by the position of the nodes on the second element and its shape functions). This situation is illustrated in Figure 3 and it can be seen that the scaling factor is given by

$$\gamma = \frac{1-\eta_k}{|\eta_j|-|\eta_k|} \tag{11}$$

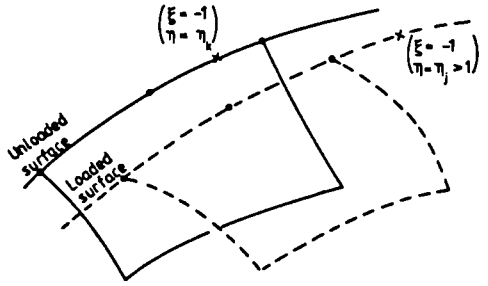


Figure 3. Sliding from one target element to another

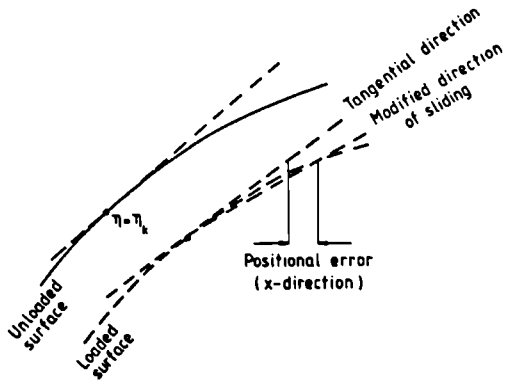


Figure 4. Computation of modified sliding direction

where η_k represents the initial point of contact on a target surface defined by $\xi = \pm 1$. η_j represents the contact point at completion of the load stage.

In the case of sliding over curved targets contactor nodes are constrained to move in directions tangential to the target surface. In cases where the curvature of the target element is shallow this linear approximation may be satisfactory if relative displacements are small. The final positional error of the contactor node may then be very small in comparison with the relative displacements. For cases where the curvature is significant or where displacements are large an improvement is necessary. This is achieved in the solution algorithm by using the positional error of the contactor node to calculate a new direction of sliding, as shown in Figure 4. Thus the displacement constraint equations are modified by adjusting the definition of the normal and tangent and a re-solution is performed. This method gives an improved estimate of sliding over curved surfaces since the modified direction represents an average tangent of the target surface over which the sliding node has moved.

RESULTS

In this section two problems illustrating the cases of sticking and sliding contact are solved using the finite element method and the results are compared with results obtained from classical methods available in the literature.

Sticking contact

To demonstrate the performance of the method in the treatment of sticking contact problems we consider the case of a pin in a hole with small radial clearance. This problem is illustrated in Figure 5 where a concentrated load is applied at the centre-line of the pin.

It is known[1] that when the arc of contact, 2α, occupies an appreciable fraction of the circumference of the hole then the classical Hertz treatment is invalid. Persson[10] provided a more representative model of the problem by using stress functions appropriate to a circular disc and to a circular hole in an infinite plate assuming frictionless contact.

A finite element model using eight-node quadrilateral elements is shown in Figure 6. The analysis was performed at a section remote from the ends of the pin.

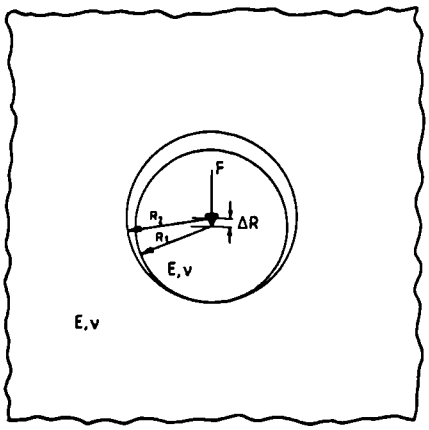


Figure 5. Pin in a hole with small radial clearance

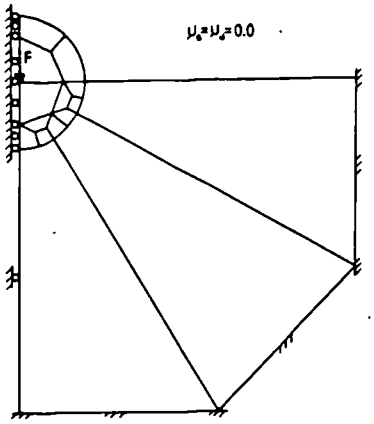


Figure 6. Finite element model of a pin in a hole

Although the problem appeared to be a wholly sticking one, with the static and dynamic coefficients of friction set as zero, sticking nodes at which tangential contact forces were generated were reset within the algorithm to sliding.

Plots of contact pressure versus contact angle are shown in Figure 7. Three load cases were considered. The full lines indicate Persson's results for $\alpha = 30°$, $\alpha = 60°$ and $\alpha = 90°$. The discrete results were obtained using finite elements with $\alpha = 30°$, $\alpha = 60°$ and $\alpha = 82°$. They represent averaged pressure per element. Comparison of the 30° and 60° results shows close similarity between the two methods and, although comparative results for 82° and 90° were not available, it can be seen that the 82° results lie in the correct region. The pin variables (Young's modulus, E, clearance, ΔR, and load, F) are plotted against contact angle in Figure 8. It can be seen that the finite element results are in very close agreement with Persson's solution despite the coarseness of the finite element mesh.

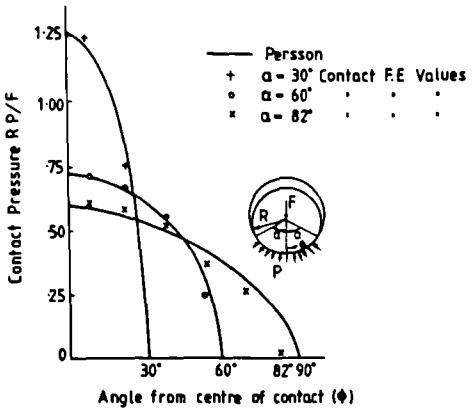


Figure 7. Comparison of contact pressures

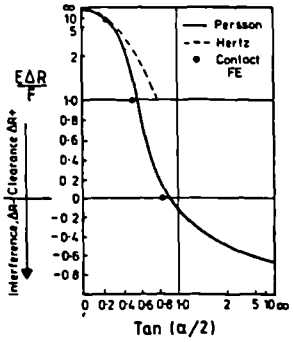


Figure 8. Comparison of predicted contact angles for different pin conditions

Sliding contact

We consider the problem of a long cylinder subject to normal and tangential forces sliding over a flat frictional surface. The problem, which is illustrated in Figure 9, would appear to be indeterminate since the cylinder is allowed to slide continuously. To overcome this problem the cylinder and surface may be inclined as shown in Figure 10. A horizontal force is then applied to the cylinder and vertical motion of the cylinder is prevented. This strategy naturally generates the normal and tangential forces P and Q and sliding of the cylinder occurs until jamming prevents further relative motion of the contactor with respect to the target. The finite element model of the system, which uses eight-node quadrilateral elements, is shown in Figure 11.

The amount of sliding between the two meshes after application of a horizontal force of 1083 N is shown in Figure 12. From Hertzian analysis of contact between a cylinder and a flat surface a semi-elliptical normal pressure distribution would be expected between the two bodies. Figure 13 shows the finite element results representing averaged pressure per contacting element under nine contactor nodes. The finite element pressure results are compared with values obtained using Hertzian theory in the figure.

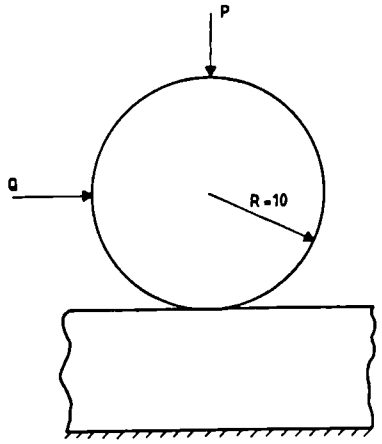


Figure 9. Sliding cylinder over flat surface

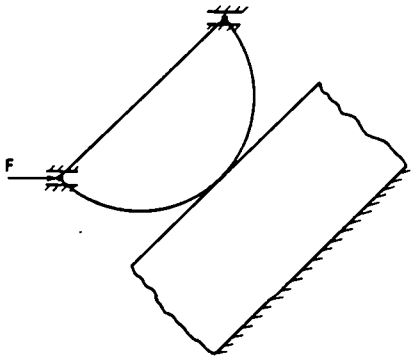


Figure 10. Inclined sliding model

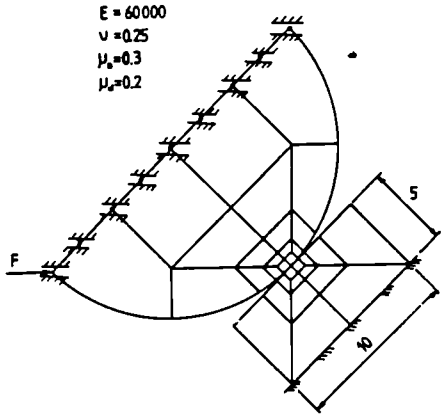


Figure 11. Finite element model of sliding cylinder

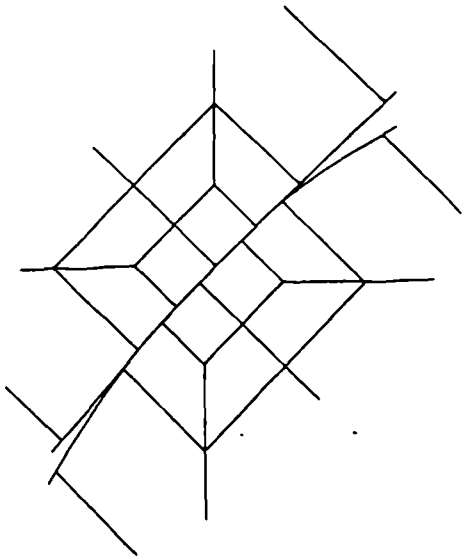


Figure 12. Magnified view of meshes in contact region after application of 1083 N.

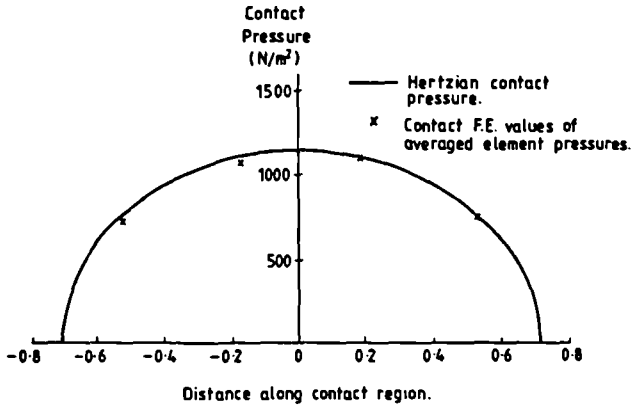


Figure 13. Comparison of contact pressures

An analytical solution exists for this sliding cylinder contact problem[1] whereby the maximum shear stresses in the target are found to be subsurface when $\mu_d < 0{\cdot}25$. Contact finite element results using the mesh shown in Figure 11 confirmed that the maximum shear stresses were subsurface but the magnitudes of the computed stresses were in error by a factor of two (over-estimated). This latter error was found to disappear with mesh refinement. A comparison of maximum shear stresses obtained analytically with the finite element stresses is given in Figure 14. In order to avoid a complete resolution using a finer mesh the finite element results were obtained using a refined target mesh alone. Prescribed nodal displacements were interpolated from the displacement results given by the coarse mesh in Figure 11 and applied to the refined target mesh. The refined target mesh used was a grid of 15×8 equal size quadrilateral eight-node elements. It can be seen that, using this approach, the finite element results were generally in close agreement

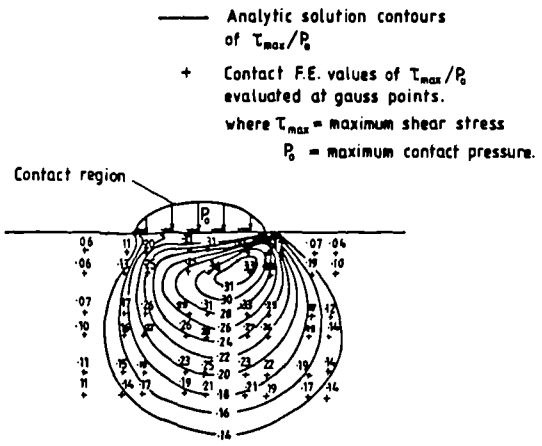


Figure 14. Comparison of maximum shear stress

with the analytical results though the highest value of the maximum shear stress was in error by about 20 per cent.

CONCLUSIONS

A finite element solution method has been presented for contact stress problems involving curved surfaces and frictional effects. Using displacement and force constraints derived from the shape functions and Coulomb's friction law a stiffness approach is employed whereby normal and tangential forces are computed at the nodes on the contact surfaces. In the case of nodal sliding the dynamic coefficient of friction is directly included in the solution matrix. Loading is performed in stages, each stage introducing a change in the boundary conditions. The final mesh displacements are given by summation of the contributions from each of the stages.

The results presented in the paper compare favourably with classical solutions although there seems to be some sensitivity to mesh refinement.

ACKNOWLEDGEMENTS

S. K. Pascoe is supported by an S.E.R.C. (C.A.S.E.) award in collaboration with the Central Electricity Generating Board.

APPENDIX I

Nodal contact displacements expressed in normal and tangential directions

Element displacements in the global x and y directions can be obtained from the nodal displacements and the shape functions. We consider the surface of an element where $\xi = \pm 1$. Then for an eight-node quadrilateral element,

$$u_x = u_{xa}(\eta(\eta-1)/2) + u_{xb}(1-\eta^2) + u_{xc}(\eta(\eta+1)/2) \tag{12}$$

$$u_y = u_{ya}(\eta(\eta-1)/2) + u_{yb}(1-\eta^2) + u_{yc}(\eta(\eta+1)/2) \tag{13}$$

The normal and tangential displacements of a contactor node, k, in terms of its x and y displacements are given by

$$u_{nk} = n_{nx} u_{xk} + n_{ny} u_{yk} \tag{14}$$

$$u_{tk} = n_{ny} u_{xk} + n_{nx} u_{yk} \tag{15}$$

where n_{nx} is the x-component of the unit vector normal to the surface, n_{ny} is the y-component of the same unit vector.

Thus, to satisfy the condition of sticking of a contactor node, the normal and tangential displacements of the contactor node must equal that of the point of contact on the target surface. Consequently the normal and tangential constraint equations become

$$\begin{aligned} n_{nx} u_{xk} + n_{ny} u_{yk} = {} & u_{xa} n_{nx}(\eta(\eta-1)/2) + u_{ya} n_{ny}(\eta(\eta-1)/2) \\ & + u_{xb} n_{nx}(1-\eta^2) + u_{yb} n_{ny}(1-\eta^2) \\ & + u_{xc} n_{nx}(\eta(\eta+1)/2) + u_{yc} n_{ny}(\eta(\eta+1)/2) \end{aligned} \tag{16}$$

$$\begin{aligned} n_{ny} u_{xk} - n_{nx} u_{yk} = {} & u_{xa} n_{ny}(\eta(\eta-1)/2) - u_{yc} n_{nx}(\eta(\eta-1)/2) \\ & + u_{xb} n_{ny}(1-\eta^2) - u_{yb} n_{nx}(1-\eta^2) \\ & + u_{xc} n_{ny}(\eta(\eta+1)/2) - u_{yc} n_{nx}(\eta(\eta+1)/2) \end{aligned} \tag{17}$$

In the case of sticking contact both constraint equations (16) and (17) are applied but in the case of sliding the tangential constraint (17) is released.

APPENDIX II

Bandedness of the 'stiffness' array

The contact constraints in this solution method are applied as extra rows and columns around the usual global finite element stiffness matrix resulting, in general, in a sparsely populated, unbanded matrix. However, by node numbering of the target and contactor bodies in a sequential manner over the two contacting surfaces as shown in Figure 15 and by matrix manipulation as

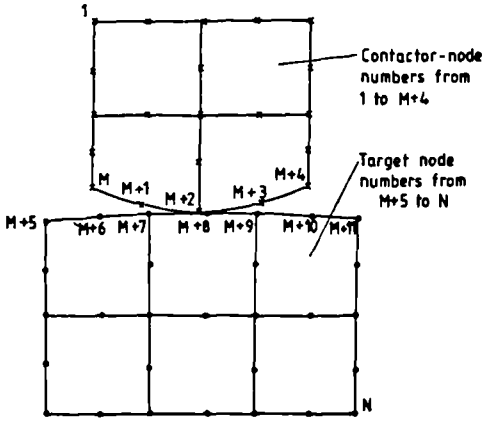


Figure 15. Contact surface node numbering sequence

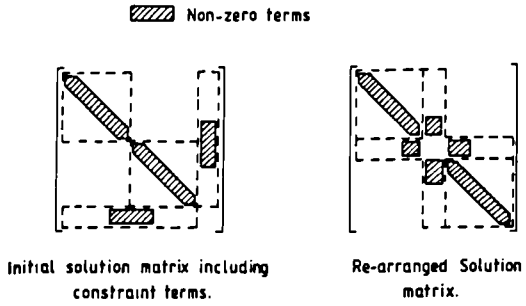


Figure 16. Showing re-arrangement of solution matrix for improved bandedness

shown in Figure 16, it is possible to produce a banded matrix of half bandwidth.

$$H_{b(contact)} = H_{b(stiffness)} + N_c$$

where

$H_{b(contact)}$ = contact half-bandwidth,
$H_{b(stiffness)}$ = stiffness matrix half-bandwidth,
$N_c = 2$ (number of sticking nodes) + (number of sliding nodes).

REFERENCES

1. K. L. Johnson, *Contact Mechanics*, Cambridge University Press, 1985.
2. M. J. Hartnett, 'A general numerical solution for elastic body contact problems', *AMD Symposium Series, Vol. 39, Solid Contact and Lubrication*, ASME Applied Mechanics Division, 1980, pp. 51–56.
3. A Francavilla and O. C. Zienkiewicz, 'A note on numerical computation of elastic contact problems', *Int. j. numer. methods eng.*, **9**, 913–924 (1975).
4. T. J. R. Hughes, R. L. Taylor, J. L. Sackman, A. Curnier and W. Kanoknukulchai, 'A finite element method for a class of contact-impact problems', *Comp. Methods Appl. Mech. Eng.*, **8**, 249–276 (1976).
5. K.-J. Bathe and A. Chaudhary, 'A solution method for planar and axisymmetric contact problems', *Int. j. numer. methods eng.*, **21**, 65–88 (1985).
6. A Chaudhary and K.-J. Bathe, 'A solution method for static and dynamic analysis of three-dimensional contact problems with friction', *Comp. Struct.*, **24**, 855–873 (1986).
7. N. M. Newmark, 'A method of computation for structural dynamics', *J. Eng. Mech. Div. ASCE*, **85**, 67–94 (1959).
8. J. T. Oden and E. B. Pires, 'Non-local and non-linear friction laws and variational principles for contact problems in elasticity', *J. Appl. Mech. ASME*, **50**, 67–76 (1983).
9. J. T. Oden and J. A. C. Martins, 'Models and computational methods for dynamic friction phenomena', *Comp. Methods Appl. Mech. Eng.*, **52**, 527–634 (1985).
10. A. Persson, 'On the stress distribution of cylindrical elastic bodies in contact', *Dissertation*, Chalmers Tekniska Hogskola, Göteburg, 1964.

Computers & Structures Vol. 32, No. 1, pp. 137–144, 1989
Printed in Great Britain.

0045-7949/89 $3.00 + 0.00

TWO NEW FINITE ELEMENT CONTACT ALGORITHMS

S. K. PASCOE and J. E. MOTTERSHEAD
Department of Mechanical Engineering, University of Liverpool, Liverpool, U.K.

(*Received* 1 *April* 1988)

Abstract—This paper details two symmetric finite element methods for the solution of two-dimensional elastic contact problems. The methods are based on variants of an unsymmetric finite element method introduced by the authors in an earlier paper. Overlapping of the meshes in the contact region is prevented by the inclusion of displacement and force constraints, which are based on the finite element shape functions and Coulomb's friction law. As a result of the application of these constraints, the stiffness matrix, displacement vector and force vector become augmented with additional terms.

Loading of the structure is in stages, with a new boundary condition included at each stage. The effects of sliding friction are included by iteration, with friction forces added to the augmented force vector in one method and normal gap terms added to the 'force' vector in the other method.

Sample problems include comparison of the two symmetric methods with the unsymmetric method and a classical solution from the literature.

1. INTRODUCTION

The solution of contact problems by the finite element method has given rise to a variety of different approaches. Papers by Böhm [1] and Chang *et al.* [2], contain good literature surveys into many of the finite element methods available. Essentially all the methods attempt to prevent overlapping of the finite element meshes and to give a satisfactory pressure distribution over the contact region.

Historically the first analytic contact solutions were developed by Hertz [3]: however these methods were restricted to non-conforming bodies and frictionless contact. The use of finite elements to solve contact problems was successfully implemented by Hughes *et al.* [4]. However, node on node contact was necessary in the contact region, preventing sliding from being modelled.

A recent area of research has involved the use of contact constraints added as extra rows and columns around the stiffness matrix, the corresponding positions in the displacement and force vectors representing contact nodal force terms and overlaps of the meshes respectively.

Bathe and Chaudhary [5], using this approach allowed node on node contact to be no longer necessary between the contacting bodies. Linear displacement constraints were applied in a Lagrange multiplier method around the standard stiffness matrix. Predefined load steps were implemented with iteration necessary to determine the displacements, forces and nodal states (sticking or sliding) in the contact region. The overall solution was obtained by summation of these results. The forces due to sliding friction were included by assuming a constant frictional traction over each contacting element and implemented as extra force terms in the force vector. The authors of this paper however noted that problems in the convergence of the iterative solution were raised by this method.

Pascoe and Mottershead [6] improved upon Bathe and Chaudhary's method, by including curved contact and sliding friction directly in the stiffness matrix. Force and displacement constraints based on the finite element shape functions were used. The solution for a contact problem was obtained by initially applying the full external forces. A checking routine was then entered to identify if any new boundary conditions had occurred, e.g. new nodal contact. If so, then the forces and displacements were scaled until the first new boundary condition just came into effect. An accurate solution with this scaled load was then obtained. The remaining force was then applied with the new boundary condition included in the stiffness matrix. The above procedure was repeated until the structure was fully loaded, with the summation of each accurate stage giving the overall solution.

A further advantage of this method was that the complete loading history of the contact zone was given. Unfortunately by including the constraints of sliding friction directly in the stiffness matrix, a non-symmetric matrix was produced. In most finite element packages investment has been concentrated in efficient routines for solving symmetric systems of equations. Unsymmetric matrices demand greater memory capacity and processing times are longer.

The current paper details two methods based on variants of the unsymmetric method described above. The important feature of symmetry of the stiffness matrix is preserved, although iteration is necessary at each load stage. The first method is a straightforward variant, where initially frictionless contact is assumed at sliding nodes. The normal forces obtained from

this solution are used to calculate the 'missing' tangential friction forces, which are then included in the following iterations as extra terms in the force vector. These tangential forces were only approximate, as the normal forces were obtained assuming frictionless contact. However, by iteration an accurate solution was obtained.

The second method is based on a less obvious variant of the unsymmetric method. In this method sliding friction forces are included directly in the stiffness matrix by applying incorrect displacement constraints. Corrective gap terms are calculated after each iteration and included in the next 'force' vector.

2. THE FINITE ELEMENT APPROACH

A typical contact problem to be solved is shown in Fig. 1, where the contacting bodies are arbitrarily denoted as contactor and target. In the contact region displacement compatibility and force equilibrium are desired. From a previous paper by the authors [6], it was shown that displacement and force constraints could be derived from the finite element shape functions. These constraints were inserted as extra rows and columns around the stiffness matrix:

$$\begin{bmatrix} \mathbf{K} & \mathbf{J} \\ \mathbf{L} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{u} \\ \mathbf{c} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix} \tag{1}$$

where the usual finite element terms are:

$\mathbf{K}$ = stiffness matrix of the unconstrained meshes,
$\mathbf{u}$ = vector of unknown nodal displacements,
$\mathbf{f}$ = vector of applied external loads;

and the contact terms are:

$\mathbf{J}$ = force constraint matrix,
$\mathbf{L}$ = displacement constraint matrix,
$\mathbf{c}$ = vector of unknown contact forces.

This system of constraints prevented any touching contactor nodes from penetrating the target surface, although target nodes could penetrate the contactor surface. However the amount of penetration was generally small.

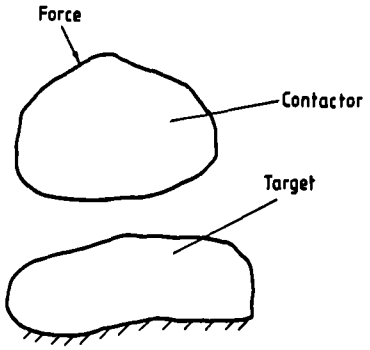


Fig. 1. Notation of contact bodies.

The state of a touching contactor node could either be sticking or sliding. The constraints included in the stiffness matrix were dependent upon these nodal states.

For each sticking contactor node, two displacement constraints representing the normal and tangential directions were incorporated as two rows in $\mathbf{L}$. For a quadratic target element, this ensured that a parabola passed through the contacting node and the three target element nodes during all stages of sticking contact, as shown in Fig. 2. The contactor nodal forces were distributed over the target element nodes in the ratio of the shape functions. This was implemented within the stiffness matrix as normal and tangential force constraints, which were symmetric to the displacement constraints, i.e.

$$[\mathbf{J}] = [\mathbf{L}]^T. \tag{2}$$

For each sliding contactor node, only one displacement and one force constraint was applied. The displacement constraint was in the normal direction, with the tangential constraint released to allow sliding. The force constraint combined the normal and tangential friction forces. Hence in all sliding problems where friction was present, the displacement and force constraints were not symmetric, i.e.

$$[\mathbf{J}] \neq [\mathbf{L}]^T. \tag{3}$$

The solution for a contact problem was initiated by defining touching contactor nodes as sticking. Appropriate constraints were included in eqn (1), and a solution was obtained. From the normal and tangential forces generated in $\mathbf{c}$, Coulomb's friction law was applied for each touching contactor node as follows,

$$|c_t| < |\mu_s c_n| \tag{4}$$

where

c_t = tangential contactor nodal force,
c_n = normal contactor nodal force,
μ_s = static coefficient of friction.

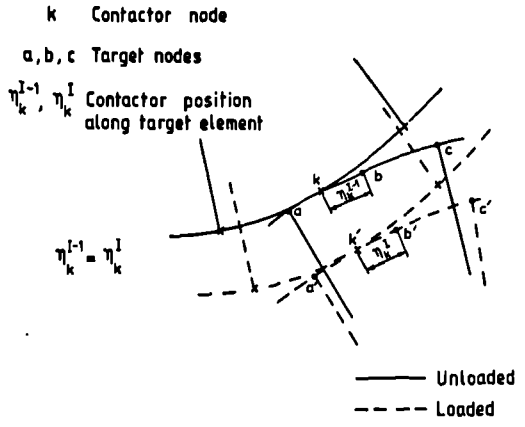


Fig. 2. Unloaded and loaded sticking contact.

If the tangential force exceeded this condition, then the nodal state was changed to sliding and the following equation was implemented by the force constraints,

$$c_t = \pm \mu_d c_n \tag{5}$$

where μ_d = dynamic coefficient of friction (sign dependent upon the sliding direction).

If after a solution any new boundary conditions were identified, for example new nodal contact. Then the forces and displacements for that stage were scaled, such that the new boundary conditions just came into effect. An accurate solution was then obtained with this scaled load. The remaining load was then applied with the new boundary conditions included in the stiffness matrix. This scaling and re-solution procedure was repeated for each new boundary condition until the structure was fully loaded. The overall displacements and forces were obtained by summation of each load stage.

In all contact problems where sliding of contactor nodes occurs, the above method will yield a stiffness matrix which is unsymmetric. This may present difficulties of implementation in some existing finite element programs.

However, two symmetric finite element methods based on variants of the unsymmetric method of eqn (1) are detailed below.

One method is termed the tangential force method. In this method, tangential friction force terms due to sliding, are included in successive iterations in the force vector. In the other method, termed the normal gap method, incorrect sliding displacement constraints are initially applied. However corrective normal gap terms are included in successive iterations in the 'force' vector.

3. THE TANGENTIAL FORCE METHOD

In this method the normal displacement constraint of L is used as the normal force constraint in the stiffness matrix as shown below,

$$\begin{bmatrix} \mathbf{K} & \mathbf{L}^T \\ \mathbf{L} & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{u} \\ \mathbf{c} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \mathbf{0} \end{Bmatrix}. \tag{6}$$

By applying this constraint to the sliding nodes, a solution is obtained in which the tangential forces are zero, i.e. frictionless contact has been imposed at the sliding nodes. However from the normal forces generated by the above equation, an estimate of the tangential friction forces are calculated using Coulomb's friction law, eqn (5). These tangential forces are then distributed in the ratio of the shape functions onto the relevant contactor and target nodes. For a quadratic element with three nodes per side, this gives tangential forces of

$$\left.\begin{aligned} c_{kt} &= +\mu_d c_{kn} \\ c_{at} &= -\mu_d c_{kn}[\eta_k(1-\eta_k)/2] \\ c_{bt} &= -\mu_d c_{kn}[1-\eta_k^2] \\ c_{ct} &= -\mu_d c_{kn}[\eta_k(1+\eta_k)/2] \end{aligned}\right\} \tag{7}$$

where η_k = local positional co-ordinate of the touching contactor node along the target element surface, with subscript k referring to the touching contactor node and subscripts a, b and c to the target element surface nodes.

The above tangential friction forces are resolved into global x and y forces and then included in the appropriate rows of the force vector. By iteration, improved estimates of the tangential forces are included until convergence occurs. The method of including sliding friction terms in the force vector is used by Bathe and Chaudhary [5] in their contact algorithm.

The complete solution for a contact problem by the tangential force method is obtained using the same loading and scaling procedure as used by the unsymmetric method.

4. THE NORMAL GAP METHOD

In this method the single sliding force constraint, coupling normal and tangential forces is used for the sliding displacement constraint in the stiffness matrix as shown below:

$$\begin{bmatrix} \mathbf{K} & \mathbf{J} \\ \mathbf{J}^T & \mathbf{0} \end{bmatrix} \begin{Bmatrix} \mathbf{u} \\ \mathbf{c} \end{Bmatrix} = \begin{Bmatrix} \mathbf{f} \\ \boldsymbol{\delta}_n \end{Bmatrix} \tag{8}$$

where,

$$[\mathbf{J}] = [\mathbf{L}_n]^T + \mu_d [\mathbf{L}_t]^T \tag{9}$$

and

$[\mathbf{L}_n]$ = normal direction displacement constraint array,

$[\mathbf{L}_t]$ = tangential direction displacement constraint array,

$\{\boldsymbol{\delta}_n\}$ = vector of contactor normal nodal gaps.

(NB. In the previous section for sliding nodes $\mathbf{L} = \mathbf{L}_n$.)

In the stiffness matrix of eqn (8), It can be seen that by using the sliding force constraint of eqn (9), an extra term of $\mu_d[\mathbf{L}_t]$ has been included in the displacement constraint, when compared with the unsymmetric method, eqn (1). This term appears because the algorithm enforces an unreal symmetry constraint. It represents a displacement equal to the product of the amount of sliding of each contactor node and the dynamic coefficient of friction. Constraining a node to slide in this manner results in it sliding along a different slope to that of the true target surface, by an

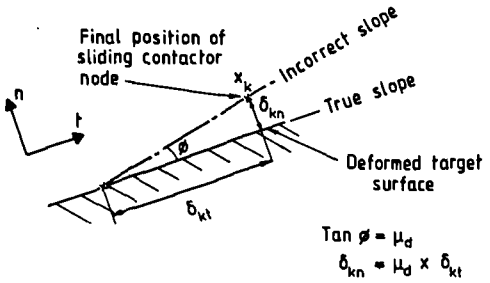


Fig. 3. Effect of incorrect displacement constraint.

angle equivalent to the dynamic coefficient of friction, as illustrated in Fig. 3. The forces generated by these constraints are approximately correct as dynamic friction is included.

Hence the initial solution yields approximately correct forces and incorrect displacements. The displacements result in normal gaps which appear between the contactor nodes and the target surface. The magnitude of the gaps is given by $\mu_d[\mathbf{L}_t]$ as indicated in Fig. 3. If the exact amount of sliding for a contactor node can be calculated, then a corrective normal gap term can be placed in the force vector and an accurate solution can be obtained. However due to inter-nodal effects of sliding along incorrect slopes, only an approximate value of the amount of sliding can be calculated. By including this approximate value in the 'force' vector, a more accurate solution is obtained from which a closer estimate of the amount of sliding can be calculated. Iteration is continued in this manner until convergence to an accurate solution is achieved.

4.1. Calculation of the normal gap terms

After the first solution involving sliding contactor nodes, positional gaps exist between the contactor nodes and the target surface as illustrated in Fig. 3. The amount of sliding, δ_{kt}, could be calculated for each contactor node as

$$\delta_{kt} = n_t \mathbf{x}_k - n_t \mathbf{x}_{k\text{sticking}} \tag{10}$$

where

n_t = unit vector in the tangential direction relative to the target surface,

$\mathbf{x}_k$ = final contactor co-ordinates of sliding node k,

$\mathbf{x}_{k\text{sticking}}$ = final nodal co-ordinates of node k, assuming node as sticking (i.e. origin of sliding).

However convergence to the accurate values of δ_{kt} was found to be slow because the positional gap was not being accounted for. The authors have found that the following method gave faster convergence, as both positional gaps and inter-nodal effects were taken into account together.

The method involves translation of the contactor body in the direction of 'free movement' (i.e. the direction it would move assuming a small force was applied). The translation distance was constrained such that the sum of the normal contactor gaps above the target surface, equalled the sum of the normal gaps beneath it, that is

$$\Sigma(+\delta_n) = |\Sigma(-\delta_n)|. \tag{11}$$

The final position of the contactor body representing an average touching position, is shown in Fig. 4. Using these translated contactor co-ordinates for $\mathbf{x}_k$, the amount of sliding can be calculated from eqn (10). These values are multiplied by μ_d and inserted in the appropriate position of the 'force' vector, (the row position in the force vector is the same as the row position of the sliding constraint in the stiffness matrix).

By successive iteration with progressively more accurate normal gap terms, convergence to the correct solution is obtained.

4.2. Decision on the direction of sliding

The decision of whether to update the state of a sticking node to sliding is made by application of Coulomb's friction law, eqn (4), to the normal and

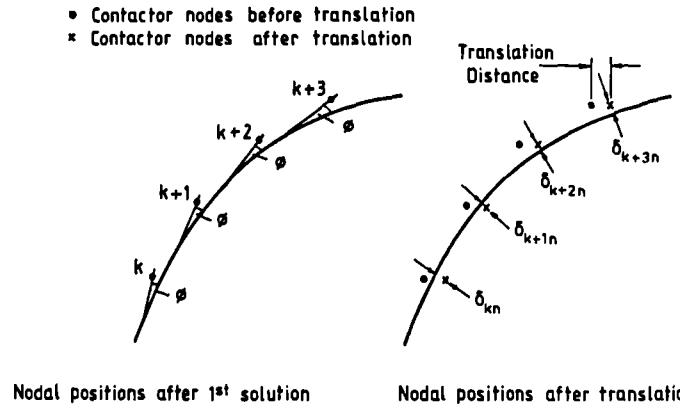


Fig. 4. Showing translation of contact nodes.

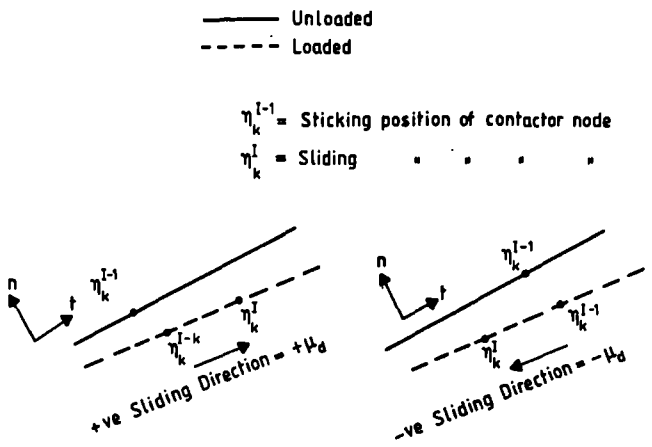


Fig. 5. Use of sticking and sliding nodal positions to decide sign of μ_d.

tangential nodal forces generated in the last solution. If a node changes to sliding, then the direction and appropriate friction forces are included directly in the stiffness matrix by the sign of μ_d. This procedure has been found to be acceptable in contact problems where only one contactor node is initially touching, with the number of contacting nodes progressively increasing with load. However, in problems where more than one contactor node is initially touching, the use of 'sticking forces' to decide the direction of sliding is not necessarily correct. This is because of inter-nodal sticking effects.

To overcome this difficulty, a checking stage has been introduced after each solution. By this method, the position of a sliding node is compared with its position if it had remained as sticking on the target. Hence, if a sliding node is in the positive tangential direction relative to the sticking position, then positive sliding is defined as occurring, as illustrated in Fig. 5, and a positive value of μ_d is inserted in the stiffness matrix for that node. Conversely, if negative sliding has occurred, then a negative value of μ_d is inserted.

This positional check to decide the sign of μ_d, is applicable even when the previous value obtained from the sticking forces was incorrect. This is because the sign of μ_d represents the direction of the tangential friction forces imposed on the system and using the incorrect sign for μ_d, actually exaggerates the true direction of sliding. This effect is illustrated in Fig. 6.

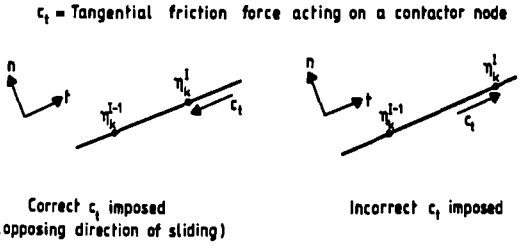


Fig. 6. Effect of wrong tangential force on nodal displacement.

5. RESULTS

In this section three contact problems are presented. The objective in the first two problems was to identify which of the two symmetric methods was fastest in terms of convergence. The best method was then used for a more complex sliding problem. Comparison of results are made with an analytical solution and results obtained using the unsymmetric method.

5.1. Flat sliding contact

The first problem analysed was of a sliding block over a flat inclined surface. The contactor and target were non-rigid with friction present along the contact surface. To allow relatively easy comparison of the results, a simple finite element mesh of four quadrilateral, eight-node elements was used as illustrated in Fig. 7. In this model, five contactor nodes are touching the target surface.

Finite element results by the two symmetric methods and the unsymmetric method are tabulated in Table 1. In this table are listed the number of iterations, the final contactor nodal states, the nodal normal forces and the displacements at node 1. This allows a full and direct comparison to be made between the three methods.

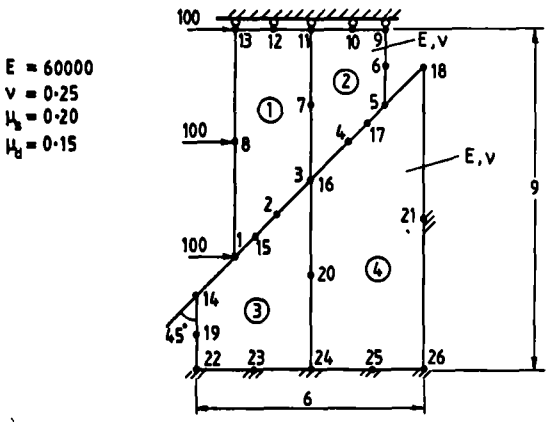


Fig. 7. Finite element mesh for flat contact.

Table 1. Comparison of finite element results for the flat sliding problem

Method	Iterations	Contactor nodal states	c_{1n}	c_{2n}	c_{3n}	c_{4n}	c_{5n}	δ_{1x}	δ_{1y}
Tangential force method	8	All positive sliding	91.053	5.1929	69.549	156.03	47.102	$0.66845E-4$	$0.49919E-4$
Normal gap method	6	All positive sliding	91.053	5.1963	69.548	156.03	47.102	$0.66847E-4$	$0.49921E-4$
Unsymmetric method	4	All positive sliding	91.054	5.1935	69.546	156.03	47.104	$0.66846E-4$	$0.49920E-4$

The final contact surface profile is shown in Fig. 8. A comparison in the rate of convergence between the two symmetric methods is shown in Fig. 9, where the normal force at node 1 is plotted for each iteration (no values are shown for iterations 1 and 2 as sticking nodes are present in these stages).

From Table 1, it can be seen that convergence to the unsymmetric method results was achieved by both of the symmetric methods. The gap method however required less iterations than the tangential force method. Interestingly, iterations 2 and 3 of the unsymmetric and gap methods used the incorrect sign of μ_d for some of the sliding nodes. However, by stage 4, these signs had been corrected by the displacement checking routine.

5.2. *Curved sliding contact*

To analyse the symmetric methods over a curved contact region, the second contact problem considered was that of a sliding block over a sharply curved contact surface. Again the contactor and target bodies were non-rigid with friction present along the contact boundary. The finite element mesh used consisted of four quadrilateral, eight-node elements as shown in Fig. 10.

Finite element results by the two symmetric methods and the unsymmetric method are tabulated in Table 2. The final contact surface profile is shown in Fig. 11.

From Table 2, it can be seen that convergence to the unsymmetric results has occurred with both of the symmetric methods. Again the gap method has converged in less iterations than the tangential force method. The surface profile has resulted in the extreme right contactor node separating from the target, as indicated by $C_{5n} = 0$. This occurred because the forces generated by the other four touching contactor nodes deformed the target downwards, resulting in separation of this node.

The normal forces produced at the contactor nodes may at first appear erratic. This is partially due to the very coarse mesh and the lack of nodes in the contact region. However, the deformed contact profiles show approximate slope continuity between adjacent elements, which is a desired feature. Also a check on the average pressure exerted by each contactor element confirms expected trends, i.e. in the flat contact problem, the maximum pressure occurs at the element 2 interface and in the curved contact problem at the element 1 interface.

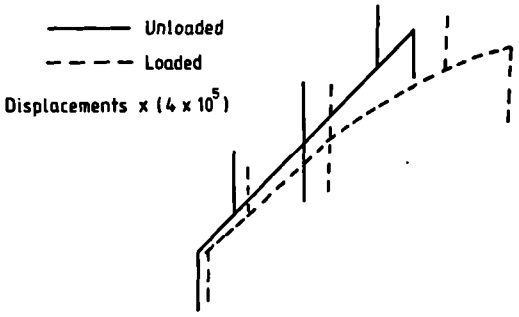


Fig. 8. Undeformed and deformed flat surface profiles.

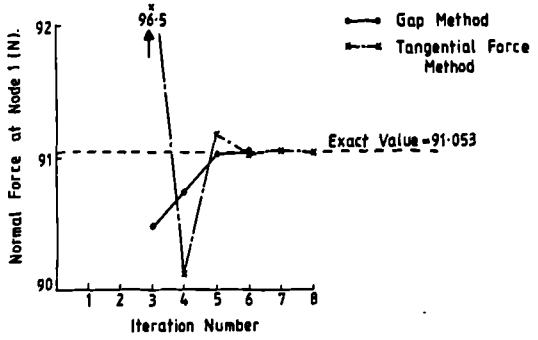


Fig. 9. Comparison of convergence of normal force of node 1.

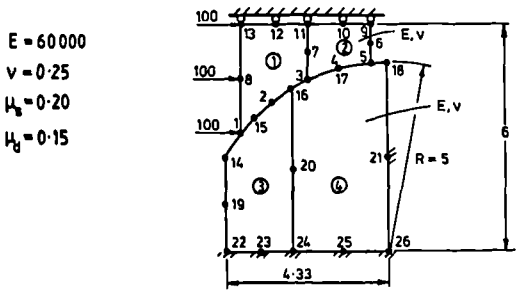


Fig. 10. Finite element mesh for curved contact.

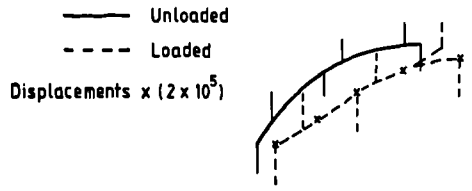


Fig. 11. Undeformed and deformed curved surface profiles.

Table 2. Comparison of finite element results for the curved sliding problem

Method	Iterations	Contactor nodal states	c_{1n}	c_{2n}	c_{3n}	c_{4n}	c_{5n}	δ_{1x}	δ_{1y}
Tangential force method	11	All positive sliding except for node 5 which has separated	73.262	193.54	107.19	89.607	0	$0.11156E-3$	$0.71023E-4$
Normal gap method	8		73.262	193.51	107.19	89.599	0	$0.11155E-3$	$0.71026E-4$
Unsymmetric method	5		73.261	193.52	107.18	89.600	0	$0.11155E-3$	$0.71016E-4$

From the above results and other models compared by the authors, the normal gap method has proved faster in convergence than the tangential force method.

5.3. *Sliding cylinder problem*

This contact problem consists of a cylinder sliding up an inclined plane as illustrated in Fig. 12. This same problem was analysed by the authors in a previous paper [6], where results by the unsymmetric finite element method were compared with a known analytic solution. To allow comparison of results by the gap method with this paper, the same finite element mesh was used as shown in Fig. 13.

Comparative results between the symmetric gap method and the unsymmetric method were obtained for the stage where nine contactor nodes were touching the target surface. Listed in Table 3 are comparative values of the number of iterations and the load applied for the ninth node to just touch the target surface, along with sample displacements in the contact region.

From these results it can be seen that the nodal displacements obtained by the two methods are almost identical, resulting in the deformed meshes as shown in Fig. 14. The shear stresses obtained by using these displacements as prescribed values over the surface of a refined target mesh (to allow more accurate stress calculation) are also identical to the unsymmetric method results. Figure 15 shows comparison of the analytic shear stress contours in the contact region with the finite element results. These shear stresses by the finite element method have confirmed the approximate position of the maximum shear stresses, although the peak value is overestimated by about 20%.

Comparison of the number of iterations between the symmetric gap method and the unsymmetric method is interesting. Although about half as many more iterations are required, the nett computer time

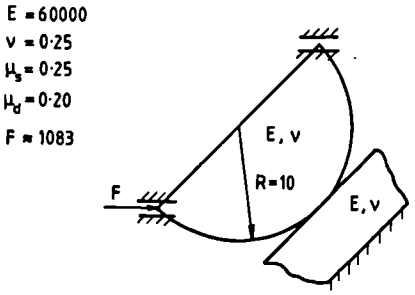


Fig. 12. Sliding cylinder contact problem.

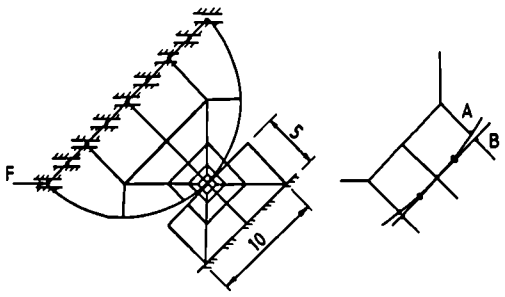


Fig. 13. Finite element mesh for sliding cylinder.

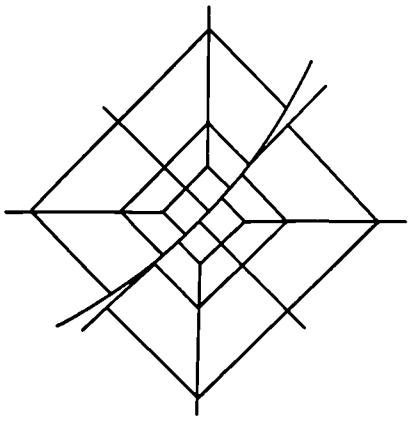


Fig. 14. Magnified view of meshes in contact region.

Table 3. Comparison of finite element results for the sliding cylinder problem

Method	Iterations	Contactor nodal states	Total load (N)	δ_{Ax}	δ_{Ay}	δ_{Bx}	δ_{By}
Normal gap method	100	All positive sliding	1082.4	$0.80001E-1$	$0.27889E-1$	$0.28272E-1$	$0.16273E-1$
Unsymmetric method	84	All positive sliding	1083.2	$0.80062E-1$	$0.27914E-1$	$0.28287E-1$	$0.16282E-1$

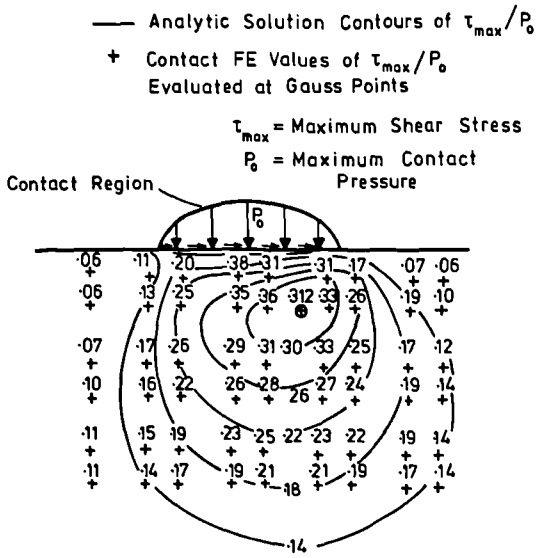


Fig. 15. Comparison of maximum shear stress.

is less as symmetry has been preserved. The number of iterations may appear high, this is due to a strict positional tolerance on contact included in the checking routines. This is necessary as the large amount of sliding of the contactor nodes in their initial tangent direction, may result in their final position not lying on the deformed target surface as illustrated in Fig. 16. However by re-solution with a modified tangent direction, the strict positional tolerance is maintained.

6. CONCLUSIONS

The solution of curved friction contact problems by two different symmetric finite element methods has been presented. Both methods were based on variants of an unsymmetric method and employed force and displacement constraints related to the element shape functions. Incorrect constraints were initially applied with corrective terms included by iteration in the force vector. One method used corrective friction force terms and the other corrective normal gap

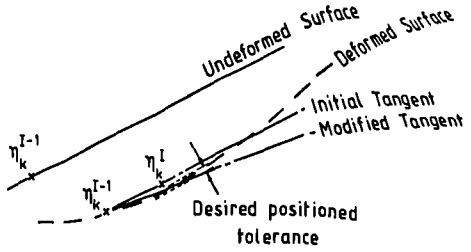


Fig. 16. Showing new tangent direction.

terms. Results showed that both of the symmetric methods and the unsymmetric method converged to the same solution, with convergence by the normal gap method being significantly faster than by the friction force method.

Further investigation is currently being undertaken into the effect of relaxing the positional tolerance on contacting nodes and thereby allowing less iterations per solution.

Acknowledgement—S. K. Pascoe is supported by an SERC (CASE) award in collaboration with the Central Electricity Generating Board.

REFERENCES

1. J. Böhm, A comparison of different contact algorithms with applications. *Comput. Struct.* **26**, 207–221 (1987).
2. T. Y. Chang, A. F. Saleeb and S. C. Shyu, Finite element solutions of two-dimensional contact problems based on a consistent mixed formulation. *Comput. Struct.* **27**, 455–466 (1987).
3. S. Timoshenko and J. N. Goodier, *Theory of Elasticity*. McGraw-Hill, New York (1951).
4. T. J. R. Hughes, R. L. Taylor, J. L. Sackman, A. Curnier and W. Kanoknukulchai, A finite element method for a class of contact–impact problems. *Comput. Meth. appl. Mech. Engng* **8**, 249–276 (1976).
5. K.-J. Bathe and A. Chaudhary, A solution method for planar and axisymmetric contact problems. *Int. J. num. Method. Engng* **21**, 65–88 (1985).
6. S. K. Pascoe and J. E. Mottershead, Linear elastic contact problems using curved elements and including dynamic friction. *Int. J. num. Meth. Engng* **26**, 1631–1643 (1988).

TABLES

Interface Materials	μ_s	μ_d
(unlubricated)		
steel on steel	0.80	0.50
steel on copper	0.22	0.20
steel on brass	0.35	0.20
steel on cast iron	0.40	0.30
(lubricated)		
steel on copper	0.15	0.10
steel on brass	0.20	0.10
steel on cast iron	0.25	0.15

Table 3.1 Coefficients of friction for some commonly used materials

'Easy'	'Difficult'
Non-varying contact areas	Varying contact areas
Sticking only or sliding only contact without friction	Mixed sticking and sliding with friction
Only one body deformable	Both bodies deformable
Two dimensional contact	Three dimensional contact
Node on node contact	Mis-aligned nodal contact
Flat contact	Curved contact
Small scale sliding	Large scale sliding

Table 4.1 Comparison of different contact features

Stage	Contactor node number						
	1	2	3	4	5	6	7
1	2	0	0	0	0	0	0
2	1	2	2	2	2	2	0
3	1	1	1	1	1	0	0
4	1	1	1	1	1	0	0
5	1	1	1	1	1	0	0

Table 10.1 Nodal contact states during the Hertz solution

Stage	Contactor node number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0	0
3	1	2	0	0	0	0	0	0	0	0	0	0	0
4	1	1	0	0	0	0	0	0	0	0	0	0	0
5	1	1	2	0	0	0	0	0	0	0	0	0	0
6	1	1	1	0	0	0	0	0	0	0	0	0	0
7	1	1	1	2	0	0	0	0	0	0	0	0	0
8	1	1	1	1	0	0	0	0	0	0	0	0	0
9	1	1	1	1	2	0	0	0	0	0	0	0	0
10	1	1	1	1	1	0	0	0	0	0	0	0	0
11	1	1	1	1	1	2	0	0	0	0	0	0	0
12	1	1	1	1	1	1	0	0	0	0	0	0	0
13	1	1	1	1	1	1	2	0	0	0	0	0	0
14	1	1	1	1	1	1	1	0	0	0	0	0	0
15	1	1	1	1	1	1	1	2	0	0	0	0	0
16	1	1	1	1	1	1	1	1	0	0	0	0	0
17	1	1	1	1	1	1	1	1	2	0	0	0	0
18	1	1	1	1	1	1	1	1	1	0	0	0	0

Table 10.2 Nodal contact states during the pin in a hole solution (1 b.c.loading per stage)

Stage	Contactor node number												
	1	2	3	4	5	6	7	8	9	10	11	12	13
1	2	0	0	0	0	0	0	0	0	0	0	0	0
2	1	2	2	2	2	2	2	2	2	2	2	0	0
3	1	1	1	1	1	1	1	1	1	0	0	0	0
4	1	1	1	1	1	1	1	1	0	0	0	0	0
5	1	1	1	1	1	1	1	1	0	0	0	0	0
6	1	1	1	1	1	1	1	1	0	0	0	0	0
7	1	1	1	1	1	1	1	1	0	0	0	0	0

Table 10.3 Nodal contact states during the pin in a hole solution (incremental loading)

	Method of Constraint Imposition					
Sample disp's	Lagrange Multipliers	Trans' Matrix	Penalty Method (α) 10^4	10^6	10^8	10^{10}
δ_{3x}	0.10069	0.10069	0.29299	0.10261	0.10071	0.10069
δ_{3y}	0.0	0.0	0.0	0.0	0.0	0.0
δ_{7x}	0.10251	0.10251	0.29481	0.10443	0.10253	0.10251
δ_{7y}	0.00444	0.00444	0.00444	0.00444	0.00444	0.00444
δ_{11x}	0.00470	0.00470	0.00470	0.00470	0.00470	0.00470
δ_{11y}	0.00278	0.00278	0.00278	0.00278	0.00278	0.00278
δ_{13x}	0.01741	0.01741	0.01741	0.01741	0.01741	0.01741
δ_{13y}	-.00941	-.00941	-.00941	-.00941	-.00941	-.00941

Table 10.4 Comparison of constraint imposition techniques

1	-2	Contactor element and surface indicator
2	-2	Contactor element and surface indicator
3	-2	Contactor element and surface indicator
4	-2	Contactor element and surface indicator
5	-2	Contactor element and surface indicator
6	-2	Contactor element and surface indicator
7	-2	Contactor element and surface indicator
8	-2	Contactor element and surface indicator
9	2	Target element and surface indicator
10	2	Target element and surface indicator
11	2	Target element and surface indicator
12	2	Target element and surface indicator
13	2	Target element and surface indicator
14	2	Target element and surface indicator
15	2	Target element and surface indicator
16	2	Target element and surface indicator
μ_s	0.2	Static coefficient of friction
μ_d	0.1	Dynamic coefficient of friction

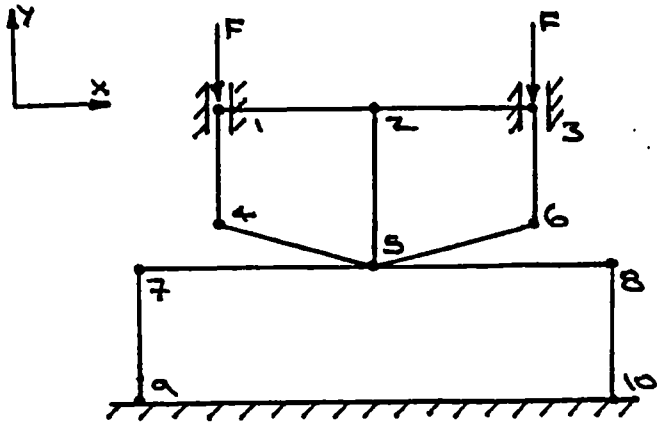
Table D.1 Extra input data for contact modelling

Node	Contact Mesh Results	Standard Mesh Results
1_x	0.0	0.0
1_y	-0.28415D-5	-0.28415D-5
2_x	0.0	0.0
2_y	-0.28415D-5	-0.28415D-5
3_x	-0.42623D-6	-0.42623D-6
3_y	-0.14208D-5	-0.14208D-5
4_x	0.42623D-6	0.42623D-6
4_y	-0.14208D-5	-0.14208D-5
5_x	-0.42623D-6	0.0
5_y	-0.14208D-5	0.0
6_x	0.42623D-6	0.0
6_y	-0.14208D-5	0.0
7_x	0.0	-
7_y	0.0	-
8_x	0.0	-
8_y	0.0	-

Table F.1 Displacement results for the 'simple' sticking problem

Node	Frictionless Sliding Results $\mu_d = 0.0$	Frictional Sliding Results $\mu_d = 0.15$
A_x	0.38342D-02	0.21024D-2
A_y	0.0	0.0
B_x	0.39297D-2	0.21202D-2
B_y	0.19187D-3	0.80863D-4
C_x	0.16418D-3	0.11639D-3
C_y	0.10332D-3	0.73849D-4
D_x	0.41648D-3	0.28347D-3
D_y	0.28326D-4	0.29090D-3
E_x	0.62567D-3	0.41838D-3
E_y	-0.33977D-3	-0.19928D-3
F_n	278.86	136.88
F_t	0.0	20.53
Final θ	8.247°	8.267°

Table F.2 Comparison of frictionless and frictional results for the 'simple' sliding problem



Iteration	Position	Disp'/Force	Overlap
1	δ_{7y} and δ_{8y}	-0.2252D-3	
"	δ_{4y} and δ_{6y}	-1.4502D-3	-0.2250D-3
"	δ_{5y}	-0.2252D-3	0.0
"	F_{4x}	0	
"	F_{4y}	0	
"	F_{5x}	0	
"	F_{5y}	5000	
"	F_{6x}	0	
"	F_{6y}	0	
2	δ_{7y} and δ_{8y}	-0.0007D-3	
"	δ_{4y} and δ_{6y}	0.2243D-3	0.0
"	δ_{5y}	-0.0006D-3	0.0001D-3
"	F_{4x}	20	
"	F_{4y}	278.5	
"	F_{5x}	0	
"	F_{5y}	4443	
"	F_{6x}	-20	
"	F_{6y}	278.5	

Table F.3 Sample displacements from the overlap model

Node	δ_x	δ_y	δ_z
1,2,3,4,5,6,7,8	0.0	0.0	0.0
9	-0.134D-6	-0.134D-6	-0.242D-6
10	-0.131D-6	-0.131D-6	-0.306D-6
11	0.134D-6	0.134D-6	-0.242D-6
12	0.131D-6	-0.131D-6	-0.306D-6
13	-0.230D-8	-0.230D-8	-0.452D-6
14	-0.544D-7	0.221D-7	-0.735D-6
15	-0.796D-7	0.796D-7	-0.688D-6
16	-0.221D-7	0.544D-7	-0.735D-6
17	0.230D-8	0.230D-8	-0.452D-6
18	0.544D-7	-0.221D-7	-0.735D-6
19	0.796D-7	-0.796D-7	-0.688D-6
20	0.221D-7	-0.544D-7	-0.735D-6
21	-0.147D-6	-0.147D-6	-0.452D-6
22	-0.132D-6	-0.197D-7	-0.735D-6
23	-0.172D-6	0.172D-6	-0.688D-6
24	0.197D-7	0.132D-6	-0.735D-6
25	0.147D-6	0.147D-6	-0.452D-6
26	0.132D-6	0.197D-7	-0.735D-6
27	0.172D-6	-0.172D-6	-0.688D-6
28	-0.197D-7	-0.132D-6	-0.735D-6
29	0.0	0.0	-0.744D-6
30	-0.429D-7	0.429D-7	-0.258D-5
31	0.0	0.0	-0.744D-6
32	0.429D-7	-0.429D-7	-0.258D-5
33	0.572D-6	0.572D-6	0.347D-6
34	-0.480D-6	0.621D-6	-0.192D-5
35	-0.159D-5	0.159D-5	-0.616D-5
36	-0.621D-6	0.480D-6	-0.192D-5
37	-0.572D-6	-0.572D-6	0.347D-6
38	0.480D-6	-0.621D-6	-0.192D-5
39	0.159D-5	-0.159D-5	-0.616D-5
40	0.621D-6	-0.480D-6	-0.192D-5

Table F.4 3-D Frictionless sliding results

FIGURES

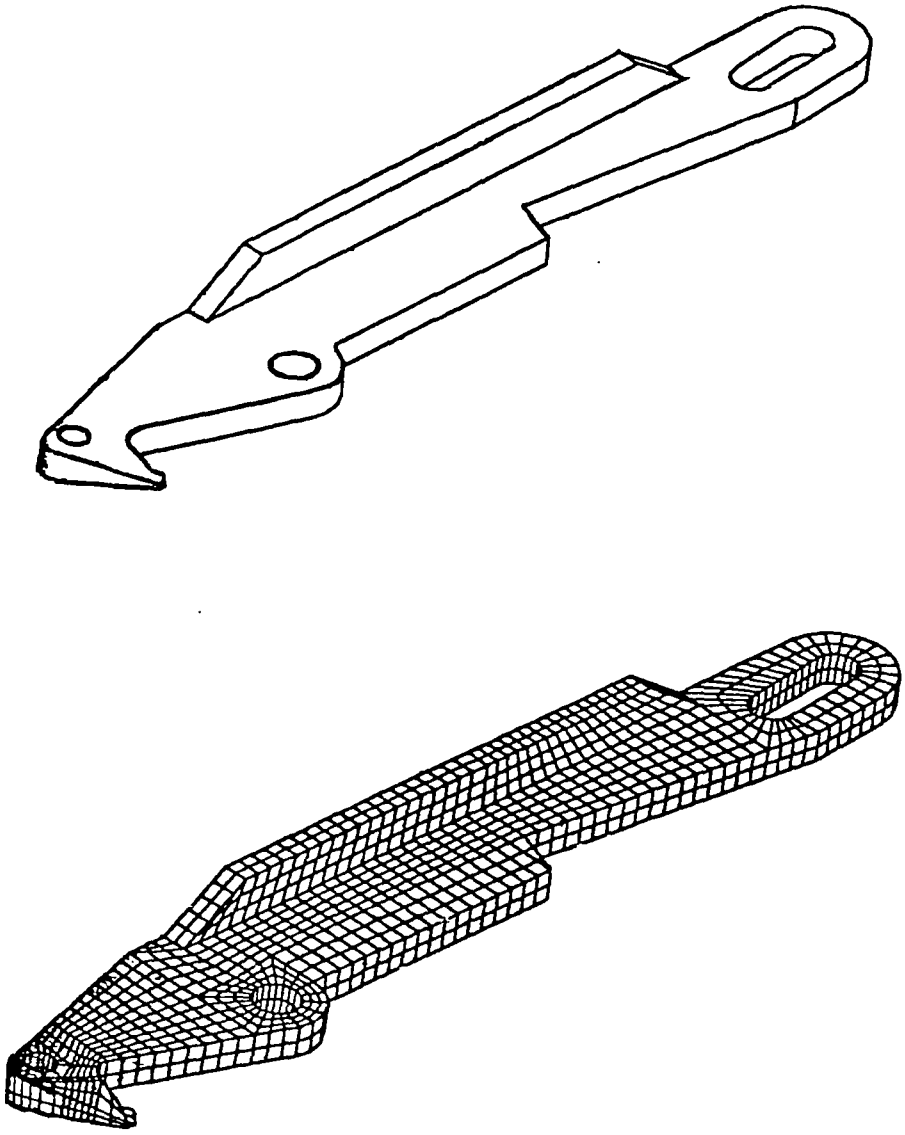


Figure 2.1 Typical finite element mesh

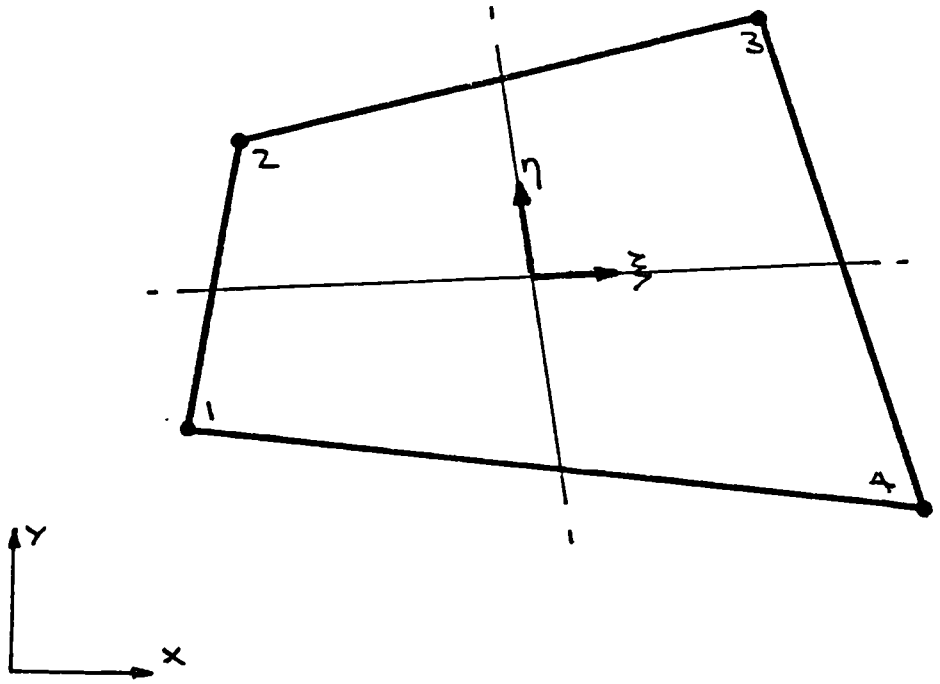


Figure 2.2 The four noded, two dimensional quadrilateral element

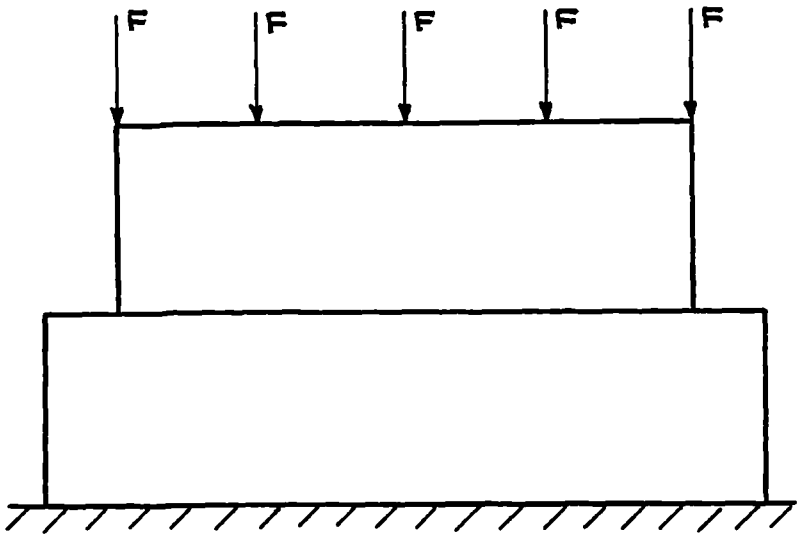


Figure 3.1 Example of conforming contact

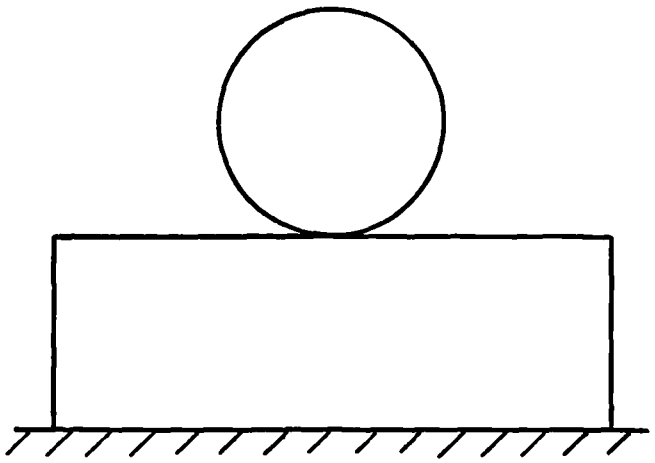


Figure 3.2a Point contact of a cylinder on a flat base

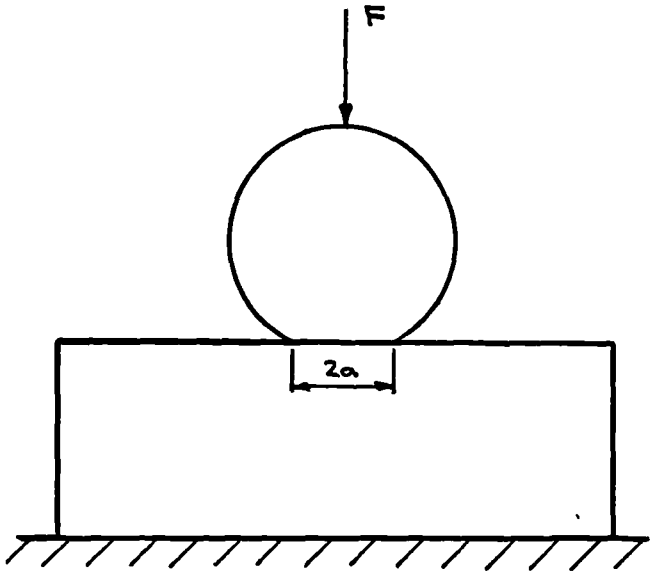


Figure 3.2b Area contact after application of load

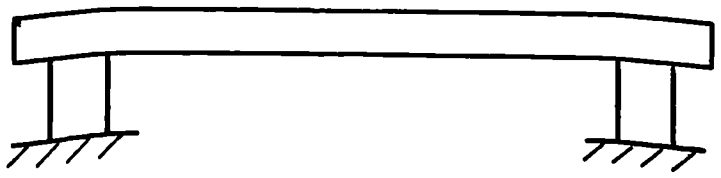


Figure 3.3a Slender beam resting on two wide supports

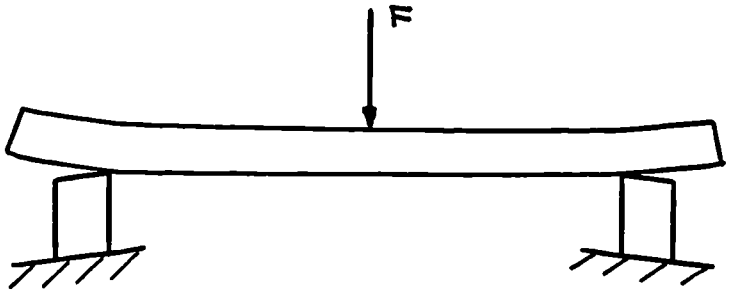


Figure 3.3b Localised contact after application of load

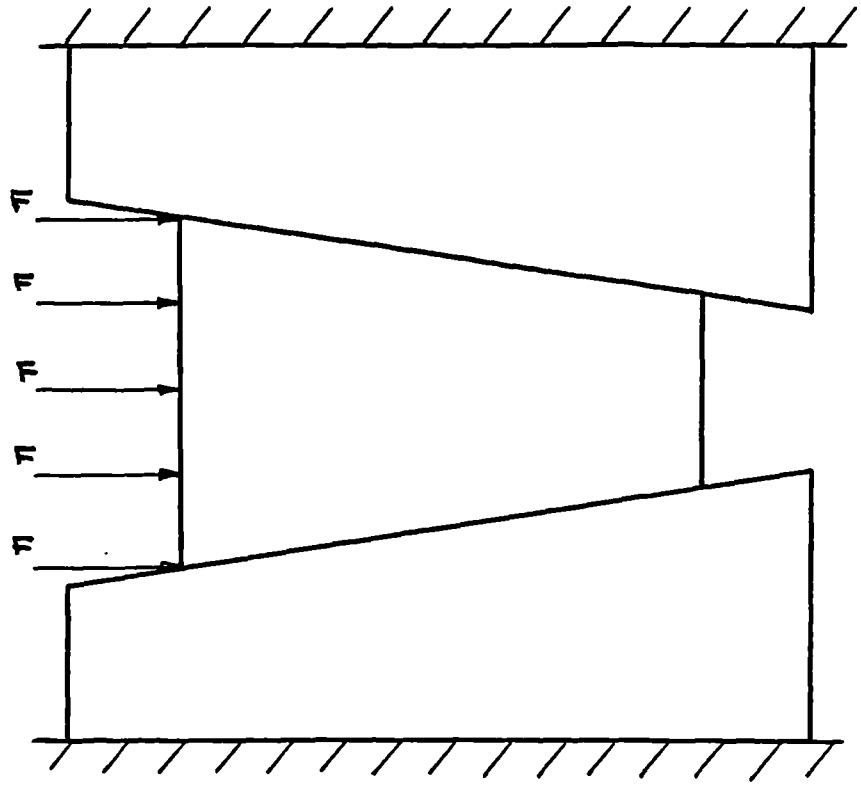


Figure 3.4 Wedge contact problem

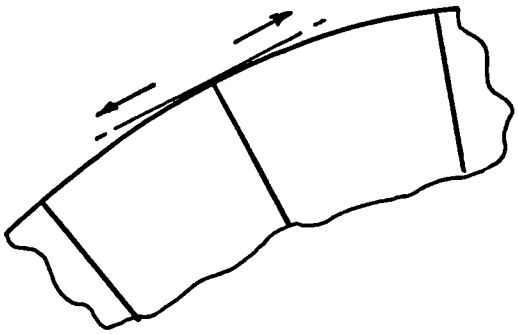


Figure 3.5a Possible sliding directions in 2-D contact

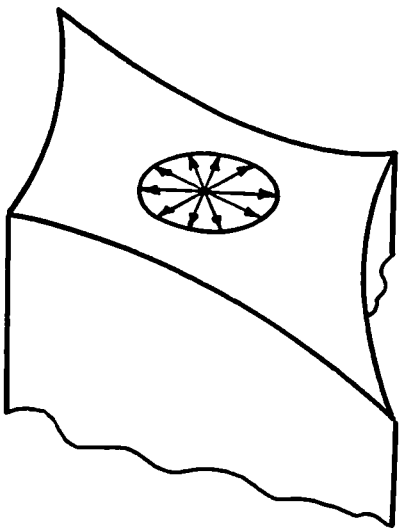


Figure 3.5b Possible sliding directions in 3-D contact

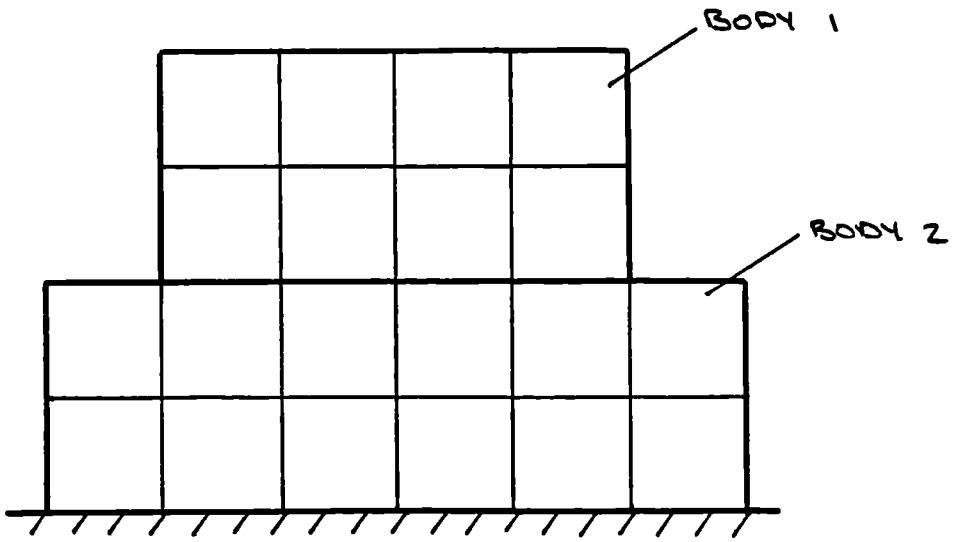


Figure 4.1 Example of aligned nodal contact

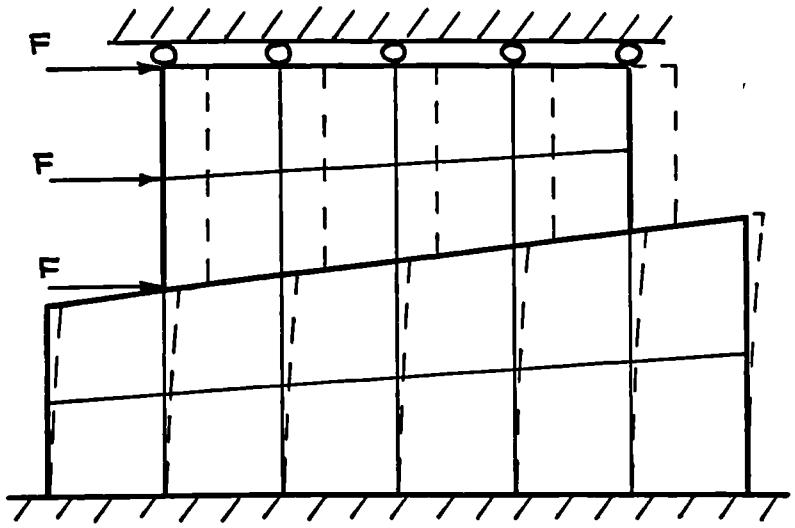


Figure 4.2 Example of nodal mis-alignment due to large scale sliding

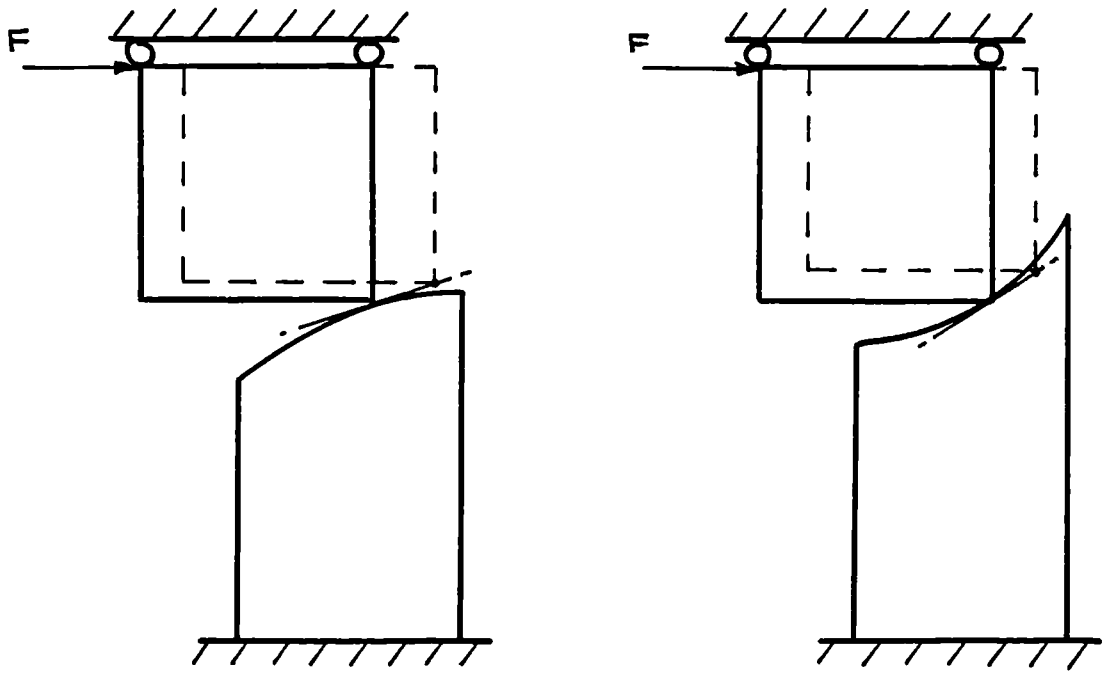


Figure 4.3 Examples of induced gaps or overlaps due to large scale sliding

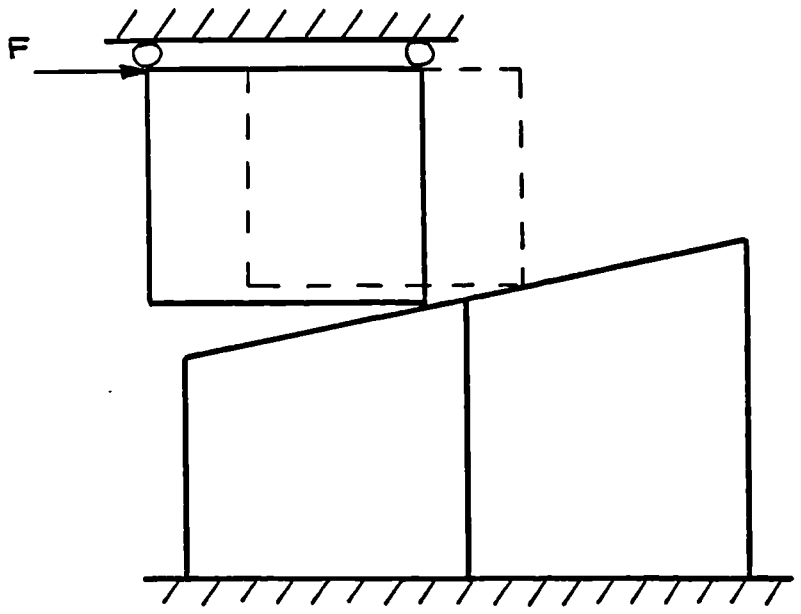


Figure 4.4 Example of sliding from one element to another

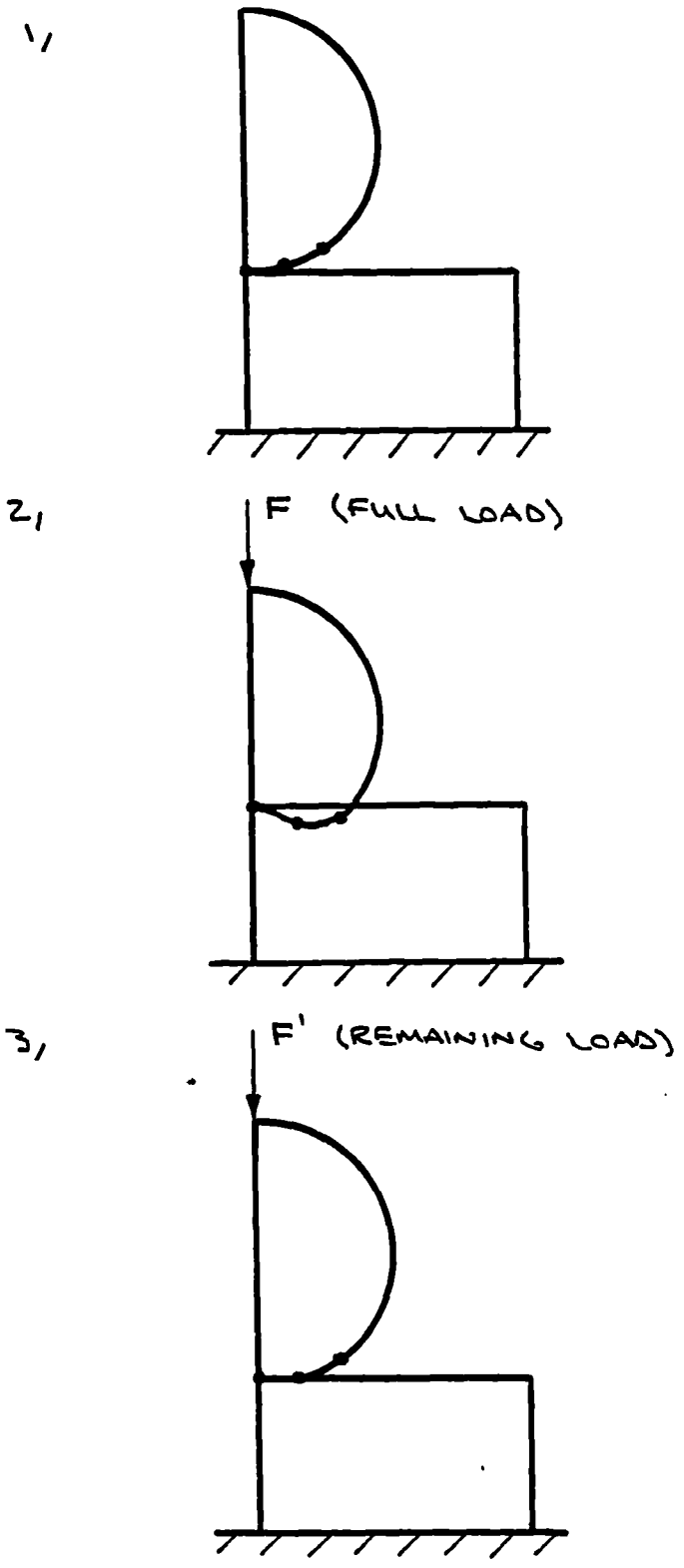


Figure 5.1 Scaling procedure for one boundary condition loading

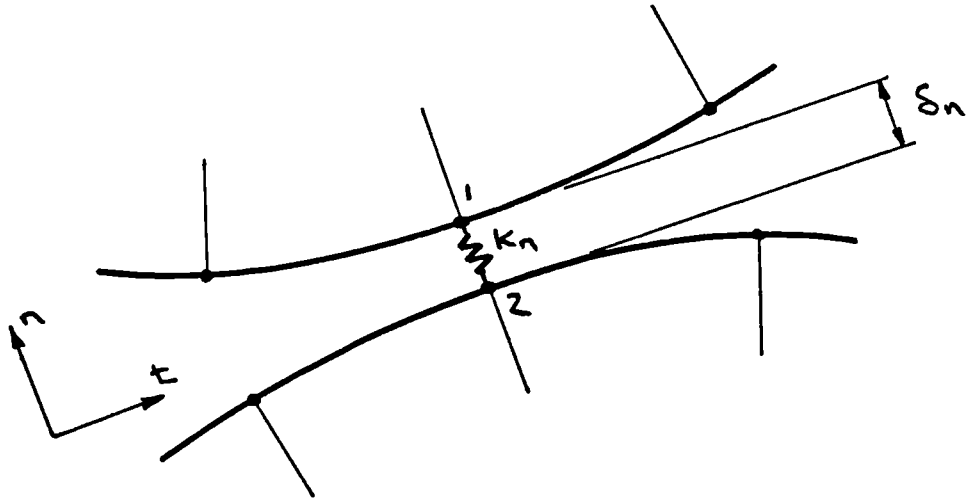


Figure 5.2 Linear gap element connected across adjacent nodes

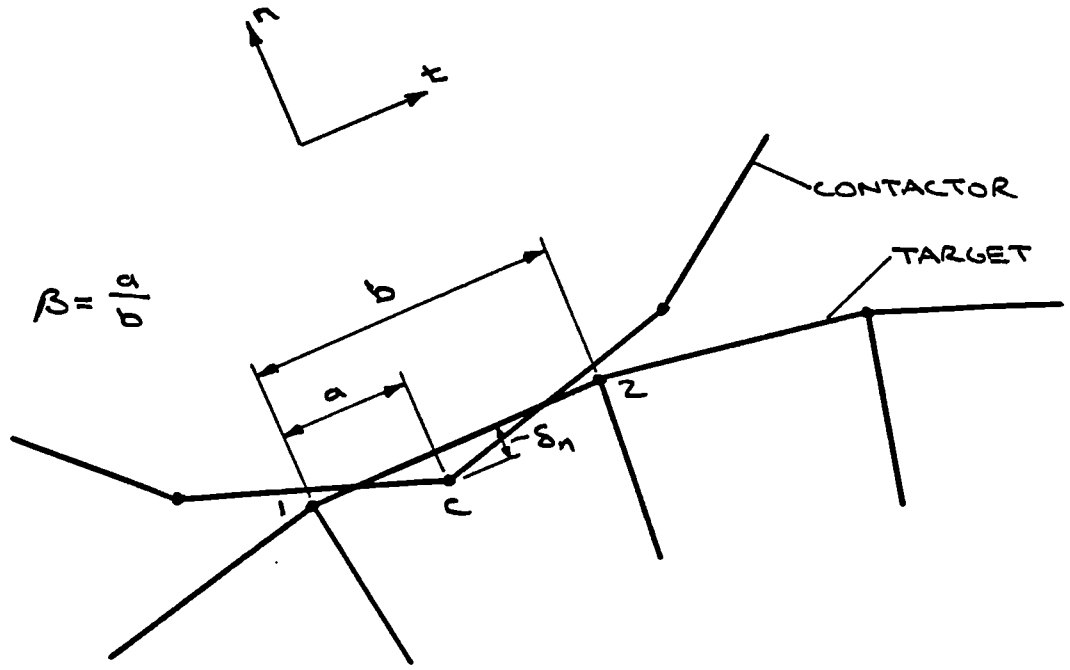


Figure 5.3 Contact of a single node

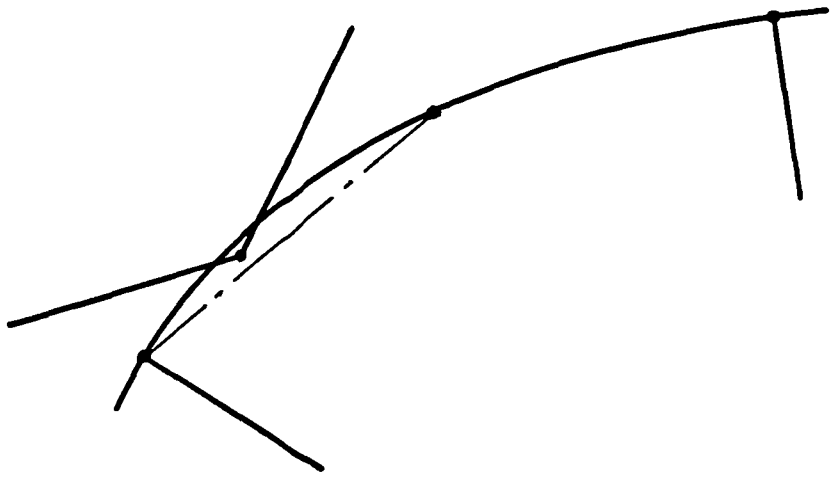


Figure 5.4 Example where linear contact identification fails

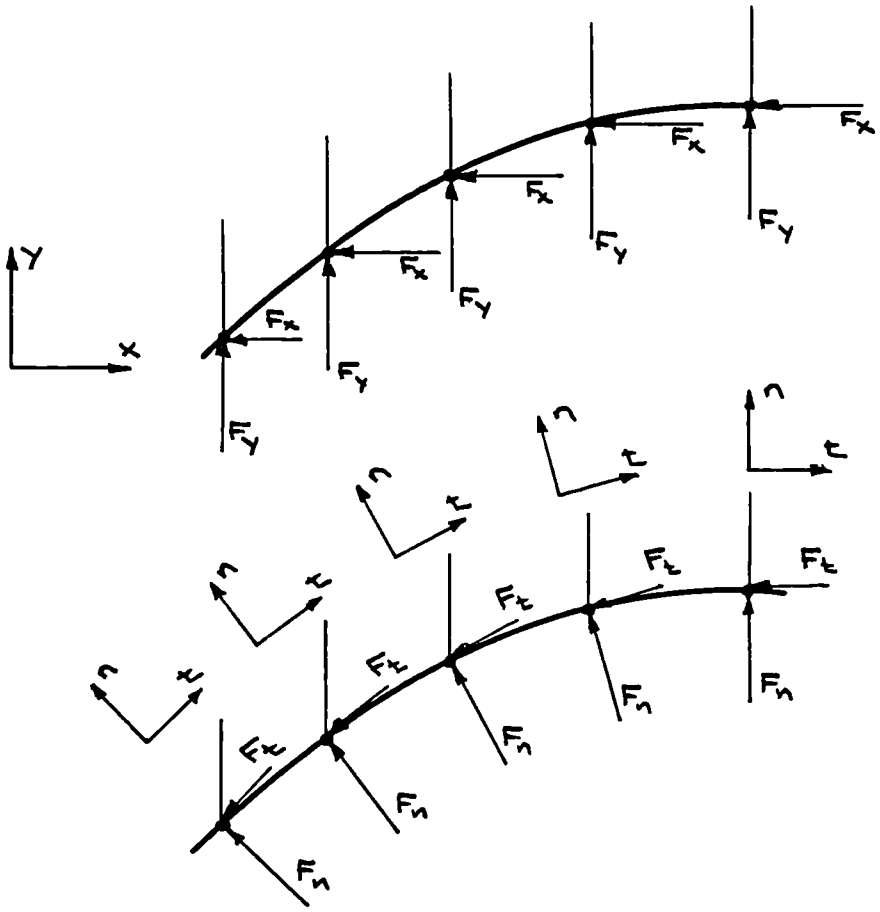


Figure 5.5 Example of conversion from global to local contact forces

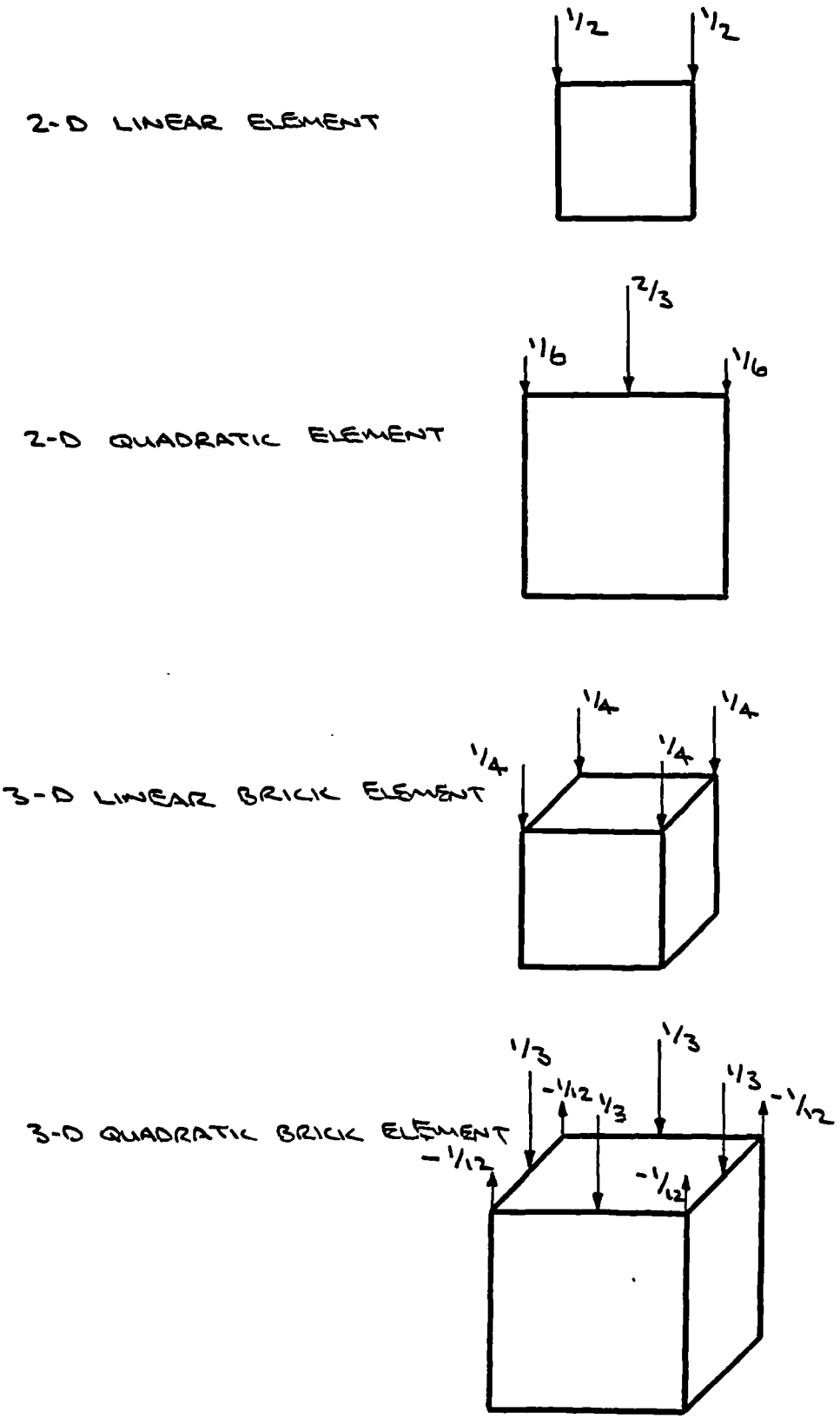
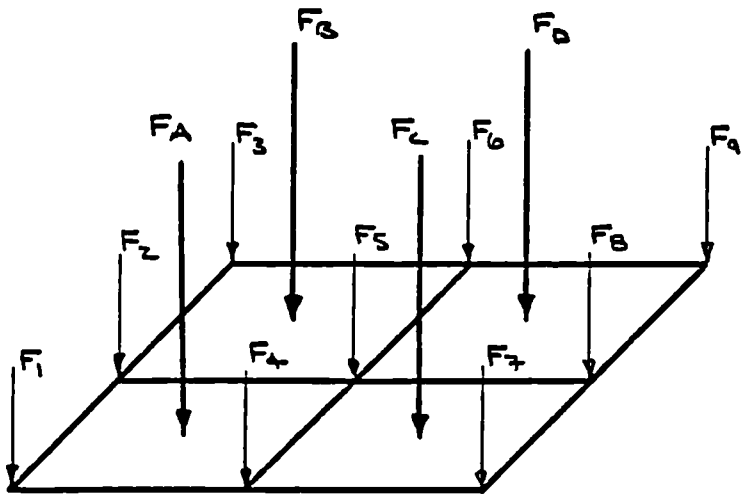


Figure 5.6 Constant pressure loading for different element types



$$F_A = F_1 + \frac{F_2}{2} + \frac{F_3}{4} + \frac{F_4}{2}$$

$$F_B = \frac{F_2}{2} + F_3 + \frac{F_6}{2} + \frac{F_5}{4}$$

$$F_C = \frac{F_4}{2} + \frac{F_5}{4} + \frac{F_8}{2} + F_7$$

$$F_D = \frac{F_5}{4} + \frac{F_6}{2} + F_9 + \frac{F_8}{2}$$

Figure 5.7 Example of force averaging over an element face

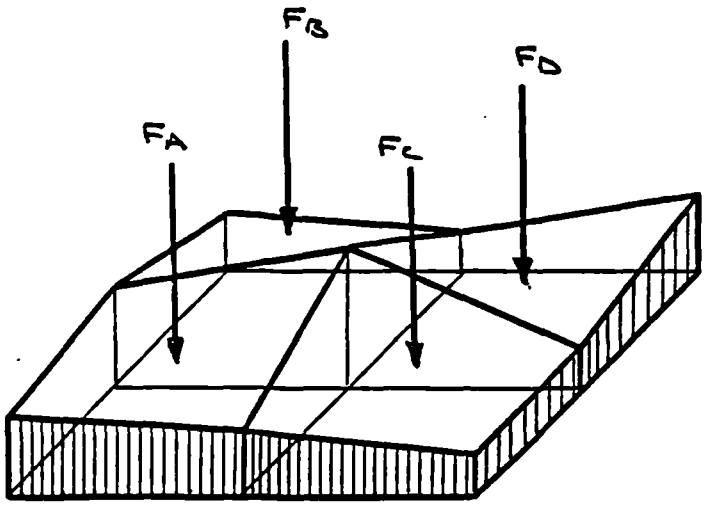


Figure 5.8 Element force calculation by integration of equivalent pressures

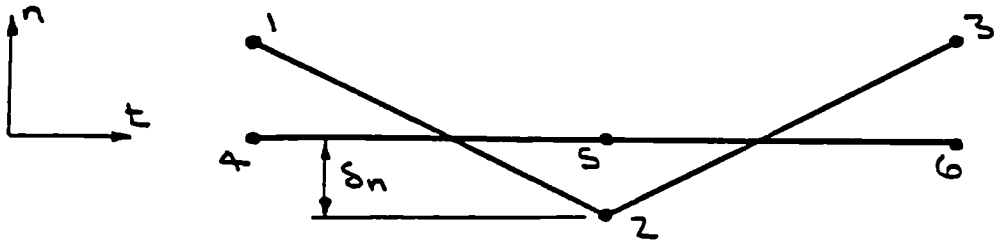


Figure 5.9 Example of simple nodal overlap

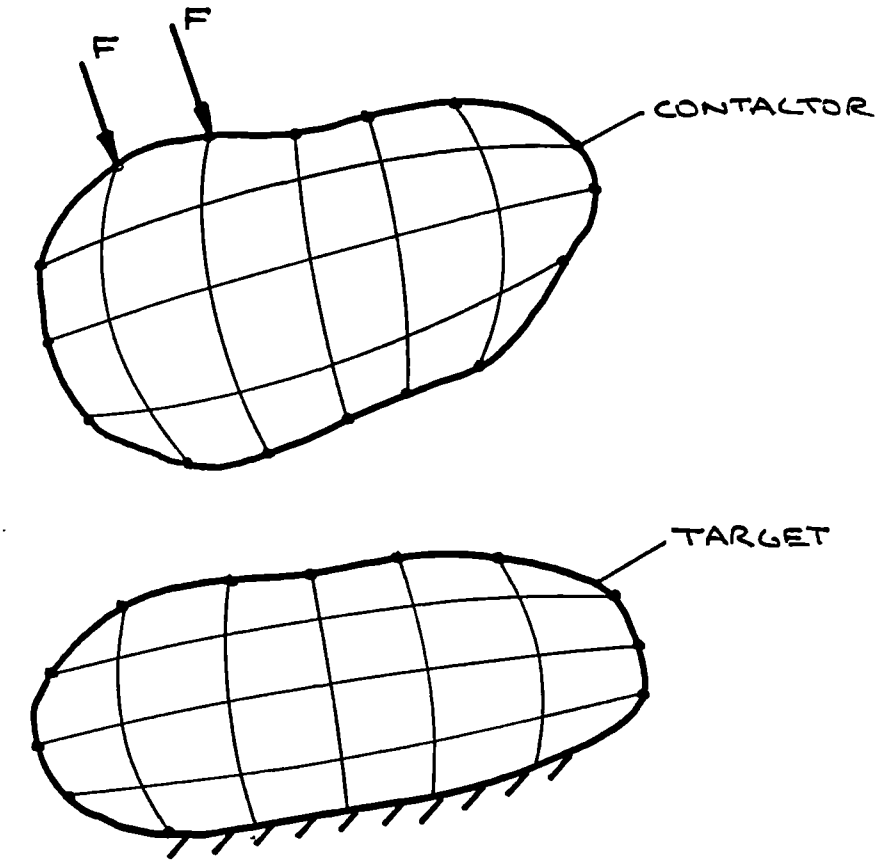


Figure 6.1 Contactor and target body definition

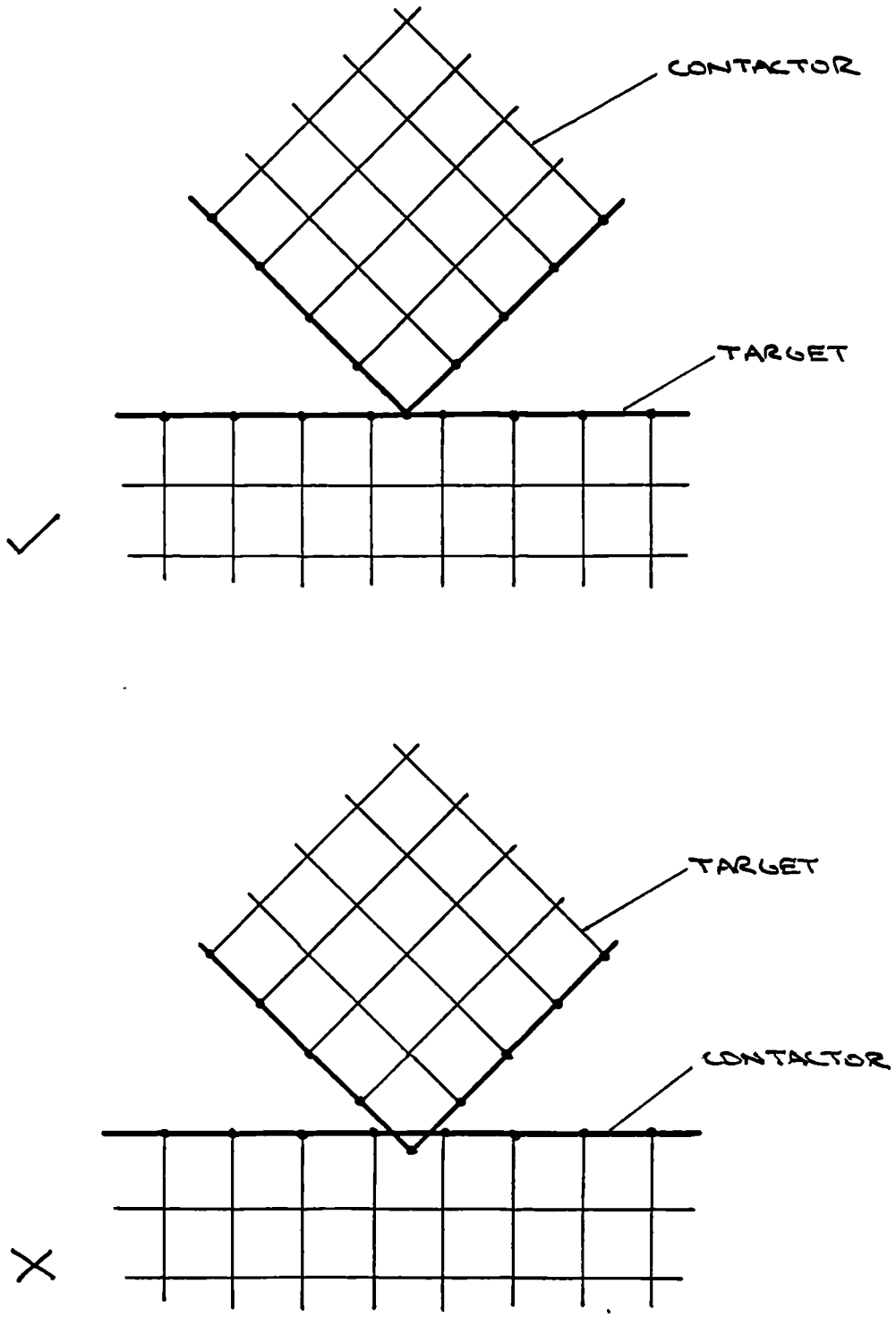


Figure 6.2 Examples of 'good' and 'poor' contactor and target definition

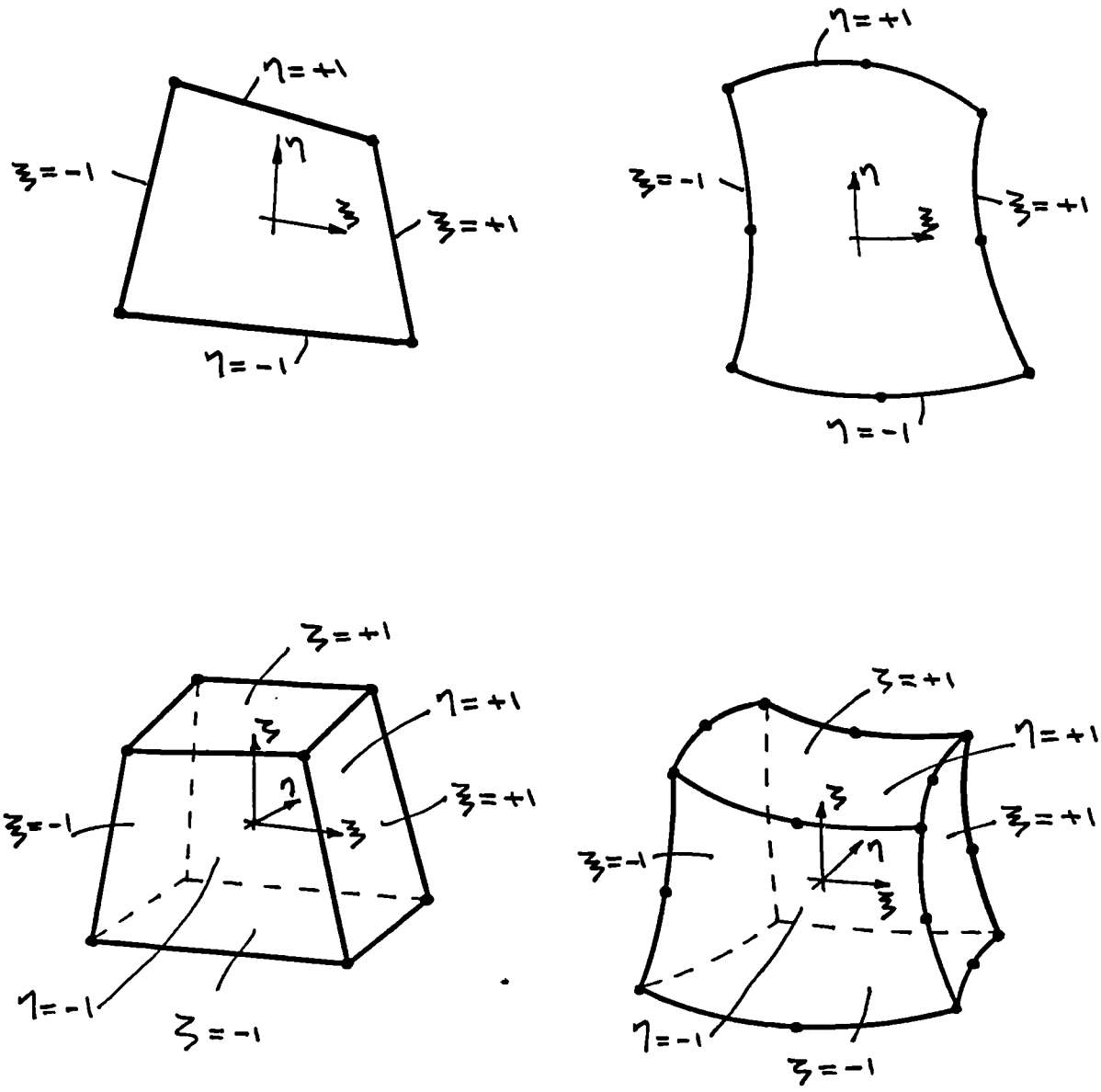


Figure 6.3 Local isoparametric coordinate systems for various elements

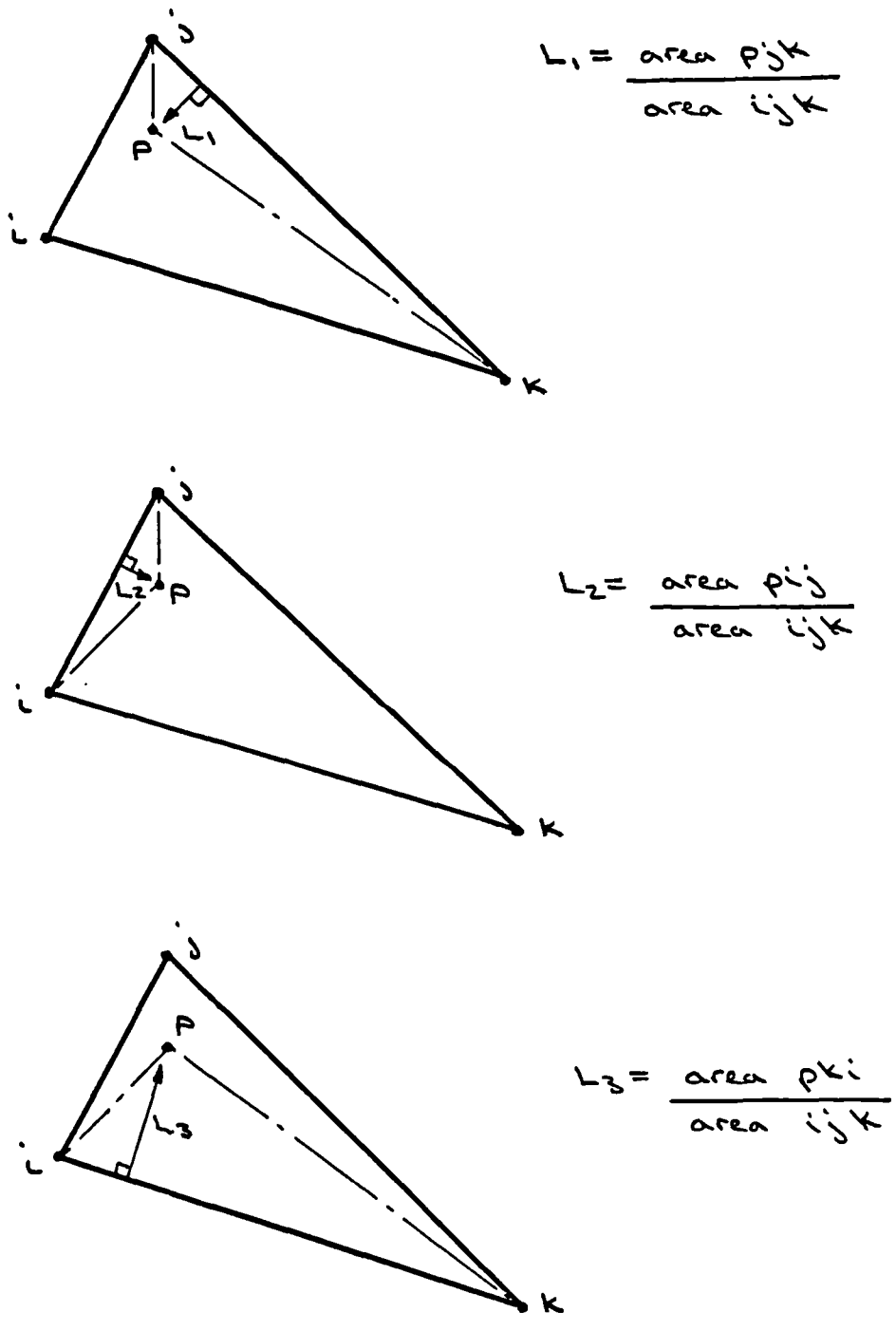


Figure 6.4 Area coordinates for triangular elements

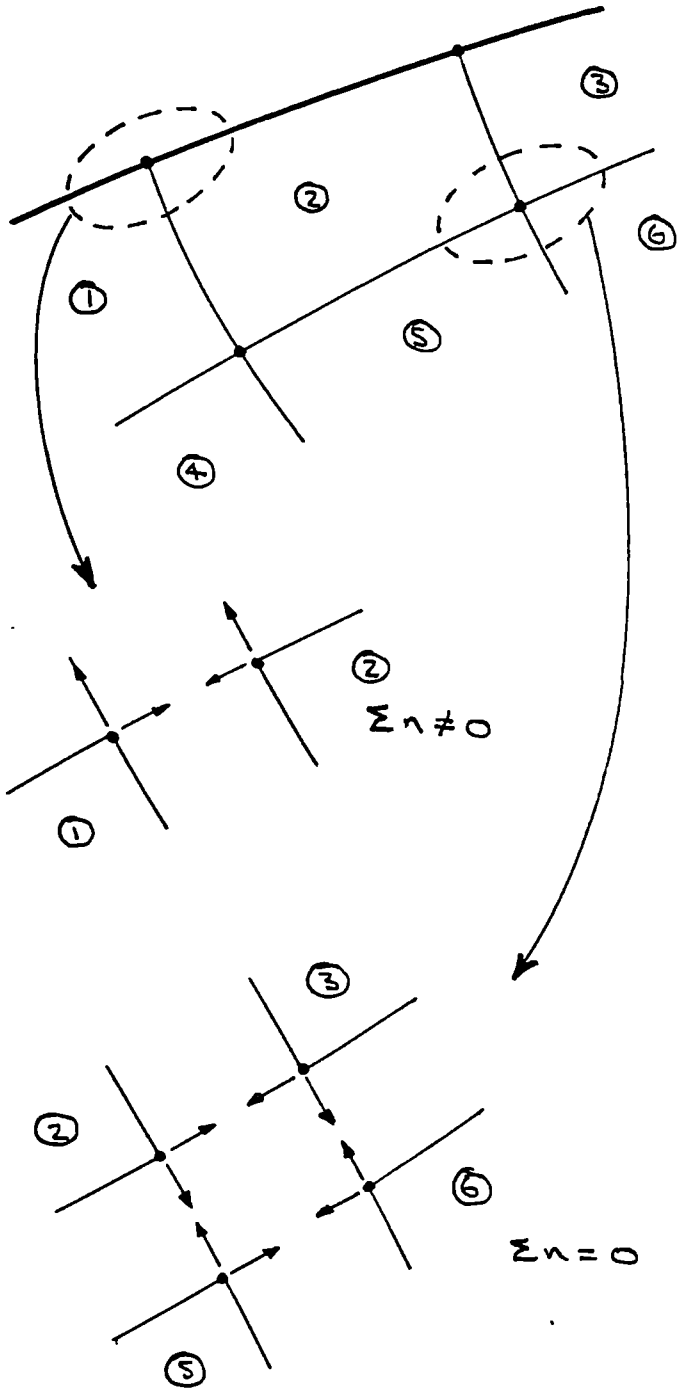


Figure 6.5 Use of element outward normals to automatically identify surfaces

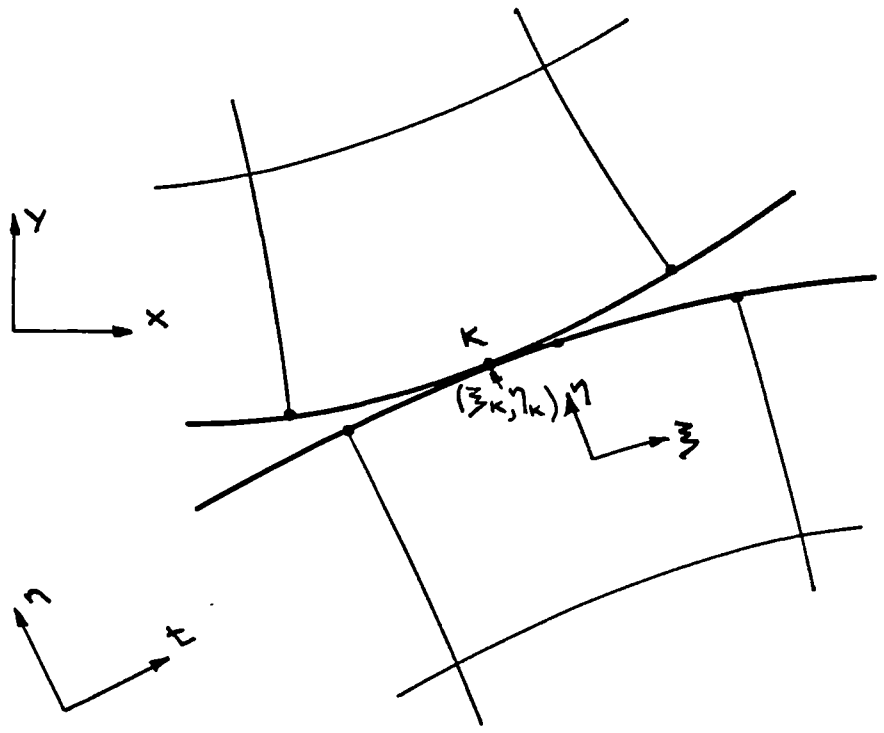


Figure 6.6 Local point of contact for a single contacting node

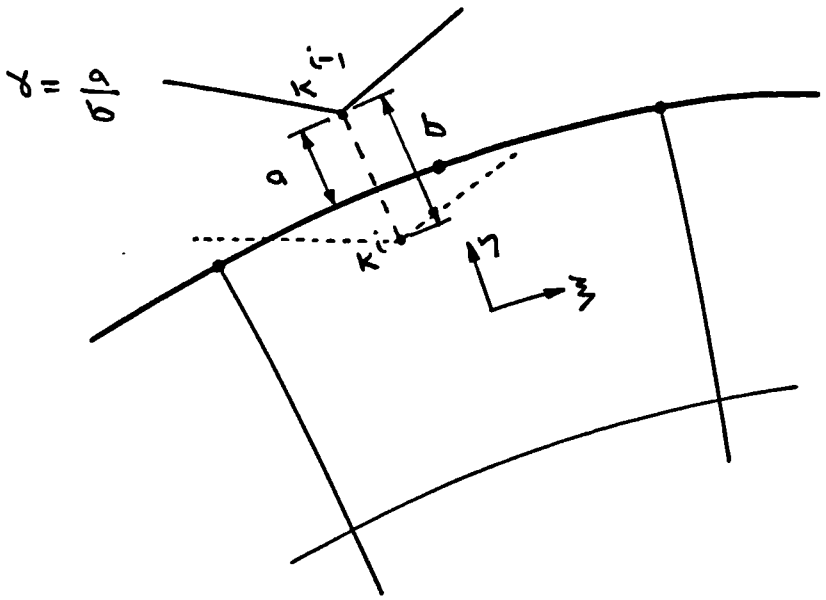


Figure 6.7 Identification of position of initial surface contact for overlapped nodes

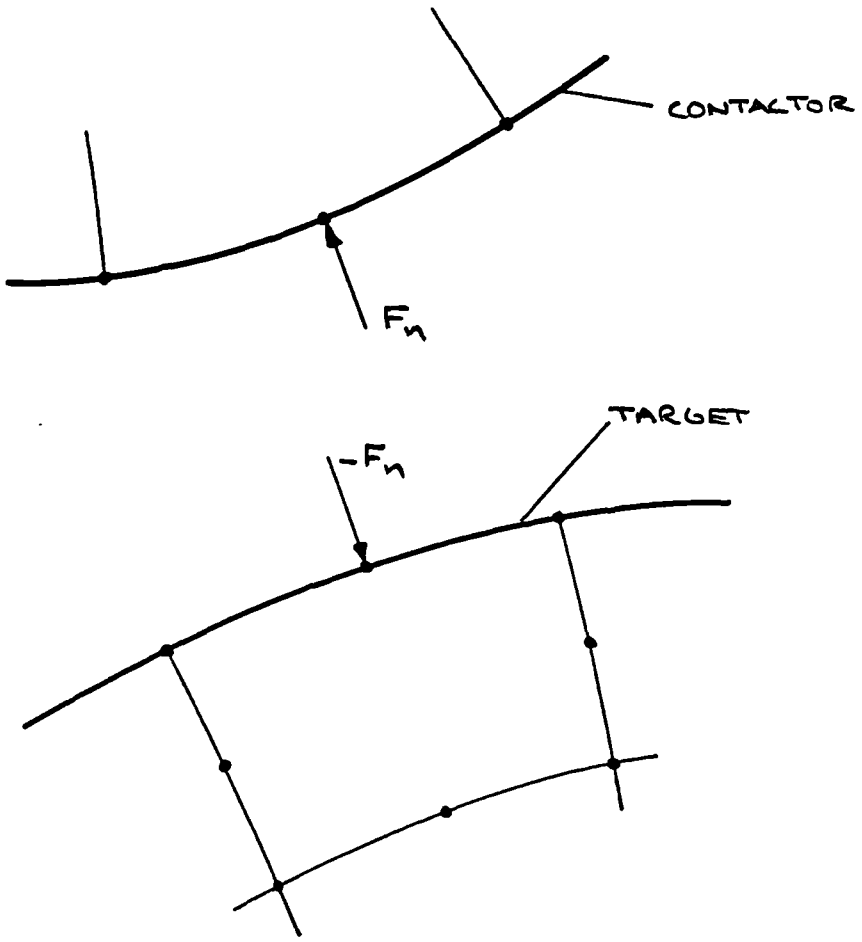


Figure 6.8 Normal contact forces between contacting bodies

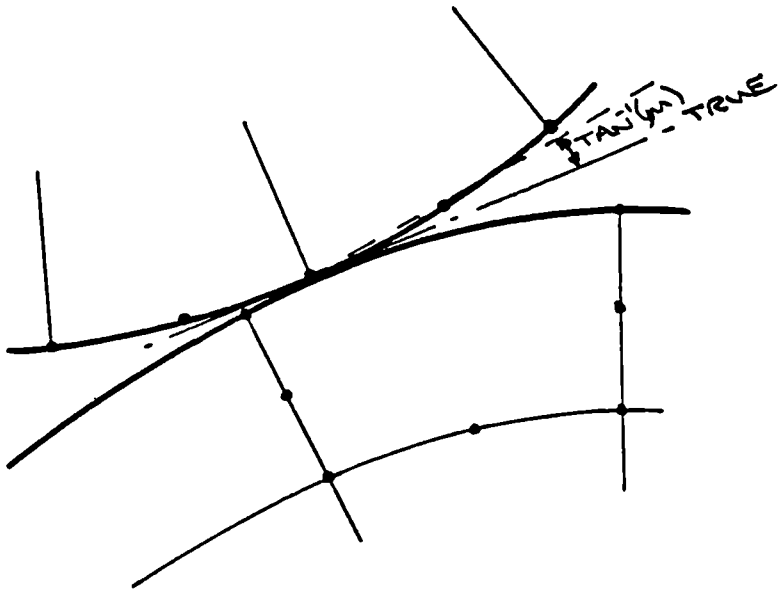
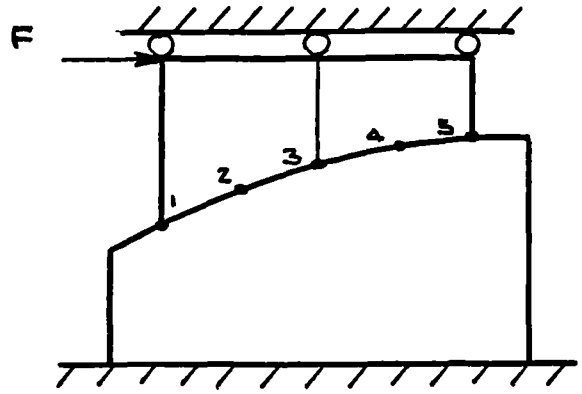


Figure 6.9 Sliding along incorrect slope due to application of [J] displacement constraint



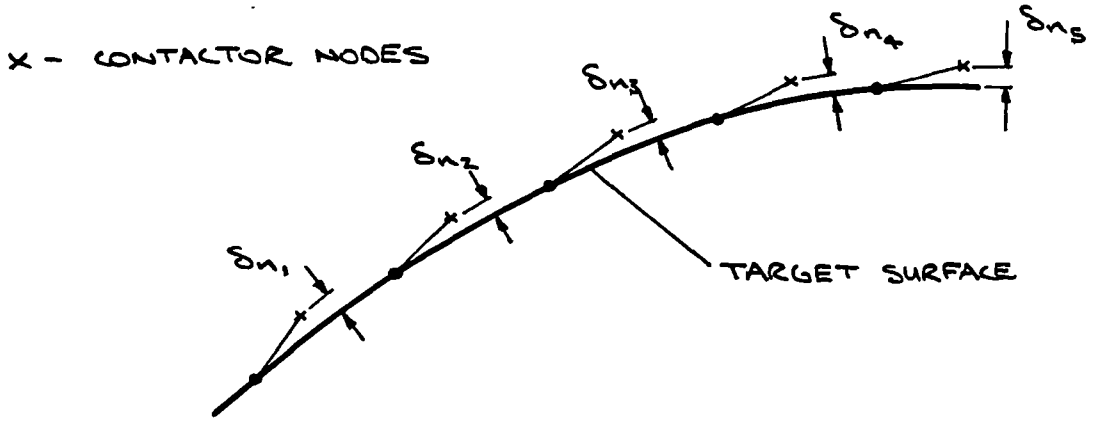


Figure 6.10 Induced normal gaps due to [J] constraint

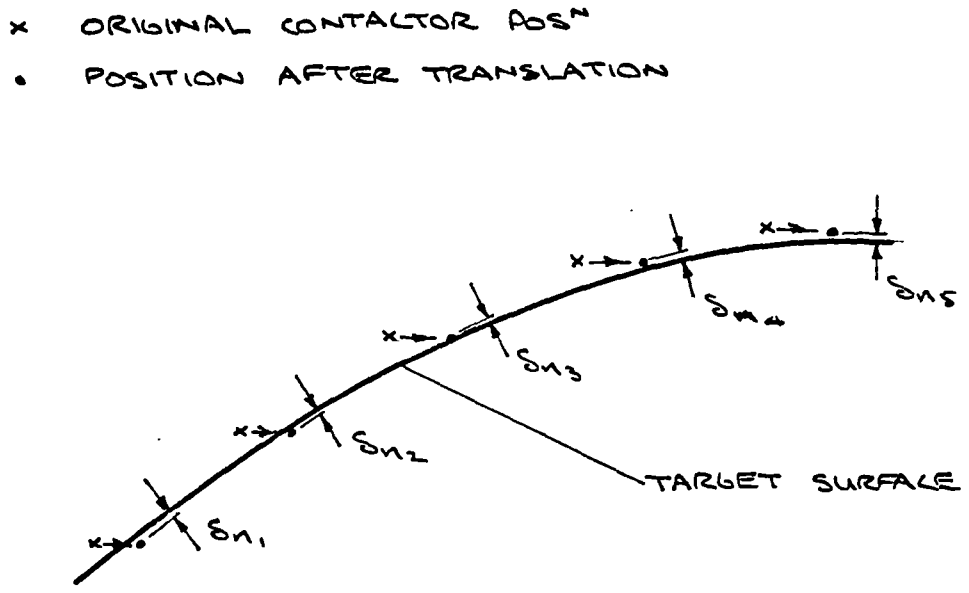


Figure 6.11 Translated contactor nodes

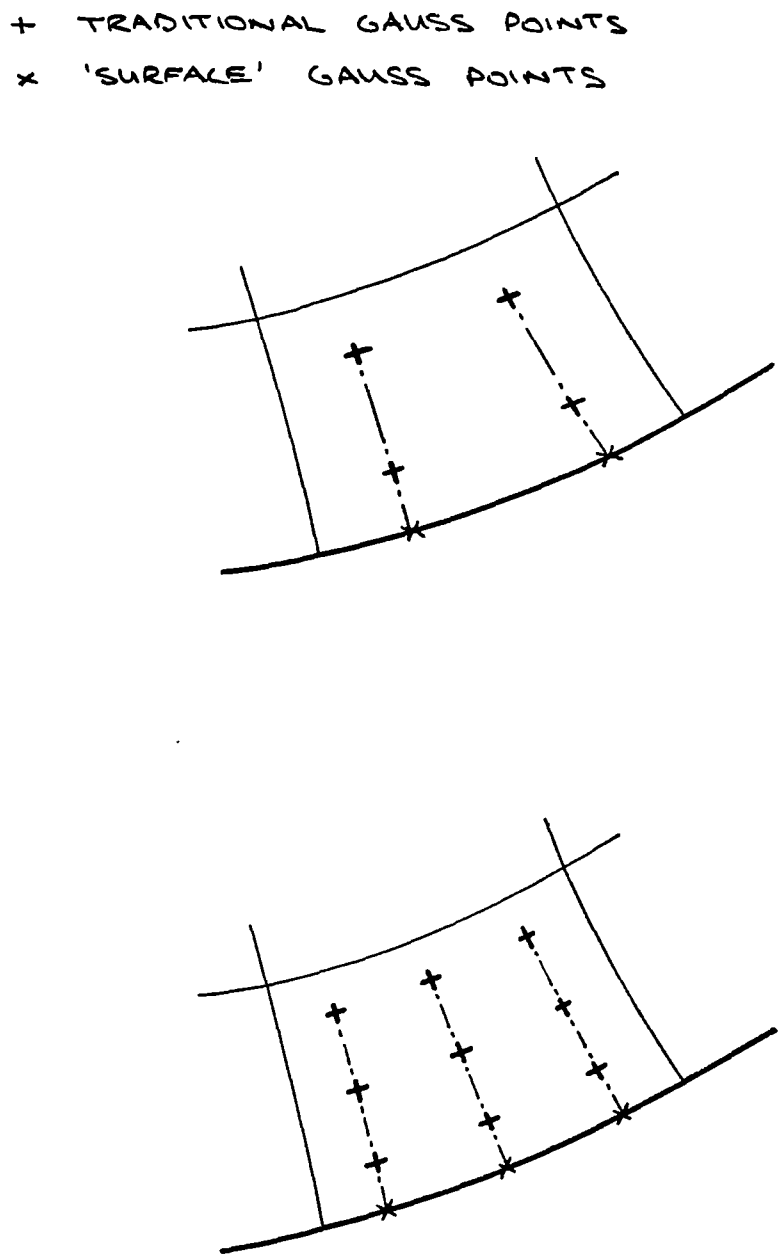


Figure 6.12 Definition of surface Gauss points

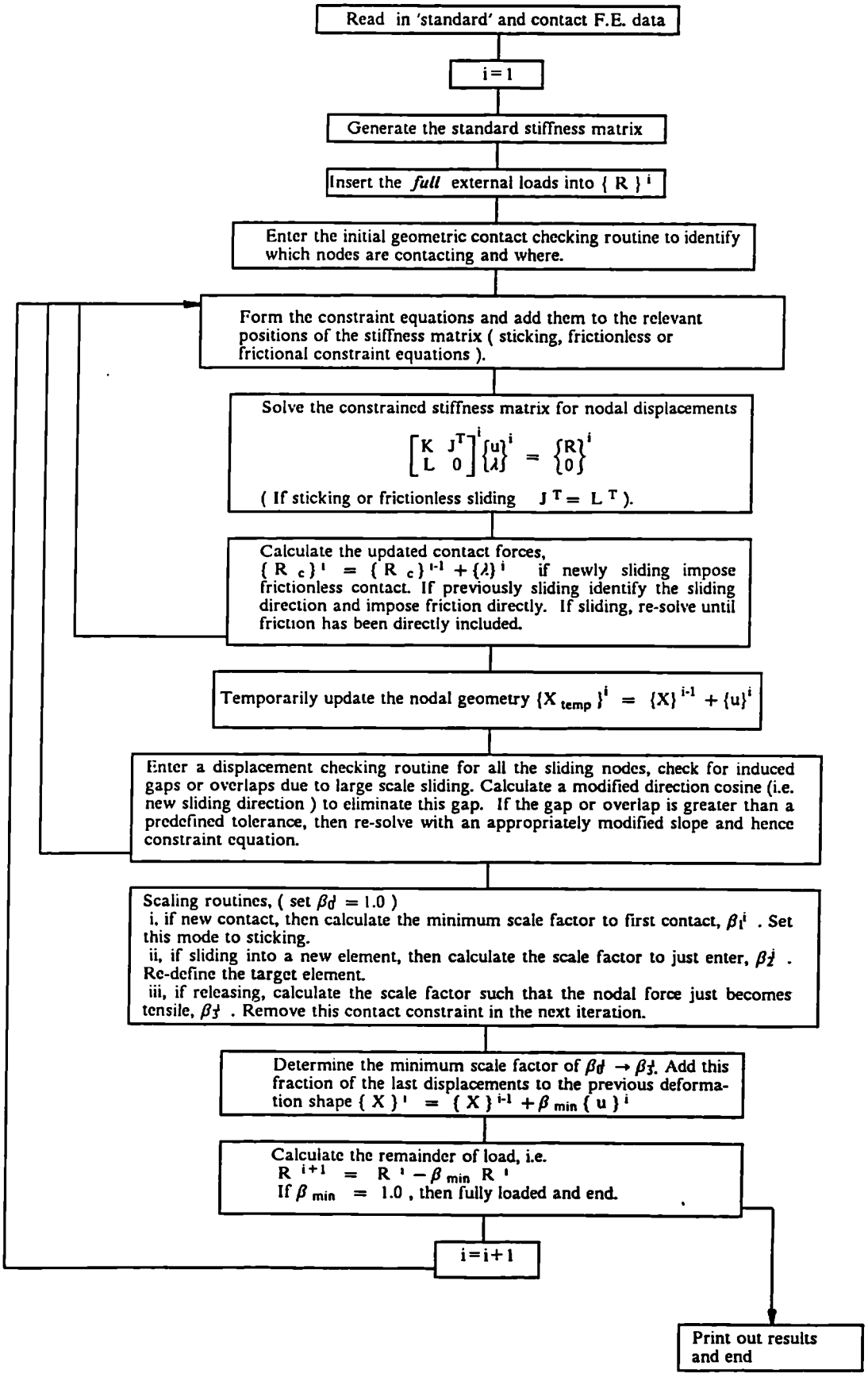


Figure 7.1 Direct inclusion of friction algorithm

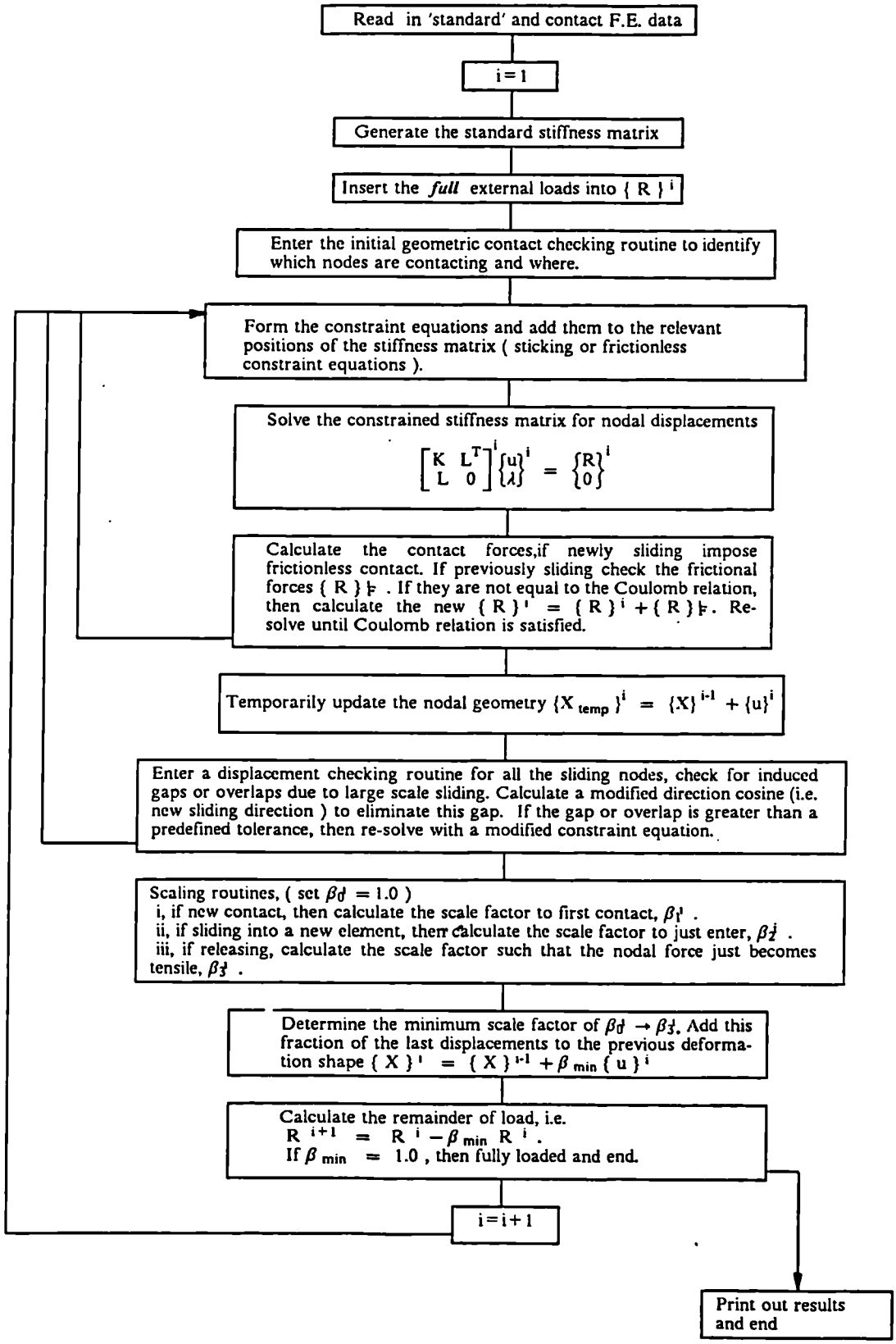


Figure 7.2 Post inclusion of friction algorithm

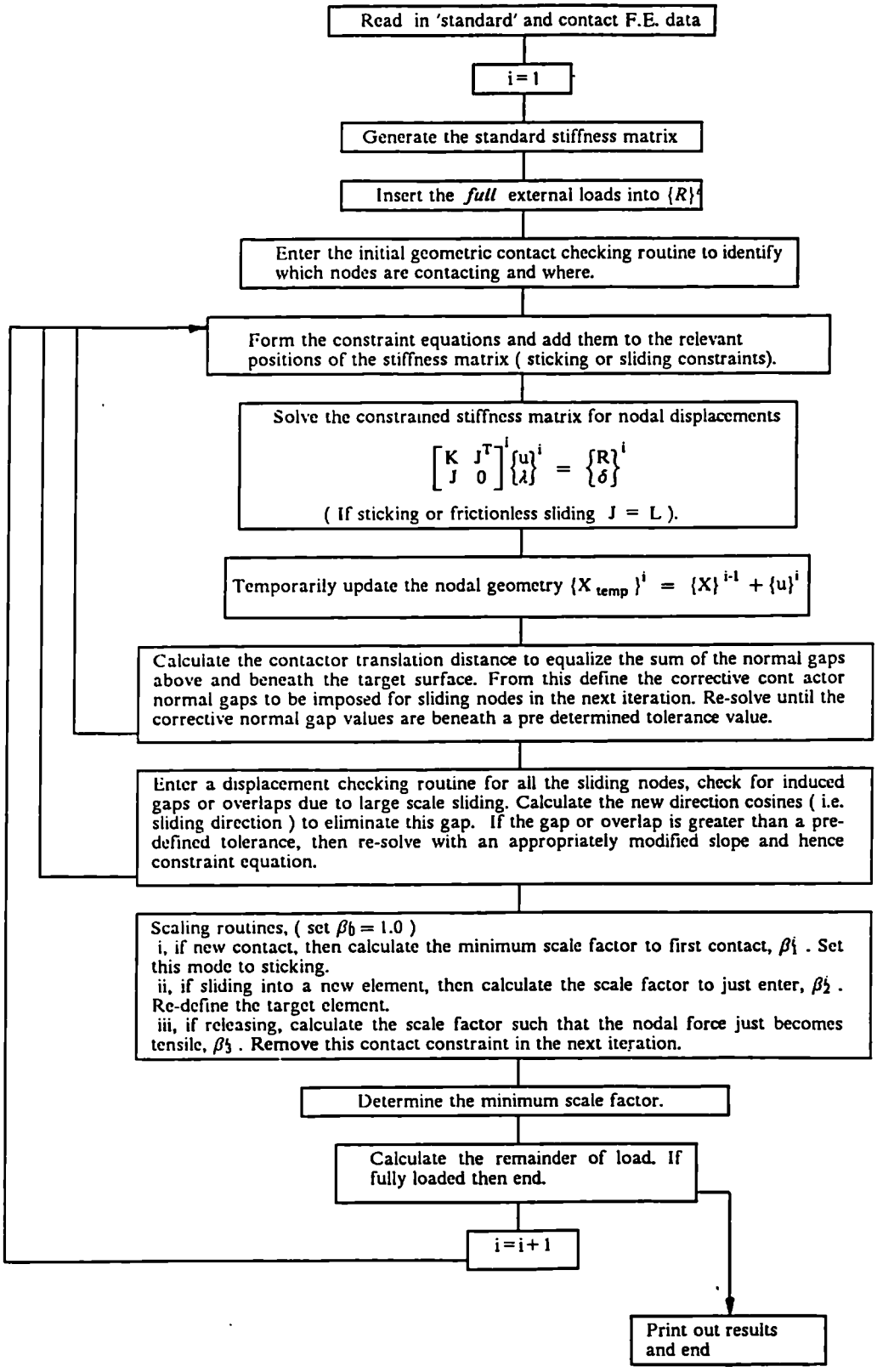


Figure 7.3 Post Inclusion of normal gaps algorithm

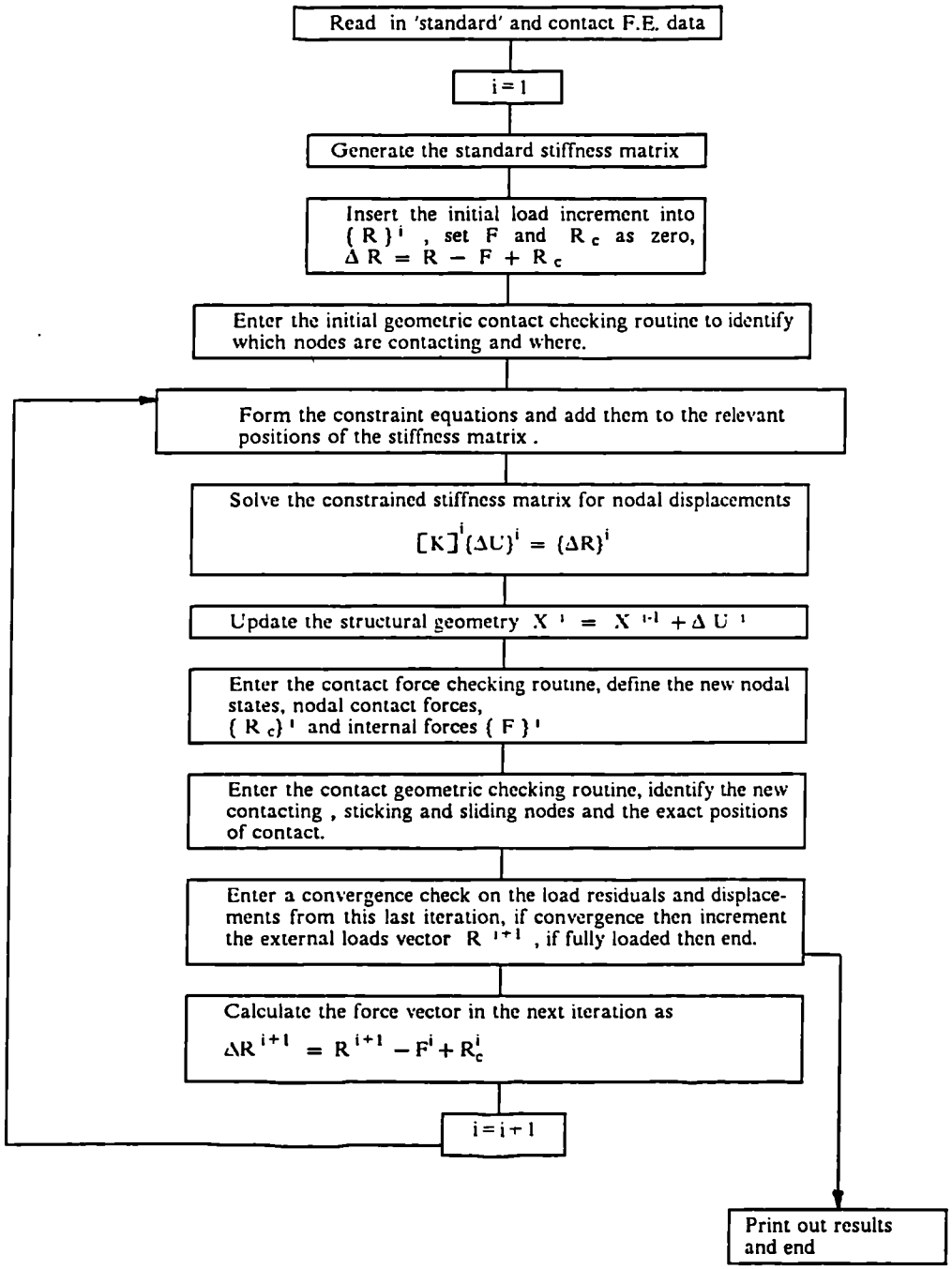


Figure 7.4 Incremental loading algorithm

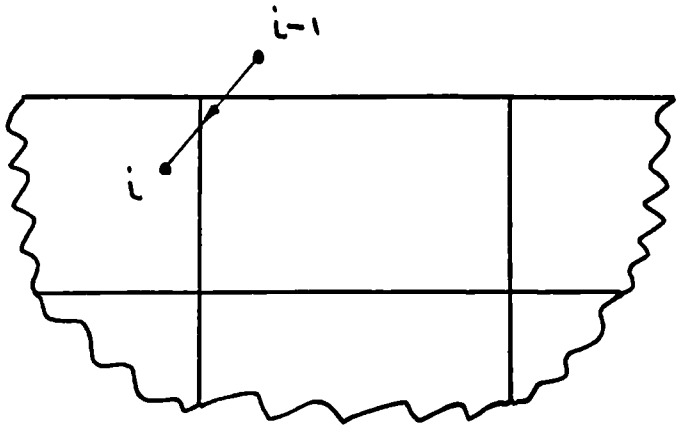


Figure 8.1a Example of contact identification difficulty

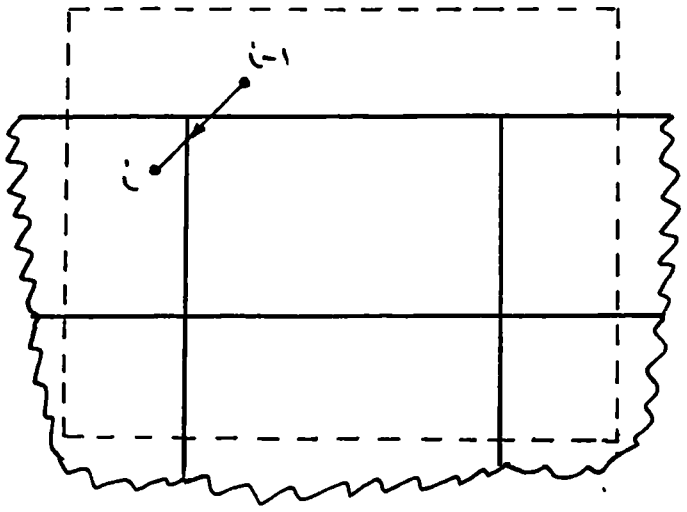


Figure 8.1b Definition of envelope to overcome contact identification difficulty

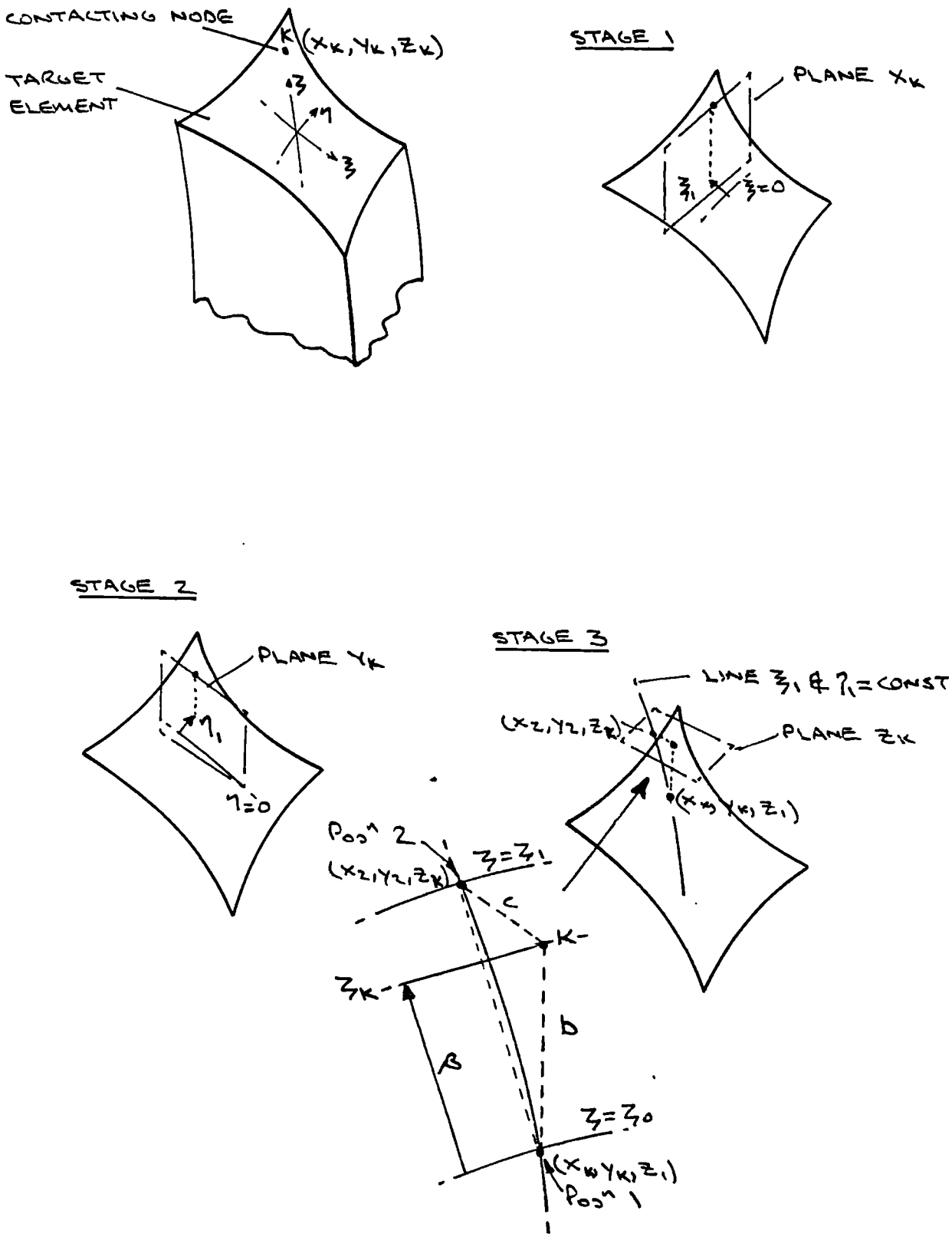
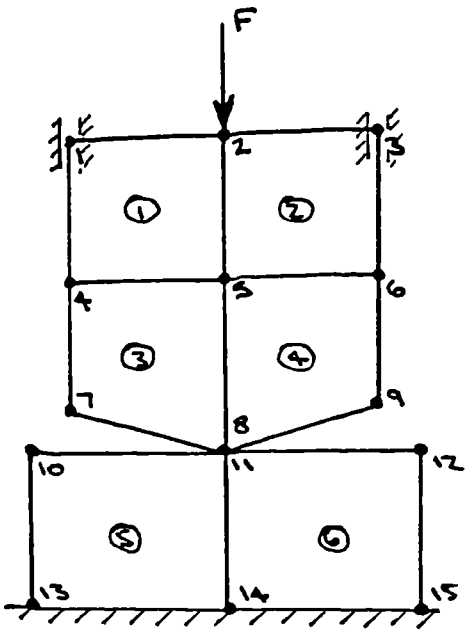


Figure 8.2 Procedure for calculation of local coordinates



dof	element number						
	1	2	3	4	5	6	7
1	*						
2	*	*					*
3		*					
4	*		*				
5	*	*	*	*			
6		*		*			
7			*				*
8			*	*			*
9				*			*
10					*		*
11					*	*	*
12						*	*
13					*		
14					*	*	
15						*	

Figure 9.1 Typical element v d.o.f. table to permit contact frontal elimination

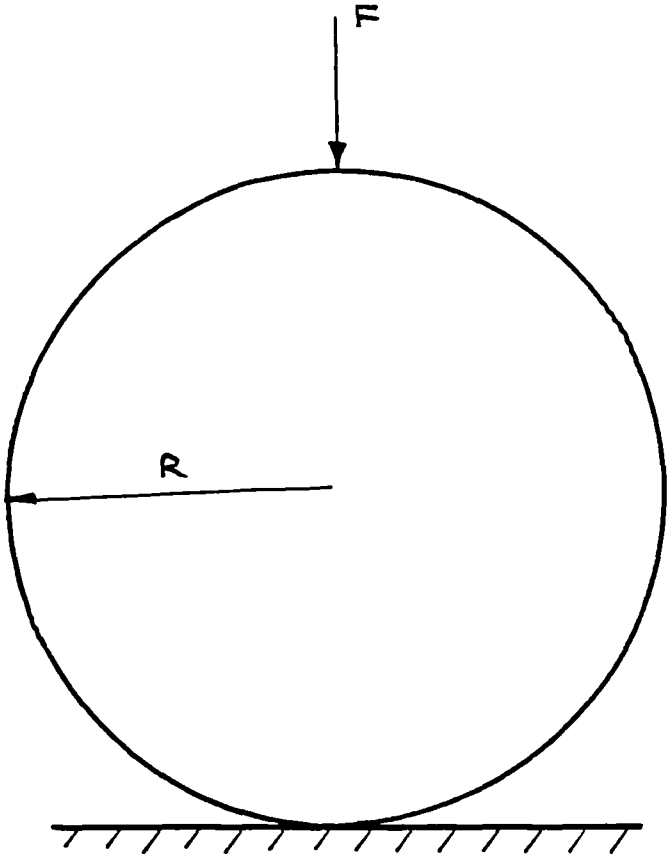


Figure 10.1 Hertz contact problem

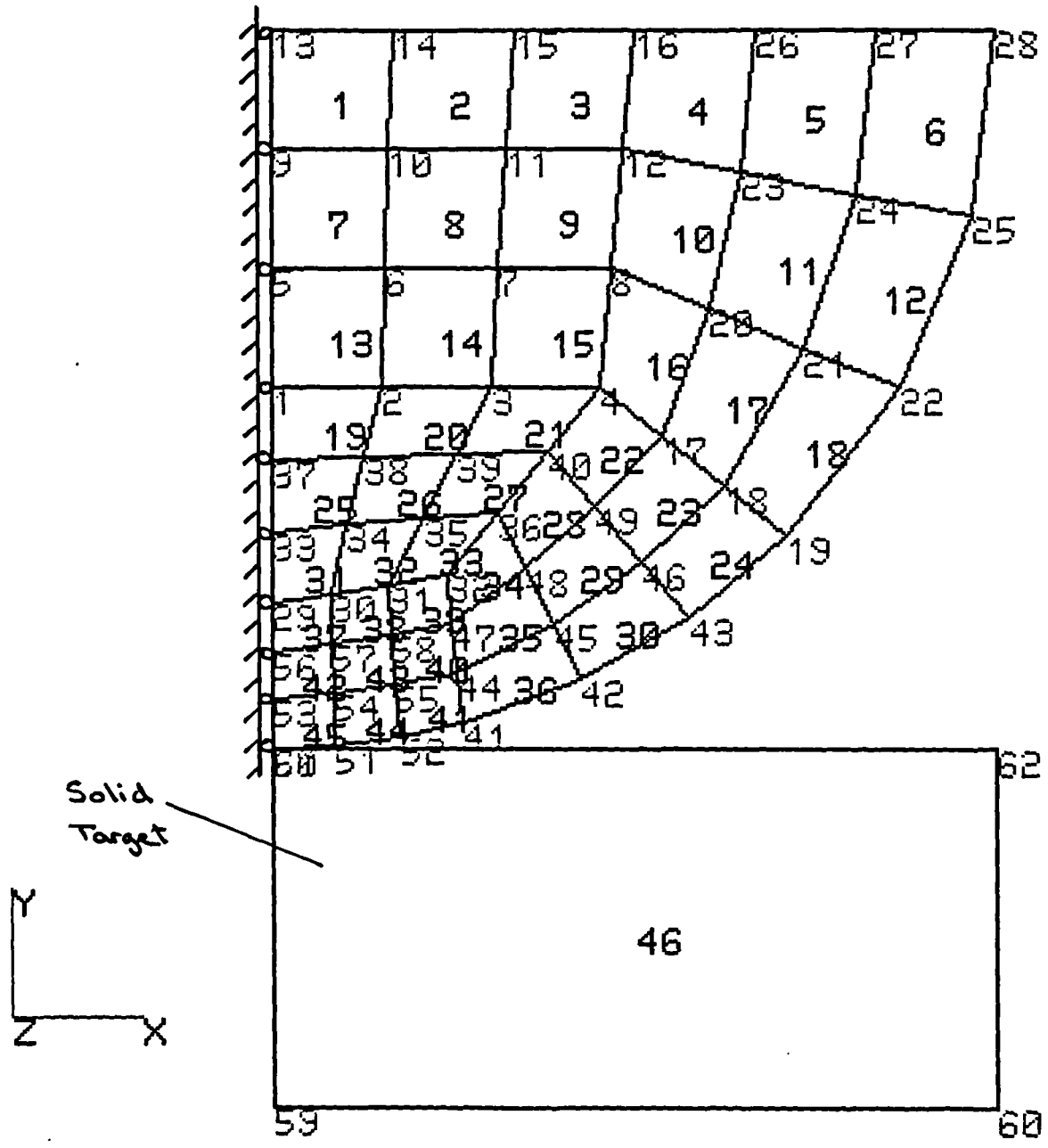


Figure 10.2 Hertz contact finite element mesh

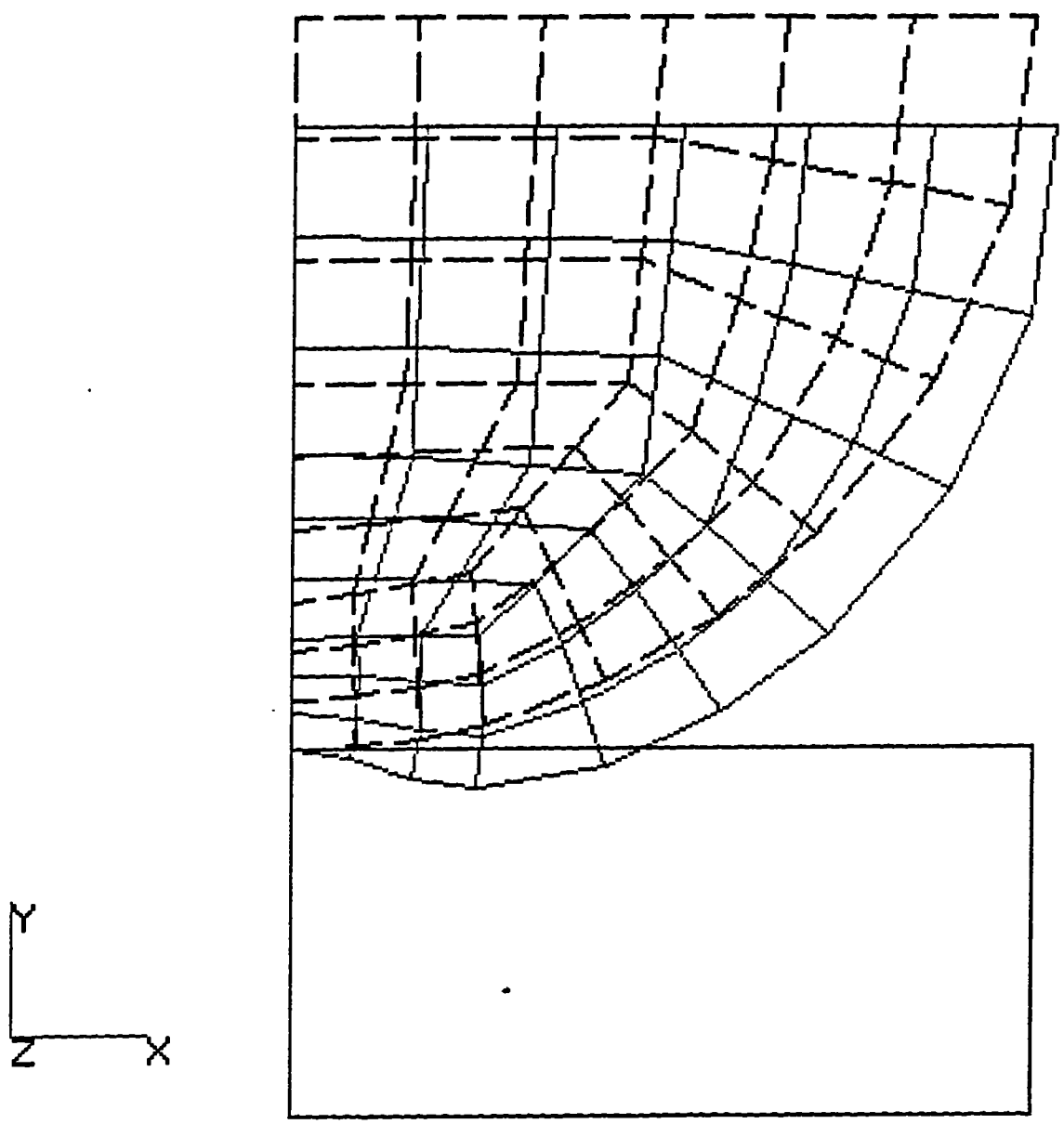


Figure 10.3 Displacement plot of the F.E. Hertz contact results

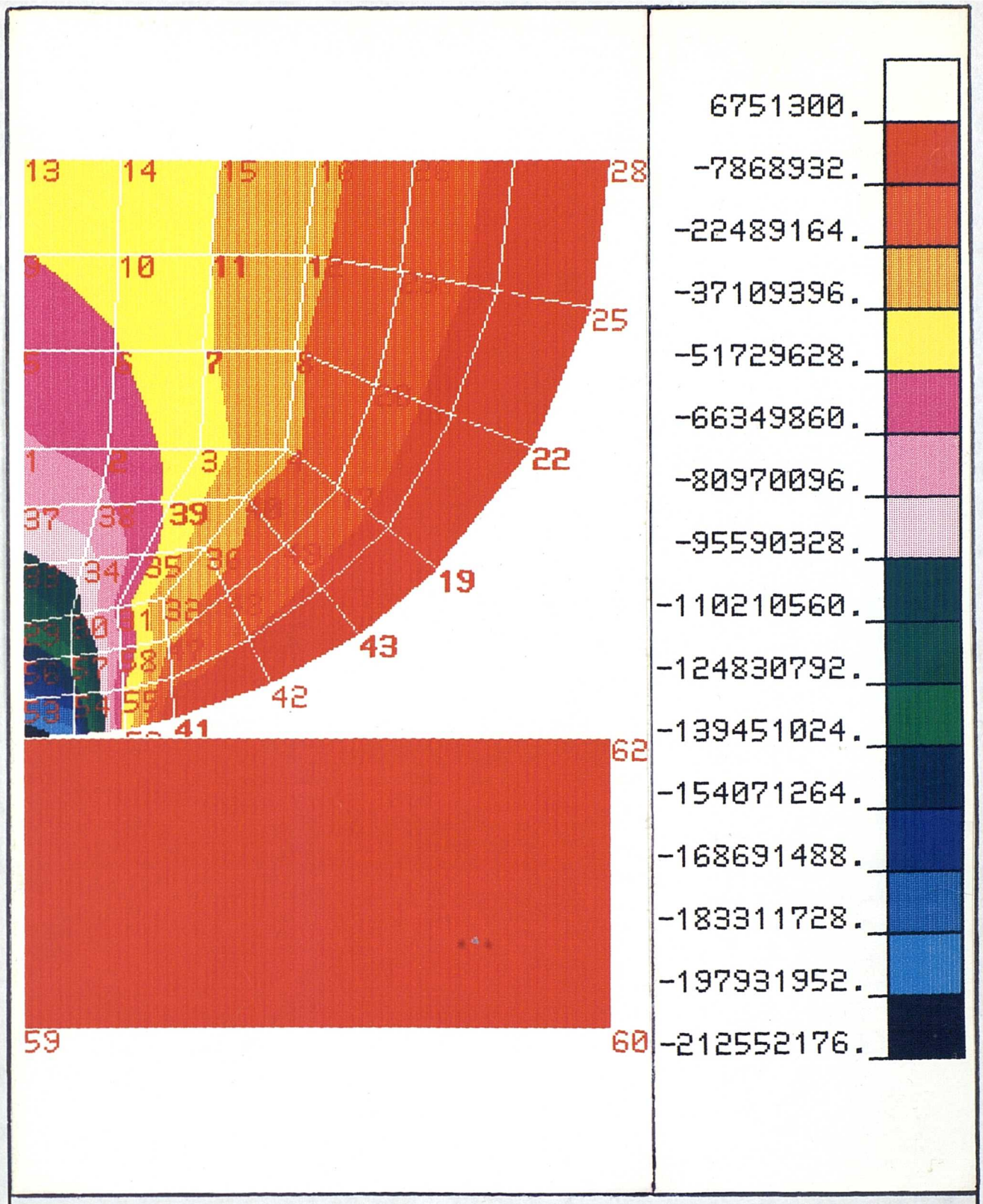


Figure 10.4 σ_y stress results for the linear Hertz mesh

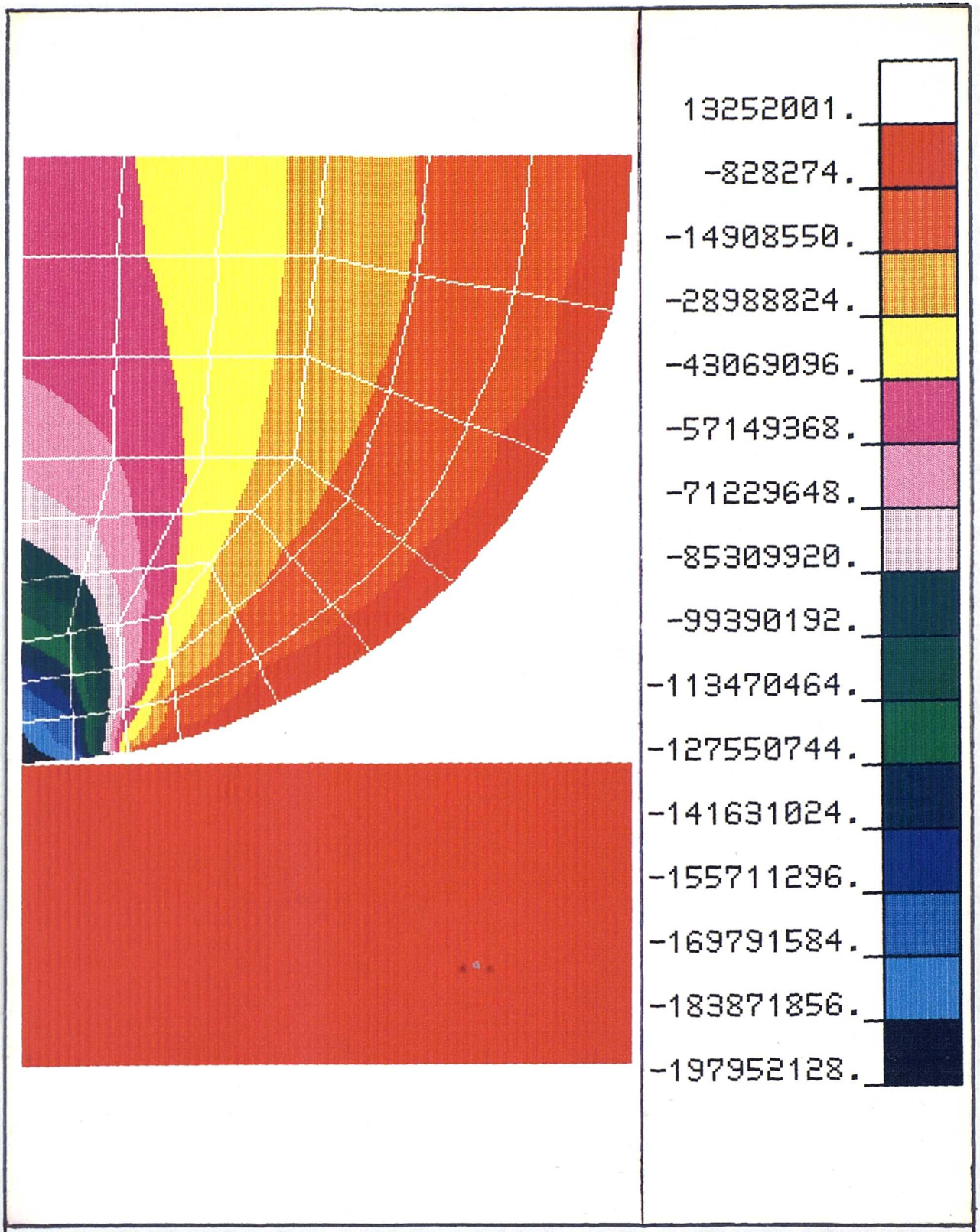


Figure 10.5 σ_y stress results for the quadratic Hertz mesh

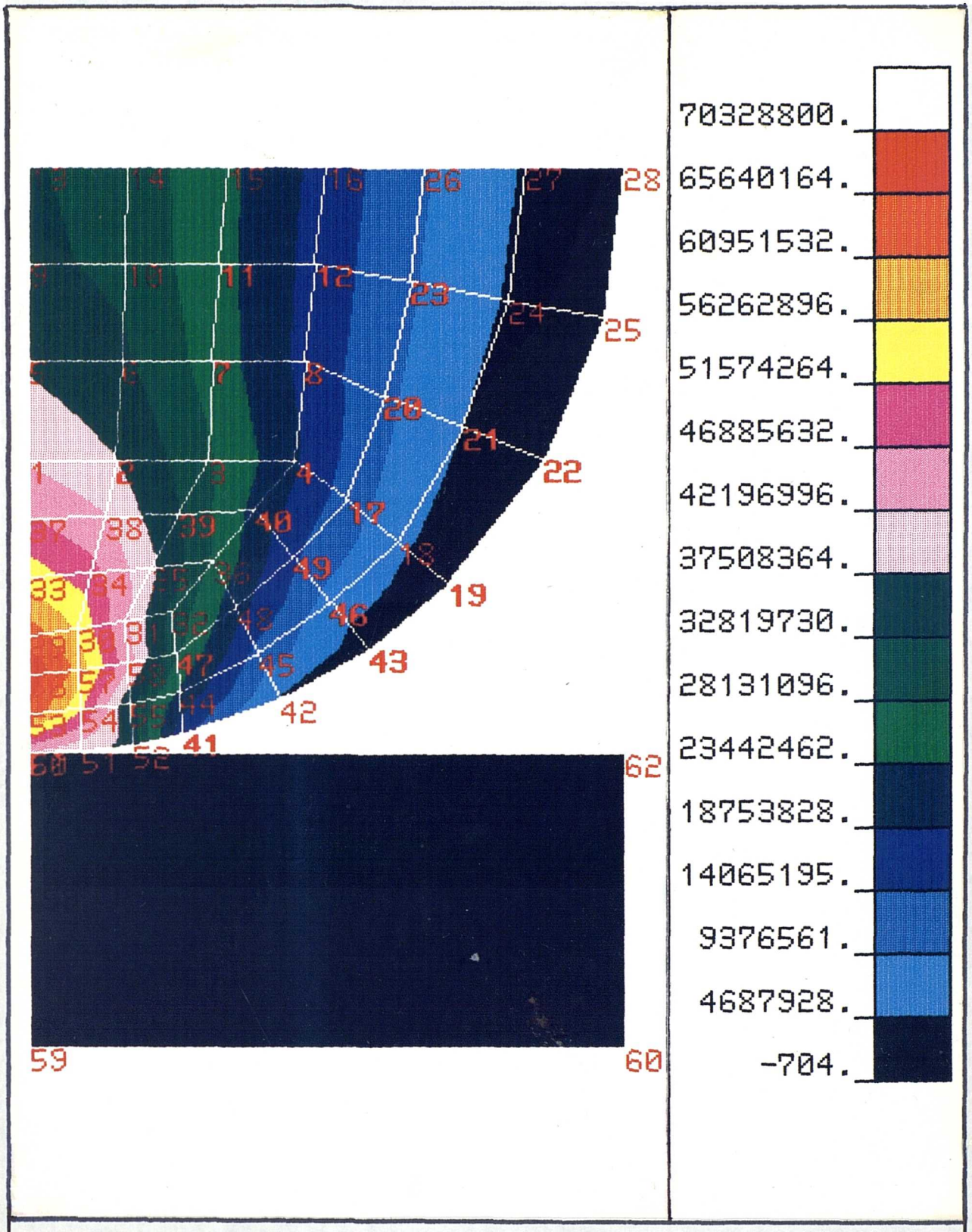


Figure 10.6 Principle shear stress results for the linear Hertz mesh

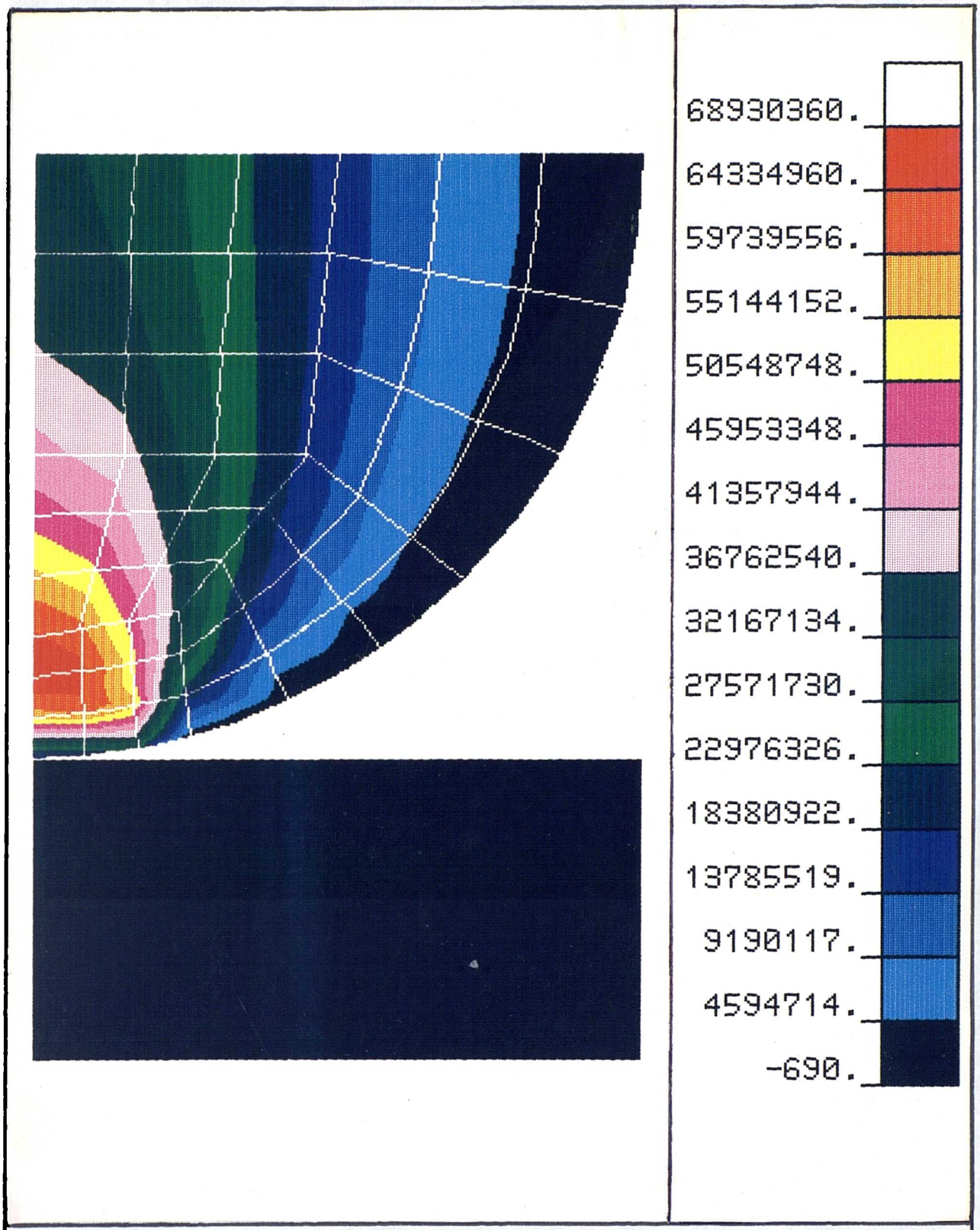


Figure 10.7 Principle shear stress results for the quadratic Hertz mesh

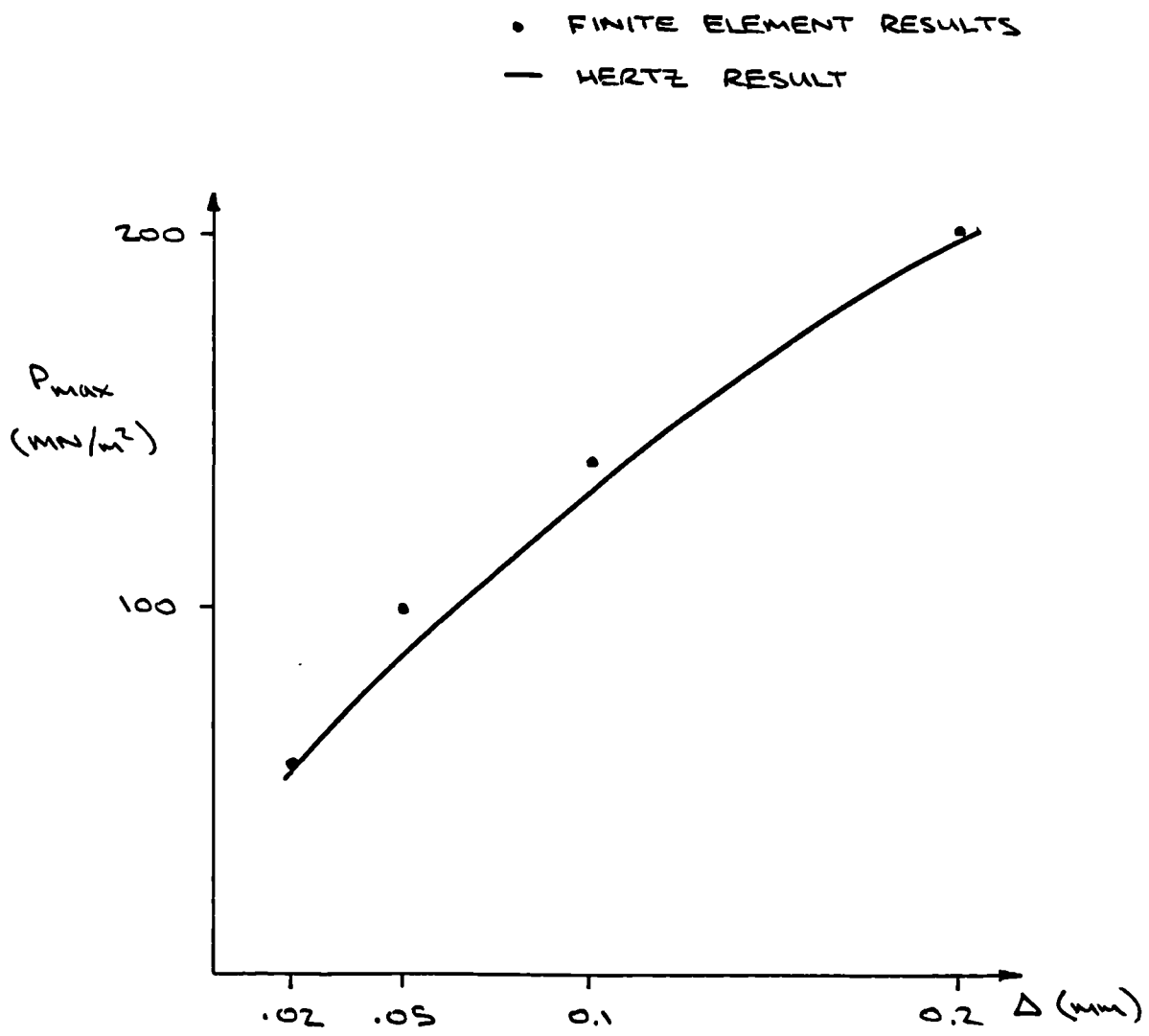


Figure 10.8 Comparison of maximum contact pressures between F.E and Hertz results

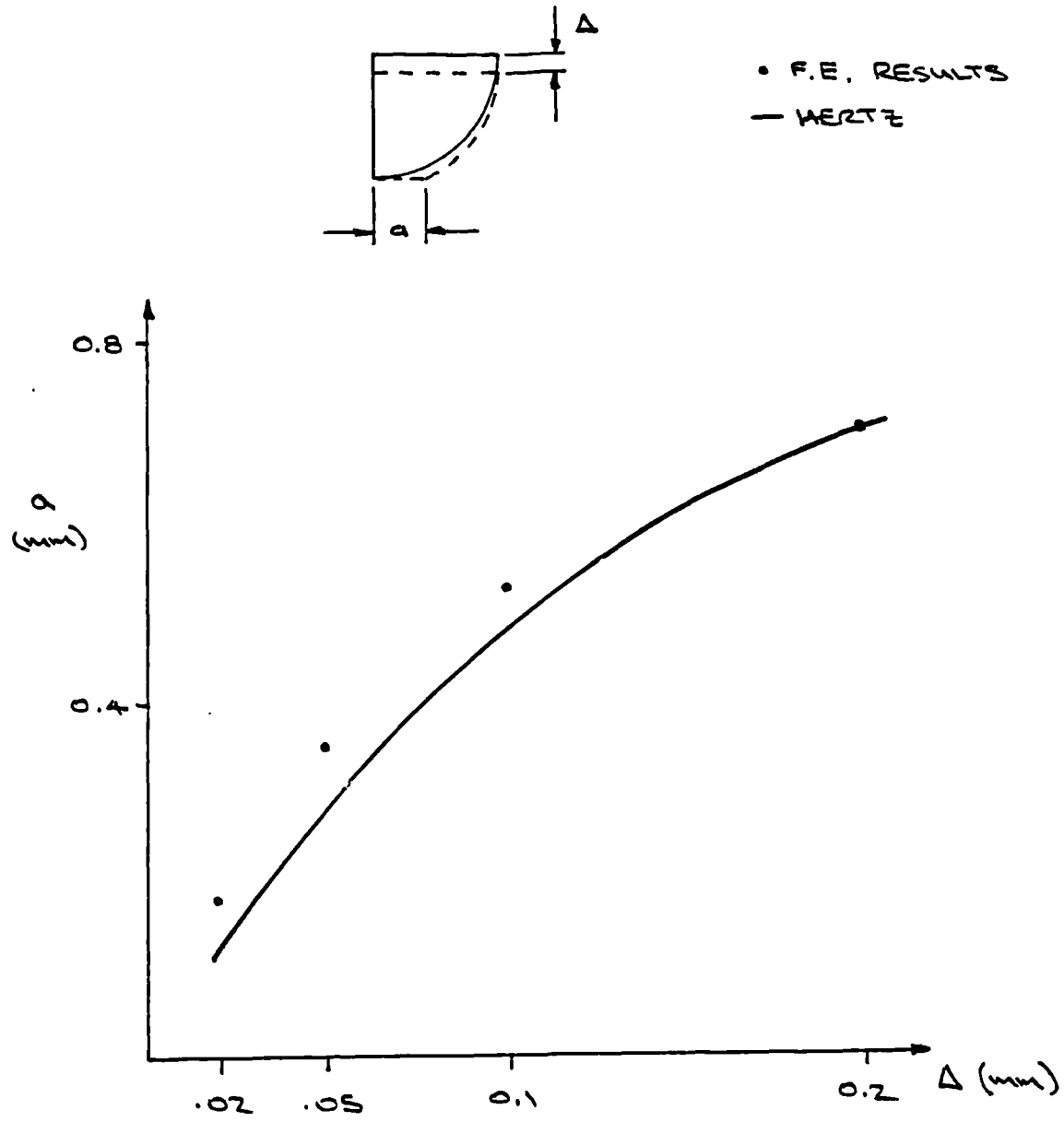


Figure 10.9 Comparison of contact areas between F.E. and Hertz results

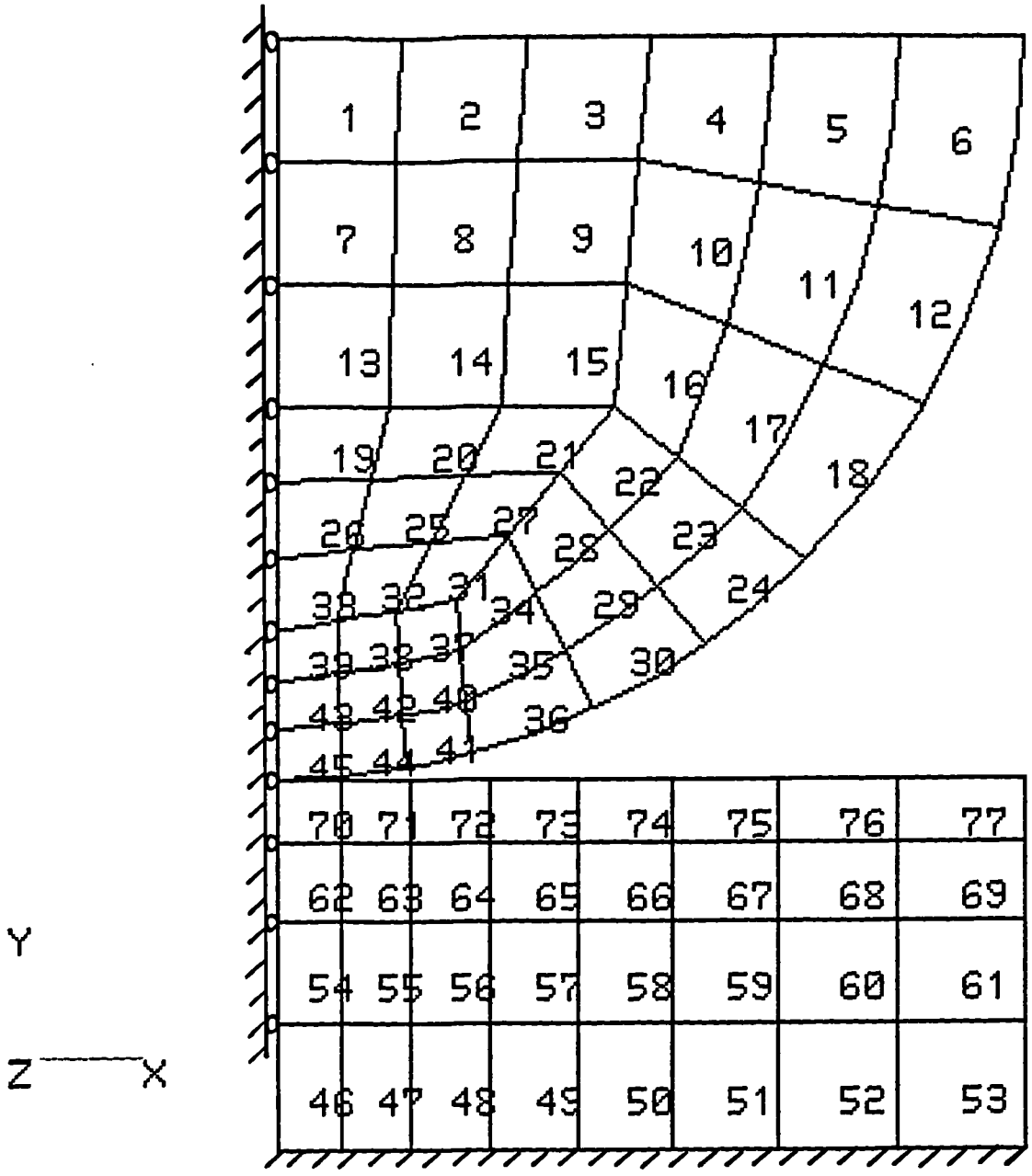


Figure 10.10 Finite element mesh for the Hertz contact problem with a non-rigid base

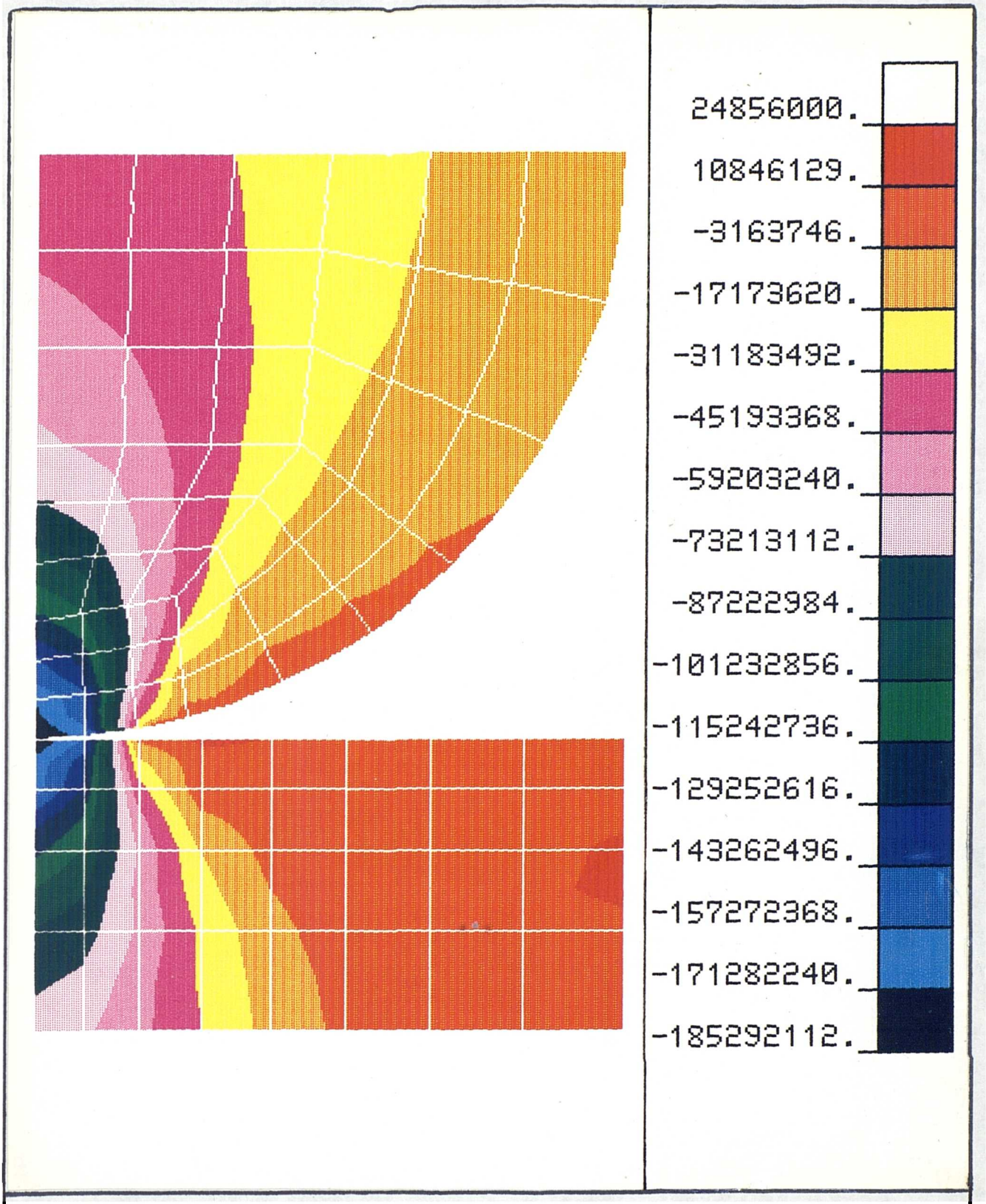


Figure 10.11 σ_y stress results for the deformable base Hertz problem

E=2060

ν=0.3

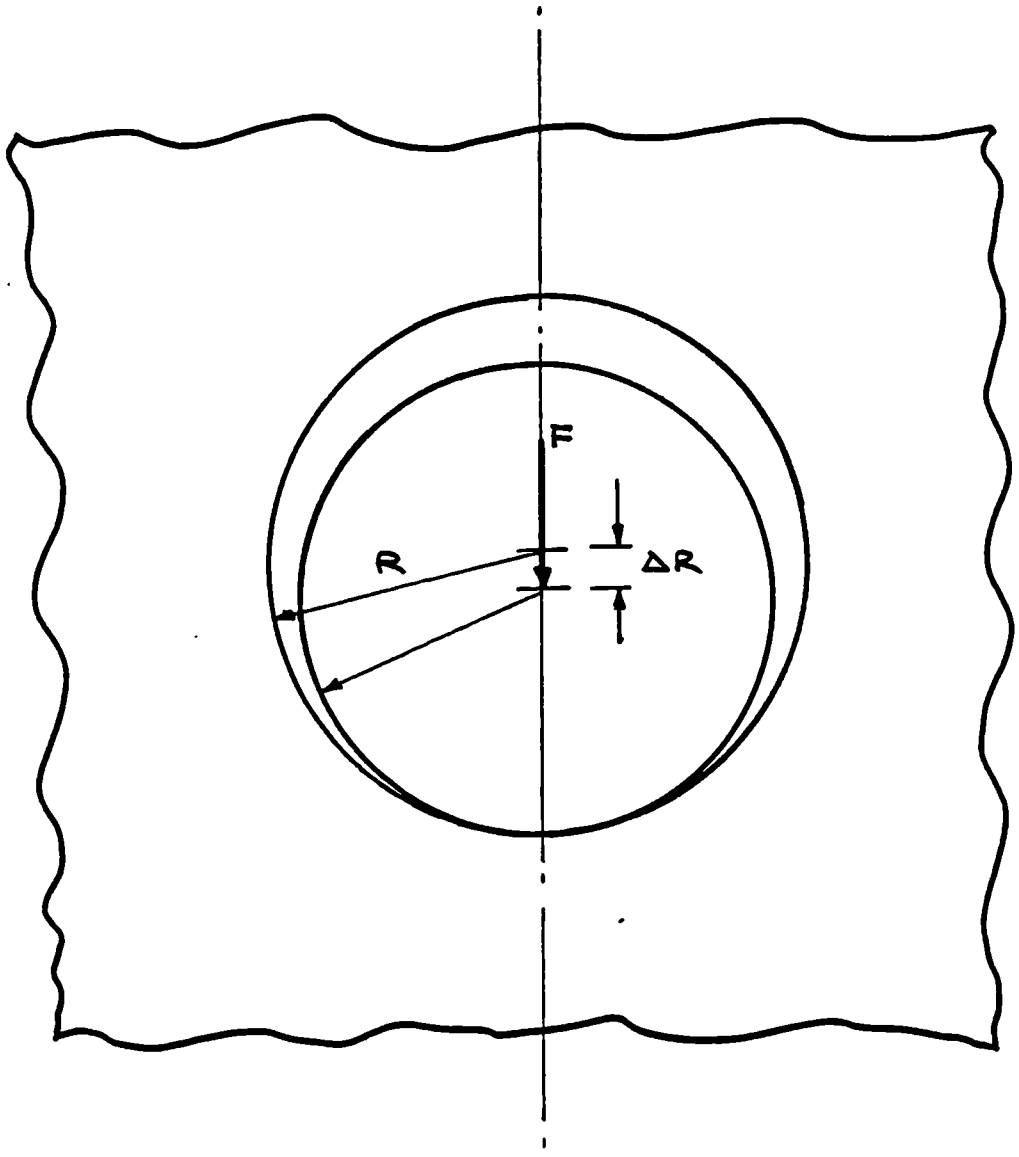


Figure 10.12 Pin in a hole contact problem

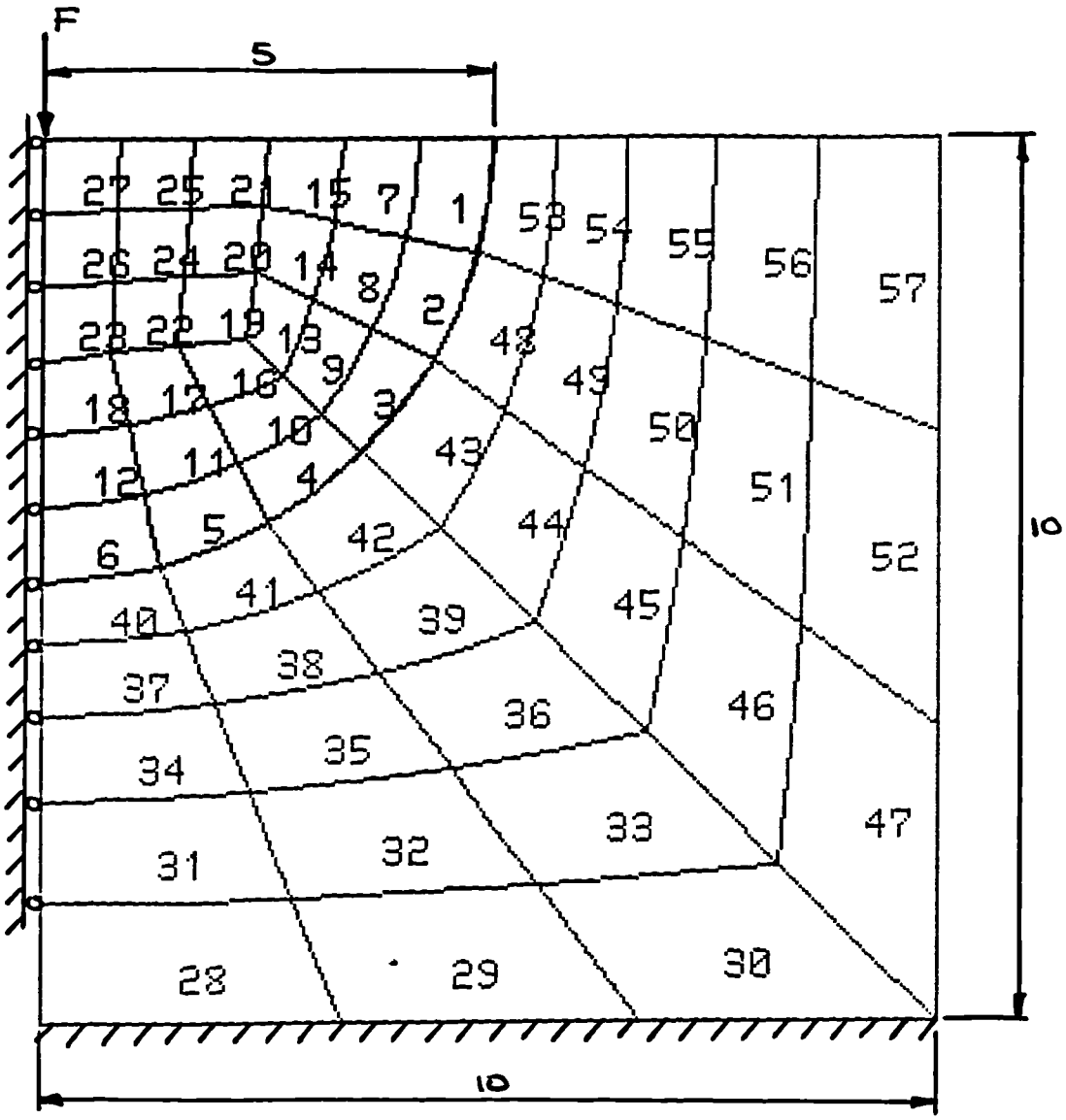


Figure 10.13 Finite element mesh for the pin in a hole contact problem

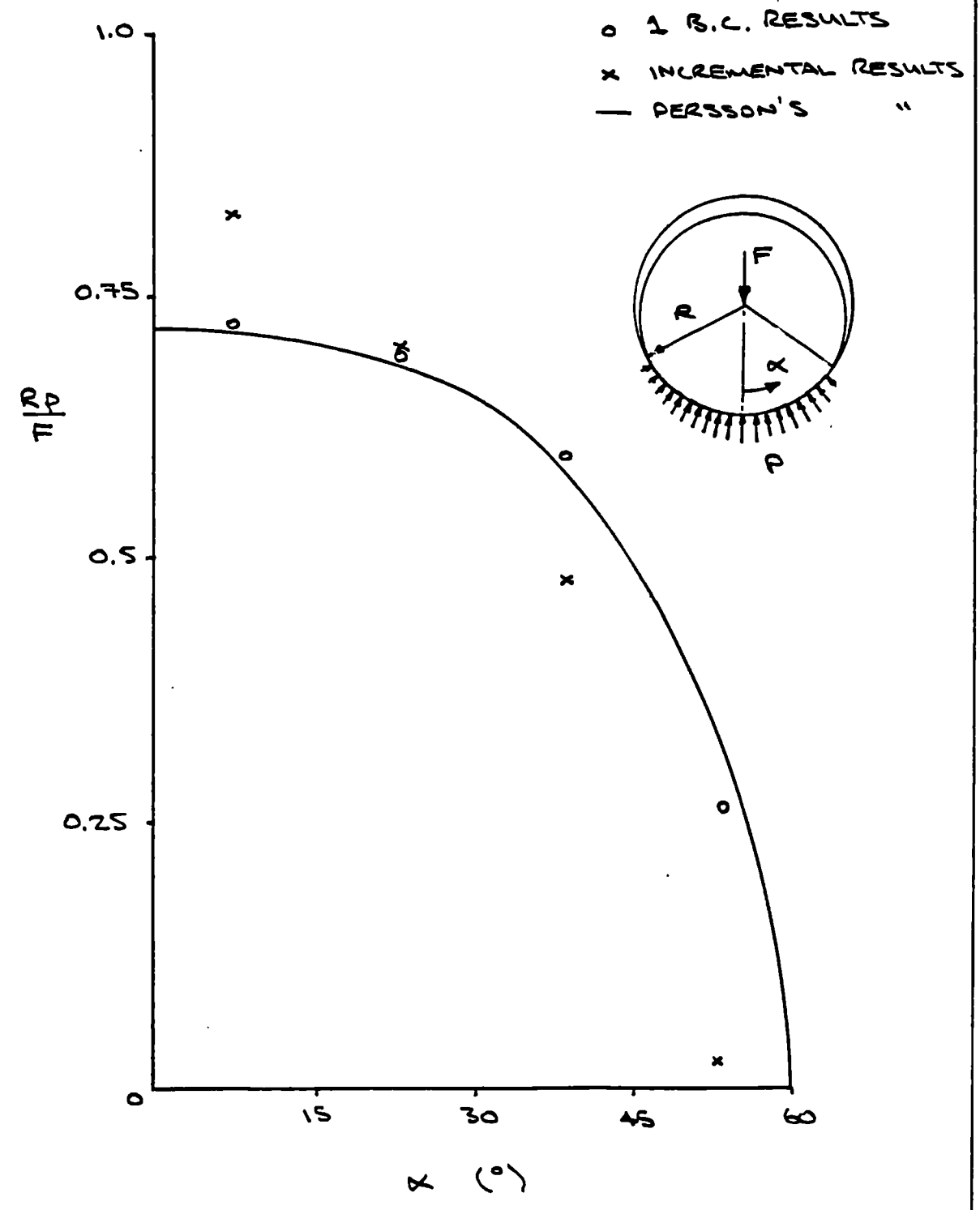


Figure 10.14 Comparison of contact pressures for pin in a hole

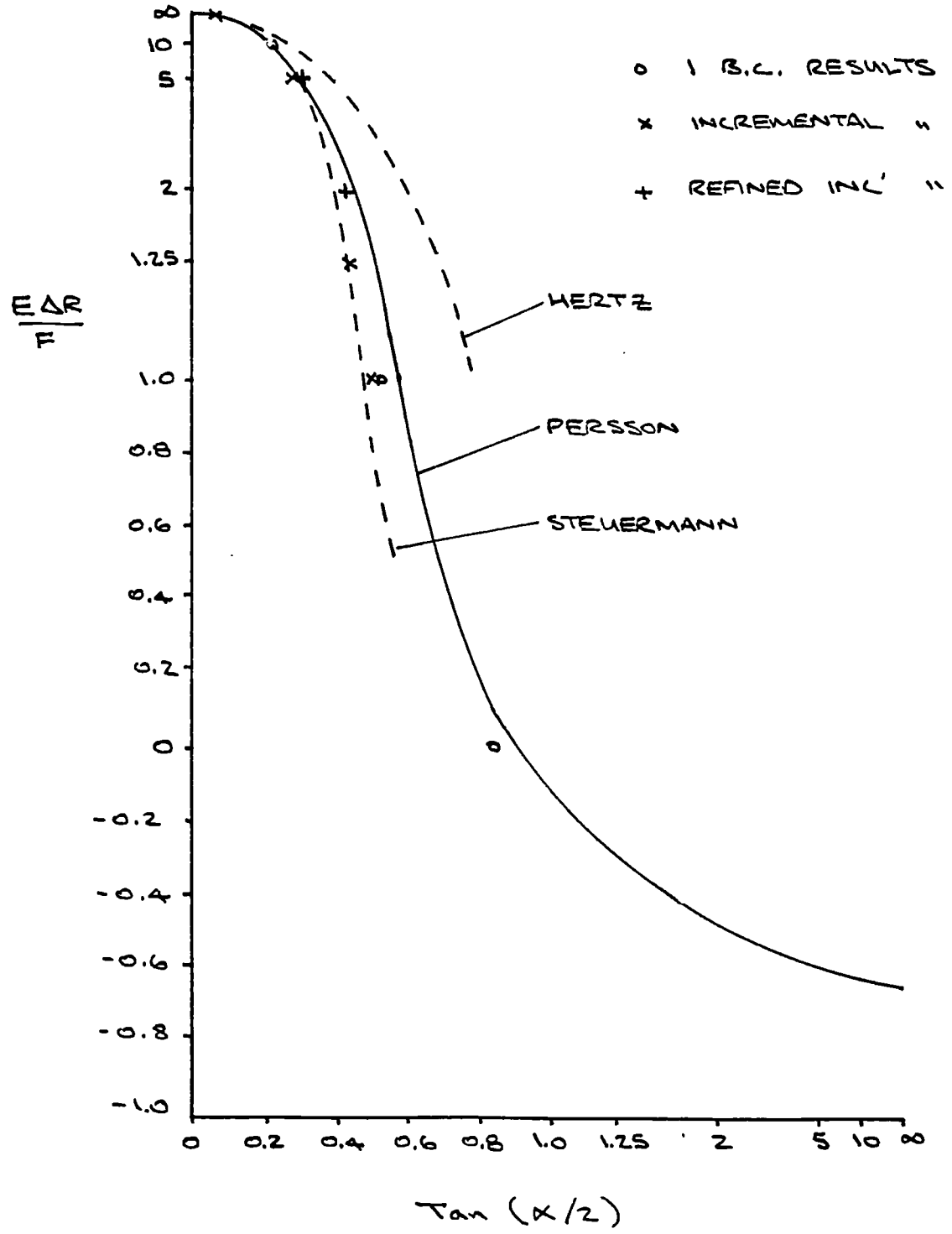


Figure 10.15 Comparison of contact areas for pin in a hole

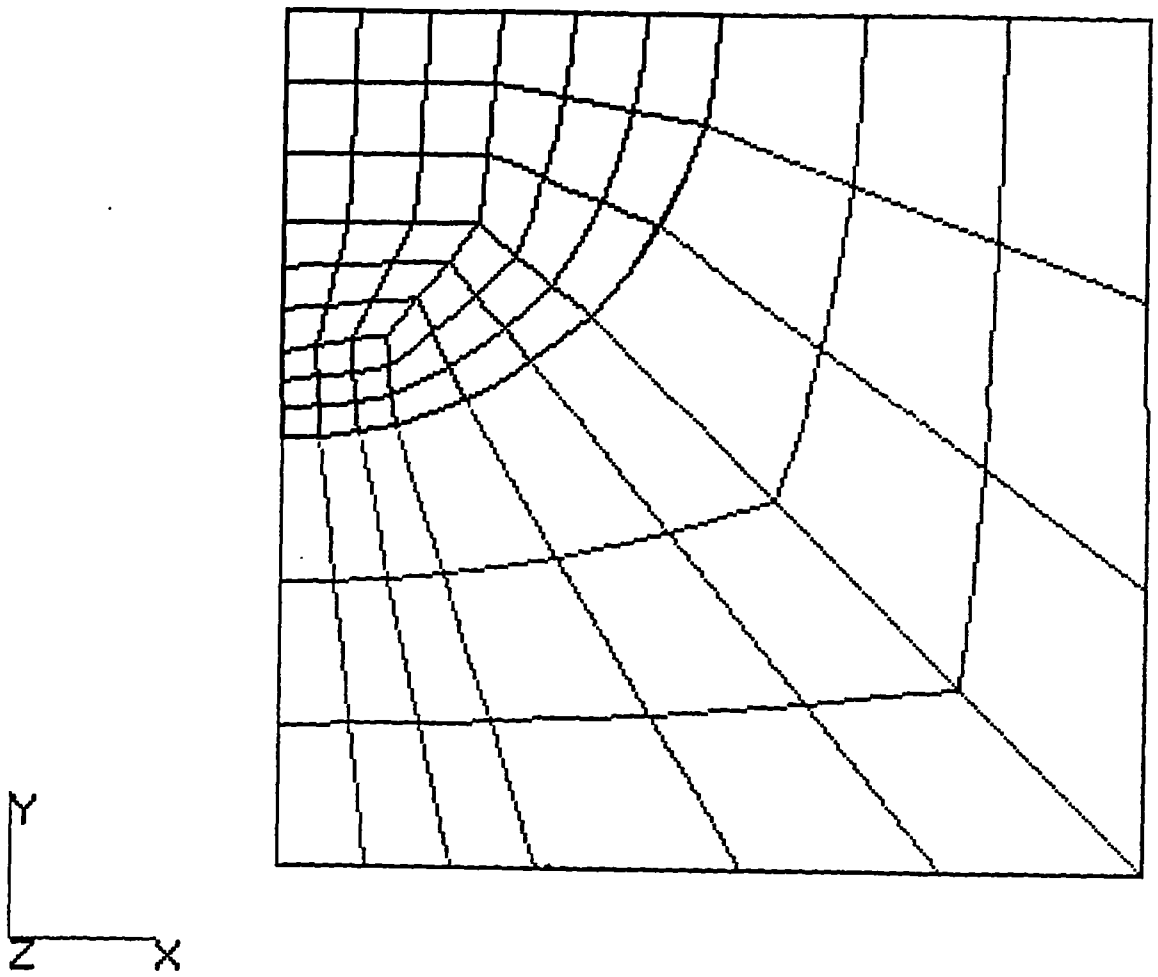


Figure 10.16 Refined finite element mesh for the pin in a hole contact problem

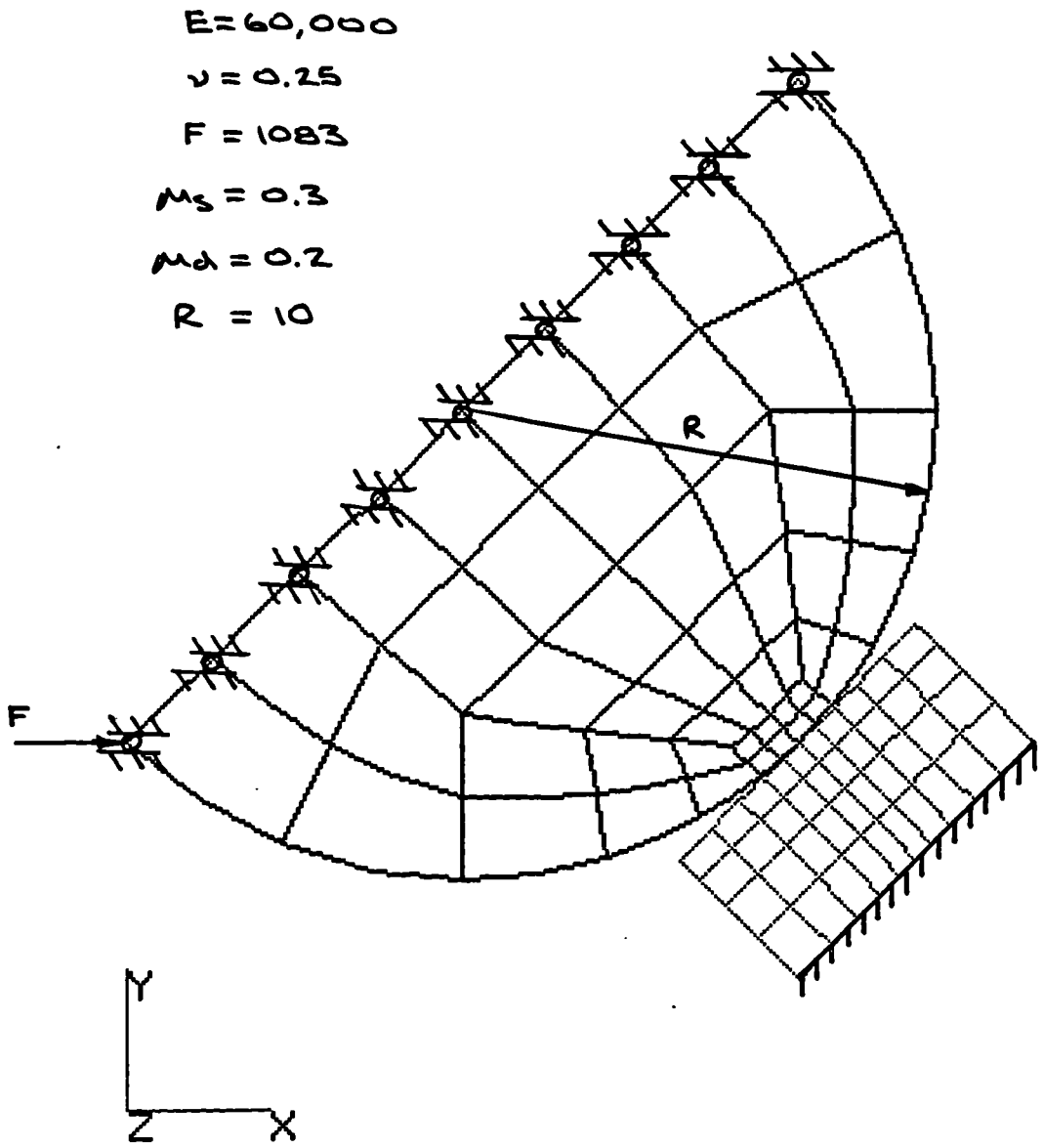


Figure 10.17 Sliding cylinder finite element mesh

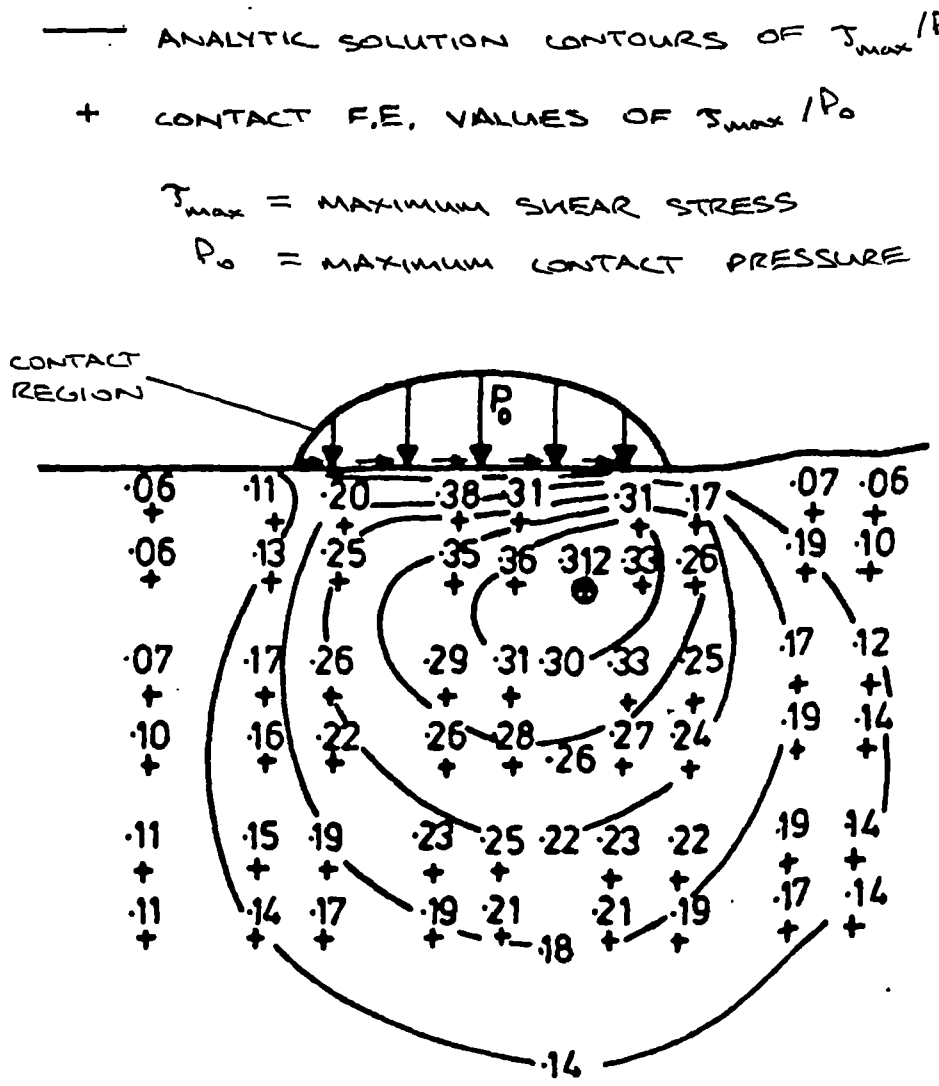


Figure 10.18 Comparison of maximum shear stress results by the one boundary condition loading method with theory

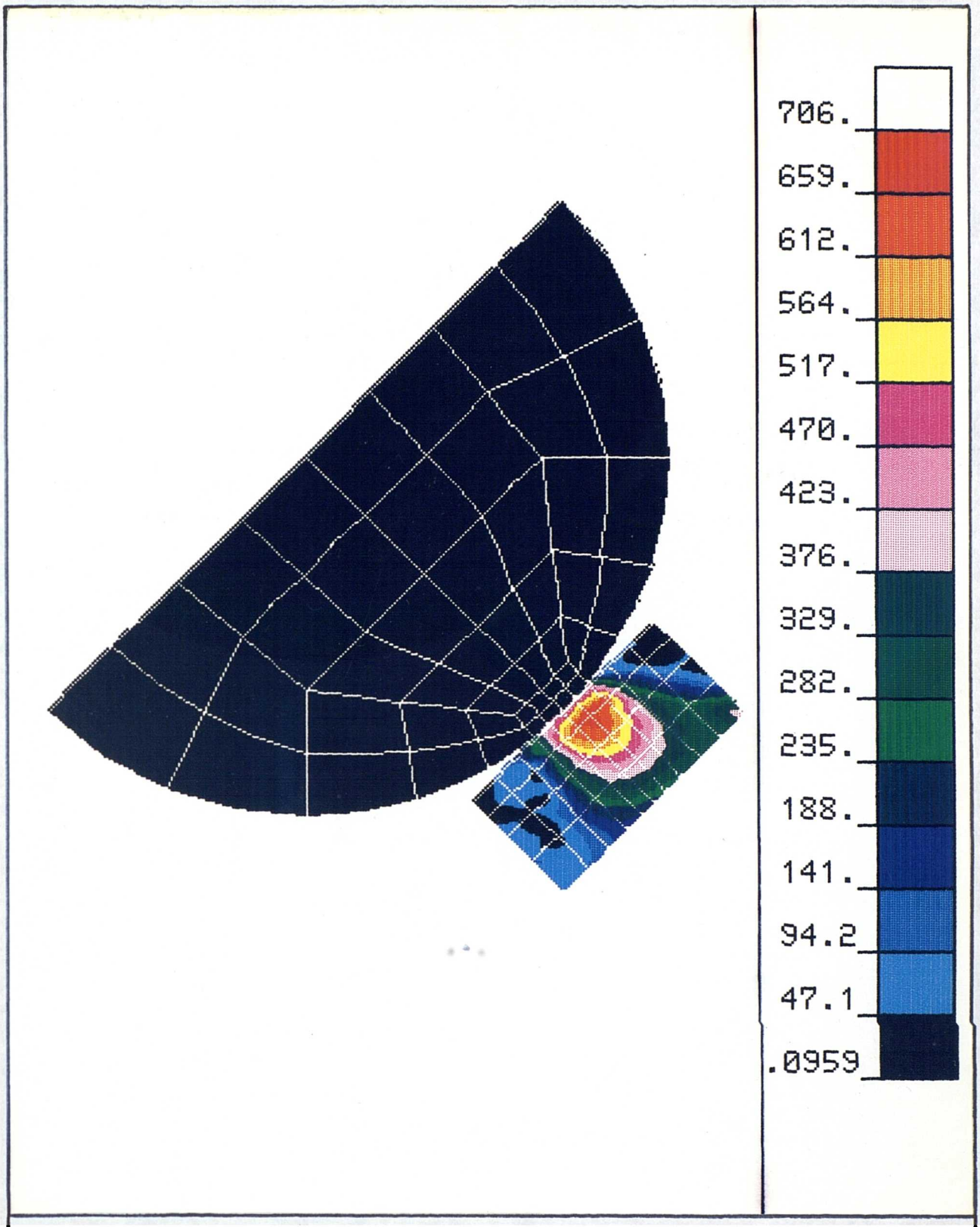


Figure 10.19 Maximum shear stress results by the incremental loading method

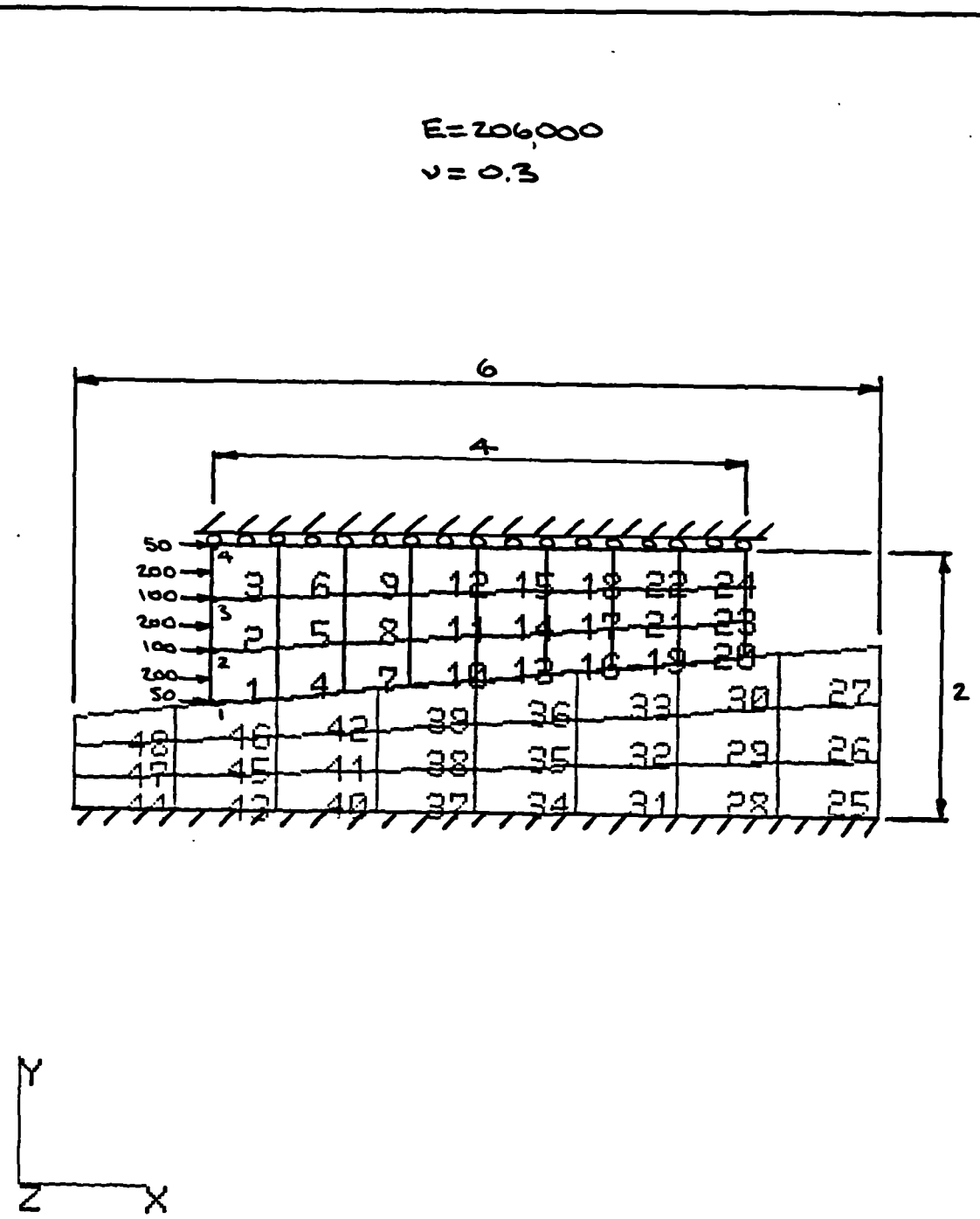


Figure 10.20 Finite element mesh for the sliding wedge contact problem

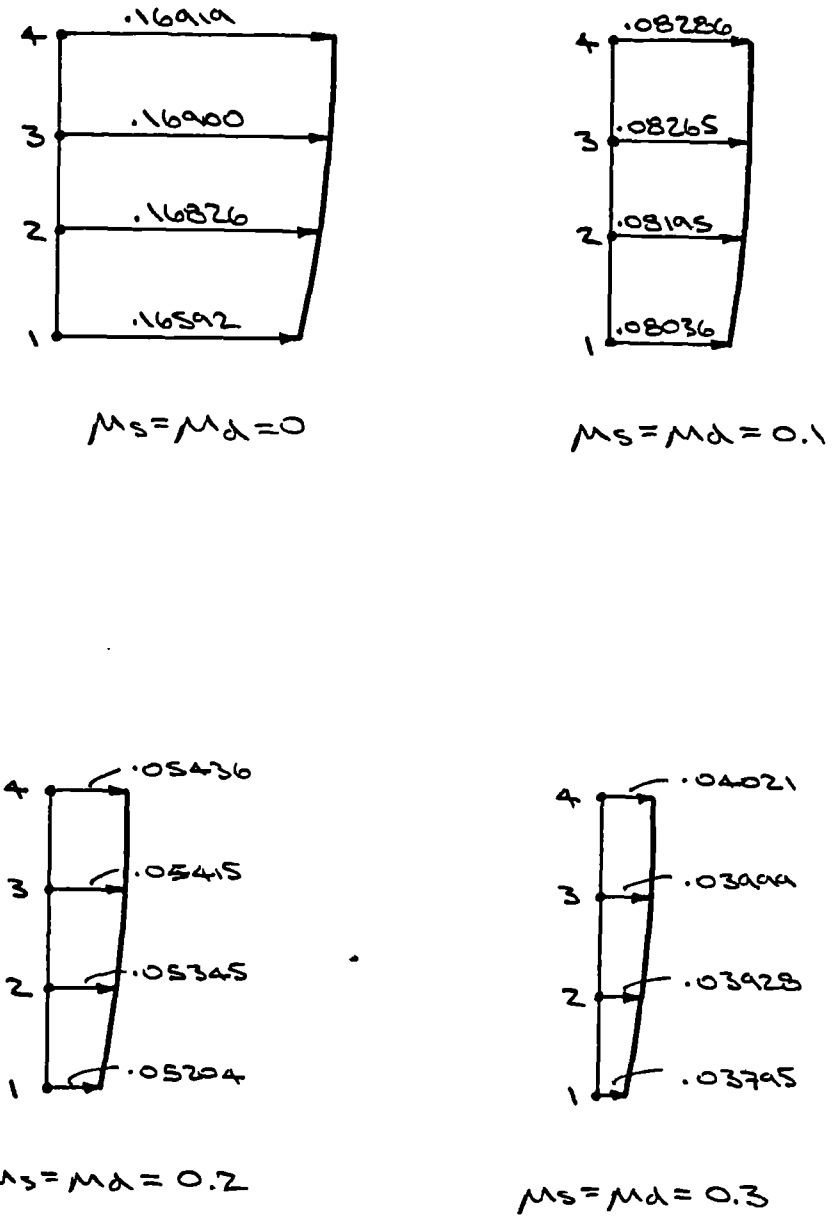


Figure 10.21 Displacement profiles with varying coefficients of friction for the wedge problem

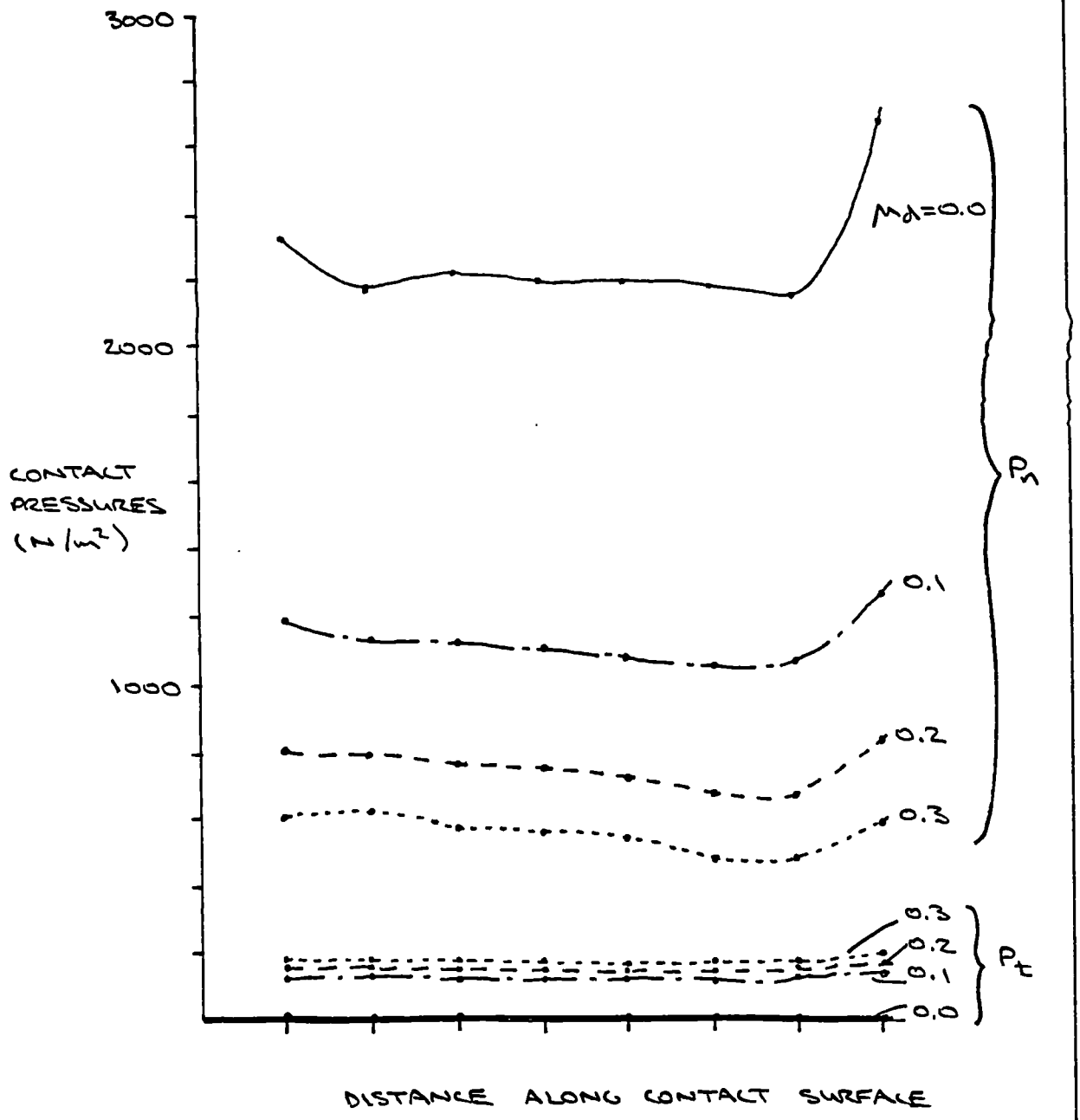


Figure 10.22 Normal and tangential contact pressures for the sliding wedge problem

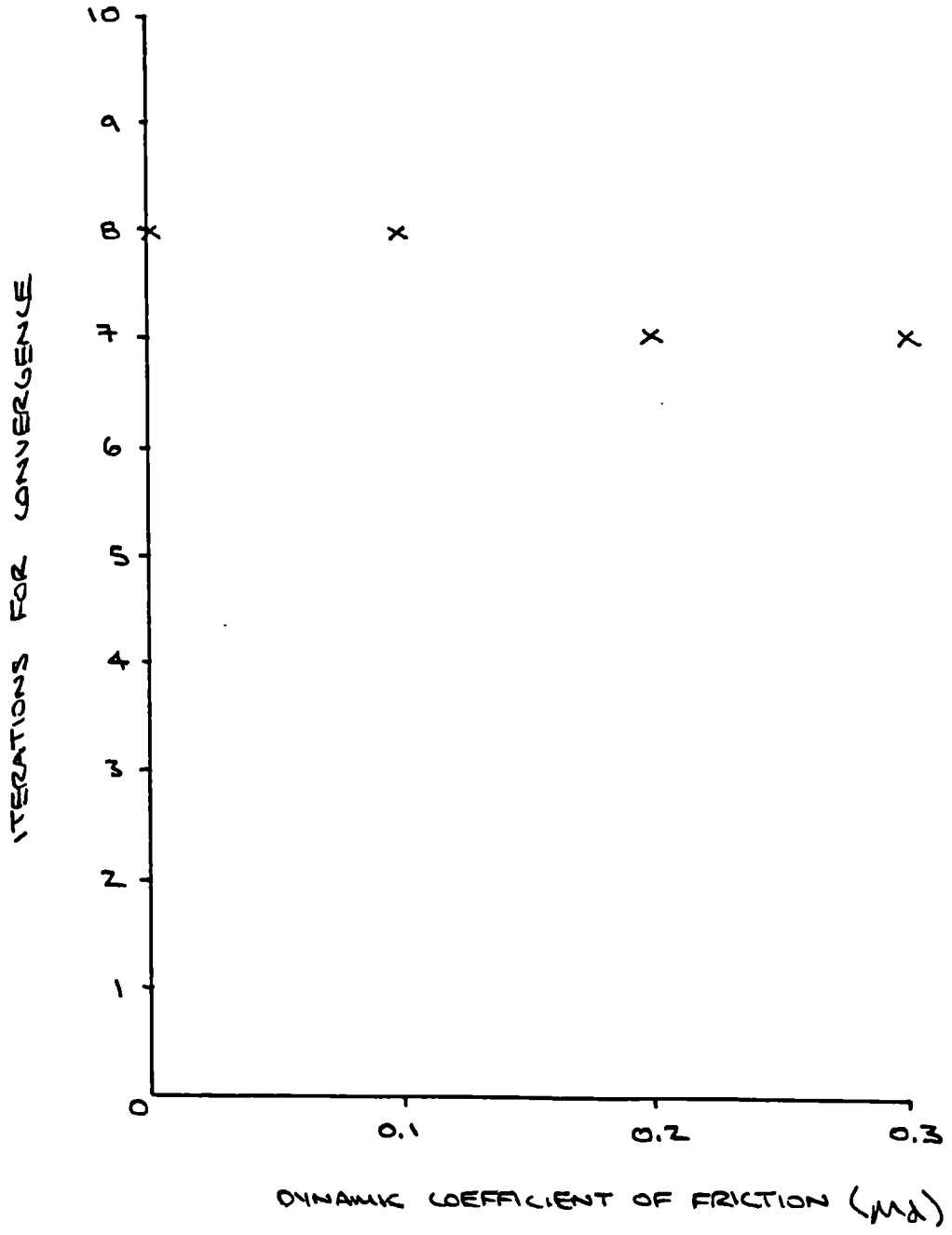


Figure 10.23 Effect of friction on iterations for convergence

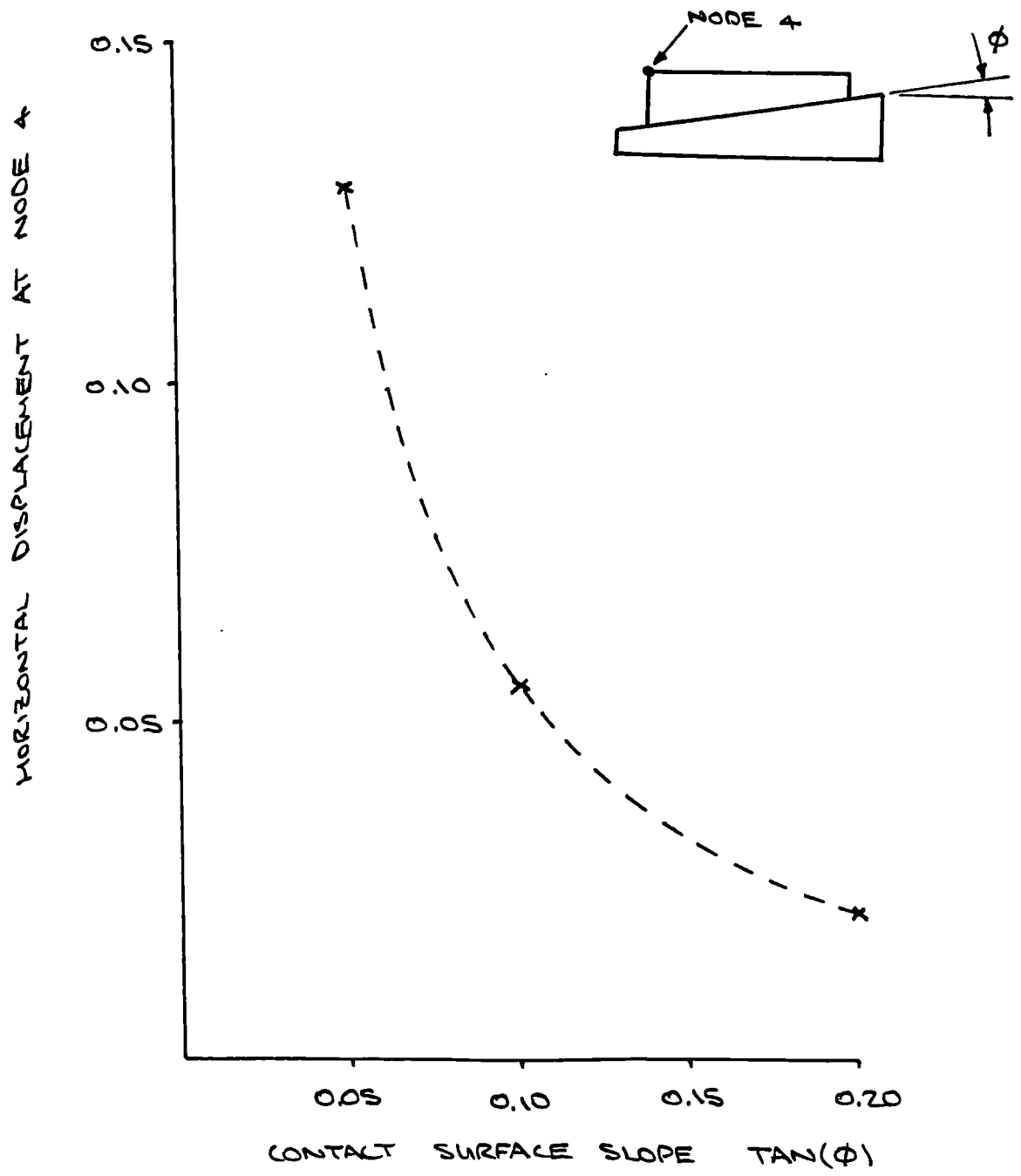


Figure 10.24 Variation of displacement with slope angle

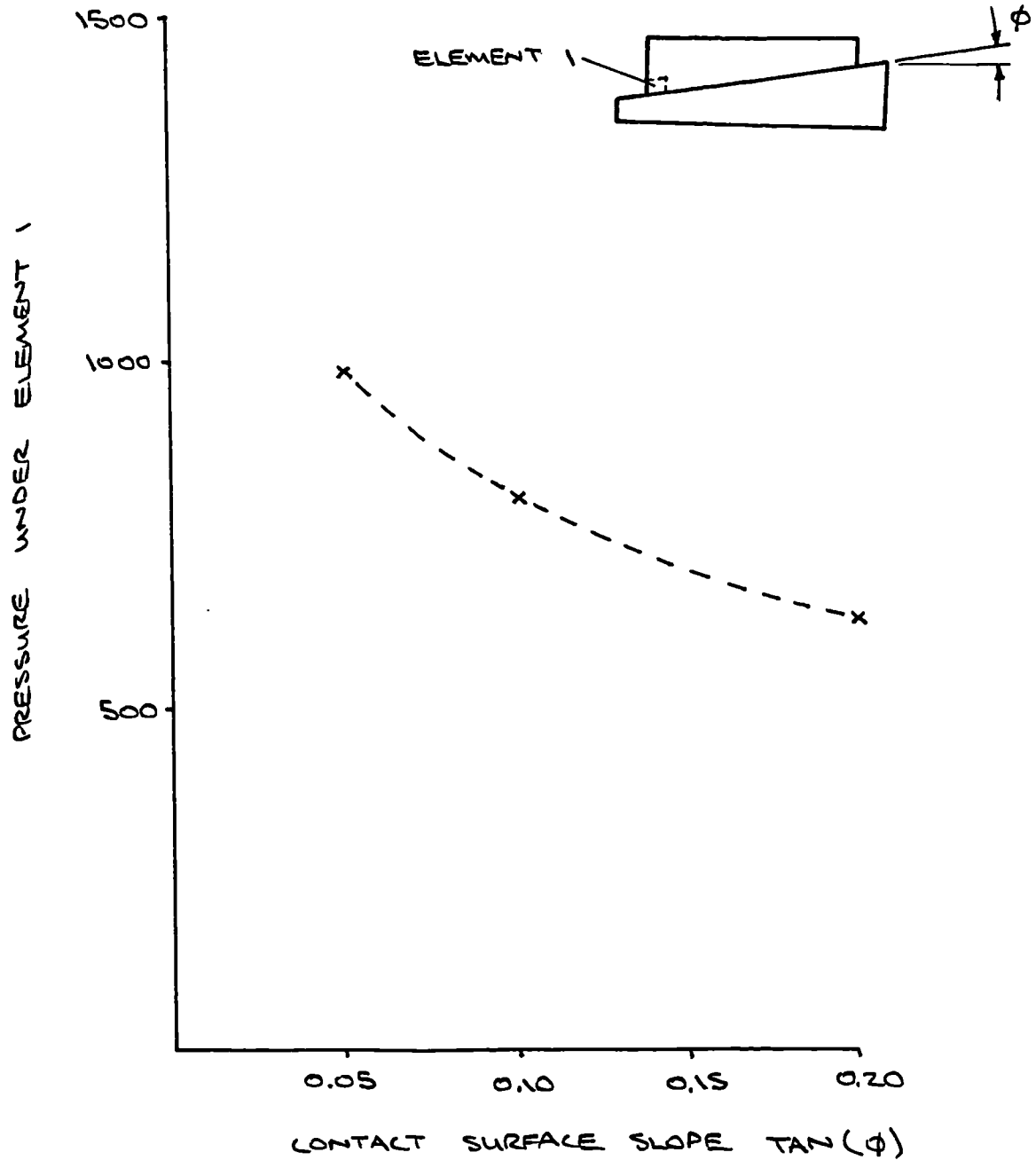


Figure 10.25 Variation of contact pressure with slope angle

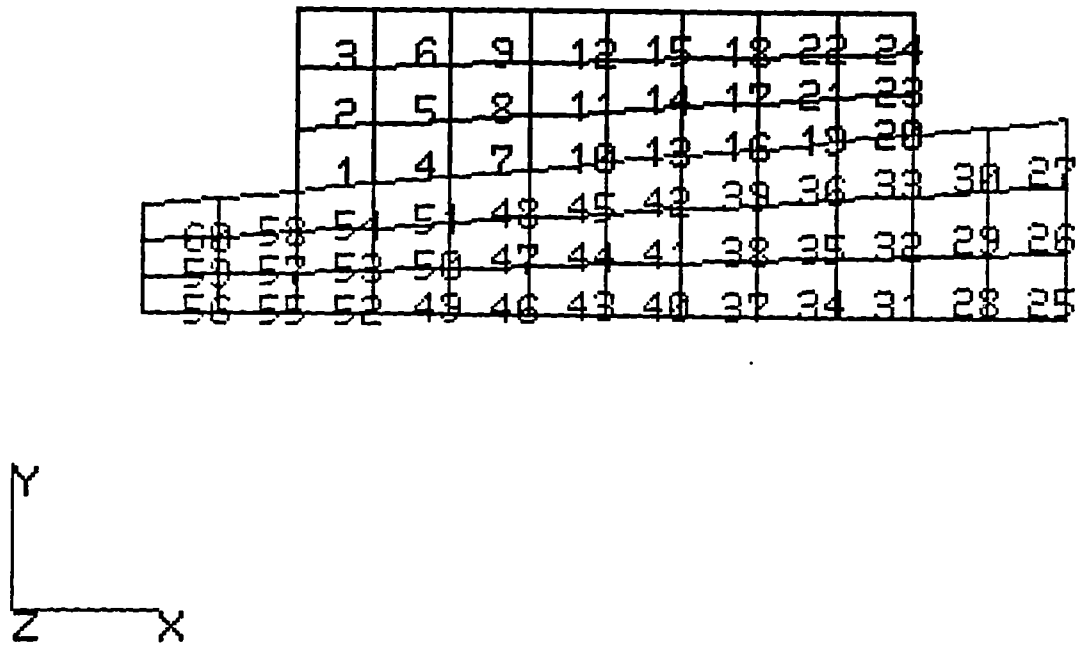


Figure 10.26 Aligned mesh for the sliding wedge problem

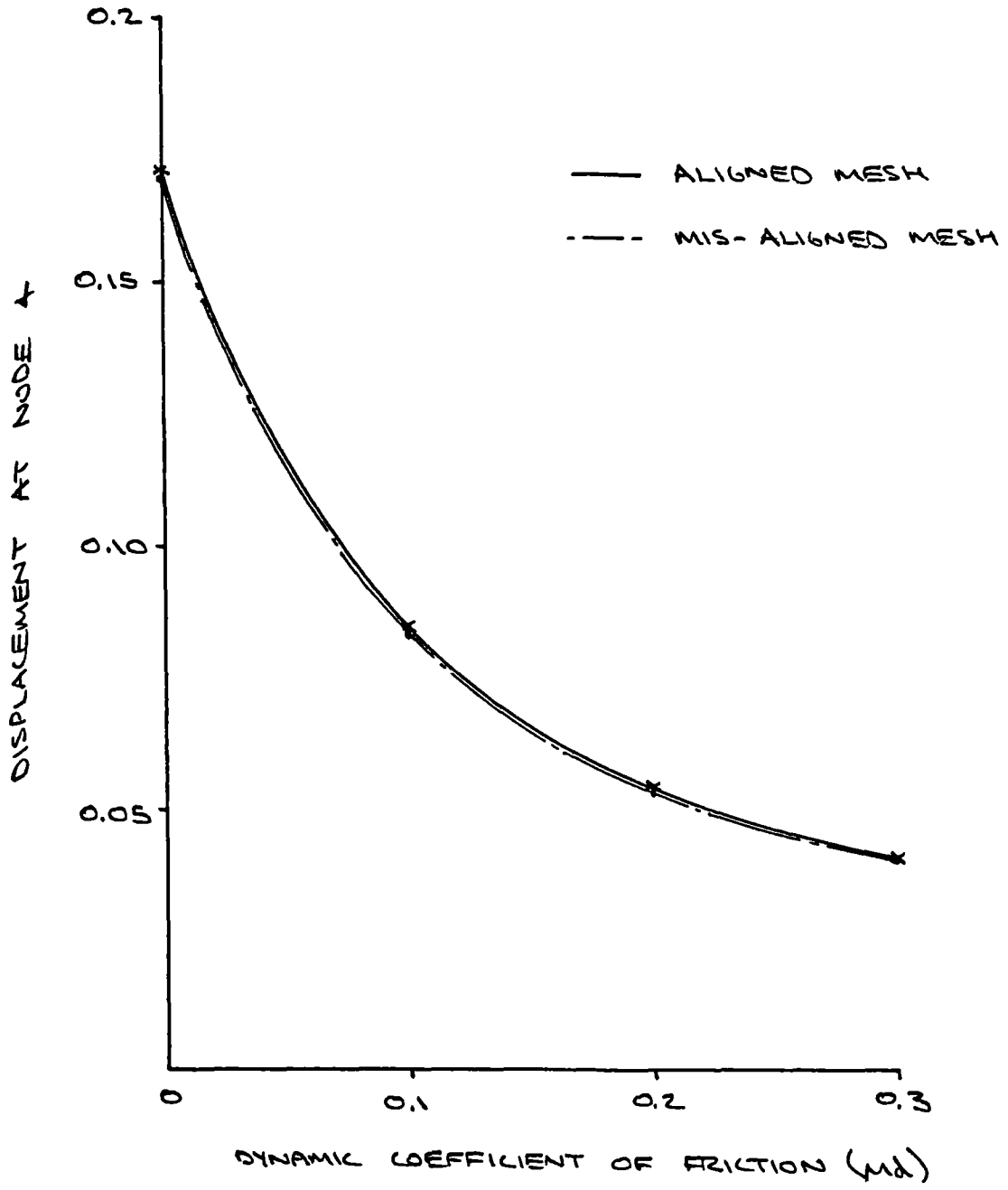


Figure 10.27 Comparison of displacements between the aligned and mis-aligned meshes

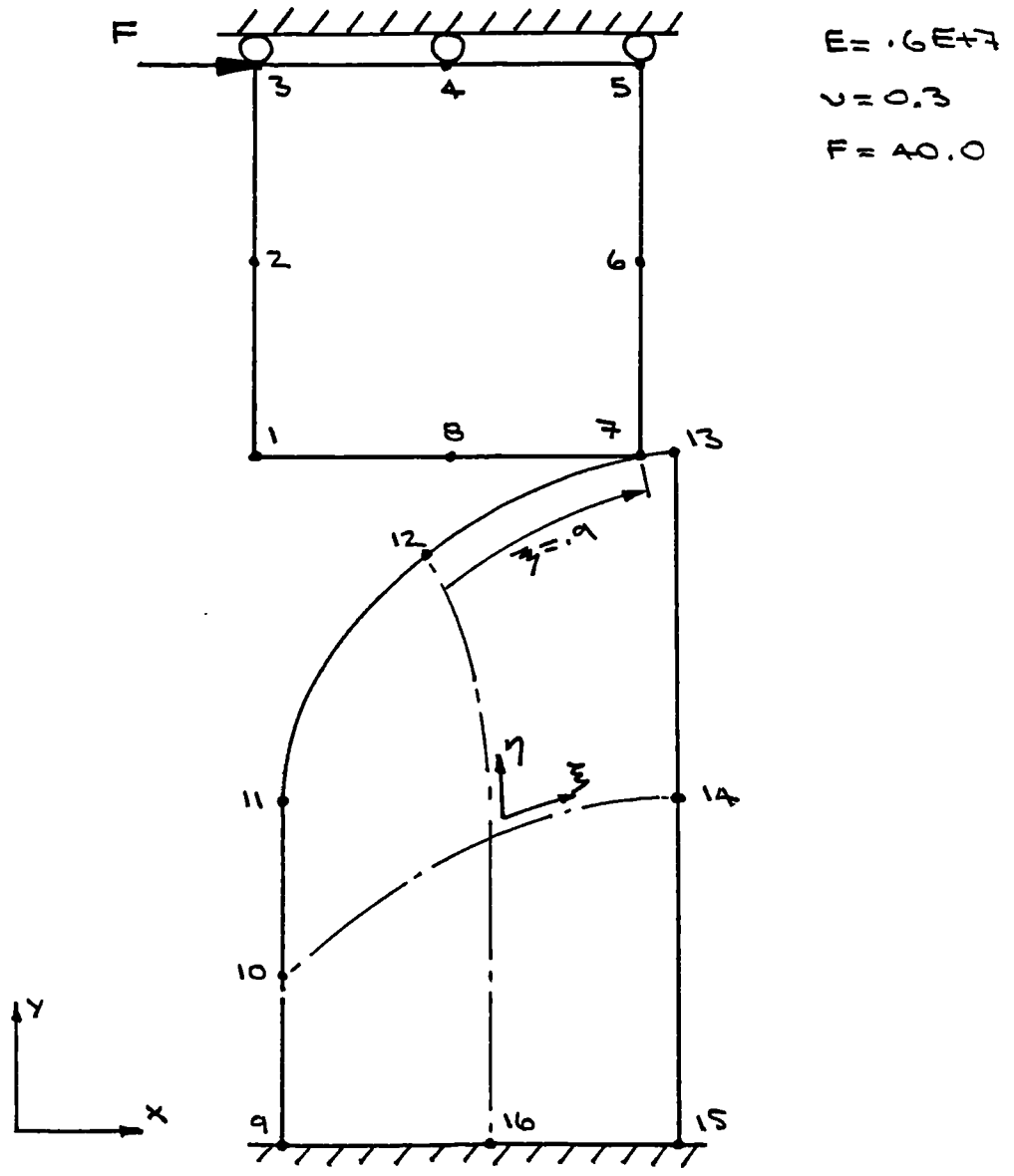


Figure 10.28 Simple finite element model for comparison of different constraint techniques

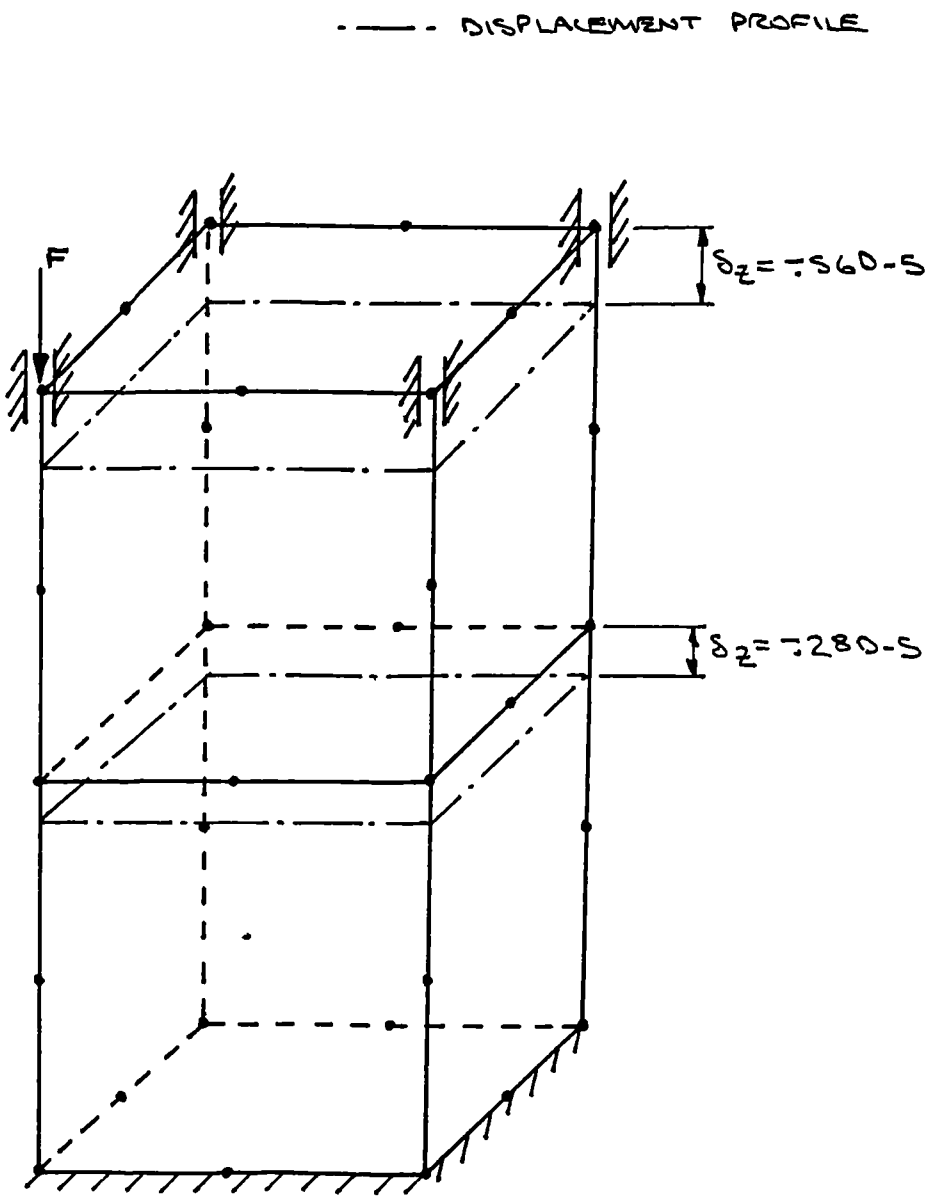


Figure 10.29 Slope constraint model including displacement results

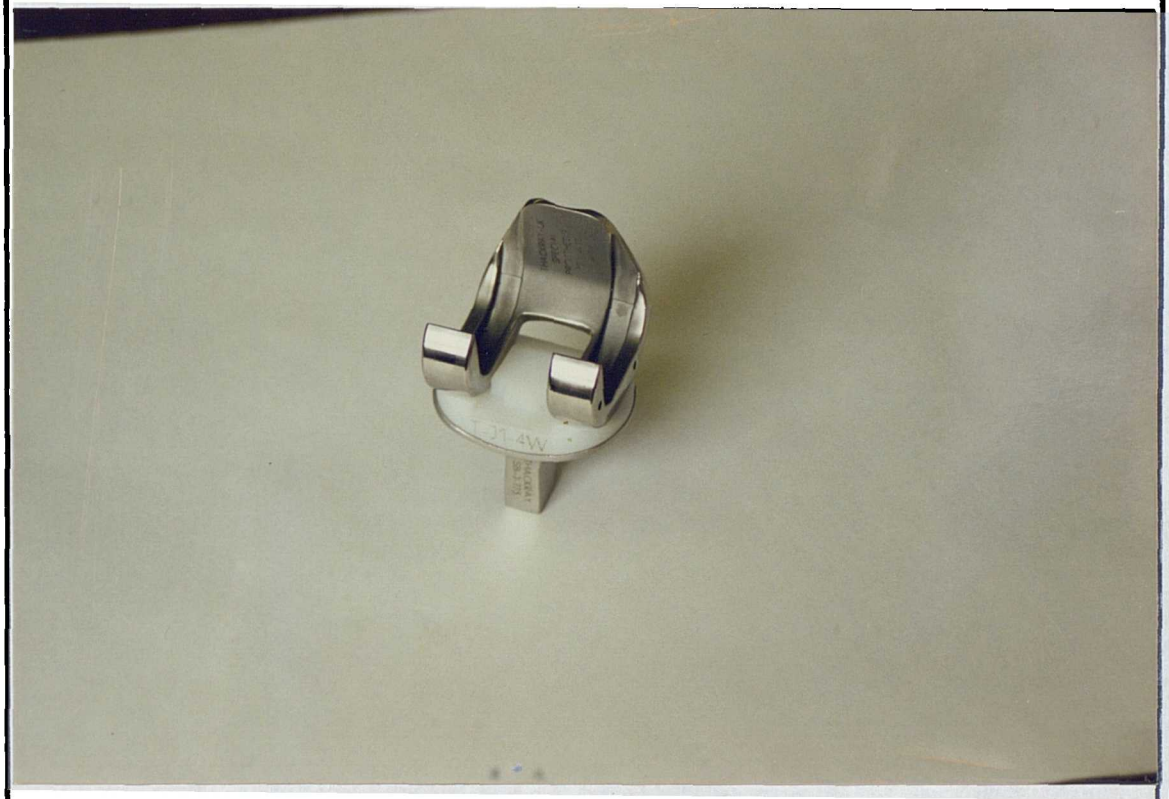


Figure 10.30 Knee Prosthesis

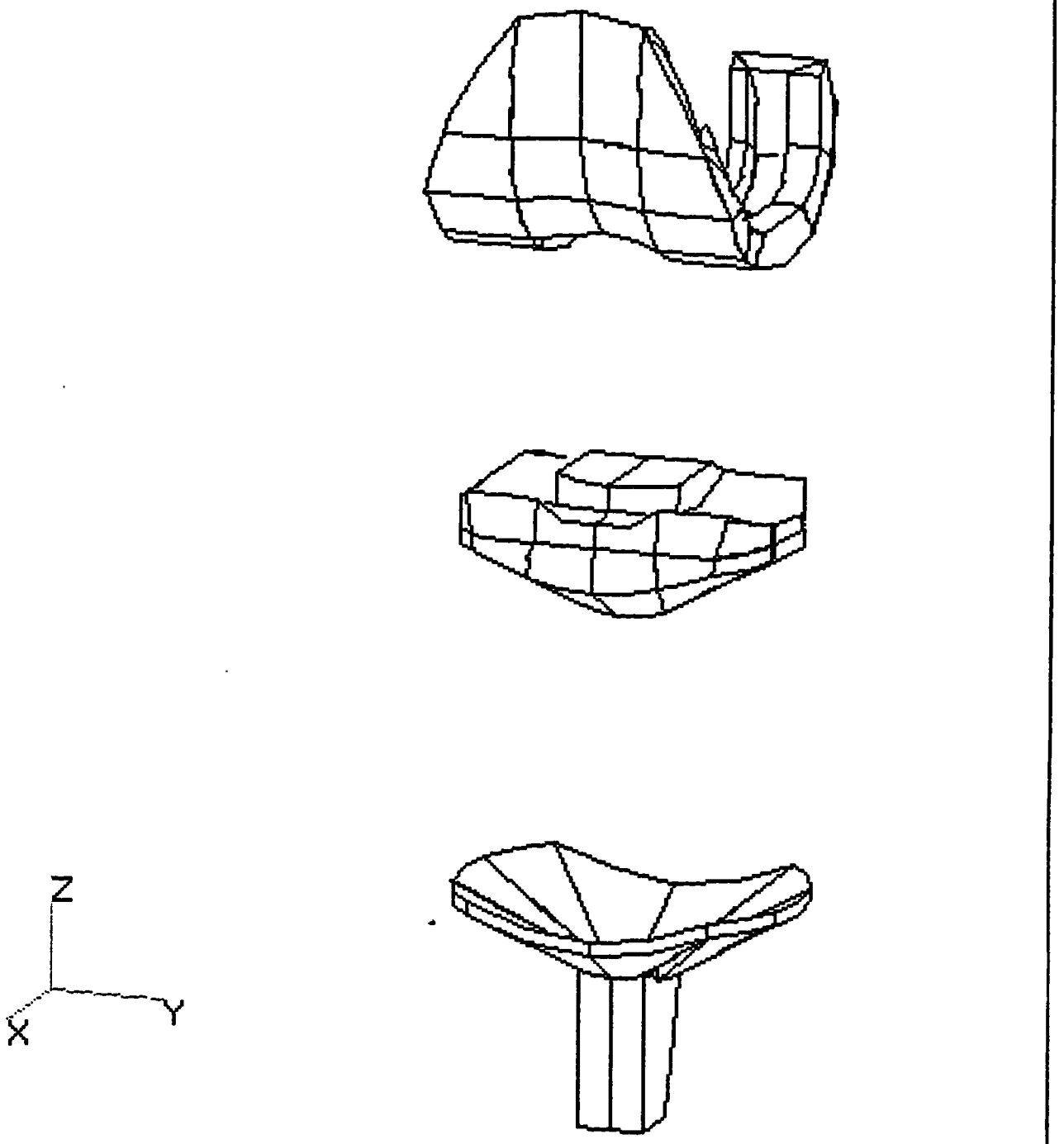


Figure 10.31 Finite element mesh for the Knee Prosthesis

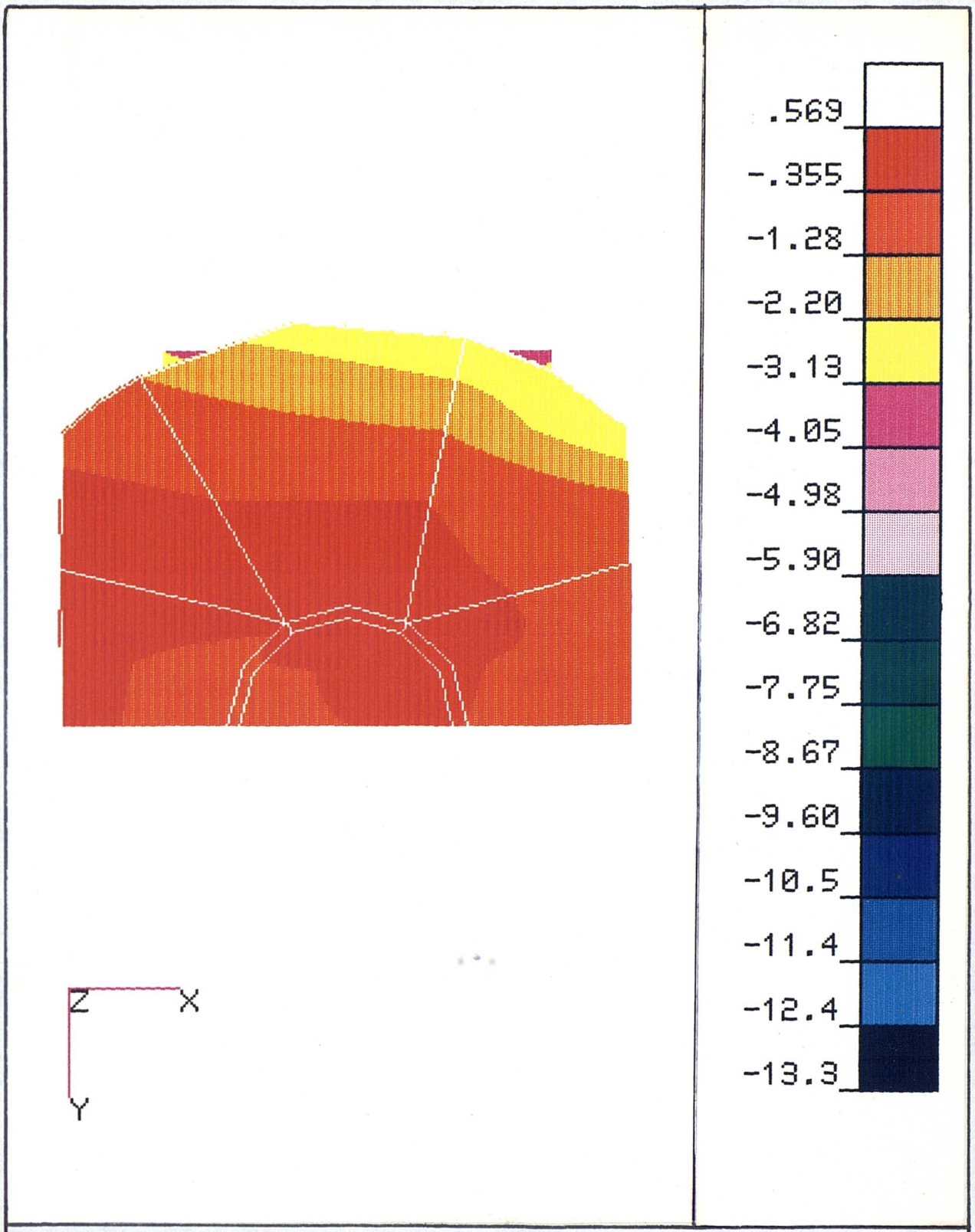


Figure 10.32 σ_z contact stresses on the underside of the meniscal component

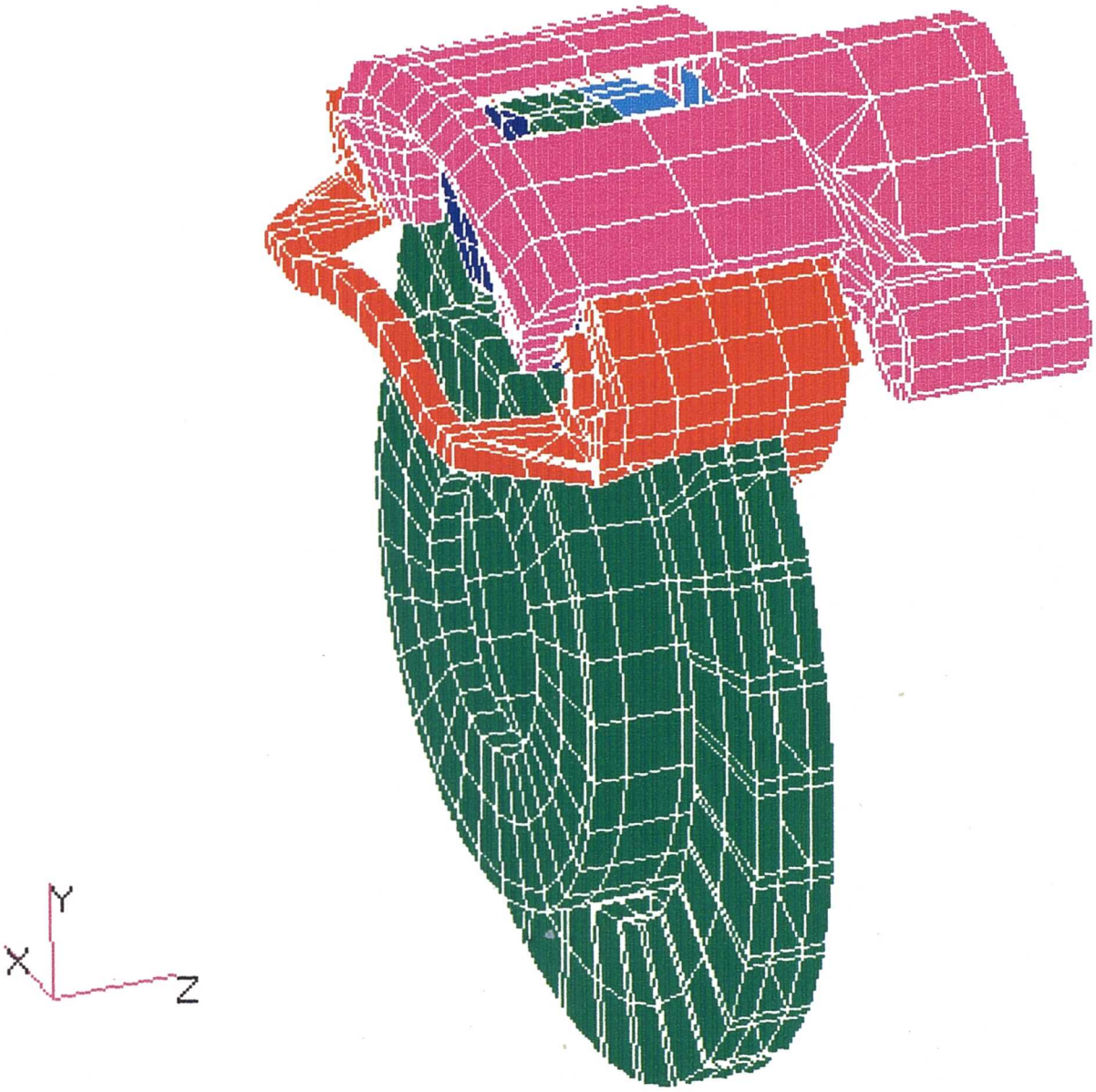


Figure 10.33 Finite element mesh of the assembled brake

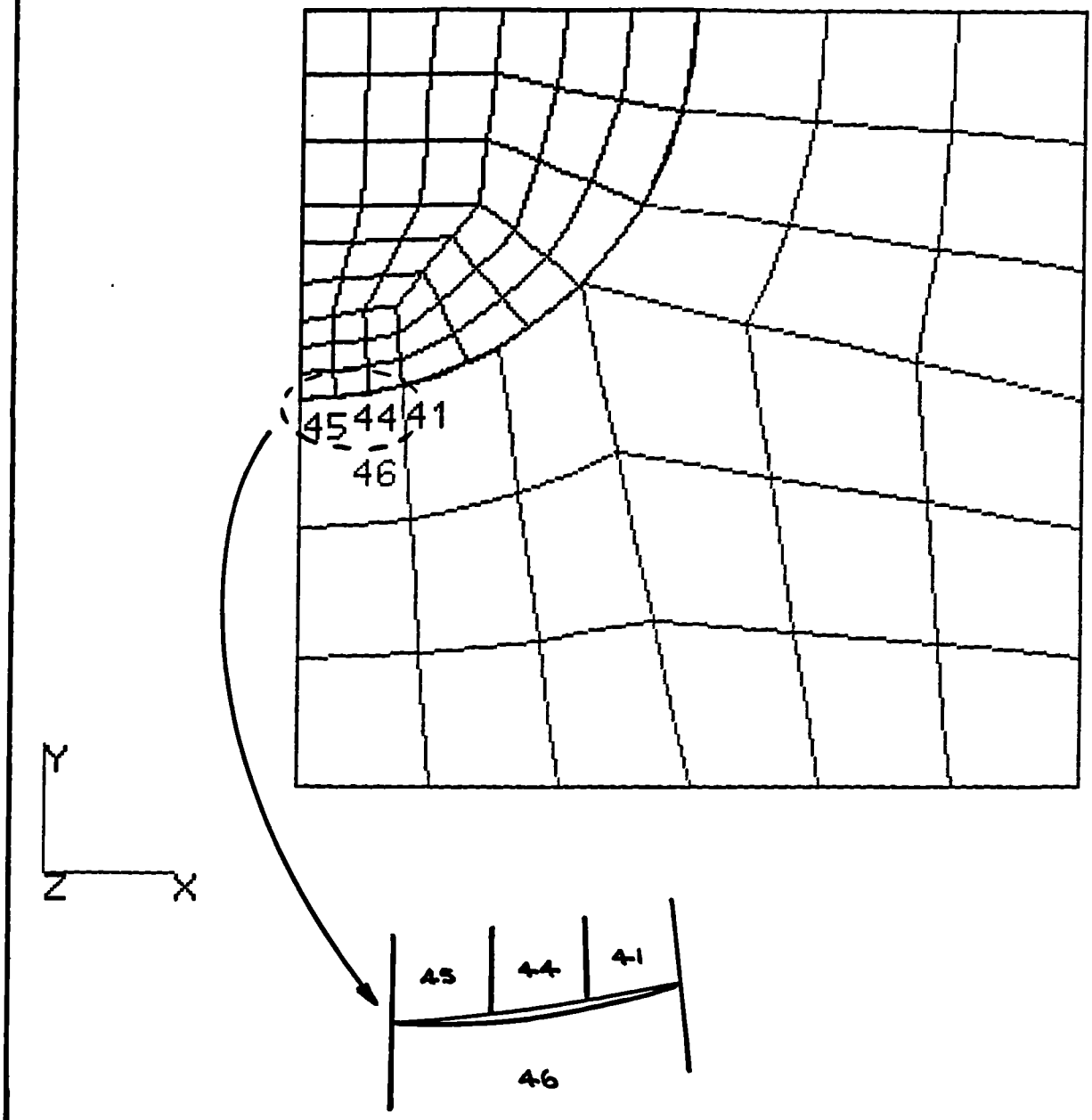


Figure 11.1 Initial refined pin a hole mesh

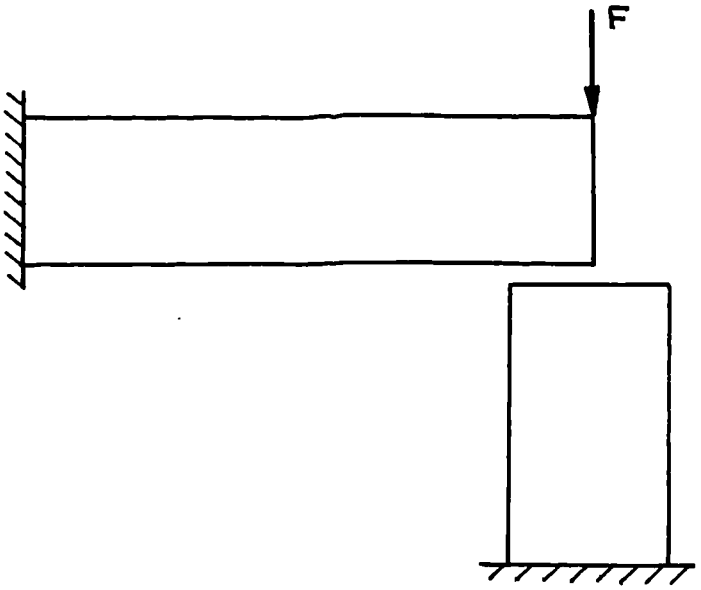


Figure A.1 Cantilever and beam contact problem

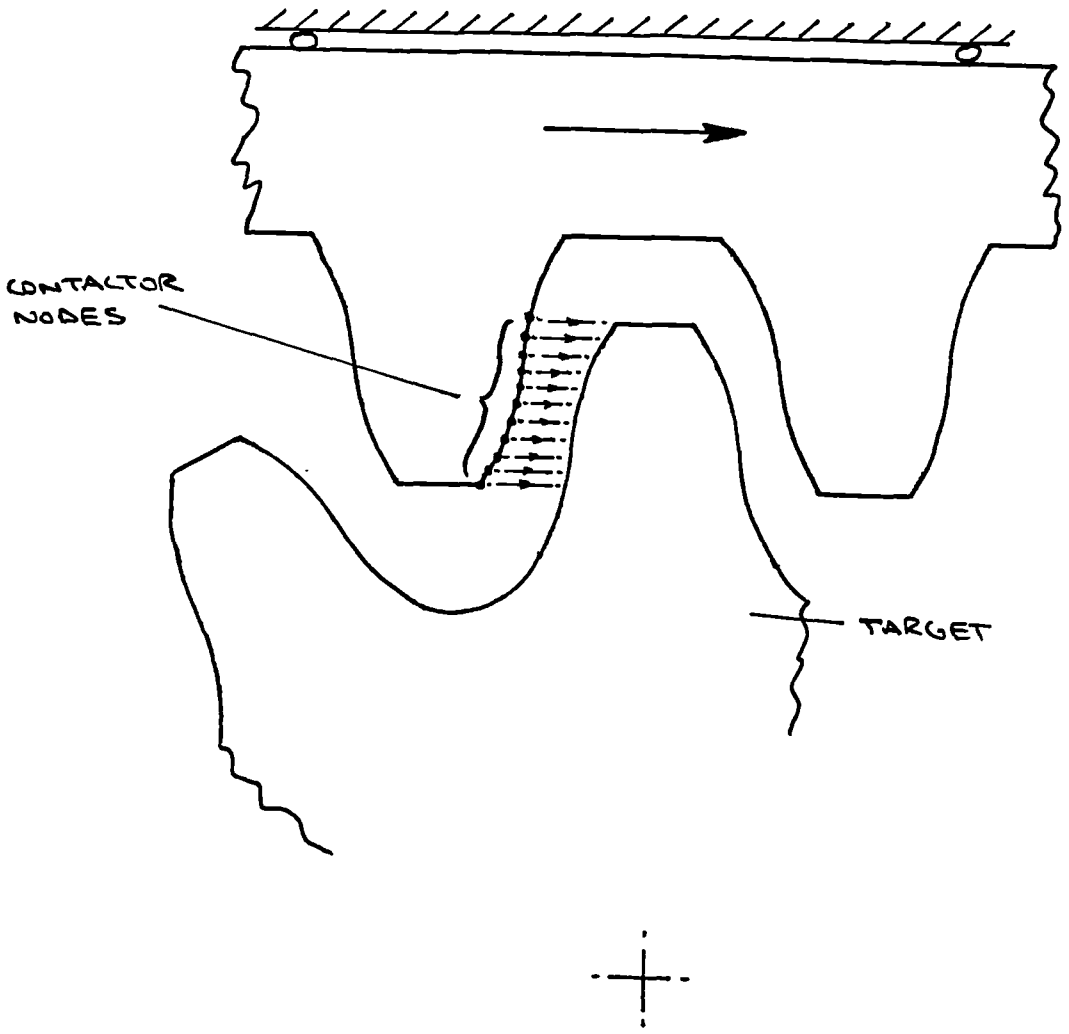


Figure A.2 Contact of complex surface geometries

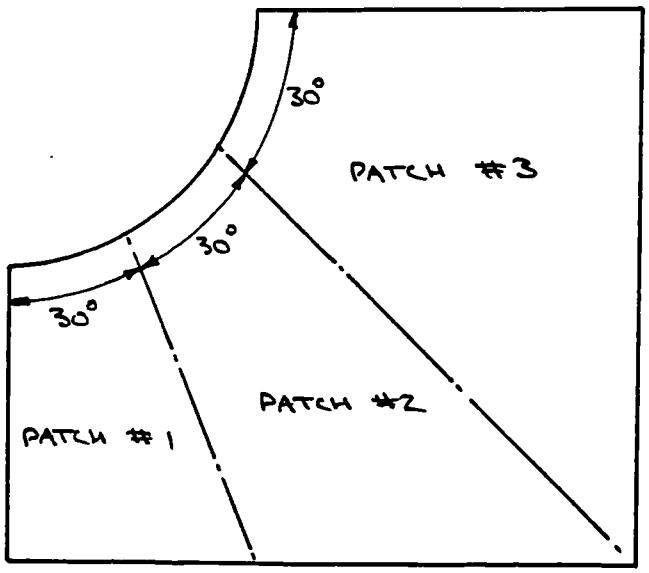


Figure B.1 Recommended minimum division for accurate modelling of a 90^{o} curved profile

E = .6+7
ν = 0.3
F = -5

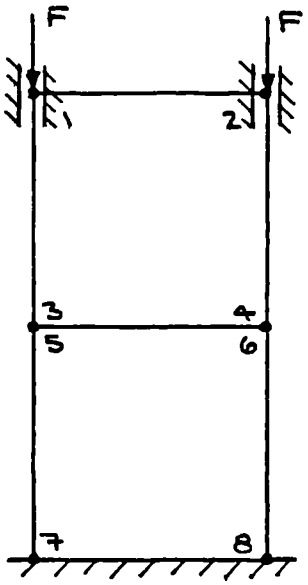


Figure F.1a 'Contact' mesh

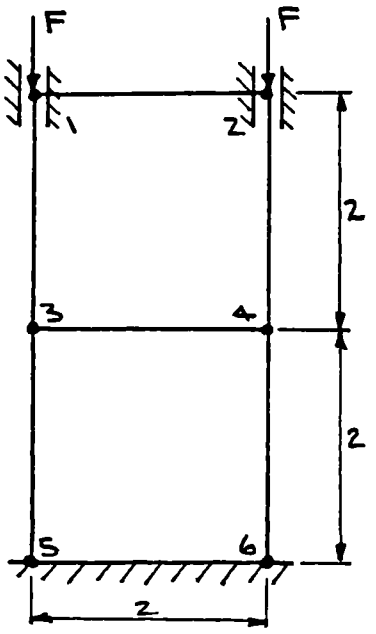


Figure F.1b 'Standard' mesh

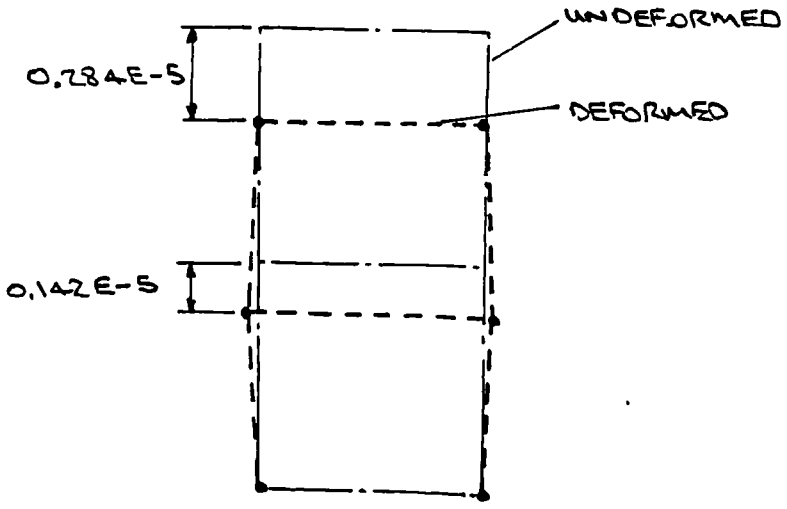


Figure F.1c Displacement results for both of the above meshes

E= .6E+7

ν= 0.3

F= 40.0

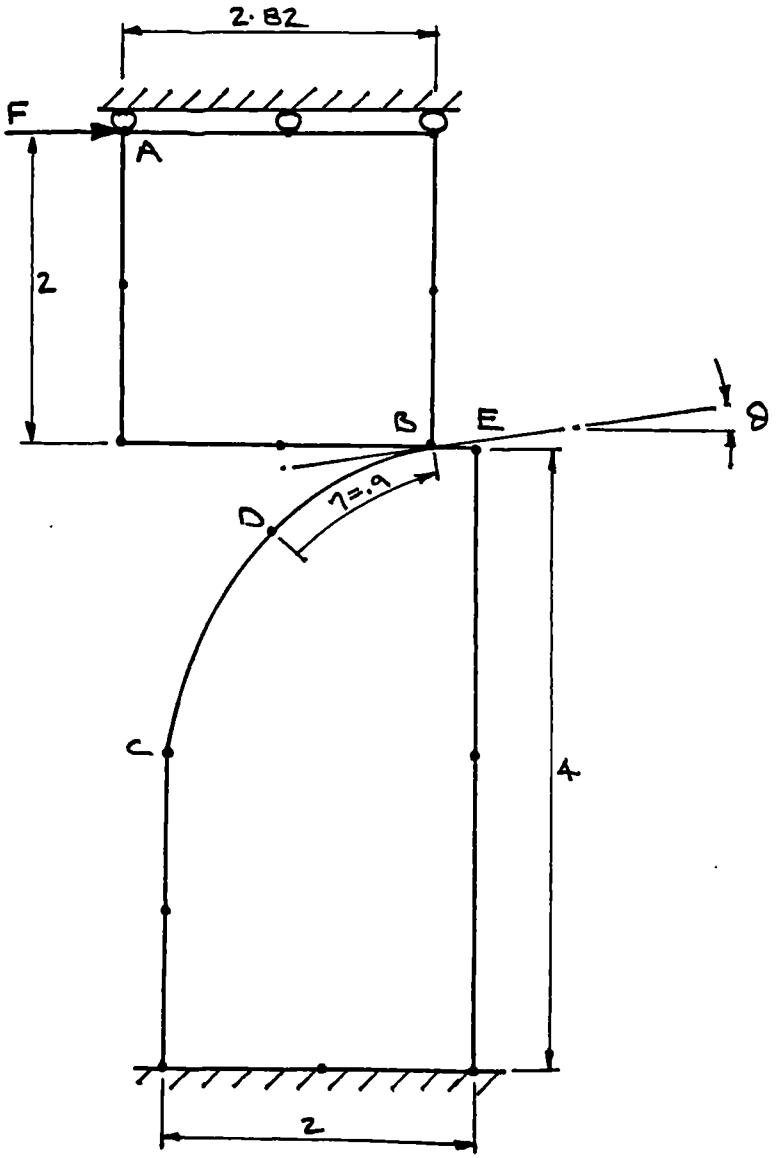


Figure F.2 Sliding contact mesh

E= .6E+7

ν = 0.3

F = -2500

$M_S = M_d = 0.05$

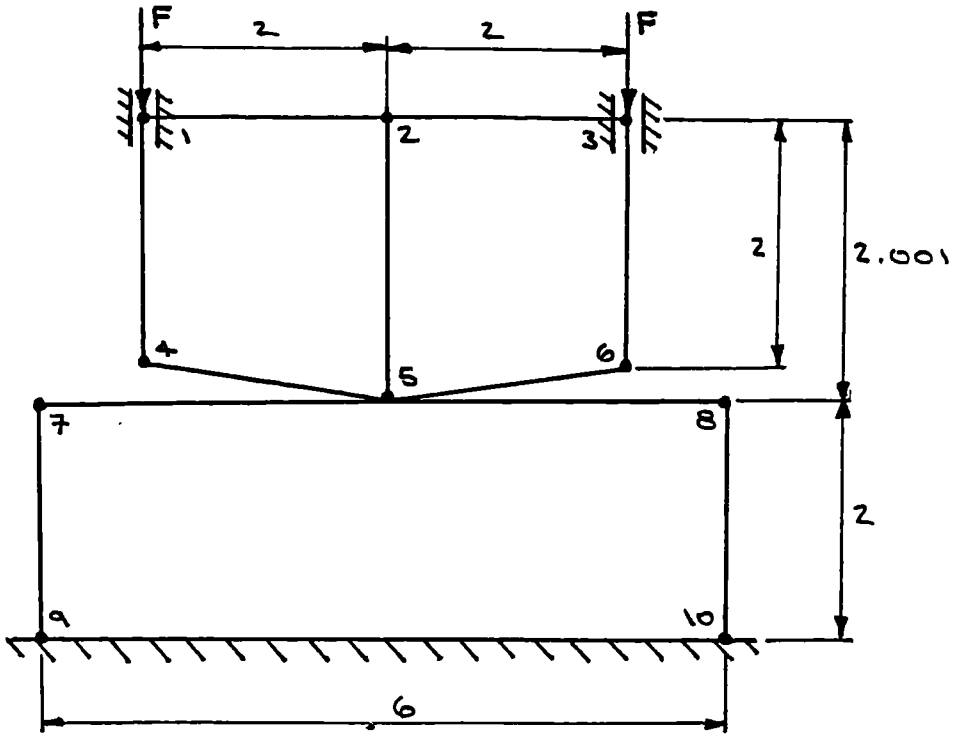


Figure F.3 Simple overlap mesh

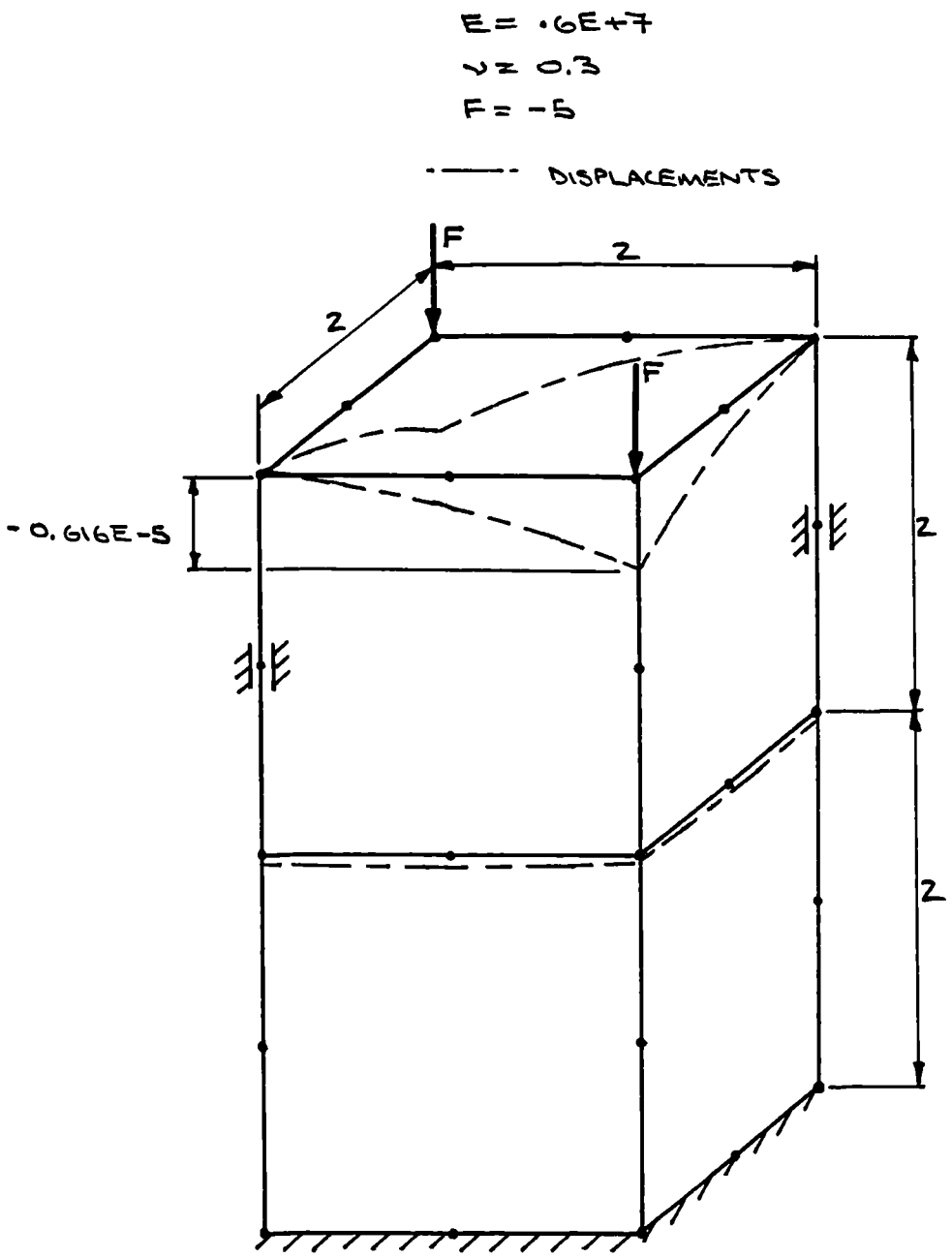


Figure F.4 Frictionless 3-D contact model

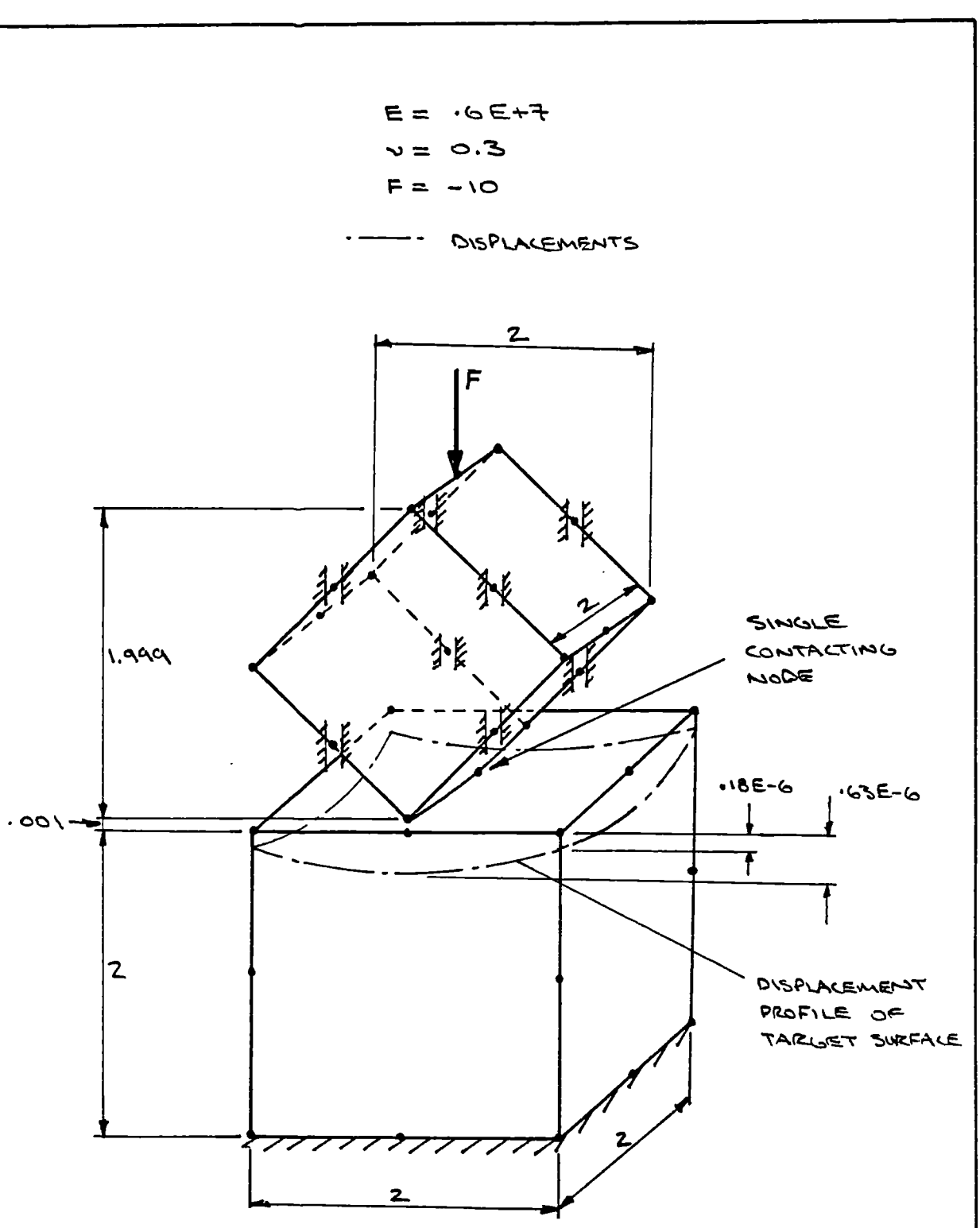


Figure F.5 3-D contact model with non-aligned mesh

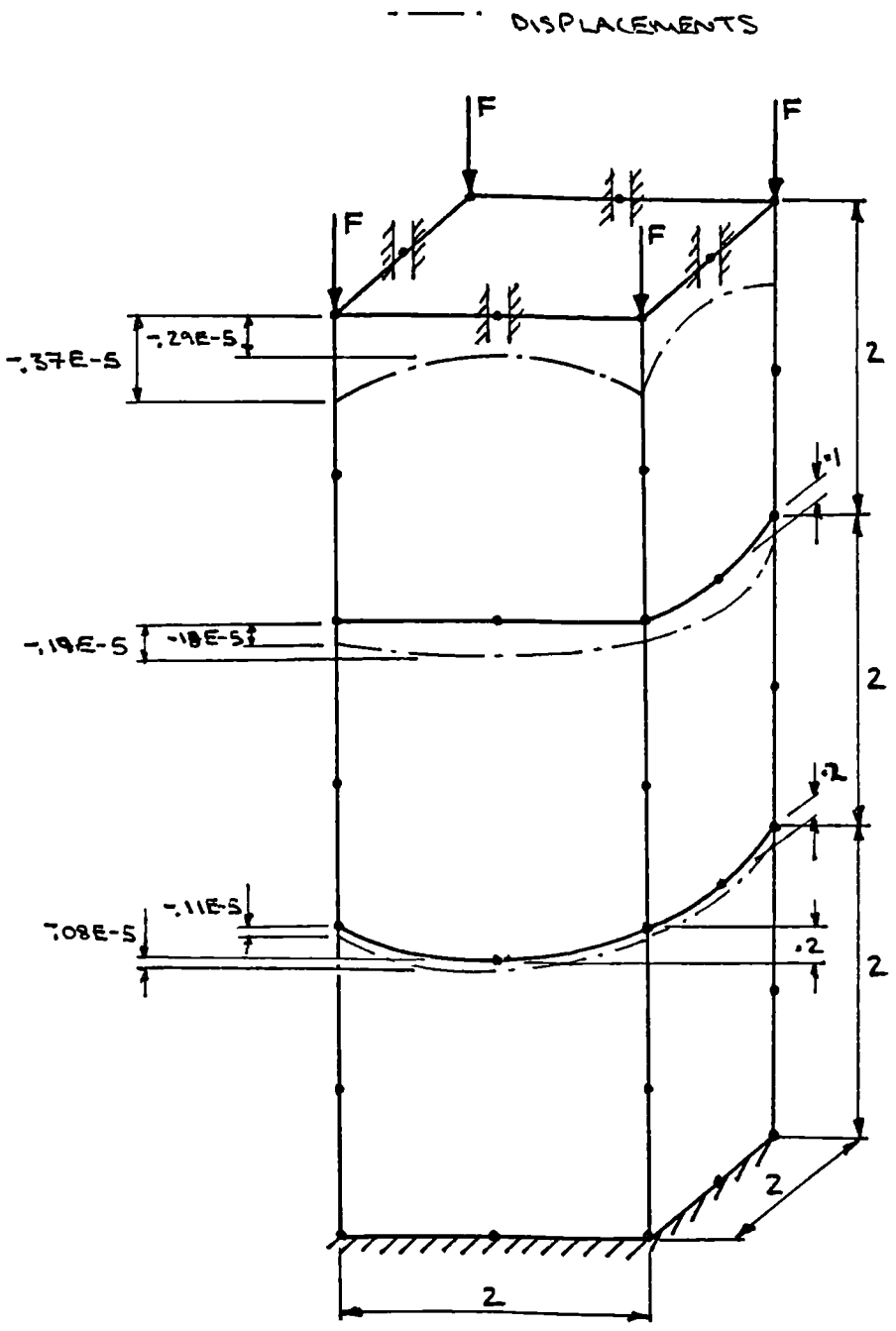


Figure F.6 Three body contact model

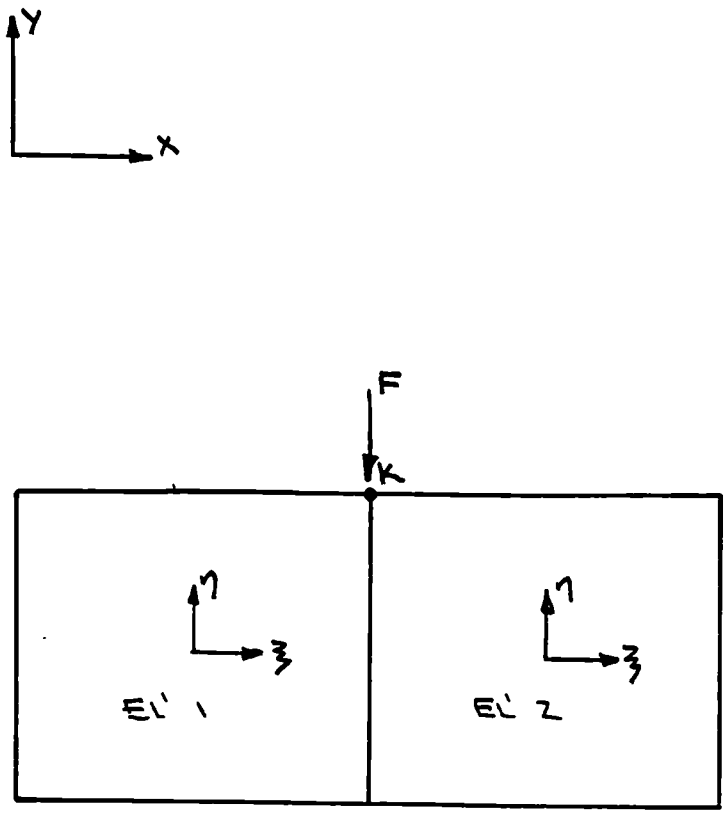


Figure G.1 Slope continuity model