INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book. These are also available as one exposure on a standard 35mm slide or as a 17” x 23” black and white photographic print for an additional charge.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Finite element analysis of elastic contact problems with friction

Jinn, Jong-Tae, Ph.D.
The Ohio State University, 1989
FINITE ELEMENT ANALYSIS OF ELASTIC CONTACT PROBLEMS
WITH FRICTION

DISSERTATION
Presented in Partial Fulfillment of the Requirements
for the Degree Doctor of Philosophy in the Graduate
School of The Ohio State University
By
Jong-Tae Jinn, B.S., M.S.

The Ohio State University
1989

Approved By
Co-Adviser, Professor
Department of Engineering Mechanics

Dissertation Committee:
Dr. J.K. Lee
Dr. S.H. Advani
Dr. S.E. Bechtel

Co-Adviser, Chairman and Professor
Department of Engineering Mechanics
ACKNOWLEDGEMENTS

The author wishes to express his sincere gratitude to his advisers Professors June K. Lee and Sunder H. Advani for their constant encouragement, support and guidance during the course of this research. The author also wishes to thank Professor Steve E. Bechtel who served as a member of his reading committee.

A note of appreciation is also due to his colleagues Hyun Moon, Chun S. Kim and Joong D. Yoo for their help in the preparation of this dissertation.

This research was sponsored by U.S. Department of Energy under METC Contract No. DE-AC21-83MC20338, National Science Foundation grant (NSF-8311643) and CRAY Research, Inc. Their financial support is gratefully appreciated. Computational supports by The Ohio State University Instruction and Research Computer Center are also appreciated.

Finally, the author is thankful to his wife and daughters for their endless patience and moral support during his study.
VITA

May 20, 1954 ................... Born in Taegu, Korea

1977 ............................ B.S., Seoul National Univ., Seoul, Korea

1979 ............................ M.S., Korea Advanced Institute of Science, Seoul, Korea

1979-1983 ....................... Assistant Professor, Ulsan University, Ulsan, Korea

1984-1989 ....................... Graduate Research Associate, Department of Engineering Mechanics, The Ohio State University, Columbus, Ohio

Publications


"Block Gauss-Seidel Method for Finite Element Analysis of Frictional Contact Problems", WCCC Abstracts, First World
Congress on Computational Mechanics, Vol. 1, Univ. of Texas, Austin, 1986 (with Lee, J.K. and Advani, S.H.)

FIELDS OF STUDY

Continuum Mechanics and Finite Element Methods
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................. ii

VITA ........................................................................ iii

LIST OF TABLES .......................................................... vii

LIST OF FIGURES ..................................................... viii

NOMENCLATURE ......................................................... xiii

CHAPTER ....................................................................... PAGE

I. INTRODUCTION .................................................... 1
  1.1. Opening Remarks ............................................ 1
  1.2. Literature Review .......................................... 8
  1.3. Outline of the Present Study ......................... 16

II. FORMULATION OF CONTACT PROBLEMS .............. 19
  2.1. Introduction ................................................ 19
  2.2. Basic Relations ............................................ 20
  2.3. Contact Conditions ....................................... 37
  2.4. Incremental Contact Boundary-Value Problem .... 57

III. FINITE ELEMENT APPROXIMATION .................. 86
  3.1. General Remarks .......................................... 86
  3.2. Discrete System Equations ............................. 88
  3.3. Evaluation of the Pairing Points and the Normal Vector ... 92
  3.4. Governing Contact Equations ......................... 97

IV. SOLUTION METHOD ............................................. 101
  4.1. Method of Embedding of Contact Conditions ... 103
  4.2. Existence and Uniqueness of Solution ............... 113
  4.3. Iterative Solution Method ............................... 117
  4.4. Computational Considerations ...................... 126
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. A comparison between Coulomb's friction model and J_2 plastic flow model with isotropic hardening</td>
<td>175</td>
</tr>
<tr>
<td>2. Admissible contact conditions for Case (2)</td>
<td>176</td>
</tr>
<tr>
<td>(a) Case 2(a): D_{nn} &lt; \mu</td>
<td>D_{nt}</td>
</tr>
<tr>
<td>(b) Case 2(b): D_{nn} &lt; \mu</td>
<td>D_{nt}</td>
</tr>
<tr>
<td>3. Admissible contact conditions for Case (3)</td>
<td>177</td>
</tr>
<tr>
<td>(a) Case 3(a): D_{nn} = \mu</td>
<td>D_{nt}</td>
</tr>
<tr>
<td>(b) case 3(b): D_{nn} = \mu</td>
<td>D_{nt}</td>
</tr>
<tr>
<td>4. Limiting values of |Q_i^I|_{2} for all possible combinations of contact status between (z_i^I) and (z_i^I)</td>
<td>178</td>
</tr>
<tr>
<td>5. A comparison between analytic and numerical solutions (frictionless cases)</td>
<td>179</td>
</tr>
<tr>
<td>6. Effects of friction (refined mesh)</td>
<td>180</td>
</tr>
<tr>
<td>7. Frictional dissipation energy vs. (\mu)</td>
<td>181</td>
</tr>
<tr>
<td>8. CPU time consumed at each step of the developed solution algorithm</td>
<td>182</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>A two-body contact problem</td>
</tr>
<tr>
<td>2.</td>
<td>The motion trajectory of a representative contact point X&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>3.</td>
<td>Comparison between (a) Coulomb's friction model and (b) Drucker-Prager plastic flow model</td>
</tr>
<tr>
<td>4.</td>
<td>Comparison between two functions: (a) a non-differentiable function</td>
</tr>
<tr>
<td></td>
<td>(</td>
</tr>
<tr>
<td></td>
<td>(b) a regularized function (\varphi_\varepsilon(u_\tau))</td>
</tr>
<tr>
<td>5.</td>
<td>Comparison between (a) the interface model and (b) the conventional contact model</td>
</tr>
<tr>
<td>6.</td>
<td>Three possible cases of transition for an incremental loading stage: (a) contact to contact</td>
</tr>
<tr>
<td></td>
<td>(b) non-contact to contact</td>
</tr>
<tr>
<td></td>
<td>(c) contact to separation</td>
</tr>
<tr>
<td>7.</td>
<td>The consistency condition and the role of normal vector: (a) a plasticity problem</td>
</tr>
<tr>
<td></td>
<td>(b) a contact problem</td>
</tr>
<tr>
<td>8.</td>
<td>A simplified time integration scheme</td>
</tr>
<tr>
<td></td>
<td>(a) the normal direction</td>
</tr>
<tr>
<td></td>
<td>(b) developed contact forces</td>
</tr>
<tr>
<td>9.</td>
<td>Evaluation of the pairing point and the normal vector: (a) the pairing point</td>
</tr>
<tr>
<td></td>
<td>(1+\beta_K)</td>
</tr>
<tr>
<td></td>
<td>(b) a new nodal configuration for body 'b'</td>
</tr>
<tr>
<td></td>
<td>(c) the normal vector (n)</td>
</tr>
<tr>
<td>10.</td>
<td>Admissible domain for (\mathcal{F}) and (\mathcal{Z})</td>
</tr>
<tr>
<td>11.</td>
<td>Feasible solution region: Case (1): (D_{nn} \triangleright \mu</td>
</tr>
</tbody>
</table>
12. Feasible solution region: Case (2): \( D_{nn} < \mu |D_{nt}| \)

13. Feasible solution region: Case (3): \( D_{nn} = \mu |D_{nt}| \)

14. A hypothetical spring system illustrating Cases (1), (2), and (3)

15. The effect of friction coefficient on the rate of convergence
   (a) an illustrative example
   (b) the governing contact equation
   (c) the rate of convergence with respect to \( e^p \)
   (d) the rate of convergence with respect to \( e^z \)

16. Representation of the relation (4.3.12)

17. Solution algorithm

18. Compression of an elastic cylinder between rigid and flat dies (Plane-strain state)

19. FEM meshes and deformed shapes
   (a) coarse mesh with displacement loading
   (b) coarse mesh with traction loading
   (c) refined mesh with displacement loading
   (d) refined mesh with traction loading

20. A comparison between analytic and numerical solutions for frictionless cases (displacement loading)

21. A comparison between analytic and numerical solutions for frictionless cases (traction loading)

22. A comparison between types of loading
   (a) depth of compression vs. resultant force on the top surface of half-cylinder
   (b) distribution of reaction force on the top surface of half-cylinder

23. A comparison between analytic and numerical solutions for frictionless cases (summary)
   (a) half contact-width vs. applied force
   (b) max. normal traction vs. applied force
24. Effects of friction (refined mesh; $\delta=0.1$; variable-increment; ERON=1.0E-15) ..................... 209
   (a) contact traction
   (b) normalized traction ratio

25. Effects of friction (refined mesh; $\delta=0.3$; variable-increment; ERON=1.0E-15) ..................... 210
   (a) contact traction
   (b) normalized traction ratio

26. Effects of friction (refined mesh; $\delta=0.5$; variable-increment; ERON=1.0E-15) ........................ 211
   (a) contact traction
   (b) normalized traction ratio

27. Effects of number of load increments (refined mesh; $\delta=0.3$; $\mu=0.2$; ERON=1.0E-15) ..................... 212
   (a) contact traction
   (b) normalized traction ratio

28. Effects of number of load increments (refined mesh; $\delta=0.3$; $\mu=0.3$; ERON=1.0E-15) ..................... 213
   (a) contact traction
   (b) normalized traction ratio

29. Effects of number of load increments (refined mesh; $\delta=0.3$; $\mu=0.5$; ERON=1.0E-15) ..................... 214
   (a) contact traction
   (b) normalized traction ratio

30. Effects of allowable error norm (coarse mesh; $\delta=0.3$; $\mu=0.5$; NINC=1) .......................... 215
    (a) contact traction
    (b) normalized traction ratio

31. Effects of allowable error norm (coarse mesh; $\delta=0.3$; $\mu=0.5$; NINC=20) ........................ 216
    (a) contact traction
    (b) normalized traction ratio

32. Effects of allowable error norm (coarse mesh; $\delta=0.3$; $\mu=0.5$; NINC=100) ........................ 217
    (a) contact traction ratio
    (b) normalized traction ratio

33. Rate of convergence vs. friction coefficient (refined mesh; $\delta=0.3$; variable-increment) .... 218
34. Frictional indentational problems .............. 219
(a) flat punch (axisymmetric)
(b) spherical punch (axisymmetric)
(c) flat punch with transverse motion (plain-strain)

35. FEM meshes and deformed shapes
(a) axisymmetric flat punch .................... 220
(b) axisymmetric spherical punch .............. 221
(c) plane-strain flat punch with transverse motion ........................................ 222

36. Distribution of normal contact traction
(axisymmetric flat punch) ....................... 223

37. Distribution of normal contact traction
(axisymmetric spherical punch) ............... 224

38. Distribution of normal contact traction
(flat punch with transverse motion u; plane-strain condition) .......................... 225

39. Distribution of normalized traction ratio
(axisymmetric flat punch) ....................... 226

40. Distribution of normalized traction ratio
(axisymmetric spherical punch) ............... 227

41. Distribution of normalized traction ratio
(flat punch with transverse motion; plane-strain condition) .......................... 228

42. Stick/slip region vs. normalized traction ratio ........................................ 229

43. The ring compression test ..................... 230

44. Deformation modes and stick/slip regions .............................................. 231 - 232

45. Distribution of contact traction with various friction coefficients ................... 233

46. Normalized traction ratio with various friction coefficients .......................... 234

47. Shear stress distribution for different friction coefficients ....................... 235
48. Shear stress distribution for different Poisson’s ratio ........................................ 236
49. The indentation test for ceramic composites .. 237
   (a) an axisymmetric model
   (b) FEM mesh
50. Distribution of normal contact traction along fiber-matrix interface ...................... 238
51. Distribution of frictional traction along fiber-matrix interface .......................... 239
52. Distribution of normalized traction ratio along fiber-matrix interface .................... 240
53. Computed load-point displacement vs. coefficient of friction ......................... 241
54. Plane-strain hydraulic fracturing model ...... 242
55. FEM mesh and details of crack tip region ...... 243
56. The effect of crack tip element size on the stress intensity factor K .................. 244
57. Stick/slip response prediction ............... 245
   (a) stress ratio vs. coefficient of friction
   (b) modulus contrast vs. coefficient of friction
58. Crack opening mode stress intensity factor ratio vs. coefficient of friction for penetrating crack geometries ................................. 246
NOMENCLATURE

Latin Alphabet Symbols

\( b \) the body force vector

\( B(\cdot,\cdot) \) a bilinear form used in Section 2.4.3.

\( B, \partial B, \bar{B} \) a body, its boundary, the closure of a body

\( b \) the body force vector

\( c, C_1, C_2 \) constants

\( D \) a frictional dissipation power function

\( D, L, U \) the diagonal, lower, upper triangular part of \( M \)

\( ds, dS \) surface elements in the deformed, referential configuration

\( dv, dV \) the volume elements in the deformed, referential configuration

\( e \) a unit vector

\( e^p, e^z \) the error norm defined in (4.3.7,8)

\( E_{ijkl} \) the elasticity tensor

\( E, G \) Young's modulus, shear modulus

\( f(\cdot) \) the slip function in (2.3.4) or the yield function

\( f \) a prescribed boundary traction

\( f(\cdot) \) a linear functional in Section 2.4.3

\( F \) the deformation gradient or the nodal force vector

\( g \) a gap vector

\( G \) the Green's function or the associated matrix

\( H \) a matrix defined in (3.2.1)
\( H^1, H^{1/2}, H^{-1/2} \) the Hilbertian Sobolev spaces

I an identity matrix

I(\cdot) the indication function defined in (2.3.12)

j(\cdot,\cdot) a frictional energy defined in Section 2.3.3.

J the Jacobian of deformation

K the stiffness matrix

L,N the interpolation function for the displacement, contact traction

L^2,L^\infty spaces of Lebesque integrable functions

M,W the matrix, vector in the governing contact equation in Section 2.4

m,M positive constants

n,N unit normal vectors

N a null space

p,p_n,p_t the contact traction, its normal and tangential components

P the nodal contact force or the 2nd Piola-Kirchhoff stress tensor

\& the admissible contact traction field

Q a matrix defined in (4.3.12)

r the normalized traction ratio

R a transformation matrix from the global to the local coordinate system

R,U,V the rotation tensor, the right,left stretch tensors for F

\&,# abstract relations, constraints

\mathbb{R}^N a N-dimensional Euclidean space

s the deviatoric part of the Cauchy stress
\( S(\cdot) \) a mollified form defined in Section 2.3.3

\( S,T \) embedding mappings or the associated matrices

\( t,T \) unit tangential vectors

\( t,\Delta t \) time, time increment

\( T \) the Cauchy stress tensor

\( \underline{u},\underline{u} \) the displacement, the prescribed displacement

\( u_p,\underline{u_p} \) parts of the displacement vector in Section 3.2.

\( \wp,\wp \) the admissible displacement fields

\( v \) the velocity vector or the virtual displacement

\( V \) a dummy vector for the evaluation of matrix \( C \)

\( X, x \) a material particle, its position vector

\( z \) the incremental relative motion

\( \dot{z} \) the convective relative velocity

**Greek Alphabet Symbols**

\( \alpha,\beta,\gamma \) parameters concerning the integration scheme in Section 2.4.1

\( \beta,\beta_1 \) a positive constant defined in (4.3.13,14)

\( \tau \) the trace map in Section 2.4.3

\( \gamma_1,\gamma_2 \) parameters defined in Section 4.1.

\( \Gamma,\partial\Gamma,\overline{\Gamma} \) the configuration of boundary, its boundary, the closure

\( \varepsilon,\sigma \) the infinitesimal strain, stress tensors

\( \varepsilon,\lambda,\rho \) dummy constants

\( \kappa \) a parameter defining the pairing points
\( \lambda, G \)  
Lame constants

\( \mu \)  
the coefficient of friction

\( \nu \)  
Poisson's ratio

\( \xi \)  
a vector-valued dummy variable

\( \Pi \)  
a potential energy

\( \tau \)  
time

\( \phi, \phi_j, \phi_G \)  
contractive mappings

\( \varphi_\varepsilon(\cdot), \omega_p(\cdot) \)  
functions used in Section 2.3.3

\( \chi, \chi_\overline{X} \)  
the motion, its extension

\( \psi \)  
a matrix defined in (4.3.18)

\( \Omega, \Omega_\overline{\Omega} \)  
the configuration of a body, its extension

\( \kappa \)  
the pairing map

**Superscripts**

0,1,2  
the undeformed, referential, deformed state corresponding to \( \tau = 0, t, t+\Delta t \)

a,b  
body 'a', 'b'

i,j, etc.  
iteration counters

p  
associated with the pairing points

\( \alpha, \beta, \gamma \)  
indices concerning the integration scheme in Section 4.1.

**Subscripts**

0,1,2  
the undeformed, referential, deformed state or vector components
a part concerning contact

index denoting the dimension

nodal number counters

the normal, tangential component

the referential normal, tangential component
CHAPTER I
INTRODUCTION

1.1. Opening Remarks

Accurate analysis of contact phenomenon associated with solids has long been of great interest in various areas of applied mechanics and engineering. When deformable bodies come into contact under the action of external loads, their behavior is subjected to various complicated mechanisms attributed to the mechanical interaction of solid surfaces as well as the constitutive relations for the respective bodies. From the vantage point of mechanics, surface tractions are developed on a contact surface resulting from the penetration of one body into the other as well as a frictional resistance to the relative motion between the two surfaces. The contact traction and the contact surface change continuously during the progressive action of external loads in accordance with the deformation of parent bodies. Moreover, friction, one of the most common mechanisms of energy dissipation, is always accompanied by heat generation and surface wear. In this respect, the subject of contact mechanics entails the construction of an
integrated mathematical model which characterizes both the constitutive properties of parent bodies and various contact phenomena in a unified manner under the framework of field theories in continuum mechanics. The development of robust methods of solution for general contact problems is also an essential component.

For convenience, general contact problems are sometimes classified into the following categories[32]: (1) conforming or non-conforming problems depending on whether the undeformed shapes of candidate contact surfaces are similar or non-similar, respectively, (2) frictional or frictionless problems, (3) elastic or inelastic problems, (4) dynamic or quasi-static problems depending on the presence of inertia effects, and so on. Rolling contact problems are often referred to as those in which one or both of the bodies in contact are subjected to a gross rotational motion. For most stress analyses of deformable bodies, the effects of heat generation and surface wear are neglected although they are of great importance in the analyses of tool wear, abrasive machining processes, etc.

Some important areas of applications may be grouped as follows. First, there are many machine elements with their performance governed by the relative motion between contact surfaces (gears, cams, bearings, ball-and-socket joints, wheels, etc.) Most problems in this group correspond to the non-conforming elastic contact problems, and so the contact
stresses are especially high since the area of contact is small. The effects of friction are often neglected because the machined surfaces of these elements are usually well-finished and well-lubricated. Secondly, stresses and flow pattern of a workpiece in metal forming processes are greatly influenced by frictional resistance and heat generation on a die-workpiece interface. Most non-steady state forming processes such as forging, sheet metal forming, etc. correspond to the non-conforming, quasi-static and inelastic contact problems. For steady state forming processes such as extrusion, rolling, etc., the contact region and contact status are stationary in space. Finally, frictional resistance also plays an important role in mechanical fastening and fitting of machine elements, mechanics of composite materials and multi-layered geological media. Most problems in this group belong to the conforming contact classification and there is a relatively small amount of slippage. Thus, the assumption of 'perfect-bonding' between bodies in contact is often employed for simplicity of analysis.

Contact problems are inherently very complicated even when other sources of non-linearities and thermal effects are ignored. Contact problems are intrinsically non-linear since the contact traction changes in a highly non-linear fashion under a progressive action of external loads. Such a non-linearity can be attributed to three different
sources: viz., changes of the contact boundary, the local normal direction of contact surfaces, and the stick/slip region in the progress of deformation. The first two sources are related to geometric changes of contacting bodies, and contact problems can thus be considered to be geometrically non-linear regardless of the underlying deformation theory of each continuum. The last source is, of course, related to friction problems with the friction laws of Coulomb type. Friction problems are also path-dependent with the instantaneous relative motion between bodies in contact strongly depending on the most recent state of the contact traction. Thus, special attention must be paid to the description of such behavior in the mathematical formulation and associated solution strategy. It is known that inaccurate evaluation of the trajectory of contact points often results in a physically unrealistic solution. In addition to the above complexities, neither the contact traction nor the contact region is known a-priori as a function of time or a time-like loading parameter. In other words, the boundary condition on a part of the boundary of each body is not specified explicitly, but is given in the form of both equality and inequality constraints imposed on each pair of contact points. The contact point pairs change continuously during the deformation. In view of their mathematical structure, contact problems belong to a class of variational
inequality problems because the kinematically and/or the kinetically admissible field(s) of a state variable(s) is constrained unilaterally. Thus, they generally require very sophisticated methods of solution.

Considerable efforts have been devoted to studying the solution methodologies for general contact problems since the pioneering work of H. Hertz, On the Contact of Elastic Solids, was presented in 1881. Complex stress functions and integral transform methods have been widely used to solve contact problems analytically. However, the closed form solution is usually limited to some idealized problems with simple geometries. On the other hand, computational mechanics has played an important role recently in advancing non-linear mechanics. For general contact problems, various numerical methods of solution have been proposed, mostly associated with the finite element methods. These methods mainly deal with the improvement of iterative solution schemes for treating unknown and non-linear contact boundary conditions iteratively. Despite the advent of various solution methods, several important questions still remain unresolved on issues such as realistic mathematical modeling of friction problems, reliability of the respective iterative scheme, and general applicability.

In the present work, a new iterative method of solution associated with the finite element method is proposed for two-dimensional elastostatic contact problems. This method
eliminates some drawbacks of existing methods dealing with the accuracy of solution for friction problems and the convergence of the associated iterative scheme. The general structure of mathematical contact problems with Coulomb's friction law is first studied under the framework of non-linear field theory in continuum mechanics. The incremental contact boundary-value problem is then formulated in accordance with the updated Lagrangian formalism so that the contact point pairs and the local normal direction of contact surfaces can be updated at each incremental stage. Instead of taking a full account of the kinematic relations between configurations, the relations are assumed constant during the incremental stage and determined simply with respect to an 'assumed' deformed configuration. Then, both the displacement of each body and the contact traction are taken as independent state variables in the present variational formulation. Therefore, the variational equation for the equilibrium of each body is in a bilateral form with respect to the unknown displacement field for the respective body. Also, the equations dealing with the contact conditions between two bodies appear in a unilateral form with respect to the unknown contact traction field. By applying the general procedures of the finite element method and manipulating the discretized equations, the inequality problem is transformed into a non-linear fixed point problem for the unknown nodal
contact force in a finite dimensional vector space. In the present study, the non-linear fixed point problem is solved for the nodal contact traction iteratively by adopting the block Gauss-Seidel iterative method. The existence and uniqueness of solution for elastostatic contact problems with friction is discussed within the finite dimensional space, and the convergence characteristics of the present iterative scheme are also discussed.

The solution algorithm appears to be very stable, and the coefficient of friction has little influence on the solution convergence. The computational efficiency and general applicability of the present method are demonstrated through a variety of numerical examples. It is also shown that the manner of evaluating the trajectory of contact nodes may have a great influence on the accuracy of solution for frictional contact problems.

Apart from the development of the solution algorithm for frictional contact problems, the general structure of constitutive equations for a class of inelastic materials is studied within the framework of continuum thermodynamics with internal state variables. It is shown that the objective stress rate for the description of constitutive equations with finite strain can be determined uniquely on the basis of the above formalism. Existing constitutive models for thermo-elasto-viscoplastic materials are then examined with the proposed theory.
1.2. Literature Review

1.2.1. Friction law

The topics concerning the interaction between solid surfaces, especially the subject of friction, may start with the understanding of complicated topographical and constitutional properties of a real solid surface. In fact, a solid surface is topographically very rough and irregular and it is almost impossible to reproduce the real topography of the surface. Instead, the topographical characteristics are usually expressed in engineering terms such as roughness, waviness, etc. In addition to the topographical complication, a real surface is not clean in the sense that it is always covered with various forms of contaminants and wear debris, and with oxide film in cases of metallic surfaces. Also, the thermomechanical properties of the material in the vicinity of the surface are generally much different from those of the substrate material.

Among many phenomena attributed to the mechanical interaction between rough surfaces, friction has long been of particular interest to both applied physicists and engineers. A great deal of experimental and theoretical research has been compiled by Bowden and Tabor [5], Rabinowicz [66], Kragelsky et al. [39], etc. Recently, Oden and Martins [60] have presented a review paper on friction theories with a focus on their adaptability to continuum models.
It may be remarked that the subject of friction termed as tribology began with the pioneering activity of Leonard da Vinci in the sixteenth century. The progress in friction theories is then largely divided into two types of work in accordance with the underlying tribological concepts.

The earlier theories of friction developed during the period from the seventeenth to the nineteenth centuries are mostly based on the idea that the origin of the friction force is attributed to the riding of rigid asperities of one surface over the other. The most important work during this period may be the foundation of the friction laws based on the concept of the coefficient of friction, first introduced by Amontons in 1699. However, the designation "Coulomb's law" is now cited after Coulomb's work on the investigation of the distinction between static and dynamic friction in 1781.

The modern theories of friction are based on the idea that the origins of the friction force are mainly attributed to the formation and shearing of junctions between asperities and plowing of softer material by hard asperities. Bowden and Tabor elaborated the well-known adhesional theory of friction during the middle of this century and a large number of investigators have to date clarified or expanded the friction theories based on this junction model. According to these theories, the significant portion of the resultant friction force is the
average shear strength of the junctions multiplied by the real area of contact which is usually only a small fraction of the apparent area of contact. Some observations on friction theories and friction laws are summarized below.

Despite the evolution in the underlying concepts of tribology, the structure of friction laws has been retained in the same form as that of Coulomb's law although the nature and form of the coefficient of friction has been much improved. In general, the coefficient of friction is independent of the applied load and the apparent area of contact. However, it strongly depends on the combination of the parent material properties and the nature of lubrication. At the present time, no single theory can explain the friction phenomena completely, and the effect of the surface roughness on the coefficient of friction has not been entirely clarified yet. In this respect, it is concluded that friction laws of the Coulomb's law type are still useful and some typical values of the coefficient of friction tabulated in open literature can be used for most engineering applications.

1.2.2. Continuum contact model with friction

A great deal of effort has been conducted in constructing a general structure of constitutive models for inelastic or materials incorporating damage theories within
the framework of continuum thermodynamics with internal state variables. Some work related to this topic is presented in Appendix A and B.

On the other hand, general contact problems have rarely been studied on the basis of the theories of continuum thermodynamics despite the similarities between plastic and friction phenomena. Up to date, most contact problems have been analyzed using purely mechanical and infinitesimal elastic deformation theory.

In formulating a continuum model for general contact problems, the first step may involve the incorporation of the phenomenological contact laws into fundamental continuum models designed for describing the responses of parent bodies. There are generally two alternative methods, namely, the interpretation of contact laws as boundary conditions based on the kinematic constraint of non-penetration, and the introduction of a special interface element between two bodies for simulating contact laws.

Herrmann [21] and Heuze and Barbour [22] used various artificial interface elements in the finite element analysis for geomechanics problems, attempting to simulate Coulomb's law of friction with the adjusted stiffness of interface elements attached between possible contact surfaces. This method seems to be applicable to a limited class of problems i.e. perfectly-bonded cases. Recently, Oden and Martins [60] developed an interface model designed to simulate some
physical properties of a real interface observed from a number of experimental results. This model is further discussed in Section 2.3.3.

Friction laws developed on the basis of tribological concepts mentioned before must be interpreted or modified in an appropriate manner in accordance with the underlying continuum concepts. In tribology, they are expressed in terms of the resultant forces against the gross sliding motion without paying much attention to the local stick/slip motion during a progressive deformation. The so-called pointwise version of Coulomb's friction law is widely used in the field of computational mechanics. Several alternative versions have also been proposed and they may be classified into two groups depending on the theoretical framework.

Several attempts [12,16,69, etc.] have been made in constructing an analytic friction model inspired by the classical theory of plasticity with the non-associated flow rule. Here, the term 'classical' is used to represent plasticity theories based on the notion of yield criteria. Recently, Cunier [12] elaborated a friction model governed by three different mechanisms called adherence, tear, and wear by analogy with elastic, kinematic-, and isotropic-hardening mechanisms, respectively.

The mathematical structure of mechanics problems with inequality constraints (typically contact and plastic
problems) has been studied rigorously with the use of extended concepts of differentiation and optimization in recent years. Several general forms of the contact laws with friction have been studied [13,38,65] in view of subdifferential boundary conditions. Oden and co-workers [57-61] have consistently strived to develop mathematically sound friction models focused on physical aspects. They proposed two versions of Coulomb's friction law contained in a single model called the non-linear and non-local friction law. Both friction models proposed by Oden et al. and Curnier are discussed in Section 2.3.3.

The final step involves the introduction of the non-linear and path-dependent nature of general contact problems in a physically sound continuum model. In a sense, the terms involving classical and incremental Coulomb’s friction laws may not be adequate because friction problems are intrinsically path-dependent and 'incremental' in nature. Therefore, a suitable technique of path-integration for general path-dependent problems has to be developed.

1.2.3. Solution method

Except for a limited number of closed form solutions [17,50], most contact problems with or without friction have been numerically evaluated mostly by the finite element method. Various approximate methods and numerical solution schemes have been proposed in the last decade.
Frictional contact problems have intrinsic mathematical complications associated with the inequality relations for the contact boundary conditions; i.e., the non-penetration condition, the slip criterion, and the requirement of non-negative frictional dissipation of energy. Each of the above inequality conditions then serves as the role of kinematical, kinetical, and constitutional inequality constraints, respectively, as discussed in Chapter 2. Due to such inherent complications, numerical solution methods are not accompanied by an appropriate variational method.

Several variational principles have been proposed for contact problems with Coulomb's friction law. Duvaut and Lions [13], Kalker [33], Hughes et al. [26], Oden and co-workers [57-61], etc. have studied several forms of variational formulations under the assumption that normal or tangential components of the contact traction is known a-priori. They reported that the primal method of the variational formulation results in a mathematically ill-defined form for Signorini's problems with Coulomb's friction law. As a modification, Oden and co-workers [57,59] have studied a variational principle with the non-local friction law instead of the pointwise version of Coulomb's friction law. In the meantime, most numerical solution methods have been developed under the assumption that the contact traction is known a-priori so that the associated variational problem can be simply replaced by the
principle of virtual work subjected to the bilateral virtual displacement field.

Various iterative solution schemes have been proposed for the effective resolution of the inequality constraints concerned with finite dimensional contact problems with or without friction. They may be largely classified into several groups depending on the manner of the resolution of the inequality constraints: e.g., trial-and-error methods with elaborate decision tables [7-9,15,30,35,45,62,68, etc.], penalty or Lagrangian methods [4,6,36,53,56,70, etc], mathematical programming techniques [29,38,41,42,76, etc.] and so on. Critical reviews of these solution methods are presented in References [30,45].

In order to improve the effectiveness of solution algorithms, Francavilla and Zienkiewicz [15] used a reduced flexibility matrix for frictionless contact problems. This technique has been extended to static and dynamic contact problems with friction in References [30,34,44,45], and efficient methods for the computation of reduced flexibility matrices have also been studied. They have also developed a convergent iterative solution method for elastostatic contact problems and then applied the methodology to several geomechanics problems.

Okamoto and Nakazawa [62] and Torstenfelt [77] have increased the externally applied load by an amount which causes a change in the contact status of only one nodal
point at a time in order to improve the accuracy of solution for path-dependent friction problems. Fredriksson [16] used a similar technique to account for a special slip rule for friction.

In addition to the above methods, the mixed finite element method [23, 79] and the boundary element method [74] have also been used instead of the displacement finite element methods. Also, an extension of solution methods to three-dimensional [10], and geometrical and/or materially non-linear problems [31, 67] has been attempted recently.

1.3. Outline of the present study

The main objectives of the present work are to develop a variational principle, incorporating the path-dependence, within the framework of continuum mechanics and to develop an efficient computational model based on the theory for two-dimensional elastostatic contact problems with friction. Of particular interest are the accuracy of solution for friction problems and the convergence of the associated iteration scheme.

A general mathematical structure of contact problems with Coulomb's friction law is studied in Chapter II. In Section 2.2., a continuum concept of the pairing map is developed for adequately describing the path-dependent nature of friction problems. In Section 2.3., the contact
conditions are expressed in the non-integrable rate form with use of a convective relative velocity. The consistency condition for friction problems is compared with that for plasticity problems.

In Section 2.4., a semi-implicit scheme is proposed for the effective numerical time-integration of the contact conditions. The incremental contact boundary value problem is formulated on the basis of the proposed time-integration scheme.

A variational principle is proposed for elastic contact problems with friction on the basis of 'mini-max' principles. Both displacement and contact traction vectors are taken as independent state variables. The mathematical equivalence between the variational and the boundary value problems is established in the context of functional analysis.

In Chapter III, the variational continuum problem is discretized into a finite dimensional contact problem by employing the standard finite element method and the concept of node-to-segment contact.

As a central part of the present numerical scheme, the method of embedding the contact conditions into the discrete governing contact equations is developed in Chapter IV. The associated non-linear surjective mapping is called the embedding map. With the help of the embedding map, the discrete problem can be represented in the form of a fixed
point problem for the nodal contact force vector. As a consequence, the existence and uniqueness of solution is studied using the Banach fixed point theorem. An iterative scheme based on the block Gauss-Seidel method is proposed for solving the non-linear fixed point problem. Convergence criteria for the iterative scheme are also studied.

The validity of the present solution method is then demonstrated by several numerical examples in Chapter V. Several factors which influence the accuracy of the solution including the number of load increments, prescribed allowance for the error norm, finite element meshes, are investigated. The numerical convergence of the iterative scheme is also examined through these examples.

Finally, conclusions and recommendations for future research are presented in Chapter VI.
2.1. Introduction

Much progress has been made in the non-linear field theories of mechanics and in the computational methodologies for approximate solutions of the respective initial/boundary value problems in recent decades. Continuum mechanics and computational mechanics are so closely related to each other that a well-posed continuum model also turns out to be a sound computational model.

A great deal of effort has been made to study the mathematical structures of continuum contact problems with Coulomb's friction law and to develop an effective solution method associated with the finite element methods. Despite the advent of various solution methods, several important questions still remain to be resolved and refined. For instance, the fundamental questions associated with friction in continuum sense, the existence of a unique solution and variational principles will remain difficult research issues for a long time to come. On the practical side of the research issues, the reliability, convergence and path-dependence of discrete solutions must also be
investigated. In a sense, the above difficulties are attributed to the intrinsic complications due to coupling effects between friction and non-penetration constraints.

An investigation of the mathematical structure of general contact problems within the framework of non-linear field theories of continuum mechanics is presented in this chapter. Basic relations between state variables of contacting bodies are studied by introducing additional variables to describe contact boundary, pairing map, and contact tractions. In section 2.3., contact conditions are represented in several different forms involving multiple-choice type boundary conditions. Also, some similarities and differences between the friction and classical theories of plasticity are discussed. Some of recent modifications of Coulomb's friction model are reviewed focusing on the motivations and consequences. In Section 2.4, time-integration schemes are discussed for the effective resolution of the frictional contact condition in the non-integrable rate form. Variational principles for incremental contact boundary-value problems are also discussed. A new variational method is proposed on the basis of 'mini-max' principles.

2.2. Basic Relations

Some fundamental concepts and terminologies established
in continuum mechanics (see general references, e.g. [46,78]) are first recapitulated briefly in this section. Additional kinematic and kinetic variables are then defined for the description of continuum contact problems.

2.2.1. Preliminary concepts and notations

A two-body contact problem is depicted in Fig. 1, where the superscripts 'a' and 'b' are used to distinguish each body. Bodies 'a' and 'b' may be chosen arbitrarily, but a deformable body is, for convenience, taken as body 'a' when the other is assumed to be a rigid body.

Each body is conceived as an infinite set $B$ of material particles $X$ endowed with not only a non-negative measure called the mass but also a material coordinate system to establish a one-to-one correspondence between each particle and an ordered $N$-tuple of real numbers ($N = 2$ or $3$). The configuration of a body is now defined as a smooth and invertible mapping of the body onto an open bounded region $\Omega$ of $N$-dimensional Euclidean space $\mathbb{R}^N$ bounded by a closed surface $\Gamma$. A one-parameter family of configurations is then referred to as the motion of a body, and a real parameter $\tau$ denoting time. The motion is usually represented by a smooth and invertible map $\chi : B \times [0, \infty) \to \Omega$ such that

$$x = \chi(X, \tau),$$

(2.2.1)

where $X \in B$, $\tau \in \mathbb{R}^+$, and the position vector $x \in \Omega \subset \mathbb{R}^N$. 
denotes the place occupied by a particle \( X \) at time \( \tau \). The inverse of the motion \( \chi \) is defined as

\[
X = \chi^{-1}(x,\tau).
\]  

\( (2.2.2) \)

The behavior of a continuum is generally described in any of three different methods: viz., the material, the referential (Lagrangian), and the spatial (Eulerian) methods of description, which are all mathematically equivalent as long as the motion \( \chi \) is smooth and invertible. Thus, the preference for a particular method generally results from the convenience of describing the respective continuum model and of formulating initial/boundary-value problems. For most solid mechanics problems, the referential method is widely used because of the convenience of describing boundary conditions and ease of formulating incremental boundary-value problems.

Two particular configurations denoted by \( ^0\Omega \) and \( ^1\Omega \) are used to represent the region of a body at two fixed instants \( \tau = 0 \) and \( \tau = t \), designating the undeformed and referential configurations, respectively. On the other hand, the configuration \( \Omega \) at a variable time \( \tau \) is, for convenience, called a deformed configuration. Here, \( t \) denotes a fixed referential time\(^\dagger\), and the term 'time' may

\( \dagger \) it is sometimes called the current time in accordance with the updated Lagrangian formalism, but it must be distinguished from the variable time \( \tau \).
also be interpreted as a time-like loading parameter. The positions of a representative particle X in each of the above undeformed, referential, and deformed configurations are denoted by \( o_x \), \( i_x \) (usually by X), and \( x \), respectively.

A tensor-valued material description \( \tilde{\phi}(X, \tau) \) defined for the body can be transformed into the referential and spatial description as follows:

\[
\tilde{\phi}(X, \tau) \equiv \phi(x^{-1}(X, t), \tau)
\]

in the referential form and

\[
\phi(x, \tau) \equiv \phi(x^{-1}(x, \tau), \tau) = \tilde{\phi} \circ x^{-1}(x, \tau)
\]

in the spatial form, where the symbol '\( \circ \)' denotes the composition of two functions. Here, and in general, the left-subscript of a state variable is used to indicate that its space domain corresponds to the referential configuration at a specified time (e.g., \( o\phi(0_x, \tau) \), \( i\phi(X, \tau) \), etc.). On the other hand, a state variable without a left-subscript indicates that it is expressed in the spatial form (e.g., \( \phi(x, \tau) \)). Also, the left-superscript is used to denote its value at a specified time (e.g., \( 1\phi(X) = 1\phi(X, t) \), \( 2\phi(X) = 1\phi(X, t+\Delta t) \), etc.). The superscripts \( 0, 1 \), and \( 2 \) then designate three representative instants \( \tau = 0, t, t+\Delta t \), respectively. For notational simplicity, however, the left-subscript \( i \), in the sequel, is omitted unless a distinction has to be made.

A bold character is usually used to denote a tensor with an appropriate order greater than zero. The index
notation and the summation convention are also adopted with respect to a fixed rectangular Cartesian system, unless otherwise specified, with \( N \) orthonormal base vectors \( \{e_i\}_{i=1}^N \) such that \( x_i = x \cdot e_i \), \( P_{ij} = e_i \cdot P e_j \), etc., where \( P = P_{ij} e_i \otimes e_j \) denotes a second order tensor and the usual summation is implied for repeated indices.

2.2.2. Contact surface

The boundary of a body, \( \partial B \) (and its deformed configuration \( \Gamma \)) is divided into three distinctive parts; viz., \( \partial B_u \) (\( \Gamma_u \)) where the displacement is prescribed, \( \partial B_f \) (\( \Gamma_f \)) where the traction is prescribed, and a possible contact segment \( \partial B_0 \) (\( \Gamma_0 \)) where neither the displacement nor the traction is prescribed. It is assumed that \( \partial B = \overline{\partial B_u} \cup \overline{\partial B_f} \cup \overline{\partial B_0} \) and the intersection of one type of boundary with another is always empty. It may also be assumed without loss of generality that the type of boundary at each boundary point does not change during the incremental deformation (say, \( \tau \in [t,t+\Delta t] \)) except for the contact boundary \( \partial B_c(\tau) \subset \partial B_0 \). Therefore, the contact boundary \( \partial B_c(\tau) \) is always treated as a variable subset of the possible contact boundary \( \partial B_0 \) with respect to time \( \tau \).

It is now assumed that there exists a unique continuous extension \( \overline{\chi} \) of the motion \( \chi \) defined in (2.2.1) such that \( \overline{\chi} : \overline{B} \times [0,\omega) \to \overline{\Omega} \), where \( \overline{B} \) and \( \overline{\Omega} \) denote the closure of the
sets $B$ and $\Omega$ (i.e., $\overline{B} = B \cup \partial B$ and $\overline{\Omega} = \Omega \cup \Gamma$), respectively. In the sequel, $X \in \partial B$ and $x \in \Gamma$ are called a boundary point and its position at time $\tau$, respectively.

The deformed configuration $\Gamma_c$ of the contact boundary at time $\tau$ can be defined as its closure $\overline{\Gamma_c}$ such that

$$\overline{\Gamma_c} = \overline{\Gamma_c^a} \cap \overline{\Gamma_c^b} = \{ x \in \mathbb{R}^N \mid x \in \overline{\Gamma_c^a} \text{ and } x \in \overline{\Gamma_c^b} \},$$

with the assumption that the non-penetration condition (i.e., $\overline{\Omega^a} \cap \overline{\Omega^b} = \emptyset$) always holds. In general, the contact surface may consist of a finite number of disconnected open regions. Thus, $\Gamma_c$ and $\partial \Gamma_c$, in the sequel, represent the totality of the open regions and their boundaries such that

$$\Gamma_c = \bigcup_{k=1}^{m} (\Gamma_c)_k, \quad \partial \Gamma_c = \bigcup_{k=1}^{m} (\partial \Gamma_c)_k, \quad \text{and } \overline{\Gamma_c} = \Gamma_c \cup \partial \Gamma_c,$$

where $m$ denotes a positive integer. It is noteworthy that the boundary of contact surface, denoted by $\partial \Gamma_c$, includes not only a geometric boundary but also the region where the contact traction vanishes. In other words, it includes the entire region where separation can occur instantaneously.

It is also assumed that each region is smooth enough for the unit outward normal vectors $\mathbf{n}^a$ and $\mathbf{n}^b$ to $\Gamma^a$ and $\Gamma^b$, respectively, to be continuous functions of position and time. In the sequel, the unit normal vector to the contact surface is simply represented by a single normal $\mathbf{n}(x,\tau)$ defined by
n(x, τ) = n^a(x, τ) = - n^b(x, τ) \hspace{1cm} (2.2.3)

for each \( x \in \Gamma_c \) and \( \tau \).

2.2.3. **Pairing map and relative motion**

Each point \( x \in \Gamma_c \) is occupied by two boundary points \( X^a \) and \( X^b \) belonging to different body sets \( B^a \) and \( B^b \). Points belonging to \( \partial B^a_c(\tau) \) and \( \partial B^b_c(\tau) \) are termed as a contact point and a pairing point, respectively, for convenience. Fig. 2 depicts a trajectory of a contact point \( X^a \) which comes into contact with its pairing points \( X^b_1 \) at \( \tau = t \) and \( X^b_2 \) at \( \tau = t + \Delta t \). In order to describe these changes of the pairing points with respect to each contact point, a pairing map

\[ \mathcal{K} : \Gamma^a_c(\tau) \to \Gamma^b_c(\tau) \]

can be defined in the referential form such that

\[ X^b = \mathcal{K}(X^a, \tau) , \hspace{1cm} (2.2.4) \]

where \( \Gamma^a_c(\tau) \) and \( \Gamma^b_c(\tau) \) denote the region of the deformed contact surface \( \Gamma_c = \Gamma^a_c = \Gamma^b_c \) at time \( \tau \) in the referential configuration of each body, respectively.

The pairing map is introduced as an auxiliary variable for the convenience of describing the frictional behavior and the local equilibrium condition, as explained later. In fact, it can be related to the motion of contacting bodies by the relation,
\begin{equation}
\kappa(X^a, \tau) = (\chi^b)^{-1} \circ \chi^a_c(X^a, \tau), \tag{2.2.5}
\end{equation}

where \( \chi^a_c \) denotes the restriction function associated with the map \( \chi^a \) for a part of its domain \( \Gamma^a_c \subset \bar{\Omega}^a \). Recall that \( \chi^a \) is the motion in the referential form defined for the closure \( \bar{\Omega}^a \) and time \( \tau \), and the symbols \(( \ )^{-1}\) and \(\circ\) denote the inverse and the composition of the designated functions, respectively.

The term 'relative motion' has physical significance when the motion between two particular particles is concerned. However, the relative motion between a particle and a (deformable or rigid) body has little meaning unless a particular particle in the body is specified. Such a particle (boundary point) is called the pairing point in the present study, which changes continuously during the deformation as described by means of the pairing map. In order to define the relative motion between such pairs of contact points, a vector-valued function, called the 'gap vector' is first introduced.

\begin{equation}
\xi(X, \tau) = \chi^a(X, \tau) - \chi^b(\kappa(X, \tau), \tau), \tag{2.2.6}
\end{equation}

\begin{equation}
= (\chi^a - \chi^b \circ \kappa)(X, \tau).
\end{equation}

According to the definition of the pairing map in (2.2.5),

\begin{equation}
\xi(X, \tau) = 0 \quad \text{for each } X \in \Gamma^a_c(\tau) \tag{2.2.7}
\end{equation}
which is a consistency condition stating that any pair of contact points must occupy the same point in space.

The material time derivative of the function \( F \) with \( X \) fixed can then be written as

\[
\dot{F}(X, \tau) = \dot{x}^a(X, \tau) - \dot{x}^b(\mathcal{R}(X, \tau), \tau),
\]

\[
= v^a(X, \tau) - v^b(\mathcal{R}(X, \tau), \tau) - F^b(\mathcal{R}(X, \tau), \tau) \dot{R}(X, \tau),
\]

where \( v \) and \( F \) are the velocity and the deformation gradient of the respective body; i.e.,

\[
v^a(X, \tau) = \frac{\partial x^a}{\partial \tau} \bigg|_{\text{fixed } X},
\]

\[
v^b(\mathcal{R}(X, \tau), \tau) = \frac{\partial x^b}{\partial \tau} \bigg|_{\text{fixed } \mathcal{R}},
\]

and \( F^b(\mathcal{R}(X, \tau), \tau) = \frac{\partial x^b}{\partial \mathcal{R}} \bigg|_{\text{fixed } \tau} \).

The superposed dot denotes the material time derivative with \( X^a \) fixed, and 'fixed \( \mathcal{R} \)' denotes a partial derivative with the pairing point \( X^b \) fixed (i.e., the same as the conventional partial derivatives defined for body 'b'). It is noteworthy that the material time derivative \( \dot{\mathcal{R}} \) can be defined only for the open region of the contact surface \( \Gamma^a_c(\tau) \).

Each of the above three parts (i.e., \( v^a, v^b, \) and \( F^b \)) of the relative velocity \( \dot{F} \) between contacting bodies is illustrated for a representative point \( X^a \) in Fig. 2. Here, \( v^a \) and \( v^b \) are the velocities of a contact point \( X^a \) and its
pairing point \( \mathbf{x}^b = \mathbf{r}(\mathbf{x}^a, \tau) \), respectively, and \( F^b \mathbf{v} \) represents the convective relative velocity. It can be shown that the normal component of the convective relative velocity always vanishes for \( \mathbf{x} \in \mathbf{\Gamma}_c^a(\tau) \) because the vector \( \mathbf{v} \) always lies on the tangent plane of the contact surface in the referential configuration and so does \( F^b \mathbf{v} \) in the deformed configuration.

It is noted that the convective relative velocity is directly related to the frictional power dissipation as discussed in Section 2.3 and it is obviously of the non-integrable form unless \( F^b \) is constant for any \( \tau \), which occurs when body 'b' is rigid and stationary. In order to satisfy the consistency condition (2.2.7) at all times, any contact point \( \mathbf{x} \in \mathbf{\Gamma}_c^a(\tau) \) must satisfy

\[
\hat{x}(\mathbf{x}, \tau) = 0 , \quad \forall \mathbf{x} \in \mathbf{\Gamma}_c^a(\tau) \tag{2.2.9}
\]

and

\[
\hat{z}_n(\mathbf{x}, \tau) \leq 0 , \quad \forall \mathbf{x} \in \partial \mathbf{\Gamma}_c^a(\tau) \tag{2.2.10}
\]

where

\[
\dot{z}(\mathbf{x}, \tau) \equiv \dot{\mathbf{v}}^a(\mathbf{x}, \tau) - \dot{\mathbf{v}}^b(\mathbf{r}(\mathbf{x}, \tau), \tau)
\]

and

\[
\dot{z}_n(\mathbf{x}, \tau) = \dot{z}(\mathbf{x}, \tau) \cdot \mathbf{n}(\mathbf{x}, \tau)
\]

Eqns (2.2.9) and (2.2.10) may be interpreted as the consistency condition (i.e., each contact point \( \mathbf{x} \in \mathbf{\Gamma}_c^a(\tau) \) remains on the contact surface at all times) and the non-penetration condition (i.e., \( \Omega^a \cap \Omega^b = \emptyset \) ) expressed in
the rate form, respectively. It is noted here that the symbol \( \dot{z} \) defines the convective relative velocity and does not mean the material time derivative of \( z \) which is not defined. More details of these conditions are discussed in Section 2.3.

In order to remove the dummy variable \( \xi \) in the sequel, the consistency condition in (2.2.9) is now rewritten as

\[
\ddot{z}(X, \tau) = F_b(K(X, r), T) \mathbf{N}(X, \tau) .
\]

(2.2.12)

From the fact that the normal component of the convective relative velocity always vanishes, the non-penetration condition (2.2.10) can also be written as

\[
\dot{z}_n(X, \tau) \leq 0 . \quad \forall X \in \Gamma_c^a(\tau)
\]

(2.2.13)

2.2.4. Contact traction

Vector-valued functions \( \mathbf{p}^a(X^a, \tau) \) and \( \mathbf{p}^b(X^b, \tau) \) are now introduced to denote the contact traction per unit area of the contact surface in the referential configuration \( \Gamma_c^a(\tau) \) and \( \Gamma_c^b(\tau) \), respectively. The contact traction vector \( \mathbf{p} \) is simply defined by

\[
\mathbf{p}(X, \tau) = \mathbf{P}(X, \tau) \mathbf{N}(X) .
\]

(2.2.14)

where the first kind of Piola-Kirchhoff stress tensor \( \mathbf{P}(X, \tau) \) (PK-1 stress for brevity, generally unsymmetric) is employed and \( \mathbf{N}(X) \) is the unit vector outward normal to \( \Gamma_c \) (i.e., the boundary of a body in the referential configuration at \( \tau = t \)), i.e., \( \mathbf{N}(X) = \mathbf{n}(X, t) \).
Using $dS$ and $ds$ to denote the surface elements for a body in the referential and deformed configurations, respectively, it can then be shown that

$$F^T(X, \tau) n(X, \tau) \, ds = J(X, \tau) \, N(X) \, dS , \quad (2.2.15)$$

where $F^T$ and $J = \det(F)$ denote the transpose and the determinant of the deformation gradient $F$, respectively. From the fact that $n^a(x, \tau) = - n^b(x, \tau)$ and $ds^a = ds^b$ for each $x \in \Gamma_c$, the relation between the referential surface elements of contact bodies $dS^a$ and $dS^b$ can be related such that

$$(\mathbf{v}\mathbf{\kappa})^T \, N^b \, ds^b = - \det(\mathbf{v}\mathbf{\kappa}) \, N^a \, dS^a \quad (2.2.16)$$

for each pair of points $X^a \in \Gamma_c^{a}(\tau)$ and $X^b = \mathbf{K}(X^a, \tau) \in \Gamma_c^{b}(\tau)$, where

$$\mathbf{v}\mathbf{\kappa}(X^a, \tau) = \left. \frac{\partial \mathbf{K}(X^a, \tau)}{\partial X^a} \right|_{\text{fixed } \tau} = (F^b)^{-1} \, F^a ,$$

and

$$\det(\mathbf{v}\mathbf{\kappa}) = \det((F^b)^{-1}) \cdot \det(F^a) = \frac{j^a}{j^b} .$$

It is now assumed that the contact tractions developed on each surface always satisfy the local (i.e., pointwise) equilibrium condition (Newton's third law) such that

$$T^a(x, \tau) = - T^b(x, \tau) , \quad (2.2.17)$$

for each $x \in \Gamma_c$ and $\tau$, where $T$ denotes the true contact traction per unit area in the deformed configuration corresponding to Cauchy stress. Since $T \, ds = p \, dS$ for each
body and $\text{d}s^a = \text{d}s^b$, (2.2.17) can also be written as

$$p^a(X^a, \tau) \text{d}s^a = - p^b(K(X^a, \tau), \tau) \text{d}s^b . \tag{2.2.18}$$

From (2.2.14), (2.2.16), and (2.2.18), the contact tractions between contacting bodies can now be related by

$$p^b(X^b, \tau) = \left[ \frac{-(vK)^T N^b \cdot N^a}{\det(vK)} \right] p^a(X^a, \tau) , \tag{2.2.19}$$

and PK-1 stresses by

$$\det(vK) p^b(X^b, \tau) = p^a(X^a, \tau) (vK)^T , \tag{2.2.20}$$

for each $X^a$ and $X^b = K(X^a, \tau)$.

As an energetically conjugate force to the convective relative velocity $\dot{z}$ defined in (2.2.11-12), the contact traction is simply represented by $p$ through the inner product defined by

$$D(\tau) = - \int_{\Gamma^a_c(\tau)} p(X, \tau) \cdot \dot{z}(X, \tau) \text{d}S . \tag{2.2.21}$$

Eqn (2.2.21) represents the frictional power dissipation at time $\tau$ which is always non-negative (see details in Section 2.3). From (2.2.18) and (2.2.19), we have

$$\int_{\Gamma^a_c(\tau)} p(X, \tau) \cdot [v^a(X, \tau) - v^b(K(X, \tau), \tau)] \text{d}S$$

$$= \int_{\Gamma^a_c(\tau)} p^a(X, \tau) \cdot v^a(X, \tau) \text{d}S + \int_{\Gamma^b_c(\tau)} p^b(X, \tau) \cdot v^b(X, \tau) \text{d}S . \tag{2.2.22}$$

where $p^b(X, \tau)$ is given by (2.2.19).
The normal and tangential components of the contact traction are defined respectively by
\[
 p_n(X, \tau) \equiv p(X, \tau) \cdot n(X, \tau),
\]
and
\[
 p_t(X, \tau) \equiv p(X, \tau) - p_n(X, \tau) n(X, \tau).
\]
respectively. For two-dimensional contact problems (e.g., plane-strain, plane-stress, axisymmetric problems), it may be convenient to introduce a unit tangential vector \( t(X, \tau) \)
\[
t(X, \tau) = e_3 \times n(X, \tau),
\]
where \( e_3 \) represents a fixed unit vector normal to the plane and the symbol 'x' denotes the cross product. Thus, for two-dimensional problems, it is convenient to express the contact traction vector \( p \) in terms of its normal and tangential components,
\[
p = p_n n + p_t t,
\]
where \( p_t(X, \tau) = p(X, \tau) \cdot t(X, \tau) \).

2.2.5. **Some comments on general contact problems**

The mechanical behavior of each continuum is usually described by a set of state variables, subjected to fundamental laws of mechanics and constitutive relations. For instance, the purely-mechanical behavior of elastic bodies can be described by two state variables; viz., the motion \( \chi \) and a symmetric stress tensor. The effect of the
mass density change may be neglected for simplicity and the body force is usually assumed to be known. Two state variables must then satisfy the static equilibrium equation and the constitutive relation. Thus, a set of nine scalar equations (for $N=3$) is provided for the nine unknown state variables defined for each material particle (i.e., each interior point of the body). The unknown variables can then be determined by solving the respective boundary-value problem as long as compatible boundary conditions are provided.

For contact problems, however, the boundary conditions on the possible contact boundary $\Gamma_0$ for each body are not explicitly available. Instead, the behavior of contacting bodies satisfies phenomenological contact laws, called the contact conditions which may be interpreted as a multiple-choice type boundary conditions as discussed in detail in the subsequent section. The contact conditions consist largely of two parts; namely, the non-penetration and friction conditions.

At this time, however, it is assumed that the contact can be described in terms of the following sets of unknown variables:

$$A = \{x^a, x^b\}, \quad B = \{p, p^b\}, \quad \text{and} \quad C = \{N, n, \Gamma_c^a\},$$

where $p = p^a$ and $n = n^a$. The set $A$ contains the motion of each body which can be determined (uniquely) by solving the respective boundary-value problem for given contact traction
data. The contact traction, a part of the problem unknown, generally changes continuously in a non-linear fashion during the progress of deformation. The set B contains the contact traction of each body which must satisfy the local equilibrium between contacting bodies as well as the contact conditions. Finally, the set C contains auxiliary variables concerned with the description of the contact conditions as discussed in the subsequent section and it is directly related to the motions of both bodies.

The above sets of unknown variables are to satisfy the following relations and constraints expressed in an abstract manner:

1. The equilibrium conditions for bodies 'a' and 'b';
   \[ \mathbf{\mathcal{A}}_1(x^a, p) = 0, \quad \mathbf{\mathcal{A}}_1(x^b, p^b) = 0. \]

2. The local equilibrium condition;
   \[ \mathbf{\mathcal{A}}_2(p^b, p, \kappa) = 0. \]

3. The kinematic relations;
   \[ \mathbf{\mathcal{A}}_3(\kappa, n, \Gamma^a, \chi^a, \chi^b) = 0. \]

4. The contact condition;
   \[ \mathbf{\mathcal{G}}(p, \chi^a, \chi^b, \kappa, n, \Gamma^a). \]

The relations for \( \mathbf{\mathcal{A}}_1 \) and \( \mathbf{\mathcal{A}}_1 \) express the abstract boundary value problems corresponding to each body for an unknown but smooth data of the contact traction. In some special cases, they can be expressed explicitly by means of Green's functions as a kernel of the respective integral operator. The general formulation of contact boundary-value
problems is discussed in detail in Section 2.4. The relation \( S_2 \) is, on the other hand, expressed explicitly in terms of the contact tractions and the pairing maps as discussed in (2.2.19) before, which is derived from the local equilibrium condition. The relation \( S_3 \) represents three kinematic relations between the motions of contacting bodies and each of the pairing map (2.2.5), the unit normal vector,

\[
n(\mathbf{x}^a, \tau) = (\pm 1) \frac{\text{grad}(\Gamma_c)}{\| \text{grad}(\Gamma_c) \|},
\]

and the contact boundary,

\[
\Gamma_c^a(\tau) = \{ \mathbf{x}^a \in \Gamma_o^a \mid \min_{\mathbf{x}^b \in \Gamma_b^a} (\| \mathbf{x}^a(\mathbf{x}^a, \tau) - \mathbf{x}^b(\mathbf{x}^b, \tau) \| ) = 0 \}\,
\]

where \( \text{grad}(\Gamma_c) \) denotes the gradient of the deformed contact surface with respect to a surface coordinate system, and \( \| \cdot \| \) denotes the Euclidean norm.
2.3. Contact Conditions

2.3.1 Phenomenological contact law

The macroscopic contact phenomena may be itemized for the continuum modeling of contact problems as follows.

(1) Non-penetration condition:

(a) any material particle of a body cannot penetrate into another,
(b) the normal component of contact traction must be compressive for each body, and
(c) the instantaneous motion which tends to separate a pair of contact points can occur only when the contact traction vanishes.

(2) Coulomb's law of friction:

(a) The magnitude of the tangential component of contact traction must be less than or equal to that of the normal component multiplied by a coefficient of friction.
(b) the instantaneous relative motion in the tangential direction for a pair of contact points can take place when the equality in (2-a) holds, and
(c) the tangential relative motion must be along the same line as the tangential component of contact traction but in the opposite sense.

The above statement of contact law can be expressed in a mathematical form by employing the variables introduced in the preceding section.

Some restrictions need to be imposed on the motion of bodies in contact so that any material particle (interior point) of a body cannot penetrate into another at all times. First, it is assumed that the deformed configuration of bodies in contact at time $\tau$ are known $\alpha$-priori as well as
the pairing map and the contact boundary at the instant. Also, they satisfy the consistency condition (2.2.7).

According to the kinematical constraint condition (1-a), the relative velocity between each pair of contact points is then restricted by the rate form of consistency and the non-penetration conditions (2.2.12) and (2.2.13).

The kinetical constraint condition (1-b) makes a distinction between the perfectly-bonded and contact problems by restricting the admissible range of the normal component of contact traction such that

$$p_n(X, \tau) \leq 0, \quad \forall \ X \in \Gamma_{c}^a(\tau) \quad (2.3.1)$$

On the other hand, the condition (1-c) plays a complementary role between the conditions (1-a) and (1-b) such that

$$\dot{z}_n(X, \tau) \cdot p_n(X, \tau) = 0, \quad \forall \ X \in \Gamma_{c}^a(\tau) \quad (2.3.2)$$

It is noted that the boundary of the contact surface $\partial \Gamma_{c}^a(\tau)$ consists of not only the geometric boundary but also the region where the contact traction vanishes. In other words, it includes the entire region where separation can occur instantaneously in accordance with (2.3.2). Also, it is always understood that

$$p_n(X, \tau) = 0, \quad \forall \ X \in (\Gamma_{o}^a - \Gamma_{c}^a(\tau)) \quad (2.3.3)$$

because the contact boundary $\Gamma_{c}^a(\tau)$ is assumed to be known a-priori at the previous configuration.

In order to describe Coulomb's friction law
effectively, a slip function is first introduced for each pair of contact points in a similar notion to the yield functions in the classical theory of plasticity. The slip function is defined by

\[ f(p) \equiv |p_t| + \mu p_n . \tag{2.3.4} \]

Here, \( \mu \) denotes the coefficient of friction which may be assumed to be assigned to each boundary point of body 'a'. Now the friction conditions (2-a,b,c) can be represented by,

\[ \forall X \in \Gamma^a_c(\tau) ; \]

\[ f \leq 0 , \tag{2.3.5} \]

\[ |\dot{z}_t| \cdot f = 0 , \tag{2.3.6} \]

and

\[ \dot{z}_t \cdot p_t \leq 0 . \tag{2.3.7} \]

Here, it is noted that at any time \( \tau \),

\[ \dot{z}_t = \ddot{z} = F^b \dot{\dot{A}} , \quad \forall X \in \Gamma^a_c(\tau) \tag{2.3.8} \]

and

\[ p_t = p = 0 , \quad \forall X \in \partial \Gamma^a_c(\tau) \tag{2.3.9} \]

provided that the non-penetration condition always holds.

Eq (2.3.5) restricts the admissible range of the slip function \( f \) and (2.3.6) represents a slip criterion, i.e., either \( \dot{z}_t = 0 \) (stick) or \( f = 0 \) (slip). Then, the inequality (2.3.7) restricts the possible slip direction so that the power associated with frictional sliding must be always non-negative in the sense of energy dissipation. In terms of
the theory of plasticity, (2.3.7) may be considered to be a non-associated plastic flow rule. Some similarities and differences between the theories of friction and plasticity are discussed in Section 2.3.3. Note that the above friction conditions are also trivially satisfied for frictionless contact problems (i.e., \( p_t = 0 \) and \( f = 0 \) for any \( z_t \)).

2.3.2. Representation of the contact conditions

The phenomenological contact law discussed previously can be expressed in several different forms in accordance with the primary variables used in the description of material responses. It is convenient to express the contact law in the form of a multiple-choice type of boundary conditions in formulating the respective boundary value problem.

The contact conditions are first summarized in terms of the convective relative velocity \( \dot{z} \) and the contact traction \( p \), for each \( X \in I_{c}^{a} (\tau) \) and \( \tau \in [t, t+At] \):

\[
\begin{align*}
\dot{z}_n & \leq 0 , & p_n & \leq 0 , & \dot{z}_n \cdot p_n = 0 , \\
f & \leq 0 , & |\dot{z}_t| f & = 0 , & \dot{z}_t \cdot p_t & \leq 0 ,
\end{align*}
\]

(2.3.10)

where

\[
\dot{z}(X, \tau) = v^a(X, \tau) - v^b(\Omega(X, \tau), \tau) ,
\]

as defined in (2.2.11) and
\[ f(p) = |p_t| + \mu p_n , \]
as in (2.3.4). The above conditions can be expressed in a rather compact form in view of the plasticity theory, such that for \( p \in \mathcal{P} \),
\[ \dot{z} = \lambda_n I(p_n) n + \lambda_t I(f) \frac{p_t}{|p_t|} , \tag{2.3.11} \]
or
\[ \dot{z} = \frac{\lambda_n}{\mu} I(p_n) \left( \frac{\partial f(p)}{\partial p_n} \right)_{\text{p} \text{ fixed}} n + \lambda_t I(f) \left( \frac{\partial f(p)}{\partial p_t} \right)_{\text{p} \text{ fixed}} , \]
where \( \lambda_n \) and \( \lambda_t \) are non-positive constants, \( I(g) \) is an indicator defined for any scalar-valued function \( g \) such that
\[ I(g) = \begin{cases} 1, & \text{if } g = 0 , \\ 0, & \text{if } g \neq 0 , \end{cases} \tag{2.3.12} \]
and a closed convex constraint set \( \mathcal{P} \) is used to represent the admissible contact traction, defined by
\[ \mathcal{P} = \{ p=(p_n, p_t) \in \mathbb{R}^N | p_n \leq 0 , |p_t| + \mu p_n \leq 0 \} . \tag{2.3.13} \]

In (2.3.11), when \( p_t = 0 \) for frictionless problems or for \( X \in \partial^\perp \Gamma_c^a(\tau) \), (2.3.11) yields an indeterminate tangential direction implying that \( \dot{z}_t \) is unconstrained. From (2.3.11), it can be shown that, \( \forall X \in \partial^\perp \Gamma_c^a(\tau) \),
\[ p \cdot \dot{z} = \lambda_t I(f) |p_t| , \tag{2.3.14} \]
\[ = \begin{cases} -\lambda_t \mu p_n , & \text{for } f = 0 \\ 0 , & \text{for } f < 0 \end{cases} \]
noting that \( p_n \cdot I(p_n) = 0 \) for any \( p_n \) and \( f = |p_t| + \mu p_n \). It can be said that \( \dot{\lambda}_t \) represents the frictional dissipation power normalized with respect to the tangential component of the contact traction.

By applying the consistency condition (2.2.12) or (2.3.8) to (2.3.11), the undetermined non-positive constant \( \dot{\lambda}_t \) can be expressed explicitly in a similar manner as in the theory of plasticity,

\[
\dot{z} = \dot{z}_t = - |F^b|^T I(f) \frac{p_t}{|p_t|},
\]

for a \( p \in \mathcal{P} \) and \( X \in \Gamma_c^a(\tau) \), or

\[
\dot{z} = \dot{\lambda}_n I(p_n) n - |F^b|^T I(f) \frac{p_t}{|p_t|},
\]

for \( p \in \mathcal{P} \) and \( X \in \Gamma_c^a(\tau) \). It is recalled that the rate of pairing map \( \dot{\mathcal{R}} \) cannot be defined for \( X \in \partial \Gamma_c^a(\tau) \) and \( I(p_n) = 0 \) (i.e., \( p_n < 0 \)) for \( X \in \Gamma_c^a(\tau) \) in (2.3.15). In order to complete the expression (2.3.15), it is assumed in (2.3.16) that there exists an extension \( \tilde{\mathcal{R}} \) of \( \mathcal{R} \) in the sense that the material time derivative \( \dot{\mathcal{R}} \) of \( \mathcal{R} \) in the sense that the

Alternatively, the contact conditions can also be expressed in terms of the convective relative velocity \( \dot{z} \) and the rate of contact traction \( \dot{p} \) such that for a known \( p \in \mathcal{P} \) and some \( \dot{p} \in \dot{\mathcal{P}} \).
\[ \dot{z} = \lambda_n I(p_n) I(p_n) n + \lambda_t I(f) I(\dot{f}) \frac{p_t}{|p_t|}. \]  

(2.3.17)

where

\[ \dot{f} \equiv \frac{p_t}{|p_t|} \dot{p}_t + \mu \dot{p}_n. \]  

(2.3.18)

and \( \mathcal{F} \) represents a closed convex constraint set defined by

\[ \mathcal{F} \equiv \{ \dot{p} | I(p_n) \dot{p}_n \leq 0, I(f) \dot{f} \leq 0 \}. \]  

(2.3.19)

As in the previous six conditions defined by (2.3.10), (2.3.17) can also be expressed in similar form:

\[ \dot{z}_n \leq 0, \quad I(p_n) \dot{p}_n \leq 0, \quad \dot{z}_n \dot{p}_n = 0. \]  

(2.3.20)

\[ I(f) \dot{f} \leq 0, \quad |\dot{z}_t| \dot{f} = 0, \quad \dot{z}_t \cdot p_t \leq 0. \]

The possible contact status and the possible transition of contact status for \( X \in \bar{\mathcal{C}}(\tau) \) for two-dimensional case are summarized as follows.

**Possible contact status (from (2.3.10) or (2.3.11))**

(a) separation:

\[ \dot{z}_n \leq 0, \quad \forall \dot{z}_t \]  

if \( p_n = 0, \quad p_t = 0 \).

(b) stick:

\[ \dot{z}_n = 0, \quad \dot{z}_t = 0 \]  

if \( p_n < 0, \quad |p_t| < \mu |p_n| \).

(c) slip:

\[ \dot{z}_n = 0, \quad \dot{z}_t \geq 0 \]  

if \( p_n < 0, \quad p_t = \mu p_n \).

(d) slip:

\[ \dot{z}_n = 0, \quad \dot{z}_t \leq 0 \]  

if \( p_n < 0, \quad p_t = -\mu p_n \).
Possible transition (from (2.3.17) or (2.3.20))

(a) impending contact: \( p_n = 0 \), \( p_t = 0 \).

(a\(\rightarrow\)a) separation:
\[
\dot{z}_n \leq 0, \quad \forall \dot{z}_t \quad \text{if} \quad \dot{p}_n = 0, \quad \dot{p}_t = 0.
\]

(a\(\rightarrow\)b) stick:
\[
\dot{z}_n = 0, \quad \dot{z}_t = 0, \quad \text{if} \quad \dot{p}_n < 0, \quad |\dot{p}_t| < \mu |\dot{p}_n|.
\]

(a\(\rightarrow\)c) slip:
\[
\dot{z}_n = 0, \quad \dot{z}_t > 0, \quad \text{if} \quad \dot{p}_n < 0, \quad \dot{p}_t = \mu \dot{p}_n.
\]

(a\(\rightarrow\)d) slip:
\[
\dot{z}_n = 0, \quad \dot{z}_t \leq 0, \quad \text{if} \quad \dot{p}_n < 0, \quad \dot{p}_t = -\mu \dot{p}_n.
\]

(b) stick: \( p_n < 0 \), \( |p_t| < \mu |p_n| \).

(b\(\rightarrow\)b) stick:
\[
\dot{z}_n = 0, \quad \dot{z}_t = 0, \quad \text{for} \quad \forall \dot{p}.
\]

(c) slip: \( p_n < 0 \), \( p_t = \mu p_n \).

(c\(\rightarrow\)b) stick:
\[
\dot{z}_n = 0, \quad \dot{z}_t = 0, \quad \text{if} \quad \dot{p}_t > \mu \dot{p}_n, \quad \forall \dot{p}_n.
\]

(c\(\rightarrow\)c) slip:
\[
\dot{z}_n = 0, \quad \dot{z}_t > 0, \quad \text{if} \quad \dot{p}_t = \mu \dot{p}_n, \quad \forall \dot{p}_n.
\]

(d) slip: \( p_n < 0 \), \( p_t = -\mu p_n \).

(d\(\rightarrow\)b) stick:
\[
\dot{z}_n = 0, \quad \dot{z}_t = 0, \quad \text{if} \quad \dot{p}_t < -\mu \dot{p}_n, \quad \forall \dot{p}_n.
\]

(d\(\rightarrow\)d) slip:
\[
\dot{z}_n = 0, \quad \dot{z}_t \leq 0, \quad \text{if} \quad \dot{p}_t = -\mu \dot{p}_n, \quad \forall \dot{p}_n.
\]

Here, the notation (\(\rightarrow\)) indicates the transition from one state to another.
It is noted that the transition from stick to other situations is not allowed instantaneously in order to prevent the jump of contact traction at any instant. In general, it is expected that each pair of contact points would satisfy one and only one of the above contact status, but not always. The existence and uniqueness of the solutions is discussed for discretized contact problems in Chapter 4.

For any contact status or transition in two dimensions, there are two equality relations and one or two inequality constraints (Go-No-Go type) expressed in terms of four scalar variables (i.e., \((p_n, p_t, z_n, z_t)\), or \((\dot{p}_n, \dot{p}_t, \ddot{z}_n, \ddot{z}_t)\)). Since two equilibrium (or rate of equilibrium) equations are available, the contact problem becomes determinate once the contact status is known. In this respect, the contact conditions may be referred to as the multiple-choice type boundary condition. For three-dimensional problems, two additional unknowns (e.g., two angles representing the directions of \(z_t\) and \(p_t\) or \(\dot{p}_t\) on the tangent plane of the contact surface) are required for the representation of contact status, but two equilibrium equations are also provided to compensate for the additional unknowns.

2.3.3. Remarks on friction laws

Coulomb's friction law was originally proposed for rigid bodies, but found a wide application in continuum
contact problems. However, mathematical difficulties arise from the pointwise application of Coulomb's friction law, and several alternative friction models have also been proposed (e.g., Curnier [12], Oden and associates [57,59,60], etc.).

In the present section, some typical features of Coulomb's friction law are investigated and compared with those of plasticity theories and alternative friction laws proposed recently.

(1) Analogy between theories of friction and plasticity:

The friction model has, in a limited sense, a similar structure to plastic flow models in plasticity based on the notion of yield criteria. Table 1 shows a comparison between Coulomb's friction model and the plastic flow model based on $J_2$-flow theory with isotropic hardening (here, the additive decomposition of the rate of deformation tensor into elastic and plastic parts is assumed). The general structure of the inelastic constitutive relations with finite strain is discussed under the framework of Continuum Thermomechanics in Appendix A.

Both frictional contact and plasticity problems belong to a class of non-linear, path-dependent, and constrained problems. In the former category, the sources of non-linearity arise from changes of the normal vector (geometric), the contact boundary, and the friction status,
which may be termed as non-linear boundary conditions. In the latter set of problems, the kinematic and the constitutive relations are non-linear (usually called geometrical and material non-linear). In both cases, the path-dependent nature is attributed to the slip (the plastic flow) and its direction depends on the most recent state of contact traction (stress). The stress field in plasticity is constrained unilaterally by a yield criterion. However, in contact problems, both the displacement (or velocity) and the contact traction are constrained unilaterally by the non-penetration condition (1-a), and both the non-penetration condition (1-b) and the slip criterion (2-a), respectively.

A major difference between the two problems is that there appears to be no equivalent normality postulate for the contact problem (see Fig.3). For plastic flow models, the rate of plastic deformation tensor $d^P$ is derived from the associated plastic potential function (i.e., $f(s) = \sqrt{J_2}$) and its occurrence is determined by the indicator $I(f)$. The Lagrange multiplier $\lambda$ [24] is then determined so that a plastic stress state always remains on the yield surface which evolves continuously according to the associated

\[\text{Section 2.4.3.}\]
hardening mechanism(s) (the consistency condition).

For Coulomb's friction model, on the other hand, the convective relative velocity \( \dot{z} \) cannot be derived from any potential function and its occurrence is determined by two indicators, acting differently on the normal and the tangential components. Specifically, either contact or non-contact is determined by \( I(p_n) \), and either slip or stick by \( I(f) \). Each pair of contact points in the slip state must satisfy two consistency conditions simultaneously: (a) such contact points must remain on the contact surface which changes continuously under the action of external loads (kinematic consistency) and (b) the contact traction must remain on the slip surface \( f \) (kinetic consistency). Of course, these two conditions are coupled so that they cannot be treated independently. Thus, the mathematical structure of frictional contact problems is generally more complicated than other mechanics problems.

The rate of work-done and the heat supplied into a frictional system by external agents are transformed not only into the changes of the internal and the kinetic energy but also the production of the frictional power dissipation. For elastic quasi-static contact problems, the work-done by external agents can be additively decomposed into one-half of the elastic strain energy stored in both bodies and the frictional energy dissipated on the contact surface. However, Coulomb's friction model itself is not equipped
with any internal evolution mechanism associated with the energy dissipation, such as hardening mechanisms for plastic flow models. Also, it has no particular mechanism for controlling the unloading behavior and permanent deformation. Thus, the original configuration of contacting bodies may or may not be recovered for a closed path of external loading, depending on whether the elastic strain energy can be released entirely during the path.

From the thermodynamics standpoint, the frictional energy may be dissipated into the forms of heat and surface energy associated with the evolution of micro-cracks and various types of wear on a solid surface. However, few attempts have been made on the quantitative description of such phenomena on the basis of continuum thermodynamics. For most structural applications, however, their effects on the overall response of a system are negligible and the amount of dissipation energy due to friction is only a small fraction of the strain energy (refer to the computational results in Section 5.1.).

(2) Some recent variations of Coulomb's friction law

In this section, some recent modifications of Coulomb's friction law are briefly reviewed focusing on their motivations and consequences. Of particular interest are non-linear and non-local friction laws, the interface model proposed by Oden and his associates [57,59,60], and also the
friction model proposed by Curnier [12]. The displacement $u$ is used rather than the convective relative velocity $\dot{z}$ for simplicity because of the manner in which the path-dependent nature is introduced here.

(a) The non-linear friction law

As a consequence of the pointwise application of Coulomb's friction law, a clear boundary separating the regions of stick and slip always exists.

Perfect stick may not be plausible, and a number of experimental results have shown that there exists a micro-slip even before a gross slip is observed. From the mathematical standpoint, on the other hand, various regularization schemes for non-differentiable functional associated with the frictional energy can produce the same effect as the Coulomb's friction model with a micro-slip. In this respect, Oden and Pires [57] proposed the non-linear friction model endowed with regularized functions. For instance, consider a non-differentiable functional

$$\psi(u) = \int_{\Gamma_c} \mu p_n |u_t| \, ds,$$

where $p_n$ is here assumed to be a known smooth function. Now, as an example, one may introduce a regularized function $\varphi_\varepsilon(u_t)$ in terms of $|u_t|$.

$$\varphi_\varepsilon(u_t) = \begin{cases} |u_t| - \frac{\varepsilon}{2}, & \text{if } |u_t| \geq \varepsilon \\ |u_t|^2 / (2\varepsilon), & \text{if } |u_t| \leq \varepsilon \end{cases}$$
where $\varepsilon$ denotes a positive constant (see Fig. 4). Then, Coulomb's friction model behaves as if the friction law is of the form

$$p_t = \mu p_n \frac{\partial \varphi}{\partial u_t} \varepsilon,$$

or equivalently,

$$p_t = \begin{cases} \mu p_n \frac{u_t}{|u_t|}, & \text{if } |u_t| \geq \varepsilon \\ \mu p_n \frac{u_t}{\varepsilon}, & \text{if } |u_t| \leq \varepsilon \end{cases}$$

Of course, the above model approaches the original Coulomb's friction model asymptotically as $\varepsilon$ approaches zero. Such a regularization scheme for Coulomb's friction law has been employed in metal forming analyses [11], in which the frictional shear stress is taken to be $\tau = \mu k$, where $k$ is yield stress in shear and $\mu$ is a constant called a friction factor.

In recent years, some phenomenological constitutive models for inelastic materials have been proposed without the introduction of yield criteria. In a sense, their behavior may be interpreted in a similar manner as above. However, each regularized evolution equation for the respective internal state variable used in the description of an inelastic constitutive model usually accounts for the respective phenomenological micro-mechanisms inspired by the theory of dislocation.

From the vantage point of computational mechanics, such
a micro-slip (micro-plastic strain) may be quantitatively trivial in terms of the computational error. Therefore, the existence of a micro-slip may have little influence on a global solution provided that the overall behavior between two models (with and without micro-slip or micro-plastic strain mechanism) is asymptotically the same after a considerable amount of slip.

(b) The non-local friction law

Some difficulties have been encountered in finding an appropriate variational formulation as a consequence of the pointwise application of Coulomb's friction law. In the primal method dealing with the variational formulation for Signorini's problem† with Coulomb's friction law [13,59,65], the term associated with the frictional energy defined as

\[ j(u; u) \equiv - \int_{\Gamma_c} \mu p_n(u) |u_t| \, ds. \]

is mathematically ill-defined because \( p_n(u) \) may not be defined for a function \( u \in H^1(\Omega) \), where \( H^1(\Omega) \) denotes the Sobolev space of square-integrable functions with square-integrable partial derivatives of order 1 in \( \Omega \) (see also in Section 2.4.3).

The issue of existence and uniqueness of solution for

† contact of an elastic body on a rigid foundation
friction problems is still unresolved except for some special cases. For instance, Necas et al. [52] showed that there exists a unique solution $u \in H^1(\Omega)$ for the problem of an infinitely long elastic strip on a rigid flat foundation subjected to an uniform displacement on the other side when the coefficient of friction $\mu$ between strip and foundation is within a certain range, i.e., $\mu < \sqrt{2 \mathcal{G}}/(\lambda + 3 \mathcal{G})$, where $\mathcal{G}$ and $\lambda$ are Lamé's constants.

Oden and Pires [57] proposed the non-local friction law in which the normal contact traction $p_n$ in Coulomb's friction law is replaced by its mollified form $S(p_n)$ defined by

$$S(p_n)(x) \equiv \int_{G_c(y)} \omega_\rho(|x - y|) \ast (-p_n(y)) \, ds(y),$$

where $'\ast'$ denotes the convolution of two functions. Here, $\omega_\rho$ represents an infinitely smooth function with a compact support (i.e., a $\delta$-sequence of test functions). For example,

$$\omega_\rho(|x|) = \begin{cases} c \exp[\rho^2/(|x|^2 - \rho^2)] , & \text{when } |x| \leq \rho \\ 0 , & \text{when } |x| \geq \rho \end{cases}$$

where the constant $c$ is determined so that

$$\int_{G_c} \omega_\rho(|x|) \, ds = 1.$$ 

The mollified traction $S(p_n)$ approaches $p_n$ asymptotically as $\rho$ approaches zero. A useful range of the parameter $\rho$ has not been suggested. By the action of the mollifier, the
frictional energy

\[ j(u; u) \equiv - \int_{\Gamma_c} \mu S(p_n(u)) |u_t| \, ds , \]

is well-defined for \( u \in H^1(\Omega) \) since \( S(p_n(u)) \in L^2(\Omega) \), where \( L^2(\Omega) \) denotes the space of square-integrable functions. Furthermore, Oden and Pires proved that Signorini's problem with the non-local friction law has at least one solution when the elasticity tensor satisfies the ellipticity condition, and it has a unique solution for a sufficiently small coefficient of friction. The physical implication of the non-local friction law is based on the argument that a realistic contact surface is formed of irregular asperities of solid surfaces and that the real area of a contact surface is only a small fraction of its apparent area.

(c) Other friction models

The following two models have been developed under the assumption that there exists an interface material region between two contacting bodies and the behavior of the interface material is governed by its own constitutive relation different from those of the parent materials. Oden and Martins [60] proposed an interface model motivated from experimental results concerning the change of gap (approach) between irregular solid surfaces under the action of normal load in a small scale. In their model, a power-law relation is adopted for the normal contact traction instead of the
usual non-penetration condition in the following way:

\[
p_n = \begin{cases} 
-C_1 (g_n + t_0)^{m_1}, & \text{when } (g_n + t_0) > 0 \\
0, & \text{when } (g_n + t_0) \leq 0
\end{cases}
\]

and

\[
g_t = \begin{cases} 
0, & \text{when } |p_t| < C_2 |p_n|^{m_2} \\
-\lambda p_t, & \text{when } |p_t| = C_2 |p_n|^{m_2}
\end{cases}
\]

where \( g \) is the gap vector between each pair of non-contacting points, \( t_0 \) is a prescribed constant representing the height of asperities, and \( C_1, C_2, m_1, \) and \( m_2 \) are material constants representing the properties of an interface element. Fig. 5 illustrates the differences between the interface model and the conventional model with the non-penetration condition discussed in the preceding section. It can be observed that a small penetration \( t_0 \) is allowed under the action of very stiff non-linear springs in the interface model.

Curnier [12] proposed a friction model inspired from the classical theory of plasticity. His model is governed by three different mechanisms called adherence, tear, and wear, by analogy with elastic, kinematic hardening, and isotropic hardening mechanisms, respectively. The gap is first decomposed additively into two parts, called adherence and slip, similar to the additive decomposition of the incremental strain into elastic and plastic parts. A cumulative slip is then introduced as a counterpart of the
effective plastic strain. As energetically conjugate forces to these three kinematic variables (adherence, slip, and cumulated slip), friction, tear, and wear forces are also introduced by analogy with the stress, the back stress, and the drag stress, respectively. The relations between these kinematic and kinetic variables are also assigned in the same manner as in plasticity. The adherence is governed by an adherence law similar to an elastic law, and the slip is governed by a slip rule analogous to the (associated or non-associated) flow rule, together with two softening laws similar to the isotropic and kinematic hardening laws in plasticity. Thus, Curnier attempted to interpret friction phenomena in a manner similar to the plasticity theory. However, the validity and utility of the model have not yet been investigated.
2.4. Incremental Contact Boundary-Value Problem

The effect of energy dissipation is usually described by means of rate dependent constitutive equations, evolution equations for the internal state variables, and the contact boundary conditions. One of the common characteristics of these descriptions is that they tend to result in stiff systems when applied to numerical integration schemes. Thus, the boundary-value problems for dissipative systems are usually solved within a small progressive (incremental) action of external loads, despite the absence of any dynamic effects. The accuracy and stability of the numerical solutions can be improved with a reasonable increment step by employing an effective integration scheme for the stiff rate equations. Accordingly, extensive investigations have been presented in the field of computational plasticity (e.g., [27,40,63,64]).

For contact problems with friction, on the other hand, a systematic study on the subject of developing effective time-marching schemes becomes much more difficult due to intrinsic complications discussed in the preceding sections. In fact, for frictional contact of deformable bodies, mathematically rigorous numerical algorithms based on a sound theory cannot be found in the open literature. This can be attributed to the mathematical ambiguity associated with the contact pairing for the pointwise application of Coulomb’s law. Coupled constraints add further difficulties
to the problem.

The overall feature of contact boundary-value problems are first investigated here for a simplified application. Numerical integration schemes for a rate form of the contact boundary conditions are then studied by employing the incremental formulation for general contact problems. A variational principle is also studied for incremental contact boundary-value problems.

2.4.1. Incremental method

Recalling the abstract representation of general contact problems in Section 2.2.5, it is now assumed that the following two relations $\mathcal{X}_{1}^{a}$ and $\mathcal{X}_{1}^{b}$ can be written explicitly:

$\mathcal{X}_{1}^{a}: \quad x^{a}(X,\tau) = \left[ G^{a}(X,Y) \cdot p^{a}(Y,\tau) \right] + w^{a}(X,\tau) \quad (2.4.1)$

$\mathcal{X}_{2}^{b}: \quad x^{b}(X,\tau) = \left[ G^{b}(X,Y) \cdot p^{b}(Y,\tau) \right] + w^{b}(X,\tau) \quad (2.4.2)$

where

$$ [G(X,Y) \cdot p(Y,\tau)] \equiv \int_{\Gamma_{0}} G(X,Y) \cdot p(Y,\tau) \, dS(Y), $$

and

$$ w(X,\tau) = o_{X} + \int_{\Omega} \hat{G}(X,Y) \cdot b(Y,\tau) \, dV + \int_{\Gamma_{f}} G(X,Y) \cdot f(Y,\tau) \, dS. \quad (2.4.3) $$

for the respective bodies. Here, $\hat{G}$ and $G$ represent the volume and surface Green's functions for each body, respectively (in fact, $G$ represents the normal gradient of
and \( \mathbf{G^0} \) is the position vector of \( \mathbf{X} \) at \( \tau = 0 \) (i.e., the undeformed state). The deformation due to the body force and applied tractions is denoted by \( \mathbf{w} \).

By assuming that the Green's functions are constant in time, the time rate of equations (2.4.1 and 2) are simply

\[
\mathbf{v^a}(\mathbf{X}, \tau) = [\mathbf{G}^a(\mathbf{X}, \mathbf{Y}) \cdot \dot{\mathbf{p}}^a(\mathbf{Y}, \tau)] + \dot{\mathbf{w}}^a(\mathbf{X}, \tau),
\]

(2.4.4)

\[
\mathbf{v^b}(\mathbf{X}, \tau) = [\mathbf{G}^b(\mathbf{X}, \mathbf{Y}) \cdot \dot{\mathbf{p}}^b(\mathbf{Y}, \tau)] + \dot{\mathbf{w}}^b(\mathbf{X}, \tau).
\]

(2.4.5)

The convective relative velocity \( \dot{\mathbf{z}} \) in (2.2.11) can be expressed in terms of the rate of contact traction by

\[
\dot{\mathbf{z}}(\mathbf{X}, \tau) = [\mathbf{M}(\mathbf{X}, \mathbf{Y}, \tau) \cdot \dot{\mathbf{p}}(\mathbf{Y}, \tau)] + \dot{\mathbf{w}}(\mathbf{X}, \tau),
\]

(2.4.6)

where

\[
[\mathbf{M}(\mathbf{X}, \mathbf{Y}, \tau) \cdot \dot{\mathbf{p}}(\mathbf{Y}, \tau)]
\]

\[
= \int_{\Gamma_C^a(\tau)} \mathbf{G}^a(\mathbf{X}, \mathbf{Y}) \cdot \dot{\mathbf{p}}^a(\mathbf{Y}, \tau) \, dS - \int_{\Gamma_C^b(\tau)} \mathbf{G}^b(\mathbf{X}, \mathbf{Y}) \cdot \dot{\mathbf{p}}^b(\mathbf{Y}, \tau) \, dS,
\]

and

\[
\dot{\mathbf{w}}(\mathbf{X}, \tau) = \dot{\mathbf{w}}^a(\mathbf{X}, \tau) - \dot{\mathbf{w}}^b(\mathbf{\mathbf{\mathbf{R}}}(\mathbf{X}, \tau), \tau).
\]

Replacing \( \dot{\mathbf{z}} \) in (2.4.6) by the contact conditions in (2.3.11),

\[
\lambda_n \mathbf{I} \mathbf{n} + \lambda_t \mathbf{I} \mathbf{f} \frac{\mathbf{p}_t}{|\mathbf{p}_t|} = [\mathbf{M}(\mathbf{\mathbf{R}}) \cdot \dot{\mathbf{p}}] + \dot{\mathbf{w}}(\mathbf{\mathbf{R}}).
\]

(2.4.7)

† Refer to (2.2.22) for the change of variables by means of the pairing map.
Here, the auxiliary variable set \( C = \{ \mathcal{K}, n, \Gamma_c^a \} \) defined in Section 2.2.5 is again used to derive the above first-order non-linear differential system equation defined for each \( X \in \Gamma_c^a(\tau) \) with respect to the contact traction \( p \). The dependence of \( M \) and \( \dot{w} \) on the pairing map is emphasized here.

The contact problem described in Section 2.2.5. can therefore be restated in definitive form if time-invariant Green's functions exist. The sets of unknown variables, i.e., \( A = \{ \chi^a, \chi^b \} \), \( B = \{ p^a, p^b \} \), and \( C = \{ \mathcal{K}, n, \Gamma_c^a \} \), must satisfy

1. the equilibrium equations (2.4.1 and 2)
2. the balance of contact forces, (2.2.19)
3. the kinematic relations between bodies, (2.2.5, 27 & 28)
4. the governing contact equations;
   (i) the contact constraints, (2.4.7) and (2.3.19)
   (ii) the consistency condition,

\[
[M(\mathcal{K}) \cdot p] + w(\mathcal{K}) = 0 .
\] (2.4.8)

where \( p \) is in the constraint set defined by (2.3.13) and

\[
w(\mathcal{K}) \equiv w^a(X, \tau) - w^b(\mathcal{K}(X, \tau), \tau) .
\]

**Representation of incremental contact equations**

The above governing contact equations in the form of non-linear first-order differential equations can now be resolved into incremental contact equations adopting various
time-marching schemes which are grossly classified into explicit and implicit methods. For an incremental loading stage from \( \tau = t \) to \( \tau = t + \Delta t \), these incremental equations can be expressed in the following form:

(i); The contact constraints (Fig. 6):

\[
[M(1+\alpha_K) \cdot \Delta p] + \Delta w(1+\alpha_K) + g(1+\alpha_K) - g(2+\alpha_K)
\]

\[
= - \left| 1+\alpha_K b \right(1+\alpha_K) \Delta K \right| I(1+\alpha_K) \frac{1+\alpha_K}{1+\alpha_K} p_t
\]

\( \forall X \in \Gamma_c^a\{tU(t+\Delta t)\} \)  

(2.4.9)

(ii); The consistency conditions:

\[
[M(1+\alpha_K) \cdot 1p] + w(1+\alpha_K) = 0 , \quad \forall X \in \Gamma_c^a(t)
\]

\[
[M(1+\alpha_K) \cdot 1p] + w(1+\alpha_K) = g(1+\alpha_K) , \quad \forall X \in \Gamma_c^a(t+\Delta t)
\]

\( 1p \in \mathcal{F} \), and \( 2p \in \mathcal{F} \) .  

(2.4.10-12)

Here,

\[
\Delta p = 1+\alpha_K p \Delta t = 2p - 1p , \quad (2.4.13)
\]

\[
\begin{bmatrix}
(2p - 1p) , \\
(2p - 0) , \\
(0 - 1p)
\end{bmatrix} , \quad \forall X \in \Gamma_c^a\{t\cap(t+\Delta t)\} \]

\[
\begin{bmatrix}
(2p - 1p) , \\
(2p - 0) , \\
(0 - 1p)
\end{bmatrix} , \quad \forall X \in \Gamma_c^a\{t-tU(t+\Delta t)\} \]

\[
\begin{bmatrix}
(2p - 1p) , \\
(2p - 0) , \\
(0 - 1p)
\end{bmatrix} , \quad \forall X \in \Gamma_c^a\{(t+\Delta t)-tU(t+\Delta t)\} \]
\[ \Delta w(1+\alpha^-) = 2w(1+\alpha^-) - w(1+\alpha^-) . \]

\[ \Delta \kappa = 1+\alpha^- \Delta t = \kappa^2 - \kappa^- , \quad (2.4.14) \]

\[ = \begin{cases} 
(2\kappa^- - 1\kappa^-) , & \forall \kappa \in 1\Gamma_c \{t\cap(t+\Delta t)\} \\
(2\kappa^- - 1+\beta \kappa^-) , & \forall \kappa \in 1\Gamma_c \{t-tU(t+\Delta t)\} \\
(1+\gamma \kappa^- - 1\kappa^-) , & \forall \kappa \in 1\Gamma_c \{(t+\Delta t)-tU(t+\Delta t)\} 
\end{cases} \]

\[ \kappa^2 = \begin{cases} 
1\kappa^- + \Delta \kappa^- , & \forall \kappa \in 1\Gamma_c \{t\cap(t+\Delta t)\} \\
1+\beta \kappa^- + \Delta \kappa^- , & \forall \kappa \in 1\Gamma_c \{t-tU(t+\Delta t)\} 
\end{cases} \quad (2.4.15) \]

\[ 1+\alpha p = (1-\alpha) \ p + \alpha \ p^2 \]

\[ = \begin{cases} 
1p + \alpha \Delta p , & \forall \kappa \in 1\Gamma_c \{t\cap(t+\Delta t)\} \\
0 + \frac{\alpha(\alpha-\beta)}{(1-\beta)} \Delta p , & \forall \kappa \in 1\Gamma_c \{t-tU(t+\Delta t)\} \\
1p + \frac{\alpha}{\gamma} \Delta p , & \forall \kappa \in 1\Gamma_c \{(t+\Delta t)-tU(t+\Delta t)\} 
\end{cases} \quad (2.4.16) \]

for \( \beta \leq \alpha \leq \gamma \).

\[ 1+\alpha^- \kappa = (1-\alpha) \ k^- + \alpha \ k^- \]

\[ = \begin{cases} 
1\kappa^- + \alpha \ \Delta \kappa^- , & \forall \kappa \in 1\Gamma_c \{t\cap(t+\Delta t)\} \\
1+\beta \kappa^- + \frac{\alpha(\alpha-\beta)}{(1-\beta)} \ \Delta \kappa^- , & \forall \kappa \in 1\Gamma_c \{t-tU(t+\Delta t)\} \\
1\kappa^- + \frac{\alpha}{\gamma} \ \Delta \kappa^- , & \forall \kappa \in 1\Gamma_c \{(t+\Delta t)-tU(t+\Delta t)\} 
\end{cases} \]

for \( \beta \leq \alpha \leq \gamma \).
\[ 1^{+\alpha_f} \equiv f(1^{+\alpha_p}) , \]

and

\[ 0 \leq \alpha \leq 1 , \]

\[ 0 \leq \beta \leq \alpha , \quad \forall X \in 1^{\overline{\alpha}}_{\overline{c}}(t-tU(t+\alpha \Delta t)) \]

\[ \alpha \leq \beta \leq 1 , \quad \forall X \in 1^{\overline{\alpha}}_{\overline{c}}((t+\alpha \Delta t)-(t+\alpha \Delta t)U(t+\Delta t)) \]

\[ 0 \leq \gamma \leq \alpha , \quad \forall X \in 1^{\overline{\alpha}}_{\overline{c}}((t+\alpha \Delta t)-tU(t+\alpha \Delta t)) \]

\[ \alpha \leq \gamma \leq 1 , \quad \forall X \in 1^{\overline{\alpha}}_{\overline{c}}((t+\Delta t)-(t+\alpha \Delta t)U(t+\Delta t)) \]

where

\[ 1^{\overline{\alpha}}_{\overline{c}}(t-tU(t+\alpha \Delta t)) \equiv 1^{\overline{\alpha}}_{\overline{c}}(t) - \{1^{\overline{\alpha}}_{\overline{c}}(t) \cup 1^{\overline{\alpha}}_{\overline{c}}(t+\alpha \Delta t)\} , \text{ etc.} \]

The parameter \( \alpha \) in the integration scheme has the conventional meaning, and the parameters \( \beta \) and \( \gamma \) are used to account for the transversality condition associated with the change of contact boundary, to be discussed later. The left-superscript is used to denote the value of a function at the specified time (e.g., \( ^1n(X) = n(X,t) \), \( ^{1+\alpha}n(X) = n(X,t+\alpha \Delta t) \), etc.), and the set-operation symbol \(-\) denotes the difference between two sets. It is noted that \( ^{1+\alpha}p \in \Phi \) when \( ^1p \in \Phi \) and \( ^2p \in \Phi \) because \( \Phi \) is a closed convex set (i.e., \( (1-\alpha)^1p + \alpha^2p \in \Phi \) for any \( \alpha \in [0,1] \) when \( ^1p, \ ^2p \in \Phi \) and \( \Phi \) is a convex set).

\[ \dagger \quad \text{e.g., } \alpha = 0 \text{ stands for the explicit method, } \alpha = 1 \text{ for the fully implicit method, and } \frac{1}{2} \leq \alpha \leq 1 \text{ for a family of implicit methods.} \]
In (2.4.11), \( ^1g \), called the referential gap vector, hereafter, represents the directional distance between two points from \( X^b = \bar{R}(X, t+\alpha t) \) to \( X \in \overline{\Gamma_c(\tau U(\tau+\Delta t))} \) at \( \tau = t \) (i.e., the reference state), that is,

\[
^1g(1+\alpha \bar{R}) = x^a(X, t) - x^b(1+\alpha \bar{R}, t), \tag{2.4.18}
\]

\[
= u^a(X, t) - u^b(1+\alpha \bar{R}, t) + \sigma_g(1+\alpha \bar{R}).
\]

where \( u \) is the displacement vector and \( \sigma_g \) represents the initial gap vector between these two points at \( \tau = 0 \) (the undeformed state). For \( \alpha = 0 \),

\[
^1g(1+\bar{R}) = 0, \quad \forall X \in \overline{\Gamma_c(t)}
\]

and

\[
^1g_n(1+\bar{R}) < 0, \quad \forall X \in \overline{\Gamma_c(t-\tau U(\tau+\Delta t))}
\]

where \( ^1g_n(1+\bar{R}) = ^1g(1+\bar{R}) \cdot n^a(X, t) \). Adding (2.4.9) to (2.4.11), we get

\[
[M(1+\alpha \bar{R}) \cdot 2p] + 2w(1+\alpha \bar{R}) - 1g(1+\alpha \bar{R}) + 1g(1+\bar{R}) - 2g(2\bar{R})
\]

\[
= -|1+\alpha_F b(1+\alpha \bar{R})| \frac{1}{\alpha f} \frac{1+\alpha_p_t}{|1+\alpha_p_t|} , \tag{2.4.19}
\]

and \( 2p \in \mathcal{G} \). It is noted that the sum of the first two terms in the left-hand part of (2.4.19) is equivalent to the deformed gap vector \( 2g(1+\alpha \bar{R}) \) between the pairs of contact points \( X \) and \( 1+\alpha \bar{R} = \bar{R}(X, t+\alpha t) \) at \( \tau = t+\Delta t \) (the deformed state).
Integration schemes and transversality conditions

In the above set of incremental contact equations, the rate of pairing map $\dot{\mathbf{R}}$, which is not defined for $X \in \mathcal{R}_0^a(\tau)$ as mentioned in (2.3.15-16), is resolved into three different parts in (2.4.14) accounting for the transversality condition associated with the change of contact boundary. They can be classified into 'contact to contact', 'non-contact to contact', and 'contact to non-contact' for all possible contact points during an incremental stage.

The contact status (or non-contact) for the possible contact points is generally determined once at a particular state for each incremental stage. On the other hand, the friction condition involves the relative motion during the interval. Thus, friction problems with a varying contact boundary generally require an additional algorithm for the estimation of initial and final pairing points denoted by $1+\beta_R$ and $1+\gamma_R$, respectively, in (2.4.14) for such transitional points, respectively. Also, the pairing points must be determined explicitly for the accurate evaluation of the friction status for these points.

For discrete contact problems using the finite element method, the size of time increment $\Delta t$ may be chosen sufficiently small so that at most one node can be in a transitional state at each incremental stage [62,77]. Yet,
there is no general rule for the selection of $\Delta t$ for the stability and accuracy of the associated numerical integration scheme. A family of implicit schemes (e.g., $\frac{1}{2} \leq \alpha \leq 1$) is recommended for incremental friction problems with a varying contact boundary. It is noted that for the fully-implicit scheme ($\alpha = 1$), the estimation of the final pairing point $\gamma, K$ for the transitional points from contact to non-contact is not required because the contact condition is imposed on the pairs of contact points at the deformed state.

The direction of the normal vector $n$ plays an important role in a manner similar to that of the yield surface in incremental plasticity theory, as shown in Fig. 7. The normal direction of a contact surface is, of course, obtained from the configuration of deforming bodies in contact. Thus, its change is directly related to the associated deformation gradient $F$ which can be decomposed into stretch and rotational parts (e.g., $F = RU = VR$ and $R^{-1} = R^T$). The stretch part includes the slip distance (i.e., $|F \Delta n| = |U \Delta n|$), on the other hand, the rotational part incorporates the change of the normal vector (i.e., $n = RN$).

The role of the pairing map $\mathcal{N}$ in the incremental contact equation (2.4.19) is in twofold. The first role is to evaluate the friction status through (2.4.9) by examining
the change of the gap between two particular points \( X^a \) and \( X^b = \mathcal{R}(X^a, t+\alpha \Delta t) \) during an incremental stage. Secondly, it considers the resolution of the local equilibrium condition (2.2.19) into the consistency condition (2.4.10) or (2.4.12).

**Implicit time-integration schemes**

In accordance with a particular choice of the parameter \( \alpha \), various explicit and implicit methods can be obtained within the above formalism. In the present work, the fully-implicit scheme (\( \alpha = 1 \)), is first discussed briefly and a simplified implicit method is then proposed for a computational efficiency.

(1) The fully-implicit scheme

By substituting \( \alpha = 1 \) and introducing the iteration counter \( i \), (2.4.19) can be written as

\[
\begin{align*}
\left[ \mathcal{W} \left( \bar{K}_i \right) - 2p^i \right] + \bar{w} \left( \bar{K}_i \right) - \bar{g} \left( \bar{K}_i \right) &= \frac{\lambda n}{2} \mathcal{I}(2p^i_n) 2n_i - |2F_b(\bar{K}_i) i \bar{K}^i| I(2f^i) \frac{2p^i_t}{|2p^i_t|}, \\
\end{align*}
\]

for \( 2p^i \in \Phi \) and \( \Delta \lambda_n \leq 0 \), \hspace{1cm} (2.4.20)

where \( 2n_i = i^n + \Delta n_i \) with \( \Delta n^o = 0 \) and \( 2n^o = i^n \). As mentioned before, \( 2g(\bar{K}) \) in (2.4.19) is now replaced by \( \frac{\Delta \lambda n}{2} \mathcal{I}(p_n) n \) with the constraint \( \Delta \lambda_n \leq 0 \). It is noted that several different solution strategies can be adopted for the
fully-implicit scheme.

Note that the above integration scheme requires two- levels of iteration within each increment. The first level is for updating $\bar{R}$ and $n$ indicated as the iteration counter $i$, and the second level for obtaining $p$ and $\Delta \bar{R}$ satisfying (2.4.20) subjected to the inequality constraints.

The iteration routine associated with $\bar{R}$ and $n$ is not always necessary even for the fully-implicit method. For instance, consider the contact of a deformable body on a rigid stationary foundation in which the convective relative velocity $\dot{z}$ can be replaced by $\dot{\bar{R}} (= v^2)$ in the integrable form as mentioned in Section 2.2.3 (the local equilibrium condition need not be considered). For this case, (2.4.20) can be rewritten as

$$[G^a \cdot 2p^i] + 2w^a - 1\bar{R} = \Delta \lambda_n I(2p^i)^2n^{i-1} - |\Delta \bar{R}| \frac{2p^i}{|2p^i|},$$

and $2p^i \in \mathcal{F}$ and $\Delta \lambda_n \leq 0$.  

(2.4.21)

by setting $G^b = 0$, $w^b = \bar{R}$, $F^b = I$, $^1g(2\bar{R}) = ^1X - 2\bar{R}$, and $^1g(1\bar{R}) = ^1X - 1\bar{R}$. It can also be written as

$$\Delta u^i + ^1X - 1\bar{R} = \Delta \lambda_n I(2p^i)^2n^{i-1} - |\Delta \bar{R}| \frac{2p^i}{|2p^i|},$$

and $2p^i \in \mathcal{F}$ and $\Delta \lambda_n \leq 0$.  

(2.4.22)

by recalling that
\[ 2X = \dot{X} + Au \]
\[ = \left[ G^a \cdot \ddot{p} \right] + 2w^a. \]

If the normal vector \( n \) does not change (e.g., a flat foundation) or its change is negligible, the iteration counter \( i \) can be removed from (2.4.21). If the contact boundary of the deformable body does not change, (2.4.22) can be further reduced to

\[ \Delta u = - \frac{\left| \Delta u \right| I^{(f)}}{|\ddot{p}_T|}, \]

and \( \ddot{p} \in \mathcal{P} \), \( (2.4.23) \)

noting that \( ^1X = ^1N \) and \( \Delta N = Au \) at all times. For a frictionless problem, on the other hand, (2.4.21) is reduced to

\[ \dot{u}_N + ^0X_N - ^0R_N = \lambda_N I^{(2p_N)} N, \]

\[ \ddot{p} \in \mathcal{P} \text{ with } p_T = 0, \text{ and } \lambda_N \leq 0, \quad (2.4.24) \]

where \( N \) denotes the constant normal vector. For a linear elastic body, (2.4.24) need not be solved in an incremental manner.

(2) A simplified time-integration scheme

A simplified implicit scheme is proposed to optimize the iteration routine designed for updating \( \overrightarrow{R} \) and \( n \). By replacing \( 2\overrightarrow{R} \) in (2.4.20) by \( 1\overrightarrow{R} \) and \( 2n \) by \( 2\hat{n} \), (2.4.20) is now
written as

\[ [\mathcal{M}(\vec{n}) \cdot \vec{p}] + 2w(\vec{R}) \]

\[ = \Delta \lambda_n I(\vec{r}_n) \vec{n} - |2F^b(\vec{R}) \Delta \vec{R}| I(\vec{r}_f) \frac{2p_f}{|2p_f|} \]

and \( \vec{p} \in \Phi \) and \( \Delta \lambda_n \leq 0 \). \hfill (2.4.25)

Here \( \vec{n} \) is the normal to the contact surface in the assumed deformed configuration of the corresponding perfectly-bonded problem with reference to the pairing map \( \vec{R} \) and it is obtained once for each incremental stage. Eq (2.4.25) is then solved for \( \vec{p} \in \Phi \) iteratively. The underlying variational principle is discussed on the basis of the above incremental formalism in Section 2.4.3.

In the proposed time-integration scheme, the local equilibrium condition is relaxed within each incremental stage (see Fig. 8), and it is then corrected at the next incremental stage successively. As mentioned before, the normal direction plays an important role in determining the admissible directions of motion satisfying the consistency condition. When the variation of the normal direction is small, falling within the range of a slip distance (i.e., \( |U \Delta R| \)) along the contact surface during an increment, the normal vector in the deformed configuration can be used in determining the admissible direction of the relative motion between the pair of contact points \( X^a \) and \( X^b = \vec{R}(X^a, t) \) during the increment, as depicted in Fig. 8. For many practical
problems, the normal direction obtained from the deformed configuration of the corresponding perfectly-bonded problem for each incremental stage may provide a reasonable estimation, instead of obtaining it in a fully-implicit manner. Detailed computational procedures associated with the finite element method are discussed in Chapter 3.

In order to discuss the general applicability of the above integration scheme, the total potential of contacting bodies at \( \tau = t + \Delta t \) can be expressed in the following form with reference to the quasi-statically equilibrated state at \( \tau = t \):

\[
2\Pi = W^a(\varepsilon^a) + W^b(\varepsilon^b) - f^a(u^a) - f^b(u^b) + \int_0^{t+\Delta t} D(\tau) \, d\tau ,
\]

(2.4.26)

where the dissipation \( D \) is as defined in (2.2.21) and the convective relative velocity \( \dot{z} \) by (2.2.11 and 12). In the above expression, \( W^{a,b} \) and \( f^{a,b} \) represent the strain energy function and the external work-done for each body, respectively, and \( D(\tau) (\geq 0) \) denotes the dissipative power due to friction at time \( \tau \) (refer to (2.3.19)). \( u^a \) and \( \varepsilon^a \) represent the displacement and strain tensor at time \( t + \Delta t \). It is, of course, assumed that the kinematic and kinetic variables in the above expression satisfy all required contact relations and constraints. Other dissipative mechanisms can be included additively in (2.4.26).
The frictional energy dissipation during the time increment $\Delta t$ from $\tau = t$ to $\tau = t + \Delta t$ is then approximated in the form

$$
\int_{t}^{t+\Delta t} D(\tau) \, d\tau \approx - \int_{\Gamma_{c}^{a}(t+\Delta t)} p(X,t+\Delta t) \cdot Az \, dS.
$$

where

$$
Az = \Delta u^{a}(X,t+\Delta t) - \Delta u^{b}(\mathcal{N}(X,t),t+\Delta t)
= F^{b}(\mathcal{N}(X,t),t+\Delta t) \Delta \mathcal{N}.
$$

The contact constraints concerning the above dissipation energy are, of course, expressed in terms of the normal direction $\hat{n}$. It is expected that a slight over- (or under-) estimation of dissipation energy may result from an increase (or decrease) of the contact boundary depending on the action of external loads. However, this is generally negligible with a reasonable choice of increment.

2.4.2. Governing field equations

The parent bodies are assumed to be linearly elastic from here on to focus our attention on the contact formulation and associated solution algorithm. Each body is assumed to undergo quasi-static and isothermal deformation. As discussed previously, contact problems intrinsically belong to a class of non-linear, path-dependent, and unilaterally constrained problems regardless of the underlying continuum model for parent bodies. Thus, the
problem considered in the sequel is based on the incremental method in accordance with the simplified implicit time-integration scheme discussed in the previous section.

We consider a representative incremental loading stage \( \tau \in [t, t+\Delta t] \). It is noted that the pairing map \( \bar{\mathbf{K}} \) and the normal vector \( \bar{\mathbf{n}} \) are determined at the beginning of each stage. Accordingly, the contact conditions are expressed in terms of the deformed gap vector \( ^2z \) and the contact traction \( ^2p \) at \( \tau = t+\Delta t \) as follows:

By replacing the left-hand terms by \( ^2z \), (2.4.25) becomes

\[
^2z = \Delta \lambda_n \mathbf{I}(^2p_n) \bar{\mathbf{n}} - |^2F \Delta \mathbf{K}| \mathbf{I}(^2f) \frac{^2p_t}{|^2p_t|} \hat{^2t},
\]

where \( ^2p_t \) denotes the unit tangential vector as defined in (2.2.25). Eq (2.4.25) can also be represented in the following form as explained in Section 2.3.2.

\[
^2z_n \leq 0.
\]  

\( ^2z_n \) denotes the unit tangential vector as defined in (2.2.25).
\[ ^2z_n^2 p_n = 0 , \]  
\[ |^2z_t| (|^2p_t| + \mu^2 p_n) = 0 , \]  
\[ - ^2z_t^2 p_t \geq 0 , \]  
and  
\[ ^2p \in \mathcal{F} = \{ p=(p_n, p_t) \mid p_n \leq 0 , f \leq 0 \} , \]  
where  
\[ ^2z(X) = ^2u^a(X) - ^2u^b(\mathbf{1}_{\mathcal{K}}) \]  
\[ = ^2u^a(X) - ^2u^b(\mathbf{1}_{\mathcal{K}}) + ^0 z , \]  
\[ ^0 z(X) = ^0u^a(X) - ^0u^b(\mathbf{1}_{\mathcal{K}}) , \]  
and \( f = |p_t| + \mu p_n \). For notational simplicity, the left-superscript \(^2\) is often omitted when obvious.

The governing field equations for each body are expressed on the basis of the infinitesimal deformation theory under the assumption\(^\dagger\) that the global equilibrium state is not affected by the change of the local distribution of the contact traction due to the deformation of the contact surface. Therefore, no distinction needs to be made between configurations except for the contact conditions. For compactness of notation, the governing

\(^\dagger\) There is no difference between the contact traction vector \( p \) per unit area in the referential configuration and the undeformed or other configurations in view of the global equilibrium state.
field equations are described with respect to the referential configuration in accordance with the description of the contact conditions. The following boundary-value problem is considered.

**PROBLEM I.**

Find $u^a$, $u^b$, and $p$ such that, employing the usual indicial notation,

\[
\begin{align*}
\sigma_{ij,j} &= 0 , & \text{in } \Omega \quad (2.4.28a) \\
\sigma_{ij} &= E_{ijkl}u^k_{,l} , & \text{in } \Omega \quad (2.4.28b) \\
\bar{u}_i &= \bar{u}_i , & \text{on } \Gamma_u \quad (2.4.29a) \\
\sigma_{ij}n_j &= \bar{f}_i , & \text{on } \Gamma_f \quad (2.4.29b) \\
\sigma_{ij}n_j &= p_i , & \text{on } \Gamma_o \quad (2.4.29c)
\end{align*}
\]

for each body 'a' and 'b', and such that all the contact conditions in (2.4.27a-f) are satisfied. □

Here, $\Omega$ and $\Gamma$ represent the respective regions in the referential configuration for each body, the comma denotes partial differentiation with respect to $X$, $\sigma = \sigma(X,t+\Delta t)$ denotes the stress tensor\(^\dagger\), and barred quantities are prescribed boundary data. The elasticity tensor $E_{ijkl}$ is assumed to satisfy the symmetry condition

\[^\dagger\text{For a finite deformation problem, it will be beneficial to employ the Piola-Kirchhoff stress tensor of the first kind.}\]
\( E_{ijkl} = E_{klij} = E_{jikl} \) \hspace{1cm} (2.4.30)

and the ellipticity condition

\[ E_{ijkl}^{u_i,j_k,l} \gtrsim m u_i, j^{u_i,j} \text{ for } m > 0. \]

Also, it is assumed that \( \exists M > 0 \) such that

\[ \max_{X \in \Omega} |E_{ijkl}(X)| \leq M \text{ for } \forall i,j,k, \text{ and } l. \]

2.4.3. Variational Principle

It is necessary to introduce additional notation and appropriate functional spaces (admissible fields) endowed with proper energy for a meaningful discussion of a variational principle associated with PROBLEM I.

The admissible displacement field, over body \( \Omega^a \) (similarly over \( \Omega^b \)), is defined by

\[ \mathcal{U}^a = \{ u^a = (u^a_1, u^a_2) \mid u^a_1 \in H^1(\Omega^a), \gamma(u^a) = \bar{u} \text{ a.e. on } \Gamma^a \} \]

and the space of virtual displacements by

\[ \mathcal{V}^a = \{ v^a = (v^a_1, v^a_2) \mid v^a_1 \in H^1(\Omega^a), \gamma(v^a) = 0 \text{ a.e. on } \Gamma^a \} \]

equipped with the norm

\[ \|v\|_\mathcal{V} = \left[ \int_{\Omega} vv \cdot vv \, dV \right]^{1/2}. \]

\[ \dagger \] \( \mathcal{U} \) is a closed linear affine manifold in \( \mathcal{V} \) (See details in Refs. [13,55,65]).
\( H^1(\Omega) \) is the Sobolev space of square integrable functions (\( L^2 \)-functions) with \( L^2 \)-partial derivatives of order 1 equipped with the usual norm

\[
\|v\|_{H^1(\Omega)} = \left[ \int_\Omega (v \cdot v + \nabla v \cdot \nabla v) \, dv \right]^{1/2}.
\]

\( \gamma(v) \) is the trace of \( v \in H^1(\Omega) \) on \( \Gamma \) such that the mapping \( v \to \gamma(v) \) is linear, continuous, and surjective from \( H^1(\Omega) \) to \( H^{1/2}(\Gamma) \) where \( H^{1/2}(\Gamma) \) is a fractional Sobolev space equipped with the norm

\[
\|\phi\|_{H^{1/2}(\Gamma)} = \inf_{u \in H^1(\Omega)} \{ \|v\|_{H^1(\Omega)} : \phi = \gamma(v) \}.
\]

The set of incremental relative motion can be defined by a closed convex subset of \( H^{1/2}(\Gamma_c) \) such that

\[
\mathcal{Z} = \{ z = (z_n, z_t) \mid z_n \in H^{1/2}(\Gamma_c), z_n \leq 0 \}.
\]

where \( z = \gamma_c(u^a) - \gamma_c(u^b) + z^0 \), \( \gamma_c \) is the restriction of the trace map \( \gamma \) to \( \Gamma_c \) for each body, and \( z^0 \in H^{1/2}(\Gamma_c) \). For notational simplicity, the trace of the displacement \( \gamma(u) \) is simply denoted by \( u \) unless otherwise necessary.

A closed set of admissible contact traction fields can be defined by

\[
\mathcal{F} = \{ p = (p_n, p_t) \mid p_i \in H^{-1/2}(\Gamma_c), p_n \leq 0, f \leq 0 \},
\]

where \( f = |p_t| + \mu p_n \) and \( H^{-1/2} \) is the dual space of \( H^{1/2} \).

The virtual internal work is defined by a bilinear map
B : \mathcal{U} \times \mathcal{V} \to \mathbb{R},

\begin{align*}
B(u,v) &= \int_{\Omega} \varepsilon_{ijkl} \ u_{i,j} \ v_{k,l} \, dV, \\
    &= \int_{\Omega} \sigma_{kl} \ v_{k,l} \, dV,
\end{align*}

for each body. It is assumed that the elasticity tensor satisfies the ellipticity condition and that

\[ E_{ijkl}(X) \in L^\infty(\Omega) \]

where \( L^\infty \) is a space of all measurable and bounded functions with the norm

\[ \| \phi \|_{L^\infty(\Omega)} = \text{ess sup} \ |\phi(X)|. \]

The bilinear form \( B(\cdot,\cdot) \) is now continuous and \( \mathcal{V} \)-elliptic so that

\[ |B(u,v)| \leq M \| u \|_\mathcal{U} \| v \|_\mathcal{V} \]

for \( a_M > 0 \)

and

\[ B(v,v) \geq m \| v \|^2_\mathcal{V} \]

for \( a_m > 0 \).

for \( \forall u \in \mathcal{U} \) and \( v \in \mathcal{V} \). Both bodies are therefore assumed as stable elastic materials.

The virtual external work is defined by

\[ f(v) = \int_{\Gamma_f} f \cdot v \, dS \]

for each body where \( f \in L^2(\Gamma_f) \). The quantity \( f(\cdot) \) denotes a continuous linear functional on \( \mathcal{V} \) such that

\[ f(v) \leq \| f \|_{L^2(\Gamma_f)} \| v \|_\mathcal{V}. \]
where \( \|f\|_{L^2(\Gamma_f)} = \left[ \int_{\Gamma_f} f \cdot f \, dS \right]^{1/2} \).

The virtual work on the contact surface \( \Gamma_c \) is defined by

\[
[p,v] = \int_{\Gamma_c} p \cdot v \, dS \quad \text{for a } p \in \mathcal{F} \text{ and } \forall v \in \mathcal{F},
\]

and the complementary work by

\[
[t,z] = \int_{\Gamma_c} t \cdot z \, dS \quad \text{for a } z \in \mathcal{Z} \text{ and } \forall t \in \mathcal{F},
\]

where \([\cdot,\cdot]\) denotes the duality pairing on \( \mathcal{F} \times \mathcal{Z} \subset H^{-1/2}(\Gamma_c) \times H^{1/2}(\Gamma_c) \).

With the above definitions, we can now define a variational problem.

**PROBLEM II.**

Find the displacements \( u^a \in \mathcal{U}^a \) and \( u^b \in \mathcal{U}^b \) for each body and the contact traction \( p \in \mathcal{F} \) such that

\[
B^a(u^a,v^a) - f^a(v^a) - [p^a,v^a] = 0 , \quad \forall v^a \in \mathcal{U}^a \tag{2.4.31}
\]

\[
B^b(u^b,v^b) - f^b(v^b) - [p^b,v^b] = 0 , \quad \forall v^b \in \mathcal{U}^b \tag{2.4.32}
\]

\[
[t - p,z] \geq [\mu t_n - \mu p_n, |z_n|] , \quad \forall t \in \mathcal{F} \tag{2.4.33}
\]

where \( z = u^a - u^b + z^0 \) and \( p = p^a = - p^b \).

**Theorem I.** Problem I and Problem II are equivalent.
Proof: The proof consists of the following two parts.

(1) If the set of variables \((u^a, u^b, p)\) satisfies Problem I,
\[
\int_{\Omega} \sigma_{ij}v_i \, dV = -\int_{\Omega} \sigma_{ij}v_i \, dV + \int_{\Gamma} \sigma_{ijn}v_i \, dS.
\]

Since \(\sigma_{ij,j} = 0\) in \(\Omega\), \(\sigma_{ijn,j} = f_i\) on \(\Gamma_f\) and \(\sigma_{ijn,j} = p_i\) on \(\Gamma_c\) for each body,
\[
\int_{\Omega} \sigma_{ij}v_i \, dV = \int_{\Gamma} \sigma_{ijn}v_i \, dS
\]
\[
= \int_{\Gamma} \sigma_{ijn}v_i \, dS + \int_{\Gamma_f} f_i v_i \, dS + \int_{\Gamma_c} p_i v_i \, dS.
\]

Thus, since \(v = 0\) on \(\Gamma_u\) for \(\forall v \in \mathcal{V}\),
\[
\int_{\Omega} \sigma_{ij}v_i \, dV = \int_{\Gamma_f} f_i v_i \, dS + \int_{\Gamma_c} p_i v_i \, dS,
\]
or
\[
B(u,v) - f(v) - [p,v] = 0, \quad (2.4.34)
\]
for each body. Now consider the inequality part
\[
\Theta \equiv [\tau - p,z] - [\mu \tau_p - \mu p_n, |z_t|]
\]
\[
= [\tau_n - p_n,z_n] + [\tau_t - p_t,z_t] - [\mu \tau_p - \mu p_n, |z_t|],
\]
where \(z \in \mathcal{Z} \subset H^{1/2}(\Gamma_c)\) as defined previously. Noting from (2.4.29) that \(z_n p_n = 0\), \(-z_t p_t = |z_t| |p_t|\), \(|z_t|(|p_t| + \mu p_n) = 0\) on \(\Gamma_c\),
\[
\Theta = [\tau_n,z_n] + [\tau_t,z_t] - [\mu \tau_n, |z_t|]
\]
\[
\geq [\tau_n,z_n] - [|\tau_t|, |z_t|] - [\mu \tau_n, |z_t|]
\]
\[
= [\tau_n,z_n] - [|\tau_t| + \mu \tau_n, |z_t|].
\]

It is further noted that \(z_n \leq 0\) since \(z \in \mathcal{Z}\) and that \(\tau_n \leq 0\)
and $|\tau_t| + \mu \tau_n \leq 0$ for $\forall \tau \in \mathcal{F}$. Therefore,

$$\Theta \geq 0, \quad \forall \tau \in \mathcal{F}$$

or $[\tau - p, z] \geq [\mu \tau_n - \mu \rho_n, |z_t|]$. \hfill (2.4.35)

From (2.4.34) and (2.4.35), part (1) is proved.

(2) Conversely, if the set of variables $(u^a, u^b, p)$ satisfies Problem II, then $u^a \in \mathcal{U}^a$, $u^b \in \mathcal{U}^b$, and $p \in \mathcal{P}$ in accordance with the postulate of Problem II. By Green's formula, for $\forall u \in \mathcal{U}$ and $v \in \mathcal{V}$ for each body,

$$B(u,v) = \int_{\Omega} \sigma_{ij} v_i dV$$

$$= - \int_{\Omega} \sigma_{ij} v_i dV + \int_{\Gamma} \sigma_{ij} n_j v_i dS,$$

where the above integrals are interpreted as the duality pairings. Then, (2.4.31-32) can now be rewritten as, for $\forall v \in \mathcal{V}$ and for each body

$$-\int_{\Omega} \sigma_{ij} v_i dV + \int_{\Gamma_f} (\sigma_{ij} n_j - f_i) v_i dS + \int_{\Gamma_c} (\sigma_{ij} n_j - p_i) v_i dS = 0.$$

From the arbitrariness of $v \in \mathcal{V}$, we deduce that, in the distributional sense,

$$\sigma_{ij} n_j = 0, \quad \text{a.e. in } \Omega$$

$$\sigma_{ij} n_j = f_i, \quad \text{a.e. on } \Gamma_f$$

and $\sigma_{ij} n_j = p_i$, \ a.e. on $\Gamma_c$. \hfill (2.4.36a)

Also, $u \in \mathcal{U}$ implies that

$$u_i = u_i, \quad \text{a.e. on } \Gamma_u$$

\hfill (2.4.36b)
Now consider the inequality

$$\theta \geq 0 \quad \forall \tau \in \mathcal{F}$$

Taking $\tau_n = p_n$ and $\tau_t = \pm \mu t_n$, from the arbitrariness of $\tau \in \mathcal{F}$, we get

$$[\mu t_n, z_t] - [p_t, z_t] \geq 0 ,$$

and

$$-[\mu t_n, z_t] - [p_t, z_t] \geq 0 .$$

Adding the above two inequalities, we get

$$-[p_t, z_t] \geq 0 ,$$

or

$$-z_t p_t \geq 0 . \quad \text{a.e. on } T_c$$

(2.4.37a)

By substituting $-[p_t, z_t] = [ |p_t|, |z_t| ]$ into the definition of $\theta$, the inequality now becomes

$$[\tau_n - p_n, z_n] + [ |p_t| + \mu p_n, |z_t| ] + [\tau_t, z_t] - [\mu t_n, |z_t|] \geq 0 .$$

(2.4.38)

It is then noted that

$$[\tau_t, z_t] - [\mu t_n, |z_t|]$$

$$\geq -[|\tau_t|, |z_t|] - [\mu t_n, |z_t|]$$

$$= -[|\tau_t| + \mu t_n, |z_t|] \geq 0 . \quad \forall \tau \in \mathcal{F}$$

(2.4.39)

Hence, the inequality (2.4.38) becomes

$$[\tau_n - p_n, z_n] + [ |p_t| + \mu p_n, |z_t| ] \geq 0 ,$$

(2.4.40)

which implies that, when $\tau_n = p_n$,

$$[ |p_t| + \mu p_n, |z_t| ] \geq 0 .$$

But, $[ |p_t| + \mu p_n, |z_t| ] \leq 0$ because $|p_t| + \mu p_n \leq 0$ for $p \in \mathcal{F}$.

Therefore, we must have
\[ |p_t| + \mu p_n, |z_t| \leq 0, \]
or \[ |z_t|(|p_t| + \mu p_n) = 0. \quad \text{a.e. on } \Gamma_c \quad (2.4.37b) \]

Hence, (2.4.40) implies that

\[ [\tau_n - p_n, z_n] \geq 0. \quad (2.4.41) \]

From which we see that

\[ [p_n, z_n] \geq 0 \quad \text{and} \quad -[p_n, z_n] \geq 0, \]

by selecting \( \tau_n = 2p_n \) and \( \tau_n = 0 \), respectively, which implies that

\[ [p_n, z_n] = 0. \]

or \( z_n p_n = 0. \quad \text{a.e. on } \Gamma_c \quad (2.4.37c) \)

Finally, substituting (2.4.37c) into (2.4.41), we get

\[ [\tau_n, z_n] \geq 0. \]

which implies that, since \( \tau_n \leq 0 \),

\[ z_n \leq 0. \quad \text{a.e. on } \Gamma_c \quad (2.4.37d) \]

Thus, from (2.4.36a,b) and (2.4.37a,b,c,d), part (2) is also proved. \( \square \)

The variational principle in the present work consists of two parts. Eq (2.4.31-32) corresponds to the principle of minimum potential energy for a fixed \( p \). Eq (2.4.33), on the other hand, expresses the principle of maximum frictional dissipation energy in the admissible contact traction field for a fixed \( z \).
For further discussion, (2.4.33) is decomposed into two parts,

\[ [\tau_n - p_n \cdot z_n] \geq 0 \quad (2.4.42a) \]
\[ [\tau_t - p_t \cdot z_t] \geq [\mu \tau_n - \mu p_n \cdot |z_t|] \quad (2.4.42b) \]

simply by relaxing the coupling effect between the transversality condition and the friction condition. Then, (2.4.42a) defines the non-penetration condition for a fixed \( z \). If \( p_n \) is fixed according to (2.4.42a), (2.4.42b) now becomes

\[ [\tau_t - p_t \cdot z_t] \geq 0 \]
\[ \forall \tau_t \in \{\tau_t \mid |\tau_t| \leq -\mu p_n \}, \text{ where } \tilde{p}_n \leq 0 \text{ is given} \]

which accounts for the principle of maximum dissipation energy due to friction \( D(p) = -[p_t \cdot z_t] \). Thus, the variational problem in the present work represents a mini-max problem.

It may be remarked that the present method is similar to the hybrid methods in the sense that the constraint conditions imposed on a boundary are treated independent of the field equations defined on the interior domain by introducing the additional dependent variable \( p \) on the contact boundary. In the present method, however, the constraint conditions (inequality relations) are treated directly.

By taking the contact traction as an additional dependent variable, the ambiguity in defining the admissible
function space for the displacement associated with the minimum potential energy can be removed as discussed in Section 2.3.3.

The existence and uniqueness of solution is discussed for the corresponding discrete equations in Section 4.2. As a necessary condition for the variational problem to possess a solution, the following compatibility condition must be satisfied:

\[ f(v) + [p,v] = 0 \quad \forall \ v \in \mathcal{N} \]

for each body, where \( \mathcal{N} \) denotes the null space of the corresponding formal adjoint operator, i.e., a space of rigid motion (rigid body translation and rotation) for each body. Thus, if there is no fixed boundary (i.e., if \( \Gamma_u = \emptyset \)) for a body, it is obvious that the induced contact traction must always be balanced with the external load \( f \). From a computational standpoint, however, some difficulties are encountered in developing solution methods using the compliance of each body due to the presence of rigid body motions for such a case. Computational strategies for the resolution of these difficulties are also discussed in the subsequent chapter.
CHAPTER III
FINITE ELEMENT APPROXIMATION

A finite element approximation of the variational continuum problem (Problem II) derived in the preceding chapter is investigated in this chapter. Essential features of the presented finite element approximation are first discussed in Section 3.1. A set of discrete equations is obtained in Section 3.2. Computational methods for the effective evaluation of the pairing map and the normal vector are studied in Section 3.3. Discrete governing contact equations are derived in Section 3.4.

3.1. General Remarks

Some intrinsic difficulties associated with the discrete contact problems must be resolved for the standard application of the finite element method.

The variational inequality can be transformed by penalizing the indicator functional and by replacing the constrained admissible function space by an unconstrained space \([36,55,56,59]\). On the other hand, a direct method may also be applied by introducing a special projection mapping from an unconstrained admissible function space to a
constrained one instead of adopting a penalization scheme. By means of such a special mapping, the inequality part can then be represented in the form of a fixed point problem. Proposed mapping method and a solution algorithm for the corresponding fixed point problem are discussed in the sequel.

The idea of the node-to-segment contact is adopted here simply because it is easy to implement and known to be efficient. Element boundaries between two contacting bodies do not conform in this case, however. In order to resolve such a non-conformity problem, the contact boundary is now considered to be a finite set of contact nodes on body 'a' and their pairing points on body 'b'. The contact traction is then taken as the nodal contact force. It is noted that an admissible interpolation function for the contact traction can be selected in a sufficiently large finite dimensional subspace which contains the Dirac delta function (distribution).

In the sequel, the following acronyms are used for the description of discrete contact problems.

NUMEL : the number of element for each body,
NUMNP : the number of nodes for each body,
NUMEQ : the number of total (active) degrees of freedom for each body,
NEN : the number of nodes in an element,
NSN : the number of nodes on a side of an element,
NCA, NCB : the numbers of nodes on the possible contact boundary \( \Gamma^a_0 \) and \( \Gamma^b_0 \) of body 'a' and 'b', respectively,
NP (= NCA) : the number of pairing points for body 'b'.
NC = NCA + NCB.

3.2. Discrete System Equations

Applying the standard procedures of the finite element method (see general references [3, 28, 54, 81, etc.]) to the variational equations (2.4.31-32), we get a set of discrete equilibrium equations for each body as follows:

\[ K^a \mathbf{u}^a = 2 \mathbf{F}^a + H^a \mathbf{p}^a , \]  
(3.2.1)

and

\[ K^b \mathbf{u}^b = 2 \mathbf{F}^b + H^b \mathbf{p}^b , \]  
(3.2.2)

where, for the respective body,

\[ K = \sum_{e=1}^{\text{NUMEL}} \int_{\Omega^e} B^T E B \, dv^e , \]

\[ F = \sum_{e=1}^{\text{NUMEL}} \int_{\Gamma_e^f} N^T f \, ds^e , \]

\[ H = \sum_{e=1}^{\text{NUMEL}} \int_{\Gamma_e^o} N^T L \, ds^e , \]

\[ u_e(\xi, \eta) = \sum_{\alpha=1}^{\text{NEN}} N^\alpha(\xi, \eta) \mathbf{u}^\alpha , \]

\[ \mathbf{x}_e(\xi, \eta) = \sum_{\alpha=1}^{\text{NEN}} N^\alpha(\xi, \eta) \mathbf{x}^\alpha , \]

\[ \mathbf{p}_e(s) = \sum_{\alpha=1}^{\text{NSN}} L^\alpha(s) \mathbf{p}^\alpha , \]

the symbol 'ξ' denotes the global assembly process of finite
elements, \( \xi \) and \( \eta \) denote the isoparametric co-ordinates, and \( s \) represents the boundary co-ordinate. Matrices \( N \) and \( L \) represent the interpolation functions, matrices \( B \) and \( E \) represent the strain-displacement relation and the elasticity tensor, respectively. The applied loads \( f \) are treated differently from the contact traction \( p \).

As discussed in the preceding section, the contact traction distribution is taken as a singular distribution. Thus, the matrix \( H \) is an identity and square matrix. In this respect, nodal contact forces \( p^a \) and \( p^b \) for each body can be used equivalently instead of the respective nodal contact traction vector \( p \). Equations (3.2.1) and (3.2.2) are now rewritten as

\[
K^a u^a = F^a + p^a, \tag{3.2.3}
\]

and

\[
K^b u^b = F^b + p^b, \tag{3.2.4}
\]

where the left-superscript \( 2 \) in the sequel, is omitted for brevity unless necessary.

**Remark**

It is sometimes convenient to represent the nodal contact force as a non-singular distribution of the contact traction for the interpretation of a physical problem. In these circumstances, there may be several alternatives: (i) the obtained contact force is regarded as a consistent nodal force vector (i.e., \( L \) is taken to be the same as \( N \)), and the nodal contact traction is then computed accordingly, or (ii) the nodal contact traction is considered simply as the characteristic area-average value of the corresponding nodal contact force. In the former, the contact traction may not satisfy the condition that \( p \in \$ \) despite the fact that
\( \mathbf{P} \in \mathcal{F} \), and an additional computation is always required. Thus, the latter scheme is computationally promising.

The reciprocal form of the above equations can be written in the form

\[
\mathbf{u}^a = \mathbf{G}^a \mathbf{P}^a + \mathbf{u}_F^a, \quad (3.2.5)
\]

and

\[
\mathbf{u}^b = \mathbf{G}^b \mathbf{P}^b + \mathbf{u}_F^b, \quad (3.2.6)
\]

where \( \mathbf{G} \) represents the matrix of the respective Green's function associated with the nodal contact forces \( \mathbf{P} \) in the same manner as described in Section 2.4.1., and \( \mathbf{u}_F \) is the nodal displacement vectors owing to the prescribed external forces \( \mathbf{F} \). In comparison with (2.4.1-2), (3.2.5-6) can also be expressed as

\[
\mathbf{x}^a = \mathbf{G}^a \mathbf{p}^a + \mathbf{w}^a, \quad (3.2.7)
\]

and

\[
\mathbf{x}^b = \mathbf{G}^b \mathbf{p}^b + \mathbf{w}^b, \quad (3.2.8)
\]

where \( \mathbf{w} = \mathbf{0} + \mathbf{u}_F \) and \( \mathbf{x} = \mathbf{0} + \mathbf{u} \) for each body. It is noteworthy that the above sets of equations (3.2.5-6) and (3.2.7-8) contain only the applicable nodes on the possible contact boundary \( \Gamma_0 \) for each body. It is also noted that the matrices \( \mathbf{G} \) are real, symmetric, and positive definite (RSPD).

In the present work, (3.2.5) and (3.2.6) are often expressed in the following 'block' component form:

\[
\mathbf{u}_I^{a_{\text{NCA}}} = \sum_{J=1}^{\text{NCA}} \mathbf{G}^a_{IJ} \mathbf{P}^a_J + (\mathbf{u}_F^a)_I, \quad 1 \leq I \leq \text{NCA} \quad (3.2.9)
\]
and
\[
\mathbf{u}_I^b = \sum_{J=1}^{NCB} \mathbf{G}_{IJ}^b \mathbf{P}_J + \mathbf{(u}_F^b)_I^b, \quad 1 \leq I \leq NCB
\]  
(3.2.10)

where \( \mathbf{u}_I^b \), \( \mathbf{P}_J \), and \( \mathbf{(u}_F^b)_I^b \) are two-by-one 'block' vectors, and \( \mathbf{G}_{IJ}^b \) is the corresponding two-by-two 'block' matrix for two-dimensional problems,

\[
\mathbf{u}_I^b = \begin{bmatrix} u_{2I-1} \\ u_{2I} \end{bmatrix}, \quad \text{and} \quad \mathbf{G}_{IJ}^b = \begin{bmatrix} G_{2I-1,2J-1} & G_{2I-1,2J} \\ G_{2I,2J-1} & G_{2I,2J} \end{bmatrix}.
\]

Here, and in the sequel, the capital indices \( I,J,K, \) etc. are used as nodal point counters which range up to the number of nodes on the possible contact boundary for the respective body, and the summation convention is assumed for the repeated indices unless specified otherwise.

The displacements \( \mathbf{u}_F^a \) and \( \mathbf{u}_F^b \) can be obtained by solving (3.2.3) and (3.2.4) for the nodal displacements resulting from the total prescribed external loads \( \mathbf{F}^a \) and \( \mathbf{F}^b \) by setting \( \mathbf{P}^a = \mathbf{F}^b = 0 \), respectively, and the displacements \( \mathbf{u}_F^a \) can be simply taken as a fraction of \( \mathbf{u}_F^b \) corresponding to the loading stage at \( \tau = t + At \). On the other hand, matrices \( \mathbf{G}^a \) and \( \mathbf{G}^b \) associated with nodes on the possible contact boundary can also be evaluated effectively as described in Section 4.4.2. For the case that body 'b' is rigid, \( \mathbf{G}^b = 0 \), and \( \mathbf{u}_F^b \) represents the corresponding rigid body displacement. Finally, it is noted that a pre-strained artificial spring element is used to remove rigid body motions when \( \Gamma_u = \phi \).
Details of such a scheme are explained in Section 5.5.

3.3. Evaluation of the Pairing Points and the Normal Vector

3.3.1. The Pairing

At the beginning of each incremental stage, the pairing map $\mathcal{K}$ is updated in accordance with the configurations of contact bodies obtained at the previous stage. For the contact candidate nodes which do not contact with body 'b' at the end of the previous stage, the pairing point (denoted by $1+\beta \mathcal{K}$ in Section 2.4.1.) is predicted as described below.

The pairing point ($1+\beta \mathcal{K}$) is assumed as a point with which a contact candidate node ($x \in T_0$) comes into contact first during the simultaneous application of both $\Delta u^a$ and $\Delta u^b$ defined by

$$\hat{u}^a = g^a (t^a + t^- \Delta t^a) + \Delta u^a_F, \quad (3.3.1)$$

where $\Delta u^a_F = t^+ \Delta u^a_F - t^a_F$ is a known value as mentioned in the preceding section, and $\Delta u^b_F$ is also defined in the same way. Since $t^a_u$ and $t^- \Delta t^a_u$, where $u_p \equiv G P$, are already computed at the previous stages for each body, a simple additional computation is required for evaluating the increment of displacement $\Delta u_p$'s owing to the change of contact force, that is,

$$\Delta u_p = t^a_u - t^- \Delta t^a_u = g (t^a_P - t^- \Delta t^a_P) \quad (3.3.2)$$

for each body.
Remark

It is noted that $\hat{\Delta u}$'s may be regarded as $\Delta u_F$ for the sake of simplicity. However, the direction of $\Delta u_F$ may be much different from that of the actual trajectory. Thus, employing such a scheme may result in inaccurate solutions especially when the coefficient of friction is large enough to maintain a stick-dominant state.

For instance, a pairing point $^1K$ paired with a contact candidate node $I$ of body 'a' and lying on a line segment between nodes $K$ and $L$ of body 'b' is determined so that the following relation is satisfied (Fig. 9).

$$X_a^i = (1-\kappa) X_K^b + \kappa X_L^b, \quad (3.3.3)$$

where

$$X_a^i = X_I^a + \beta \Delta u_I^a,$$

$$X_K^b = X_K^b + \beta \Delta u_K^b,$$

$$X_L^b = X_L^b + \beta \Delta u_L^b,$$

$$0 < \kappa < 1, \text{ and } 0 < \beta < 1 + \epsilon.$$

Here $\beta$, as explained in Section (2.4.1), denotes an increment parameter during $\tau \in [t, t + (1 + \epsilon)\Delta t]$ where $\epsilon$ is a small positive value, and $\kappa$ is an interpolation parameter representing the relative location of the pairing point $^1+\beta K$ to its adjacent nodes $K$ and $L$ such that

$$^1+\beta K = (1-\kappa) X_K^b + \kappa X_L^b. \quad (3.3.4)$$

The parameter $\beta$ is increased step-by-step (say, 0.2, etc.) and then condition (3.3.3) is imposed within an allowable error bound for $0 \leq \kappa \leq 1$. 
The parameter $\kappa$ is evaluated for all pairing points $^1\mathbb{R}(X_i^a)$ which are in contact or separated at $r = t$. The values for these pairing points can be estimated from the values at their adjacent nodes $K$ and $K+1$ by means of the following linear interpolation relations:

\[ u^p_I = (1-\kappa_I) u^b_K + \kappa_I u^b_{K+1}, \quad (3.3.5) \]

and

\[ (u^p_F)_I = (1-\kappa_I) (u^p_F)_K + \kappa_I (u^p_F)_{K+1}, \quad (3.3.6) \]

where $u^p_I$ denotes the displacement of a pairing point $^1\mathbb{R}(X_i^a)$ of body 'b' and $\kappa_I$ does its value for $^1\mathbb{R}(X_i^a)$. In accordance with a new nodal system for body 'b' whose fictitious nodes (denoted by $^1\mathbb{R}$) are paired with the contact candidate nodes of body 'a', (3.2.10) is now written as

\[ u^p_{\text{NCA}} = \sum_{J=1}^{\text{NCA}} G^p_{IJ} P^p + (u^p_F)_I, \quad 1 \leq I \leq \text{NCA} \quad (3.3.7) \]

where $G^p$ represents the matrix of Green's function associated with the nodal contact force corresponding to the new nodal pairing. Except for the case of a rigid body (i.e., $G^b = 0$), $G^p_{IJ}$ is evaluated approximately by

\[ G^p_{IJ} = (1-\kappa_J) [(1-\kappa_I) G^b_{KL} + \kappa_I G^b_{K+1,L}] 
+ \kappa_J [(1-\kappa_I) G^b_{K,L+1} + \kappa_I G^b_{K+1,L+1}], \quad (3.3.8) \]

where the adjacent nodes for the fictitious nodes $^1\mathbb{R}(X_i^a)$ and $^1\mathbb{R}(X_j^a)$ are denoted by nodes $K$ and $K+1$, and nodes $L$ and $L+1$, respectively, as depicted in Fig. 9(b). In (3.3.8), it is
assumed that the contact force $p^p_j$ at the fictitious node $\text{R}(X_j)$ can be decomposed into its contributions at the adjacent nodes $L$ and $L+1$ such that

$$p^b_L = (p^b_L) + (1-\kappa_j) p^p_j,$$

and

$$p^b_{L+1} = (p^b_{L+1}) + \kappa_j p^p_j.$$  \hspace{1cm} (3.3.9)

It can be shown from (3.3.8) that the symmetry of $G^p$ is maintained, i.e., $G^p_{IJ} = G^p_{JI}$ when $G^b_{IJ} = G^b_{JI}$.

3.3.2. The normal vector

The normal for each pair of contact points is evaluated from an 'assumed' deformed configuration in the following manner (see also Fig. 9(c)). It is first assumed that each pair of contact points $X^a$ and $\text{R}(X^a)$, which are determined previously, is bonded together during application of the current increment of external loads. The deformed position vector $x_I = (x^p_I, y_I)$ for a pair of contact points $X^a_I$ and $\text{R}(X^a_I)$ is then determined approximately such that the following relations are satisfied, with reference to a fixed rectangular co-ordinate system,

$$x^p_I = x^p_I + (x^a_I - x^p_I) \frac{(G^p_{II})_{xx}}{(G^a_{II})_{xx} + (G^p_{II})_{xx}},$$

and

$$(3.3.10)$$
\[ y_I = y_I^p + \left( y_I^a - y_I^p \right) \frac{(c_{II}^p)_{yy}}{(c_{II}^a)_{yy} + (c_{II}^p)_{yy}}. \]

where

\[ x_I^a = (x_I^a, y_I^a) = x_I^a + \Delta x_I^a, \]
\[ x_I^p = (x_I^p, y_I^p) = x_I^p + \Delta x_I^p \]
\[ = (1 - \kappa_I) x_K + \kappa_I x_K^b, \]

and

\[ [G]_{II} = \begin{bmatrix} G_{xx} & G_{xy} \\ G_{xy} & G_{yy} \end{bmatrix}_{II} = \begin{bmatrix} G_{2I-1,2I-1} & G_{2I-1,2I} \\ G_{2I,2I-1} & G_{2I,2I} \end{bmatrix}. \]

A line connecting these points \( x \) is now regarded as the 'assumed' deformed configuration of the contact boundary. The normal direction is then taken as an outward normal to a line connecting the center points of two adjacent sides for each pair with respect to body 'a' as depicted in Fig. 9(c).

Remark

The 'assumed' deformed configuration obtained in this way is not the same for the respective perfectly-bonded problem. In order to obtain the configuration corresponding to the latter problem, the linear equation \( MP + W = 0 \) from (3.4.4) is first solved for \( P \) and (3.2.5) is then solved for \( u^a \). If the matrix \( M \) is diagonal-dominant, the above method may be used effectively.

A transformation matrix \([R]_I\) for each pair is now defined by

\[ [R]_I = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}, \]

where the angle \( \theta \) is measured from the x-axis to the normal
n-axis in a counter-clockwise sense. Here and in the sequel, (x,y) and (n,t) denote global and local Cartesian co-ordinate systems, respectively. Thus, any tensor can be expressed in terms of its components either in the global or the local co-ordinate system interchangeably. For instance,

\[
\{P^a\}_I = \begin{bmatrix} p_n^a \\ p_t^a \end{bmatrix} = [R]_I \begin{bmatrix} p_x^a \\ p_y^a \end{bmatrix},
\]

and

\[
[G^a]_{II} = \begin{bmatrix} G_{nn}^a & G_{nt}^a \\ G_{nt}^a & G_{tt}^a \end{bmatrix}_{II} = [R]_I \begin{bmatrix} G_{xx}^a & G_{xy}^a \\ G_{xy}^a & G_{yy}^a \end{bmatrix}_{II} [R]_I^T,
\]

where the repeated indices do not imply summation.

3.4. Governing Contact Equations

The deformed gap vector for the I-th pair of contact points is now written as

\[
z_I = x_I^a - x_I^p
= u_I^a - u_I^p + \delta X_I^a - \delta X_I^p,
\]

where \( x_I \) and \( \delta X_I \) denote the position vectors of I-th node (or fictitious node) at \( \tau = t + \Delta t \) and \( \tau = 0 \), respectively. In place of the local equilibrium condition explained in Chapter 2, the nodal contact force vector must satisfy the following relation.

\[
P_I = p_I^a = - p_I^p,
\]

for each pair of contact points at the current incremental
stage.

By substituting (3.2.9), (3.3.7), and (3.4.2) into (2.4.1), the governing contact equations can be derived in the same form as (2.4.10) in Section 2.4.1, such that

$$z_I = \sum_{J=1}^{NCA} M_{IJ} P_J + W_I, \quad 1 \leq I \leq NCA \quad (3.4.3)$$

where

$$M_{IJ} = C_{IJ}^a + C_{IJ}^p,$$

and

$$W_I = \left( u^a_F \right)_I - \left( u^p_F \right)_I + \left( x^a_I - x^p_I \right).$$

The term \( \left( x^a_I - x^p_I \right) \) is often called the initial gap vector for the \( I \)-th pair. Eq (3.4.3) characterizes the behavior of the gap between each pair of contact points in response to the action of contact forces at the current incremental stage. Eq (3.4.3) provides \( 2 \times NCA \) equations in terms of \( 4 \times NCA \) unknowns of \( z \) and \( P \) subjected to the contact conditions.

A discrete form of the variational problem discussed in Chapter II can be stated in the following compact form.

**PROBLEM III.**

For a given \( W \), we solve a set of the governing contact equations

$$z = M P + W \quad (3.4.4)$$

for \( P = \{P_1, \cdots, P_I, \cdots, P_N\} \in \mathbb{R}^N \) and \( z = \{z_1, \cdots, z_I, \cdots, z_N\} \in \mathbb{Z}^N \).
subjected to the following contact conditions (3.4.5 abcd), for each $I$,

(a) separation (non-contact): \( z_n \leq 0 \) \( \Rightarrow \) \( z_t \) if \( P_n = 0 \) and \( P_t = 0 \).

(b) stick: \( z_n = 0 \), \( z_t = 0 \) if \( P_n < 0 \) and \( |P_t| < \mu |P_n| \).

(c) slip: \( z_n = 0 \), \( z_t \geq 0 \) if \( P_n < 0 \) and \( P_t = \mu P_n \).

(d) slip: \( z_n = 0 \), \( z_t \leq 0 \) if \( P_n < 0 \) and \( P_t = -\mu P_n \),

where \( P_I \in \Phi = \{(P_n, P_t) \mid P_n \leq 0 \text{ and } |P_t| + \mu P_n \leq 0\} \) and \( z_I \in Z = \{(z_n, z_t) \mid z_n \leq 0\} \).

The contact conditions presented in (3.4.5 a,b,c and d) are equivalent to the following:

\[ z_n \leq 0 \text{, } P_n \leq 0 \text{, } z_t P_n = 0 \text{, } |P_t| + \mu P_n \leq 0 \text{, } z_t(|P_t| + \mu P_n) = 0 \text{, and } z_t P_t \leq 0. \]

The domains of the constraint sets \( \Phi \subset \mathbb{R}^2 \) and \( Z \subset \mathbb{R}^2 \) are depicted in Fig. 10. For the case of conforming contact problems, \( H^a = H^b \) in (3.2.1-2) since the contact surfaces for each body occupy the same region at each incremental stage, and (3.4.4) simply becomes

\[ z = M H p + W. \]
where \( M = G^a + G^b \) and \( p \) denotes the contact traction here. For frictionless contact problems, i.e., \( \mu = 0 \), PROBLEM III can be viewed as a complementary problem described by

\[
z = M P + W.
\]

subjected to \( z_i \leq 0, P_i \leq 0 \), and the complementary condition \( z_i P_i = 0 \) for each \( 1 \leq i \leq N \). Here, the above set of equations includes only those quantities associated with the normal component. Therefore \( M \) is now a rank-\( N \) matrix. The solutions for this problem have been rigorously studied for the general cases of \( M \) and \( W \) in Reference [49]. If \( M \) is a P-matrix (a matrix whose principal minors are positive), the problem has one and only one solution for any \( W \). Thus, the discrete problem can be shown to have a unique solution for any \( W \).
CHAPTER IV
SOLUTION METHOD

In this chapter, an iterative solution method for PROBLEM III is studied. Once the contact status for each pair is known, the problem can be solved for the contact force directly because each contact condition possesses two equality relations. Conversely, all solutions for the problem can be found by solving the linear equations $4^{NCA}$ times because the contact status for each pair is classified into four different kinds in the present work. Various solution methods have been proposed, ranging from trial-and-error methods with decision tables to mathematical programming techniques in optimization.

Many practical problems, unlike general mathematical problems, have several nice properties such as symmetry, positive definiteness, diagonal dominance, uniqueness of solution, etc., which can be conveniently used in developing an effective solution method. For instance, if the values of (block) diagonal components of matrix $\mathbf{M}$ in (3.4.4) are sufficiently larger than those of the off-diagonal components, the contact status for a pair of contact points may not be influenced considerably by the variation of
contact forces of other pairs. In this situation, the contact status and the corresponding contact force for a pair of contact points can be evaluated by solving two nodal equations subjected to the contact conditions with the updated contact forces for the other pairs in a successive manner. In most practical problems, moreover, the conditions for the uniqueness of solution (to be discussed later) are satisfied so that the nodal equations can be solved in a much simpler manner without examining all other possible solutions.

Also, various iterative solution methods have been studied for a large system of linear equations such as the Jacobi method and the Gauss-Seidel method. These methods are sometimes superior to direct methods (see general references [3,25,58,73, etc.]), especially when the system matrix is big and sparse. The matrix $M$ in the present work is neither very large nor sparse. However, such iterative solution methods for a linear system can be applied effectively by embedding constraint condition(s) into the iteration routine. In the present work, such an idea is applied to solve PROBLEM III.

The embedding scheme for the contact conditions into the governing contact equations (3.4.4) is discussed in Section 4.1. An embedding map is designed for projecting an arbitrary element in $\mathbb{R}^2 \times \mathbb{R}^2$ into the solution space $\mathbb{Z} \times \mathbb{F} \subset \mathbb{R}^2 \times \mathbb{R}^2$ which satisfies both the contact conditions and
the nodal contact equations. With the help of the embedding map, PROBLEM III is then represented in the form of a fixed point problem in Section 4.2. The existence and uniqueness of the solution is discussed using the Banach fixed point theorem. In Section 4.3., an iterative solution scheme based on the block Gauss-Seidel method and the convergence criteria for the iterative scheme are investigated. Finally, the overall solution algorithm for a class of linear elastic contact problems is summarized in Section 4.4.1. A numerical scheme for the effective evaluation of the discrete Green's functions is explained in Section 4.4.2.

4.1. Method of Embedding of Contact Conditions

The matrix $M$ can be decomposed into diagonal, lower and upper triangular parts in the following block form:

$$M = D - L - U ,$$  \hspace{1cm} (4.1.1)

where

$$[D] = \begin{bmatrix} [M]_{11} & [M]_{22} & 0 \\ 0 & \ddots & \ddots \\ 0 & \ddots & [M]_{NN} \end{bmatrix},$$

$$[L] = - \begin{bmatrix} [0]_{11} & \ddots & 0 \\ [M]_{21} & \ddots & \ddots \\ \vdots & \ddots & \ddots \\ [M]_{N1} & \ddots & [M]_{N,N-1}[0]_{NN} \end{bmatrix}.$$
\[ [U] = - \begin{bmatrix} [0]_{11} & [M]_{12} & \cdots & [M]_{1N} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & [M]_{N-1,N} & [0]_{NN} \end{bmatrix} \]

and \( L = U^T \) since \( M \) is symmetric. Here, \([M]_{IJ}\) represents a 2x2 block matrix whose components are expressed in terms of the local co-ordinate system \((n-t)\) such that

\[ [M]_{IJ} = \begin{bmatrix} M_{nn} & M_{nt} \\ M_{nt} & M_{tt} \end{bmatrix}_{IJ} \]

By substituting (4.1.1) into (3.4.4), the governing contact equations can be written as

\[ z = D P + \bar{z}, \quad (4.1.2) \]

where \( \bar{z} = W - L P - U P \). Also, in the block component form for the \( I \)-th pair, (no summation for index \( I \))

\[ z_I = D_{II} P_I + \bar{z}_I, \quad (4.1.3) \]

or in an explicit form,

\[ \begin{bmatrix} z_n \\ z_t \end{bmatrix}_I = \begin{bmatrix} D_{nn} & D_{nt} \\ D_{nt} & D_{tt} \end{bmatrix}_{II} \begin{bmatrix} P_n \\ P_t \end{bmatrix}_I + \begin{bmatrix} \bar{z}_n \\ \bar{z}_t \end{bmatrix}_I, \]

where \( \bar{z}_I = W_I - L_{IJ} P_J - U_{IJ} P_J \). \( (I>J), \quad (I<J) \) \quad (4.1.4)

Each contact condition in (3.4.5a,b,c, and d) consists of two equality relations as a complementary part to (4.1.3) and inequality relation(s) expressed as a 'Go-No-Go' type criterion. Thus, the solutions of (4.1.3) can be expressed in terms of \( \bar{z}_I \) by substituting two equality relations into
(4.1.3) for each contact status and then examining whether the result satisfies the corresponding inequality relation(s). The number of solution for a fixed $\overline{z}_I$ depends not only on the entries of matrix $D_{II}$ and the friction coefficient $\mu$ but also the data $\overline{z}_I$.

Solutions of the nodal contact equations

A feasible solution satisfies the contact conditions in (3.4.5 a,b,c, and d) and

$$
\begin{bmatrix}
  z_n \\
  z_t
\end{bmatrix} =
\begin{bmatrix}
  D_{nn} & D_{nt} \\
  D_{nt} & D_{tt}
\end{bmatrix}
\begin{bmatrix}
  p_n \\
  p_t
\end{bmatrix} +
\begin{bmatrix}
  \overline{z}_n \\
  \overline{z}_t
\end{bmatrix},
$$

(4.1.5)

or

$$
\begin{bmatrix}
  p_n \\
  p_t
\end{bmatrix} = \frac{1}{|D|}
\begin{bmatrix}
  D_{tt} - D_{nt} & -D_{nt} \\
  -D_{nt} & D_{nn}
\end{bmatrix}
\begin{bmatrix}
  z_n - \overline{z}_n \\
  z_t - \overline{z}_t
\end{bmatrix},
$$

(4.1.6)

where $|D| = D_{nn}D_{tt} - D_{nt}^2$, and $|D| > 0$. □

By substituting two equality relations in (3.4.5 a,b,c, and d) into (4.1.5 or 6), the inequality part can now be written for each respective contact status as follows:

(a) separation: $p_n = p_t = 0$ ;

$$
z_n = \overline{z}_n \leq 0 , \quad \forall \overline{z}_t ,
$$

(4.1.7)

(b) stick: $z_n = z_t = 0$ :

$$
p_n = \frac{1}{|D|} (-D_{tt} \overline{z}_n + D_{nt} \overline{z}_t) < 0 ,
$$

(4.1.8a)

$$
p_t = \frac{1}{|D|} (D_{nt} \overline{z}_n - D_{nn} \overline{z}_t) ,
$$

$$
|D_{nt} \overline{z}_n - D_{nn} \overline{z}_t| < \mu (D_{tt} \overline{z}_n - D_{nt} \overline{z}_t) ,
$$

(4.1.8b)
(c) slip: $z_n = 0$, $P_t = \mu P_n$:

$$(D_{nn} + \mu D_{nt}) P_n = -z_n; \quad P_n < 0, \quad (4.1.9a)$$

$$z_t = (D_{nt} + \mu D_{tt}) P_n + \overline{z_t}; \quad z_t > 0. \quad (4.1.9b)$$

(d) slip: $z_n = 0$, $P_t = -\mu P_n$:

$$(D_{nn} - \mu D_{nt}) P_n = -\overline{z}_n; \quad P_n < 0, \quad (4.1.10a)$$

$$z_t = (D_{nt} - \mu D_{tt}) P_n + \overline{z_t}; \quad z_t < 0. \quad (4.1.10b)$$

In order to examine the solutions all possible cases are classified into three parts, namely, $D_{nn} >$, $=,$ or $< \mu |D_{nt}|$.

It is noted that $D_{nn} > 0$, $D_{tt} > 0$, and $|D| > 0$. For notational compactness, $\gamma_1$ and $\gamma_2$ are defined by

$$\gamma_1 = \frac{D_{nt} + \mu D_{tt}}{D_{nn} + \mu D_{nt}}, \quad \text{and} \quad \gamma_2 = \frac{D_{nt} - \mu D_{tt}}{D_{nn} - \mu D_{nt}}.$$ 

Case (1): $D_{nn} > \mu |D_{nt}|$:

(1-a): when $D_{nt} > 0$, then $\gamma_2 < \frac{D_{nt}}{D_{nn}} < \gamma_1 < \frac{D_{tt}}{D_{nt}}$.

(1-b): when $D_{nt} < 0$, then $\frac{D_{tt}}{D_{nt}} < \gamma_2 < \frac{D_{nt}}{D_{nn}} < \gamma_1$.

(1-c): when $D_{nt} = 0$, then $\gamma_2 < \frac{D_{nt}}{D_{nn}} < \gamma_1$.

If a matrix is positive definite, all principal submatrices are positive definite, and so all principal minors are positive. Since the matrix $G$ for each deformable body is RSPD (real, symmetric, and positive definite), $M$ is also RSPD, and so is $D_{II}$ for every $I$. 

\[\text{If a matrix is positive definite, all principal submatrices are positive definite, and so all principal minors are positive. Since the matrix G for each deformable body is RSPD (real, symmetric, and positive definite), M is also RSPD, and so is D_{II} for every I.}\]
Case (2): $D_{nn} < \mu |D_{nt}|$

(2-a): when $D_{nn} < \mu D_{nt}$ & $D_{nt} > 0$, then $\frac{D_{nt}}{D_{nn}} < \gamma_1 < \frac{D_{tt}}{D_{nt}} < \gamma_2$,

(2-b): when $D_{nn} < -\mu D_{nt}$ & $D_{nt} < 0$, then $\gamma_1 < \frac{D_{tt}}{D_{nt}} < \gamma_2 < \frac{D_{nt}}{D_{nn}}$.

Case (3): $D_{nn} = \mu |D_{nt}|$

(3-a): when $D_{nn} = \mu D_{nt}$ & $D_{nt} > 0$, then $\frac{D_{nt}}{D_{nn}} < \gamma_1 < \frac{D_{tt}}{D_{nt}}$,

(3-b): when $D_{nn} = -\mu D_{nt}$ & $D_{nt} < 0$, then $\gamma_1 < \frac{D_{tt}}{D_{nt}} < \gamma_2 < \frac{D_{nt}}{D_{nn}}$.

Here, the following relations are used to obtain the above comparisons between values.

\[
\gamma_1 - \gamma_2 = \frac{2\mu |D|}{(D_{nn} - \mu D_{nt})(D_{nn} + \mu D_{nt})},
\]
\[
\gamma_1 - \frac{D_{nt}}{D_{nn}} = \frac{\mu |D|}{D_{nn}(D_{nn} + \mu D_{nt})},
\]
\[
\frac{D_{nt}}{D_{nn}} - \gamma_2 = \frac{\mu |D|}{D_{nn}(D_{nn} - \mu D_{nt})},
\]
\[
\frac{D_{tt}}{D_{nt}} - \gamma_1 = \frac{|D|}{D_{nt}(D_{nn} + \mu D_{nt})},
\]
\[
\frac{D_{tt}}{D_{nt}} - \gamma_2 = \frac{|D|}{D_{nt}(D_{nn} - \mu D_{nt})},
\]
\[
\frac{D_{tt}}{D_{nt}} - \frac{D_{nt}}{D_{nn}} = \frac{|D|}{D_{nn} D_{nt}}, \text{ etc.}
\]

Now, the admissible domain of \( \bar{z} \) containing feasible solutions are examined with reference to the above results for each contact status.
Case (1-a): $D_{nn} \geq \mu |D_{nt}|$ and $D_{nt} > 0$:

(i) separation: $\overline{z}_n < 0$, $\forall \overline{x}_t$.

(ii) stick:

(4.1.8a) $\overline{z}_t < \frac{D_{tt}}{D_{nt}} \overline{z}_n$, and

(4.1.8b) either one of the following three cases:

- when $\overline{z}_t < \frac{D_{nt}}{D_{nn}} \overline{z}_n$, then $\overline{z}_t > \gamma_2 \overline{z}_n$,
- when $\overline{z}_t > \frac{D_{nt}}{D_{nn}} \overline{z}_n$, then $\overline{z}_t < \gamma_1 \overline{z}_n$, or
- when $\overline{z}_t = \frac{D_{nt}}{D_{nn}} \overline{z}_n$, then $\overline{z}_t < \frac{D_{tt}}{D_{nt}} \overline{z}_n$.

Let $\overline{z}_n > 0$, then $\gamma_2 \leq \frac{\overline{z}_t}{\overline{z}_n} \leq \gamma_1$ is admissible.

Let $\overline{z}_n \leq 0$, then no solution is admissible.

(iii) slip:

(4.1.9a) $\overline{z}_n > 0$, and (A.1.9b) $\overline{z}_n > 0$.

(iv) slip:

(4.1.10a) $\overline{z}_n > 0$, and (A.1.10b) $\overline{z}_n < \gamma_2$.

Thus, Case (1-a) always possesses one and only one solution for each $\overline{z} \in \mathbb{R}^2$ because each contact status occupies the respective distinctive region.

Cases (1-b) and (1-c) are different only in the stick status from Case (1-a).
Case (1-b): $D_{nn} > \mu \left| D_{nt} \right|$ and $D_{nt} < 0$:

(ii) stick:

(4.1.8a) $\bar{z}_t > \frac{D_{tt}}{D_{nt}} \bar{z}_n$, and

(4.1.8b) $\rightarrow$ either

when $\bar{z}_t < \frac{D_{nt}}{D_{nn}} \bar{z}_n$, then $\bar{z}_t > \gamma_2 \bar{z}_n$,

when $\bar{z}_t > \frac{D_{nt}}{D_{nn}} \bar{z}_n$, then $\bar{z}_t < \gamma_1 \bar{z}_n$, or

when $\bar{z}_t = \frac{D_{nt}}{D_{nn}} \bar{z}_n$, then $\bar{z}_t > \frac{D_{tt}}{D_{nt}} \bar{z}_n$.

Let $\bar{z}_n > 0$, then $\gamma_2 \leq \frac{\bar{z}_t}{\bar{z}_n} \leq \gamma_1$ is admissible.

Let $\bar{z}_n \leq 0$, then no solution is admissible.

Thus, Case (1-b) is the same as Case (1-a).

Case (1-c): $D_{nn} > \mu \left| D_{nt} \right|$ and $D_{nt} = 0$:

(ii) stick:

(4.1.8a) $\rightarrow \bar{z}_n > 0$, and

(4.1.8b) $\rightarrow$ either

when $\bar{z}_t > 0$, then $\frac{\bar{z}_t}{\bar{z}_n} < \mu \frac{D_{tt}}{D_{nn}} = \gamma_1$,

when $\bar{z}_t < 0$, then $\frac{\bar{z}_t}{\bar{z}_n} > \gamma_2$, or

when $\bar{z}_t = 0$, then $\forall \frac{\bar{z}_t}{\bar{z}_n}$.

Thus, $\bar{z}_n > 0$, and $\gamma_2 \leq \frac{\bar{z}_t}{\bar{z}_n} \leq \gamma_1$. 
This case is again the same as Case (1-a) or (1-b).

Therefore, Case (1) always possesses a unique solution for each \( \vec{z} \in \mathbb{R}^2 \) which is presented in (4.1.11-14). Cases (2-a and b) and (3-a and b) can be examined in the similar manner, and their results are summarized in Tables 2(a,b) and 3(a,b), respectively. The feasible solution regions for each contact status are illustrated in Fig's 11-13 for each case.

Summary of the solution(s) for each case

Case (1): \( D_{nn} > \mu |D_{nt}| \).

There exists one and only one solution for each data \( \vec{z}_I = (\vec{z}_n, \vec{z}_t) \) so that

(a) when \( \vec{z}_n \leq 0 \), the separation status is satisfied;

\[
\begin{bmatrix}
\bar{p}_n \\
\bar{p}_t
\end{bmatrix} =
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\vec{z}_n \\
\vec{z}_t
\end{bmatrix},
\]

(4.1.11a)

\[
\begin{bmatrix}
\vec{z}_n \\
\vec{z}_t
\end{bmatrix} =
\begin{bmatrix}
1 & 1 \\
1 & 1
\end{bmatrix}
\begin{bmatrix}
\vec{z}_n \\
\vec{z}_t
\end{bmatrix}.
\]

(4.1.11b)

(b) when \( \vec{z}_n > 0 \) and \( \gamma_2 \leq \frac{\vec{z}_t}{\vec{z}_n} \leq \gamma_1 \), the stick status is satisfied;

\[
\begin{bmatrix}
\bar{p}_n \\
\bar{p}_t
\end{bmatrix} = \frac{1}{|D|}
\begin{bmatrix}
-D_{tt} & D_{nt} \\
D_{nt} & -D_{nn}
\end{bmatrix}
\begin{bmatrix}
\vec{z}_n \\
\vec{z}_t
\end{bmatrix},
\]

(4.1.12a)

\[
\begin{bmatrix}
\vec{z}_n \\
\vec{z}_t
\end{bmatrix} =
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
\vec{z}_n \\
\vec{z}_t
\end{bmatrix}.
\]

(4.1.12b)
(c) when \( z_n > 0 \) and \( \frac{z_t}{z_n} > \gamma_1 \), the slip(c) status is satisfied:

\[
\begin{align*}
\begin{bmatrix}
\frac{P_n}{P_t}
\end{bmatrix} &= \frac{1}{D_{nn} + \mu D_{nt}} \begin{bmatrix}
-1 & 0 \\
-\mu & 0
\end{bmatrix} \begin{bmatrix}
\frac{z_n}{z_t}
\end{bmatrix}, \\
\begin{bmatrix}
z_n \\
z_t
\end{bmatrix} &= \begin{bmatrix}
0 & 0 \\
-\gamma_1 & 1
\end{bmatrix} \begin{bmatrix}
\frac{z_n}{z_t}
\end{bmatrix}.
\end{align*}
\]

(4.1.13a)

(4.1.13b)

(d) when \( z_n > 0 \) and \( \gamma_2 < \frac{z_t}{z_n} \), the slip(d) status is satisfied:

\[
\begin{align*}
\begin{bmatrix}
\frac{P_n}{P_t}
\end{bmatrix} &= \frac{1}{D_{nn} - \mu D_{nt}} \begin{bmatrix}
-1 & 0 \\
\mu & 0
\end{bmatrix} \begin{bmatrix}
\frac{z_n}{z_t}
\end{bmatrix}, \\
\begin{bmatrix}
z_n \\
z_t
\end{bmatrix} &= \begin{bmatrix}
0 & 0 \\
-\gamma_2 & 1
\end{bmatrix} \begin{bmatrix}
\frac{z_n}{z_t}
\end{bmatrix}.
\end{align*}
\]

(4.1.14a)

(4.1.14b)

Case (2): \( D_{nn} < \mu |D_{nt}| \).

There exists one or more but a finite number of solutions depending on the data \( z_I \) as shown in Table 2 (a) and (b) (Fig. 12) when \( D_{nt} > 0 \) and \( D_{nt} < 0 \), respectively.

Case (3): \( D_{nn} = \mu |D_{nt}| \).

There exists one or more solutions depending on the data \( z_I \). Moreover, the problem has infinitely many solutions for some \( z_I \) as revealed in Table 3 (a) and (b) (Fig. 13) for \( D_{nt} > 0 \) and \( D_{nt} < 0 \), respectively.

A hypothetical spring system in Fig. 14 for these three cases illustrates the above cases. However, most practical problems belong to Case (1) because \( D_{nn} \) is usually much
greater than $D_{nt}$ and $\mu$ is small (say, $\mu < 5$). The values of $D_{nn}$ and $D_{nt}$ for each pair of contact points represent changes of the gap vector in the normal and tangential directions, respectively, under the action of a unit normal contact force at the point. The reason why the uniqueness of solution depends on both $D_{nn}$ and $D_{nt}$, but not $D_{tt}$, can be explained from the frictional contact conditions in which $P_n$ involves the constraints in both directions, while $P_t$ only depends on the constraints in the tangential direction. As expected, the frictionless problems ($\mu = 0$) always belong to Case (1) regardless of the entries of matrix $D_{II}$ and the data $z_I^{-1}$. A sufficient condition that (4.1.4) has at least one solution for each data $z_I^{-1}$ is that $D_{II}$ is RSPD. Thus, the existence of solution is guaranteed for most practical problems.

For the iterative solution scheme, in conjunction with the embedding strategy, the contact force for each pair can be updated immediately after evaluating $z_I^{-1}$ according to (4.1.4). Even for Cases 2 and 3, which are scarcely encountered, the contact force can be determined uniquely for some $z_I^{-1}$ as listed in Tables 2 and 3, and it may be chosen arbitrarily for all feasible solutions with other values. It is noted that the separation status is commonly satisfied for all the non-unique solution regions. In order to remove the arbitrariness of the indicated selection, in the present work, it is stipulated that the contact force
satisfying the separation condition is always chosen among other feasible solutions. Finally, the embedding method explained in the present section is summarized by the following two definitions.

Definition 1. Method of Choice:
The separation condition is always chosen in preference to other feasible solutions if any.

Definition 2. Embedding Map:
A nonlinear surjective mapping \( T: \mathbb{R}^2 \to \mathbb{R}^2 \) defined by, for each \( 1 \leq i \leq N (=\text{NCA}) \),
\[
P_i = T_{II} \bar{z}_I , \quad \text{(no summation)} \quad (4.1.15)
\]
where \( \bar{z}_I \in \mathbb{R}^2 \) and \( P_i \in \mathbb{R} \), is called an embedding map. Here, the 2x2 block matrix \( T_{II} \) represents the matrix presented in (4.1.11a-14a) corresponding to \( z_I \) and the method of choice. Also, the associated mapping \( S: \mathbb{R}^2 \to \mathbb{R} \subset \mathbb{R}^2 \) can be defined by
\[
z_I = S_{II} \bar{z}_I , \quad \text{(no summation)} \quad (4.1.16)
\]
where \( S_{II} = D_{II} T_{II} + I \) and \( I \) is the 2x2 identity matrix. The matrix \( S_{II} \) can be obtained by substituting (4.1.15) into (4.1.3) as presented in (4.1.11b-14b).

4.2. Existence and Uniqueness of Solution
Sufficient conditions for the existence and uniqueness of solution for PROBLEM III are now investigated from the embedding map defined previously and the following theorem known as the contraction mapping theorem or the
Theorem 2. (Contraction Mapping Theorem):

Let \((X, \|\cdot\|)\) be a Banach space and \(\phi: X \rightarrow X\) be a contraction mapping; that is, there exists a constant \(K, 0 \leq K \leq 1\), such that for every \(x, y \in X\),
\[
\|\phi(x) - \phi(y)\| \leq K \|x - y\|.
\]
Then, \(\phi\) has a unique fixed point \(x\) in \(X\) such that
\[
x = \phi(x).
\]
Moreover, a sequence \(\{x_i\}\) converges to the point \(x\) in the limit for any starting point \(x_0 \in X\); that is,
\[
x = \lim_{i \rightarrow \infty} x_i,
\]
where \(x_i = \phi(x_{i-1})\) and \(i = 1, 2, \ldots \ldots \).

(See references [45, 55, etc] for proof.)

PROBLEM III is now represented in the form of a fixed point problem as follows. Recalling (4.2.1) and the embedding map in (4.1.16),
\[
z_I = S_{II} \overline{z}_I, \quad \text{(no summation)} \quad (4.1.16)
\]
where \(S_I = D_{II} T_{II} + I\) and \(I\) is the 2x2 identity matrix.

The results obtained in the previous section are summarized in the following two lemmas.

Lemma 1.

The matrix \(S_{II}\) always exists for each \(I\) under the conditions that \(D_{II}\) is RSDP and the method of choice is applied to Cases (2) and (3). In other word, there exists at least one solution for every \(\overline{z}_I \in \mathbb{R}^2\) satisfying both the nodal contact equation (4.1.3) and the corresponding contact condition.
Lemma 2.

The matrix $S_{II}$ is uniquely determined for each $I$ without use of the method of choice under the conditions that $D_{II}$ is RSPD and $D_{nn} > \mu |D_{nt}|$. That is, there exists one and only one solution for each $z_I$.

Eq (4.1.16) is now expressed collectively by

$$z = S \bar{z}, \quad (4.2.1)$$

where

$$[S] = \begin{bmatrix} [S]_{11} & 0 \\ [S]_{22} & \ddots \\ 0 & \ddots & [S]_{NN} \end{bmatrix},$$

and $N$ is NCA. By substituting (4.2.1) into (4.1.2), (4.1.2) can be expressed in terms of $P$ only such that

$$D P = (I - S)(L P + U P - W), \quad (4.2.2)$$

where $I$ denotes the $N \times N$ identity matrix.

Eq (4.2.2) is then represented in the conventional form of fixed point problems by introducing an appropriate iteration function. For instance, (4.2.2) can be written as

$$P = \phi_J(P), \quad (4.2.3)$$

or

$$P = \phi_G(P), \quad (4.2.4)$$

where

$$\phi_J(P) = D^{-1}(I - S)(L P + U P - W),$$

and

$$\phi_G(P) = (I - D^{-1}(I - S)L)^{-1}D^{-1}(I - S)(U P - W).$$
It is noted that \((I - D^{-1}(I - S)L)^{-1}\) always exists since \(D^{-1}(I - S)L\) is a lower triangular matrix. Here, \(\phi_J\) and \(\phi_G\) represent the 'block' Jacobi and the 'block' Gauss-Seidel iteration functions, respectively, and they are, in the sequel, denoted by \(\phi\) collectively.

Eqs (4.2.3-4) generate a sequence in a natural way such that

\[
P^i = \phi(P^{i-1}) \quad (4.2.5)
\]

where \(i = 1, 2, \ldots\), denotes an iteration counter.

**Definition 3. Contraction:**

The iteration function \(\phi\) is called a contraction in a normed space \((\mathcal{F}^N, \|\cdot\|)\) if there is a constant \(K, 0 < K < 1\), such that for every integer \(m, n, m > n\), and \(P^m, P^n \in \mathcal{F}^N\),

\[
\|\phi(P^m) - \phi(P^n)\| \leq K \|P^m - P^n\| \quad (4.2.6)
\]

where \(\|\cdot\|\) represents any norm. By assuming the existence of a fixed point \(P\), (4.2.6) can also be written as

\[
\|P^i - P\| = \|\phi(P^{i-1}) - \phi(P)\| \leq K \|P^{i-1} - P\| \quad (4.2.7)
\]

for every \(i, i = 1, 2, \ldots\).

Some properties of the iteration function \(\phi_G\) are discussed in Section 4.3.2. It is noted that if the sequence \(\phi\) in (4.2.5) converges, then \(\phi\) is a contraction, and the converse is also true because the space \((\mathcal{F}^N, \|\cdot\|)\) is a complete space.

Finally, we have the following theorems on the existence and uniqueness of solution for PROBLEM III.
Theorem 3. (Existence):

PROBLEM III has at least a solution \((z, P) \in Z \times \mathbb{R}^N\) for each data set \(\mathbb{W}\) if the matrix \(M\) is RSPD and the associated iteration function \(\phi\) is a contraction.

Proof: From Lemma 1, it is shown that an iterative function \(\phi\) always exists if \(M\) is RSPD. Since \(\phi\) is a contraction and the space \((\mathbb{R}^N, \| \cdot \|)\) is a Banach space, the conditions in Theorem 2 are all satisfied. Hence, \(\phi\) has a unique fixed point \(P \in \mathbb{R}^N\) which is a solution of PROBLEM III with the corresponding \(z \in Z\), that is, \(z = S z\) always belongs to \(Z\). However, the problem may have other solutions if it belongs to Case 2 or 3, because different methods of choice produce different fixed points. □

Theorem 4. (Uniqueness):

If PROBLEM III has a solution, it is unique if \(D_{nn} > \mu |D_{nt}|\) for every \(D_{II}\) in addition to the conditions in Theorem 2.

Proof: It is sufficient for illustrating the proof that the above condition satisfies Lemma 2. □

Corollary. (Frictionless problems):

For frictionless problems, that is, \(\mu = 0\), a solution is unique if it exists.

Proof: If \(\mu = 0\), \(D_{nn} > \mu |D_{nt}|\) always, since \(D_{nn} > 0\). □

4.3. Iterative Solution Method

The embedding scheme explained previously can be incorporated with various iteration schemes for linear or non-linear problems. A particular preference in the selection of an underlying iterative scheme may depend on the convergence of incorporated iterative schemes, ease of computer implementation, cost-effectiveness, and so on.
Another important fact in the considerations may be related to a judicious choice concerning the evaluation of full entries of the matrix $\mathbf{M}$ because the embedding scheme requires its diagonal parts only in addition to $\mathbf{z}$. In fact, $\mathbf{z}$ can be evaluated directly from the displacements of each body for updated contact forces which may be obtained separately with use of the existing solution routine for stiffness problems.

For instance, the Jacobi method requires the updated values of $\mathbf{z}$ once at the beginning of each iteration step. On the other hand, in the Gauss-Seidel method, $\mathbf{z}$ needs to be updated continuously, yet partially, during an iteration step. Thus, the former technique may be a more effective method than the latter method when the diagonal part of the matrix $\mathbf{M}$ is to be evaluated.

In this respect, an attempt has been made in developing a computational scheme for the effective evaluation of the matrix $\mathbf{M}$. Of course, most of computing time is expended for the evaluation of the compliance matrix $\mathbf{G}$ for each body. According to the present method, explained in Section 4.4.2., the procedure for evaluating the full-entries of a symmetric matrix $\mathbf{G}$ consumes the same order of computing time as that for obtaining the displacement solution. This additional computational time for evaluating the full-entries of matrix $\mathbf{M}$ may be acceptable for the desired solution accuracy. Thus, the present work is focused on the
Gauss-Seidel method, but the Jacobi or some other methods have similar characteristics.

The present iterative scheme, based on the block Gauss-Seidel method, is first explained in Section 4.3.1., and the associated convergence criteria are then studied in Section 4.3.2.

4.3.1. Gauss-Seidel iteration method

In order to explain the block Gauss-Seidel iteration method, a system of linear equations is written as

\[ \mathbf{M} \mathbf{P} + \mathbf{W} = \mathbf{0} , \]  

(4.3.1)

where \( \mathbf{P} \) is considered as an unknown \( \mathbf{N} \times 1 \) vector, \( \mathbf{M} \) is a \( \mathbf{N} \times \mathbf{N} \) matrix, and \( \mathbf{W} \) is a given \( \mathbf{N} \times 1 \) vector. In fact, (4.3.1) is the same as (4.1.2) except that (4.3.1) does not have the additional unknown \( z \). Following Section 4.1., (4.3.1) can also be written as

\[ \overline{\mathbf{D}} \mathbf{P} + \overline{\mathbf{Z}} = \mathbf{0} , \]  

(4.3.2)

where \( \overline{\mathbf{Z}} = \mathbf{W} - \mathbf{L} \mathbf{P} - \mathbf{U} \mathbf{P} \). Applying the block Gauss-Seidel iteration scheme to (4.3.2), we get

\[ \mathbf{D} \mathbf{P}^i + \overline{\mathbf{Z}}^i = \mathbf{0} , \]  

(4.3.3)

where \( \overline{\mathbf{Z}}^i = \mathbf{W} - \mathbf{L} \mathbf{P}^i - \mathbf{U} \mathbf{P}^{i-1} \), and \( i \) denotes an iteration counter. In the block component form, (no summation for \( I \))

\[ \mathbf{D}_{II} \mathbf{P}^i + \overline{\mathbf{Z}}^i_I = \mathbf{0} , \]  

(4.3.4)
where \( \bar{z}_I^i = W - L_{IJ}P_J^i - U_{IJ}P_J^{i-1} \). In view of the sequence of iteration for a fixed \( i \), \( \bar{z}_I^i \) is always known at the instant of evaluating \( P_I^i \) because \( L \) and \( U \) are triangular matrices. The Jacobi method is different from the Gauss-Seidel method only in the term \( \bar{z}_I^i \), that is, \( \bar{z}_I^i = W - L_{IJ}P_J^{i-1} - U_{IJ}P_J^{i-1} \).

We introduce the additional unknown vector \( z \) in (4.3.2), and apply the iteration scheme for both \( P \) and \( z \) such that (no summation)

\[
Z_I = D_{II}P_I^i + Z_I^i \quad (4.3.5)
\]

where \( Z_I^i \) is the same as that in (4.3.4). With the help of the embedding map, both \( Z_I^i \) and \( P_I^i \) can be obtained simultaneously for a known \( \bar{z}_I^i \) such that (no summation)

\[
z_I^i = S_{II}\bar{z}_I^i \quad (4.3.6)
\]

\[
P_I^i = T_{II}\bar{z}_I^i
\]

where \( S_{II} \) and \( T_{II} \) are the same as in (4.1.16) and (4.1.15), respectively. It is noted that \( z_I^i \) need not be computed except for computing the error because it is not used for evaluating \( z \).

Once \( \bar{z}_I^i \) is obtained, \( P_I^i \) can be evaluated with a few multiplication operations because the criteria for the uniqueness of solution are already examined when matrix \( M \) is evaluated. Thus, no additional computation is virtually required compared to the corresponding linear problem as in
(4.3.4). The above embedding process is then repeated for every pair at each iteration, and the iteration is continued until an appropriate error norm is obtained within an allowable limit. It is worthy of note that the sequence of iteration, i.e., \( I = 1, 2, \cdots, \text{NCA}, \) or \( I = \text{NCA}, \cdots, 1, \) etc., has a minor influence on the rate of convergence for most problems.

There may be two ways to define the associated error norm,

\[
e^i_P = \frac{\|P^i - P^{i-1}\|_2}{\|P^{i-1}\|_2} ,
\]

or

\[
e^i_Z = \|z^e\|_2 ,
\]

where

\[
z^e = \{z^e_1, z^e_2, \cdots, z^e_i, \cdots, z^e_N\} ,
\]

\[
z^e_i = \begin{cases} 0 & \text{if } \hat{z}^i_s \text{ satisfies the contact condition corresponding to } P^i_s , \\ z^e_i & \text{otherwise}, \end{cases}
\]

\[
\hat{z}^i_s = M_{ij} P^i_j + W_i ,
\]

and \( \| \cdot \|_2 \) denotes the Euclidean norm. In the present work, the error norm \( e^z \) is employed. However, both norms yield almost the same rate of convergence, as shown in Fig. 15. The initial value of the contact force \( t + \Delta t P^0 \) at the beginning of each incremental stage is assigned as the same value of \( tP^m \) obtained at the end of the previous stage, and
Finally, it is noted that the contact force $P$ in the present iterative scheme always satisfies the condition $P \in \mathcal{N}$.

### 4.3.2. Convergence criteria

We recall the governing contact equation (4.1.2) and the associated iterative equation (4.3.5) based on the block Gauss-Seidel method such that

$$z = D P + \overline{z},$$  \hspace{1cm} (4.1.2)

and

$$z^i = D P^i + \overline{z}^i,$$  \hspace{1cm} (4.3.9)

where $\overline{z} = \overline{W} - L P - U P$, and $\overline{z}^i = \overline{W} - L P^i - U P^{i-1}$.

Here, (4.3.5) is expressed in the form of a matrix equation like (4.3.9). By subtracting (4.1.2) from (4.3.9), we obtain

$$z^i - z = D (P^i - P) + (\overline{z}^i - \overline{z}),$$  \hspace{1cm} (4.3.10)

where

$$\overline{z}^i - \overline{z} = - L (P^i - P) - U (P^i - P).$$

Also, recall (4.2.1) and (4.3.6) such that

$$z = S \overline{z} ,$$  \hspace{1cm} (4.2.1)

and

$$z^i = S^i \overline{z}^i ,$$  \hspace{1cm} (4.3.11)

where $S$ denotes the embedding matrix as explained before, and the iteration counter $i$ on $S^i$ is used to denote the
value corresponding to \( \bar{z}^i \) since \( S \) depends on \( \bar{z} \). The relation between \( (z^i - z) \) and \( (\bar{z}^i - \bar{z}) \) is now represented by

\[
(z^i - z) = Q^i(\bar{z}^i - \bar{z}) \quad (4.3.12)
\]

where

\[
[Q] = \begin{bmatrix}
[Q]_{11} & 0 \\
[Q]_{22} & \ddots \\
0 & \ddots & [Q]_{NN}
\end{bmatrix},
\]

and each block matrix \( Q_{II} \) depends on both \( \bar{z}_I \) and \( \bar{z}_I \) as illustrated in Fig. 16.

The matrix \( Q^i \) in (4.3.12) represents a nonlinear and surjective mapping. It is now shown that the matrix \( Q^i \) is bounded such that for every \( i \) and \( I \),

\[
\|Q^i_{II}\|_2 \leq \beta_I \quad (4.3.13)
\]

and so,

\[
\|Q^i\|_2 \leq \beta \quad (4.3.14)
\]

where \( \|\cdot\|_2 \) denotes the subordinate matrix norm associated with the Euclidean norm, and \( \beta_I \) and \( \beta \) are positive constants defined as below. Although it is not easy to express \( Q^i_{II} \) explicitly for all cases, the constant \( \beta_I \) can be easily found by examining all possible contact combinations for any \( \bar{z}^i_I \in \mathbb{R}^2 \) and \( \bar{z}_I \in \mathbb{R}^2 \) as summarized in Table 4. Thus,

\[
\beta_I = \max \{ \sqrt{1+\gamma_1^2}, \sqrt{1+\gamma_2^2} \} \quad (4.3.15)
\]

and

\[
\beta = \max_{1 \leq I \leq N} \{ \beta_I \} \quad (4.3.16)
\]

where \( \gamma_1 \) and \( \gamma_2 \) are defined in Section 4.1., that is,
\[ \gamma_1 = \frac{D_{nt} + \mu D_{tt}}{D_{nn} + \mu D_{nt}}, \quad \text{and} \quad \gamma_2 = \frac{D_{nt} - \mu D_{tt}}{D_{nn} - \mu D_{nt}}. \]

Substituting (4.3.12) into (4.3.10), we get

\[ p^i - p = \psi^i(p^{i-1} - p) , \quad (4.3.17) \]

where

\[ \psi^i = (I - D^{-1}\tilde{L}^i)^{-1} D^{-1} \tilde{U}^i , \quad (4.3.18) \]

\[ \tilde{L}^i = (I - Q^i) L , \]

and

\[ \tilde{U}^i = (I - Q^i) U . \]

**Lemma 3. (Convergence I):**

A sequence \{\( P^i \)\}, where \( P^i \in \mathcal{F}^N \subset (\mathbb{R}^2)^N \) and \( i = 1, 2, \ldots \), generated by the present iterative method in (4.3.17) converges to \( P \in \mathcal{F}^N \) for all initial vector \( P^0 \in \mathcal{F}^N \) if there exists a constant \( K \) such that \( \|\psi^i\|_2 \leq K \) and \( 0 < K < 1 \) for every \( i \).

**Proof:** From the recurrence formula in (4.3.17),

\[ p^i - p = \psi^1\psi^{i-1} \cdots \psi^1(p^0 - p) , \quad (4.3.19) \]

and

\[ \|p^i - p\|_2 \leq \|\psi^i\|_2\|\psi^{i-1}\|_2 \cdots \|\psi^1\|_2\|P^0 - p\|_2 \leq (K)^i\|P^0 - p\|_2 . \quad (4.3.20) \]

Thus,

\[ \lim_{i \to \infty} \|p^i - p\| = 0 , \quad \text{for any } P^0 \quad (4.3.21) \]

since \( \lim_{i \to \infty} (K)^i = 0 \). Therefore, the sequence \{\( P^i \)\} converges to \( P \in \mathcal{F}^N \) with respect to the Euclidean norm for all \( P^0 \in \mathcal{F}^N \) because \( \mathcal{F}^N \) is a closed subset of \((\mathbb{R}^2)^N\), and so \((\mathcal{F}^N, \|\cdot\|_2)\) is a Banach space. □
It is noted that the above lemma can be stated employing other equivalent norms in the same manner.

From (4.3.18), we get

\[ \| \Psi^i \|_2 \leq \| (I - D^{-1} \tilde{L}^i)^{-1} \|_2 \| D^{-1} \tilde{U}^i \|_2 , \]  
(4.3.22)

\[ \leq \frac{\| D^{-1} \tilde{U}^i \|_2}{1 - \| D^{-1} \tilde{L}^i \|_2} , \]

\[ \leq \frac{\| \omega^i \|_2 \| D^{-1} \tilde{U}^i \|_2 \| U \|_2}{1 - \| \omega^i \|_2 \| D^{-1} \tilde{U}^i \|_2 \| U \|_2} , \]

where

\[ \omega^i = I - Q^i . \]

Here, it is assumed that

\[ \| D^{-1} \tilde{L}^i \|_2 < 1 , \]

noting that

\[ \| (I - A)^{-1} \|_2 \leq \frac{1}{1 - \| A \|_2} , \quad \text{for any } \| A \|_2 < 1 . \]

By examining the relation

\[ (\tilde{z}^i_I - \tilde{z}^i_I) - (z^i_I - z_I) = \omega^i (\tilde{z}^i_I - \tilde{z}^i_I) , \]  
(4.3.23)

for all cases with use of Fig. 16 and Table 4, in the same manner as before, it can be shown that for every i,

\[ \| \omega^i \|_2 \leq \beta^i , \]  
(4.3.24)

and

\[ \| \omega^i \|_2 \leq \beta , \]  
(4.3.24)

where \( \beta^i \) and \( \beta \) are the same as those defined in (4.3.15) and (4.3.16), respectively. Therefore, for every i,

\[ \| \Psi^i \|_2 \leq \frac{\beta \| D^{-1} \tilde{U}^i \|_2 \| U \|_2}{1 - \beta \| D^{-1} \tilde{U}^i \|_2 \| U \|_2} . \]  
(4.3.25)
Theorem 4. (Convergence II):

For a symmetric matrix $M$, i.e., $\|U\|_2 = \|L\|_2$, the sequence $\{P^i\}$ generated by the present iterative method in (4.3.17) converges to $P ∈ ℜ^N$ for any data $W ∈ (ℜ^2)^N$ and any initial vector $P^0 ∈ ℜ^N$ if

$$\beta \|D^{-1}U\|_2 ≤ \beta \|D^{-1}\|_2 \|U\|_2 < \frac{1}{2} .$$

(4.3.26)

Proof: From (4.3.25), it is obvious that the condition in (4.3.26) satisfies the convergence criterion in Lemma 3, that is, $\|Ψ^i\|_2 < 1$ for every $i$. Since $β$ is independent of $W$ as shown in (4.3.15-16), the condition is also independent of $W$. □

According to (4.3.22), the rate of convergence for the present iterative scheme depends on the norm $\|ω\|_2$ which corresponds to the overall contact status. For instance, the rate of convergence becomes the same as that for the respective linear problem in (4.3.1); i.e., $M P + W = 0$, because $ω = I$ or $Q = 0$ when all pairs of contact points satisfy the stick condition. Also, as a trivial case, the sequence converges at once because $ω = 0$ or $Q = I$ when all the pairs satisfy the separation condition. Fig. 15 illustrates the degree of dependence of the rate of convergence on the coefficient of friction for a simple problem.

4.4. Computational Considerations

4.4.1. Solution algorithm
The overall solution algorithm for the present iterative solution method can be divided largely into four different phases; viz., preliminary, incremental loading, iteration, and final phases. The associated flow chart for linear elastic problems is presented in Fig. 17. Also, it is noted that an arbitrary rigid body motion for the cases \( \Gamma_u = \Phi \) has been removed for each deformable body, if necessary, by means of artificially designed pre-strained spring elements explained in Section 6.2. Thus, the global stiffness matrices \( K^a \) and \( K^b \) in (3.2.3-4) have a sufficient rank compared to the number of degrees of freedom for the respective body, and so \( G^a \) and \( G^b \) can be evaluated from the respective stiffness matrix in the manner explained in the subsequent section.

Some details of the algorithm are now described step-by-step as follows.

**Phase 1. (Preliminary stage):**

Step 1. Evaluate the stiffness matrices for each body and decompose them into triangular factors in accordance with the strategy of the active column solver (refer to [3] for details). This step is one of the common procedures for finite element analyses.

Step 2. (1) Evaluate \( W^a \) and \( W^b \) from the displacement solutions corresponding to only the external loads \( F^a \) and \( F^b \), respectively, and store their fractions \( \Delta W^a \) and \( \Delta W^b \) corresponding to each increment of external loads. Usually, \( \Delta W = W/M \) where \( M \) denotes the number of increments. For the case when body 'b' is rigid, \( W^b \) has the same assigned value as the rigid body displacement.

(2) Evaluate the matrices \( G^a \) and \( G^b \) in (3.2.5-6) in
Phase 2. (Incremental stage):

Step 1. Update the pairing map and the normal direction (Refer to Section 3.3.)
(1) Update the pairs of contact points by finding the associated pairing point on body 'b' in contact with a possible contact node on body 'a' with reference to the deformed configuration obtained at the end of the previous incremental stage. For the case when a possible contact node does not come into contact at the end of the previous stage, the associated pairing point is predicted as a point satisfying the relation in (3.3.3). Then, compute the ratio of segment $k$ between adjacent nodes for each pairing point.
(2) Update the normal direction for each pair with reference to the assumed deformed configuration for the current incremental stage obtained in the manner explained in Section 3.3.2.

Step 2. Update $\mathbf{M}$ and $\mathbf{W}$ in the governing contact equation (3.4.4) for the current incremental stage.
(1) The global components of $\mathbf{M}_{ij}$ are first updated according to (3.3.8) and (3.4.3), and they are then transformed into the local components, as explained in Section 3.3.2. The above procedure is repeated for every $I$ and $J \geq I$ since $\mathbf{M}$ is symmetric. At this time, we may examine whether the matrix $D_{II}$ for each pair satisfies the condition $D_{nn} > \mu |D_{nt}|$.
(2) Similarly, the global components of $\mathbf{W}$ are first updated according to (3.4.3), and then they are transformed into the local components.

Step 3. Go to Phase 3 in order to solve iteratively the governing contact equation for the contact force $P$ at the current stage.

Step 4. Repeat the above steps if another load increment is needed. Otherwise, go to Phase 4.

Phase 3. (Iteration stage):

Step 1. For every pair of contact points, solve (4.1.3) by using the embedding method. (Refer Section 4.1)
(1) Compute $Z^I_i$ according to (4.1.4).
(2) Determine the contact status from the computed value $Z^I_i$ in accordance with three cases explained in Section 4.1., and obtain the contact force from the
corresponding embedding map presented in (4.1.11a)–(4.1.14a).
(3) Repeat (1) and (2) for every \( I \), \( 1 \leq I \leq NCA \), then go to Step 2.

Step 2. (1) Compute the error norm defined as either (4.3.7) or (4.3.8).
(2) Repeat Step 1 until the computed error norm reaches within a prescribed limit, or the iteration counter runs up to a maximum allowable number, then go to Step 3. Stop if the iteration diverges.

Step 3. (1) The local components of contact force obtained in Step 1 (2) are transformed into the corresponding global components.
(2) The nodal contact force for body 'b' is then evaluated according to (3.4.2) and (3.3.9).
(3) Return to Step 3 in Phase 2.

Phase 4. (Final stage):

Compute the displacement owing to both the external force \( F \) and the obtained contact force \( P \) for each body according to Eqn (3.2.3 or 4), and compute stresses, strains, etc.

4.4.2. Evaluation of compliance matrices

In order to evaluate the matrix of Green's function \( G \) for each body associated with the nodal contact forces effectively, it is recommended that a global node numbering for the possible contact nodes for each body follows that for non-contact nodes as long as the band-width of the respective stiffness matrix \( K \) does not change appreciably. It is now assumed that each stiffness matrix has already been decomposed into the triangular factors; i.e., \( K = LD L^T \) for each body, in accordance with the strategy of the active column solver, using the same notation as in Reference [3].
The displacement \( u \) corresponding to a solution of the set of active equations \( K u = R \) can then be obtained by executing the following two steps; that is,

- the forward reduction: \( L V = R \), and
- the backward substitution: \( L^T u = D^{-1} V \),

where \( V \) is a dummy vector for computation. In accordance with the nodal numbering scheme as mentioned above, the vectors \( u, V, \) and \( R \) can be represented in the form

\[
\begin{align*}
  u &= \begin{bmatrix} u_{nc} \\
                     \vdots \\
                     u_c \end{bmatrix}, \\
  V &= \begin{bmatrix} V_{nc} \\
                     \vdots \\
                     V_c \end{bmatrix}, \text{ and } \\
  R &= \begin{bmatrix} R_{nc} \\
                     \vdots \\
                     R_c \end{bmatrix},
\end{align*}
\]

where subscripts \( c \) and \( nc \) denote contact and non-contact nodes, respectively. Here, the number of active degrees of freedom for a body is denoted by \( N \), and that involving the possible contact nodes is denoted by \( NC \) (note \( NC \) is not always twice the value of \( NCA \) or \( NCB \) for two dimensional problems).

The components \( G_{ij} \), \( i,j = 1, \cdots, NC \), of a symmetric matrix \( G \) can then be evaluated as follows. For the \( i \)-th row, for instance, the vector \( V_c \) is first computed by executing the forward reduction from the \( i \)-th component to the last one with the vector \( R \) defined by \( R_{nc} = 0 \) and \( (R_c)_j = \delta_{ij} \) such that

\[\delta_{ij}\]

\( \dagger \) The components of vector \( R_c \) prior to the \( i \)-th one are all zeros.
\[ \{ R \} = \{ 0 : 0 \cdots 0 \ 1 \ 0 \cdots 0 \}^T . \]

\[ 1 \cdots i \cdots NC \]

\[ \{ R_{nc} : \quad R_c \}^T \]

The vector \( u_c \) is then obtained by executing the backward substitution from the last component to the \( i \)-th one for the vector \( V \) because components of the vector \( u_c \) prior to the \( i \)-th one are unnecessary for evaluating \( G_{ij} = (u_c)_j \), where \( j = i, \cdots, NC \). The above procedure is repeated for every \( i \), \( i = 1, 2, \cdots, NC \).

The number of operations for the evaluation of \( G \) for a body can be estimated as follows.

\[
2(\text{NC})(\text{HB})((\text{NC})-(\text{HB})) + \frac{1}{2}(\text{NC})((\text{NC})+2(\text{HB})+1)
+ \frac{1}{3}(\text{HB})((\text{HB})^2-3(\text{HB})-1) , \quad \text{when } (\text{NC}) \geq (\text{HB})
\]

and

\[
\frac{1}{6}(2(\text{NC})^3+3(\text{NC})^2+(\text{NC})) , \quad \text{when } (\text{NC}) \leq (\text{HB})
\]

where \( (\text{HB}) \) is the half-bandwidth of the respective stiffness matrix \( K \), and multiplication or division is counted as one operation which is almost always followed by addition or subtraction. It is noted that the actual number of operations may be less because the above estimation is based on the assumption of a constant column height regardless of the profile of the 'skyline' [3].

For comparison purposes, the number of operations required for obtaining the solution \( u \) of the corresponding linear problem; that is \( K u = R \), is estimated to be
\[2(N)(HB)+(N)-(HB)^2-(HB),\]

for both the forward reduction and backward substitution. For instance, we consider the ring-compression problem which is described in Section 6.1. with \(N=121\), \(HB=25\), and \(NC=22\) as depicted in Fig. 40. Then, the number of operations required for the evaluation of \(G\) is 3795, and the number required for obtaining a displacement solution \(u\) is 5521. With the numbering scheme for achieving the minimum half-bandwidth; that is, \(HB=15\), the latter number becomes 3511. Roughly speaking, the procedure for evaluating the matrix \(G\) consumes computing time almost as much as that for obtaining a displacement solution for the respective linear problem in this specific case.
CHAPTER V
NUMERICAL EXAMPLES

The present solution method is implemented into the FAST code (Finite element Analysis of Stress, moisture, and Temperature with phase-change effects) developed in the Department of Engineering Mechanics, the Ohio State University. The 'Contact' portion, which consists of 23 subroutines including the central subroutine CONACT, is connected with the underlying finite element code FAST through the argument list of subroutine CONACT associated with the stiffness matrix for each body, mesh and boundary data, and so on. Several numerical examples are then analyzed with the developed Fortran program on IBM 3081-D computer system.

Two representative contact problems are first studied for calibration using analytic and finite element solutions in Sections 5.1. and 5.2. Several factors which may have a significant influence on the accuracy of the finite element solution for friction problems are investigated, and the effectiveness of the present solution method is also discussed. Three field problems are also analyzed for practical applications in Sections 5.3.-5.5.
In the present work, the nodal contact traction is estimated from the obtained nodal contact force by dividing it by the respective mean area allocated according to the mesh configuration. As discussed in Section 3.2., the contact traction distribution along the contact boundary is, in the present mathematical model, assumed as a singular distribution in order to handle the non-conformity along the contact boundary between two bodies. Thus, the relation between the nodal contact force and traction does not follow the interpolation rule associated with the conventional formulation of a consistent force vector, that is, for the consistent force vector,

\[ P = [N^TN] \ p \]

and

\[ P = [N^TL] \ p = I \ p \]

for the contact force vector in the present work, where \( P \) and \( p \) are the nodal contact force and traction vectors, respectively, \( N \) and \( L \), defined in Section 3.2., are the interpolation functions for the displacement and the contact force vectors, respectively, \( I \) is the identity matrix, and

\[ \sum_{e=1}^{\text{NUMEL}} \int_{T^e_0} \cdot dS^e. \]

In the sequel, the following notation is used for compactness:
\( P_n \) is the normal component of contact traction (normal contact traction),
\( P_t \) is the tangential component of contact traction (tangential contact traction or friction traction),

\[
r = \frac{P_t}{\mu |P_n|}
\]

is the normalized traction ratio (ratio),

\( \text{NINC} \) is the number of load increments divided equally (e.g., NINC=1, 10, etc.), or
Variable-increment designates the incremental scheme adjusting the size of increment at each stage so that no more than one node can be changed its contact status at each stage, and

\( \text{ERON} \) is a prescribed limit of allowable error norm (e.g., ERON=10^{-7}, 10^{-10}, etc).

where \( \mu \) is the coefficient of friction, and \( r=\pm 1 \) and \( -1<r<1 \) denote the slip and stick status, respectively.

5.1. Compression of an Elastic Cylinder between Rigid and Flat Dies

Problem description:

Fig. 18 depicts an infinitely long elastic cylinder compressed between rigid and flat dies. The problem is characterized by the following dimensionless quantities:

radius: \( R = 8. \),
Young's modulus: \( E = 1000. \),
Poisson's ratio: \( \nu = 0.3 \),
depth of compression: \( 2 \delta \),
applied force: \( F \),
contact-width: \( 2b \), and
maximum normal contact traction: $p_{\text{max}}$.

Both the plane-strain condition and the infinitesimal deformation theory are underlined.

**Analytic solution (frictionless cases):**

The closed form solution is summarized for frictionless cases as follows [17]:

$$p_n(\theta) = \frac{(\cos^2 \theta - \cos^2 \theta_0)^{1/2}}{\eta(1+\cos^2 \theta_0)}, \quad 0 \leq \theta \leq \theta_0 \quad (5.1.1)$$

$$p_{\text{max}} = p_n(0) = \frac{\sin \theta_0}{\eta(1+\cos^2 \theta_0)} \quad (5.1.2)$$

$$F = \int_0^{\theta_0} p_n(\theta) \ R \ d\theta$$

$$= \frac{2R}{\eta(1+\cos^2 \theta_0)} \ [E(\sin \theta_0) - (\cos^2 \theta_0)K(\sin \theta_0)]. \quad (5.1.3)$$

$$\delta = R \left\{1 - 2\frac{E(\cos \theta_0) - (\sin^2 \theta_0)K(\cos \theta_0)}{1+\cos^2 \theta_0}\right\}, \quad (5.1.4)$$

and

$$b = \frac{2R \sin \theta_0 - R \varphi[E(\sin \theta_0) - (\cos^2 \theta_0)K(\sin \theta_0)]}{1+\cos^2 \theta_0}, \quad (5.1.5)$$

where

$$\eta = (1-v^2)/E,$$

$$\varphi = (1-2v)/[2(1-v)],$$

$K$ and $E$ are complete elliptic integrals of the first and second kinds, respectively, and

$\theta_0$ is the angle of contact measured in the undeformed
configuration, as depicted in Fig. 17. It is noted that \( b \) in (5.1.5) represents the half contact-width in the deformed configuration, i.e.,

\[
b = R \sin \theta_0 + v(\theta_0) ,
\]

where \( v \) is the radial displacement expressed in terms of \( \theta \).

For a given \( \delta \), the non-dimensional contact length \( k = \sin \theta_0 \) is obtained from (5.1.5) in the following manner: We rewrite (5.1.5) in the form

\[
f(k) = \frac{\delta}{R} - \left\{ 1 - 2 \left( \frac{E' - k^2 K'}{2 - k^2} \right) \right\} = 0 , \quad (5.1.6)
\]

where

\[
E' = E(\cos \theta_0) = E(\sqrt{1-k^2}) ,
\]

and

\[
K' = K(\cos \theta_0) = K(\sqrt{1-k^2}) .
\]

By applying the Newton-Raphson method to (5.1.6), the following recurrence formula is obtained so that \( k \) can be evaluated iteratively.

\[
k^{i+1} = k^i - \frac{f(k^i)}{(df/dk)} ,
\]

where

\[
\frac{df}{dk} = \frac{2}{(2-k^2)^2} \left[ (2-k^2) \left( \frac{dE'}{dk} - 2kk' - k^2 \frac{dk'}{dk} \right) + 2k(E' - k^2) \right] ,
\]

\[
\frac{dE'}{dk} = \frac{k}{1-k^2} (K' - E') , \quad \text{and}
\]

\[
\frac{dk'}{dk} = \frac{1}{1-k^2} (k^2 K' - E') .
\]
Finite element meshes and deformed shapes:

From the symmetry of the problem depicted in Fig. 18, only a quadrant of the cross-section of the cylinder is considered. Two different types of finite element meshes, composed of isoparametric linear elements with 3 or 4 nodes, are used to investigate the effects of mesh size; namely, coarse mesh: \( \text{NUMEL} = 113, \text{NUMNP} = 100, \) and \( \text{NCA} = 25, \)
and refined mesh: \( \text{NUMEL} = 220, \text{NUMNP} = 178, \) and \( \text{NCA} = 49. \)
In the present work, three noded elements are degenerated by collapsing two adjacent nodes of a quadrilateral element. Fig. 19 (a) and (b) show both the undeformed and a typical deformed \((\delta=0.3 \text{ and } \mu=0.0)\) configurations for the coarse and the refined meshes, respectively.

In addition to the above displacement type loading, results for a uniform normal traction applied on the top surface of half-cylinder are compared. Fig. 19 (c) and (d) show both the undeformed and a typical deformed \((F=240 \text{ and } \mu=0.0)\) configuration for the coarse and the refined meshes, respectively.

Comparison between the analytic solution and numerical results (frictionless cases):

Table 5 summarizes the computed values of \( F, \ p_{\text{max}}, \) and \( b' \) for 5 different values of \( \delta \) and compares these results with the corresponding analytic solution. For the refined
mesh cases, mesh type A represents a finite element mesh whose nodes are allocated so as to conserve the total volume (area), while, mesh type B designates nodes located on the periphery of cylinder. For all cases, the maximum normal contact traction $p_{\text{max}}$ and the (equivalent) applied force $F$ from the numerical results are always larger than those from the analytic solution, and the half contact-width $b'$ is always smaller. This response characterized by a stiffer finite element model, physically agrees with the inherent behavior of displacement finite element solutions in elasticity.

Figs. 20 and 21 show the distribution of the normal contact traction $p_n$ along the contact surface $X$ for displacement and traction loading cases, respectively. Fig. 22(a) shows the relation between $\delta$ and $F$ for both the analytic solution and numerical results for displacement loading, and Fig. 22(b) shows the distribution of the reaction force acting on the top surface of half-cylinder in comparison with the uniform distribution for traction loading. Figs. 23 (a) and (b) compare analytical and numerical values of $b$ and $p_{\text{max}}$ in terms of the applied force $F$.

Effects of friction:

Table 6 summarizes numerical values of $p_{\text{max}}$, stick vs contact length, and $F_n$ and $F_t$ (the normal and tangential
components of $F$) for several different values of $\delta$ and $\mu$. All the results in the table are for the displacement loading case with the refined mesh, the variable-increment scheme, and $ERON=10^{-15}$ in term of $e^Z$. The number of increments for the variable-increment scheme is about twice the number of contact nodes for a given $\delta$. The size of increment for each incremental stage is determined so that a contact-candidate node first comes close to body 'b', but is not in contact. Its pairing point is then determined at the subsequent increment stage.

Figs. 24-25 (a) and (b) show the distributions of the contact traction $p_n$ and $p_t$ in (a) and the ratio $r=p_t/\mu|p_n|$ in (b) for three different values of $\delta$, respectively.

Table 7 summarizes computed results of the frictional dissipation energy and the external work-done for different values of $\delta$ and $\mu$. The frictional dissipation energy during each incremental stage is obtained by multiplying the nodal friction force $P_t$ by the respective relative displacement $z_t$ and is then accumulated for every stage. The percentage of the dissipation energy to the work-done varies from 0.00017\% ($\delta=0.5$ and $\mu=0.5$) to 0.0418\% ($\delta=0.1$ and $\mu=0.1$).

**Effects of the number of load increments (NINC):**

Figs. 27 through 29 show the effects of the number of load increment on the distributions of the contact traction (a) and the ratio (b) for three different cases. These
effects are compared with different mesh configurations, i.e., the refined (Fig. 27) and the coarse (Fig. 28) meshes, for the same values of $\delta$ and $\mu$, i.e., $\delta=0.3$ and $\mu=0.2$. They are also compared for different friction coefficients, i.e., $\mu=0.2$ (Fig. 28) and $\mu=0.5$ (Fig. 29).

For all three cases, the results with NINC=1 (i.e., not updating the pairing map) are much different from those of other values of NINC, and they always have a 'kink' on the distribution of frictional traction. This unrealistic response, however, disappears rapidly as NINC is increased, NCA is decreased, or the friction coefficient becomes smaller. For frictionless cases, of course, there is no influence of NINC on the results, and the effects of NINC become significant as $\mu$ increases. According to the numerical results in the present work, there is almost no difference between computed results when NINC becomes greater than or equal to 50.$^\text{F}$.

It is worth noting that the effects of NINC become less significant in the accuracy of solution as NCA becomes smaller, i.e., a coarser mesh is used. This may be

---

$^\text{F}$ The number is about twice of the number of contact nodes, i.e., 26 for the refined mesh and 13 for the coarse mesh. The results with NINC>50 are thus almost the same as those with the variable-increment scheme.

$^\dagger$ The pairing point for each contact candidate node is initially regarded as the closest point on the die surface from the node in the present problem.
explained by the argument that for a given $\delta$, the normal component of contact traction and the number of contact nodes (or b) are almost the same for different values of $\mu$ and NINC when compared with changes of the frictional traction. Such a system may be compared to a mass-spring system having a number of blocks (i.e., NCA) connected with linear springs between them in a row on a frictional surface and subjected to fixed normal forces for each block. When one end is fixed and the other end is subjected to a given tangential force, the solution for the equilibrium state of the spring system is less influenced by a slight perturbation of the location of blocks (i.e., the pairing map) as the number of blocks decreases. Similarly, the present problem becomes less constrained to the tangential motion as NCA increases for given values of $\delta$ and $\mu$. The accuracy of numerical solutions for friction problems is also closely related to the selected values of ERON coupled with those of NINC as discussed below.

Effects of the allowable limit of error norm (ERON):

Figs. 30 through 32 show the effects of prescribed values of ERON on the distributions of the contact traction (a) and the ratio (b) for three different values of NINC. For all three cases, other factors are set to the same except NINC and ERON. Also, ERON, in the present work, is evaluated in term of $e^Z$ defined in (4.3.8). For the present
problem, the initial error norm \( \|e^Z\|_0 \) prior to the iteration routine is 1.4628 for \( \mu=0.0 \) (NINC=1), and it is about 0.03-0.04 at the beginning of each incremental stage for \( \mu\neq0.0 \) (variable-increment).

According to the present results, the effects of ERON become more significant on the accuracy of solution as NINC increases and/or \( \mu \) becomes bigger. This implies that the accumulated error may result in an inaccurate evaluation of the pairing map especially when such pairs are expected to be in the stick status. For frictionless cases (NINC=1), the value of ERON has a negligible effect on computed results as long as \( \text{ERON} \leq 10^{-4} \). For \( \mu=0.5 \), on the other hand, this range of ERON is approximately \( \text{ERON} \leq 10^{-10} \). Thus both NINC and ERON must be improved simultaneously in order to improve the accuracy of solution for friction problems.

**Rate of convergence:**

For a representative case (refined mesh; \( \delta=0.3 \); and variable-increment), the dependence of \( \mu \) on the rate of convergence is illustrated for three different values of \( \mu \) in Fig. 33. The normalized error norm is defined by

\[
\text{Normalized error norm} = \frac{\|e^Z\|_i}{\|e^Z\|_0},
\]

where 'i' is an iteration counter for an incremental stage, and \( \|e^Z\|_0 \) is again the initial error norm having

\( \|e^Z\|_0 = 1.4628 \) for \( \mu=0.0 \).
\[ = 0.0368 \text{ for } \mu = 0.2, \]  
\[ = 0.0329 \text{ for } \mu = 0.5. \]

It is noted that the rate of convergence remains almost the same for any incremental stage.

For the same case, the CPU time for the execution of the present solution method with the IBM 3081-D computer is examined for each part of the algorithm as listed in Table 8. Thus, the total CPU time for the present problem (refined mesh: NUMNP=178, NUMEL=220, and NCA=49) is estimated approximately to be

\[
(0.746 + 3.311) + 0.988 + \text{NINC} \times [(0.007 + 0.036) + (\text{no. of iterations}) \times \{(\text{no. of contact nodes}) \times (6.88 \times 10^{-4} \times 2)\}] \\
+ 0.157 + \text{(additional time for the computation of strains, and stresses, etc).}
\]

The above estimation may be subjected to an error of ±10%.

For the case of \( \delta = 0.3 \) (no. of contact nodes = 26), \( \mu = 0.0 \), NINC=1, and ERON=10\(^{-15}\) (no. of iterations=210), the total CPU time is 14.70 (sec), and it is 5.65 (sec) for the corresponding linear problem with the direct solution scheme. For the case of \( \delta = 0.5 \) (no. of contact nodes = 36), \( \mu = 0.5 \), NINC=70 (variable-increment), and ERON=10\(^{-15}\), the total CPU time becomes 349.90 (sec). It is noted that almost one-half of the above CPU time is spent for the computation of error norm in order to study the convergence characteristics of the present solution algorithm. This time estimation can therefore be reduced by a large amount.
5.2. **Indentation of Rigid Punches into a Semi-Infinite Elastic Medium**

**Problem description:**

Fig. 34 illustrates three different types of frictional indentation problems considered in the present work. The first two types are axisymmetric problems with flat and spherical shaped punches, respectively. The third one is a plane-strain problem with a flat punch indenting into an elastic medium with a transverse motion. For all three cases, the following dimensionless quantities are used:

- Young's modulus: $E = 1000$.
- Poisson's ratio: $\nu = 0$.
- Radius of the flat punch (a): $R = 1$.
- Radius of the spherical punch (b): $R = 3$.
- One-half of the width of flat punch (c): $b = 1$.

An elastic half-space medium is then idealized as a medium with a finite size of $10 \times 10$ representing the radius (one-half of width) and depth.

Spence [71,72] has studied the above three frictional indentation problems analytically and presented numerical results obtained by solving the resulting coupled Volterra equations iteratively. The following normalized contact traction and distance are used for comparison between
Spence's results and the present finite element solutions.

Normalized contact traction \( \left[ \frac{G}{1-\nu} \right] \left[ \frac{\delta}{b} \right] p \),

normalized distance from the centerline \( X = \frac{x'}{b} \),

where \( G \) is the shear modulus, \( b \) is the radius of contact or one-half of the contact length, \( \delta \) is the maximum depth of indentation, and \( x' \) is the radial or horizontal coordinate.

In the third problem, the translation of the rigid indentor in the horizontal and vertical directions are denoted by \( u \) and \( v \), respectively, and the resultant forces acting on the indentor in these directions are denoted by \( P \) and \( Q \).

Finite element meshes and deformed shapes:

From the symmetry, only one-half of the space domain is discretized into finite element meshes:

NUMNP = 99, NUMEL = 108, and NCA = 21,

for the first two axisymmetric problems.

On the other hand, the full domain is discretized into finite element meshes:

NUMNP = 186, NUMEL = 212, and NCA = 41,

for the plane-strain problem with the transverse motion.

Fig. 35 (a), (b), and (c) show the undeformed and deformed meshes for the respective problems where

(a): \( \mu = 0.2063, \delta = 0.075, \text{NINC} = 1, \text{and ERON} = 10^{-10} \),

(b): \( \mu = 0.2986, \delta = 0.1, \text{NINC} = 10, \text{and ERON} = 10^{-10} \),

and

(c): \( \mu = 0.38, u = 0.01 \) and \( v = 0.025 \) (equivalent to
the case of \( Q/\mu P = 0.6809 \), \( NINC = 1 \), and \( ERON = 10^{-10} \).

Comparison between Spence's semi-analytic and the present finite element solutions:

(a) The distribution of the normalized normal contact traction

Figs. 36 through 38 show the distribution of the normalized normal contact traction along the distance normalized by the contact length. For three typical friction coefficients, Spence's and the present numerical results are compared in Figs. 36 and 37 for the axisymmetric problems with flat and spherical punches, respectively. The present finite element solutions always give larger values of the contact traction than that of Spence's, but both results agree favorably in a global comparison. For the spherical punch problem, the normal direction of the contact surface is regarded as the global vertical direction of the problem at all times in Spence's results. However, the local normal direction is updated with \( NINC = 20 \) in the present results. Because of the similarity in the effects of \( NINC \) on the solutions of Hertzian type contact problems, this comparison of numerical results is omitted.

Fig. 38 illustrates the nature of the distribution of the normalized normal contact traction for several different values of the horizontal motion of the plane flat punch. It also shows the symmetrical characteristics of the finite
element solutions for \( u = \pm 0.005 \), and \( u = 0.000 \).

In Figs. 36 and 38, it is shown that the distribution of the normal contact traction on the flat contact surface has a 'kink' along the intersection line of regions in a stick and slip status. Also, the contact traction for flat punch problems becomes infinite along the boundary of the contact surface. It is noted that the contact traction fluctuates in sign infinitely as \( X \to \pm 1 \) when both the punch and the elastic medium are completely bonded. Johnson [32] said that such an anomalous nature of the singularity arises from the inadequacy of the linear theory of elasticity to handle the high strain gradients in the region of the singularity.

(b) The distribution of the contact traction ratio

Figs. 39 through 41 illustrate the distribution of the contact traction ratio (i.e., the regions of stick and slip) for each of the above three problems. The present solutions for both the axisymmetric and plane-strain flat punch problems again compare favorably with the corresponding results presented by Spence. The discrepancy in the stick/slip regions between the two solutions for the spherical punch problem is attributed to the differences arising from consideration of the normal direction of the contact surface, as mentioned previously. The two solutions agree well for the case of \( \text{NINC} = 1 \).
(c) The change of stick/slip regions due to the transverse motion

Fig. 42 depicts the effect of the transverse motion on the boundary of stick/slip regions. Here, the ratio $Q/\mu P$ is computed from the sums of the normal and tangential contact traction corresponding to the solution of the respective displacement-controlled problem. For $Q/\mu P=0$ (i.e., a symmetric loading), both Spence's and the present results agree exactly. However, the stick region in the present results becomes shallower and is shifted nearer to the edge of punch than that in Spence's results as the ratio $Q/\mu P$ becomes bigger.

5.3. Simulation of the Effects of Friction on Deformation Modes of an Elastic Ring under Compression

Problem description:

The ring compression test [43] provides a convenient way to investigate the frictional behavior of an inelastic material paired with a rigid die surface in metal forming industries. The change in the inner diameter of a ring is measured experimentally for a given amount of compression, and it is then fitted into the calibration curves obtained theoretically to find the friction coefficient. In many theoretical analyses for the simulation of the ring compression test such as the upper bound methods, the finite
element methods, etc., it is often assumed for simplicity that there exists a neutral radius along which the direction of slip changes.

In the present work, however, the deformation modes of an elastic ring are simulated for a variety of friction coefficients. The geometric and material properties of a standard ring are listed in Fig. 43. The primary focus of the present work is on the effects of friction on the slip (flow) direction and the stick region at the die-workpiece interface. According to the present numerical results invoking elastic analysis, a stick zone is formed instead of the neutral radius and the zone expands as the friction coefficient becomes bigger.

**Finite element model:**

From the symmetry, the upper right quadrant of the ring is discretized with axisymmetric four noded isoparametric linear finite element meshes having

\[ \text{NUMNP} = 66, \text{NUMEL} = 50, \text{and NCA} = 11, \]

as depicted in Fig. 44.

The present problem is not affected by the number of load increments. Thus, we select \( \text{NINC} = 1 \) for all cases. For all values of \( \mu \), the number of iterations for \( \| e_\| \leq 10^{-7} \) varies from 13 (\( \log(\| e_{\|}^{13}/\| e_{\|}^{0} \) = -8.4 for comparison with the case depicted in Fig. 33) for \( \mu=0.01 \) to 149 (\( \log(\| e_{\|}^{149}/\| e_{\|}^{0} \) = -7.95) for \( \mu=0.5 \).
Deformation modes and stick/slip regions:

Fig. 44 shows both the undeformed and deformed shapes of the ring, and the circle on the deformed shape designates a node in the stick condition. For a small value of $\mu$, say, $\mu = 0.01$, both the inner and outer diameters increase, and so all points of the ring at the interface slide outward. As $\mu$ increases, the inner diameter (e.g., at $z = 0$) decreases, and both the inner and outer surfaces of the ring bulge inward and outward, respectively. This behavior is attributed to friction at the die-workpiece interface.

The effect of Poisson's ratio is also examined in Fig. 44. For a given $\mu$ (i.e., $\mu=0.2$), the stick zone decreases as $\nu$ increases (i.e., from $\nu=0.3$ to $\nu=0.48$). This implies that the incompressibility condition or plastic flow can play an important role in the determination of the deformation mode and the stick/slip region.

Distribution of the contact traction and the shear stress:

Figs. 45 and 46 reveal the distribution of the contact traction and the normalized traction ratio, respectively. For a given Poisson's ratio, the normal contact traction increases slightly as $\mu$ increases. For a given $\mu$, on the other hand, it increases significantly as $\nu$ increases.

Figs. 47 and 48 depict the effects of $\mu$ and $\nu$ on the distribution of the shear stress $\tau_{rz}$, respectively. The numbers on the plots designate the appropriate value of $\tau_{rz}$. 
as shown in the 10 segments at the bottom of the figures. It is shown that friction has a great influence on the distribution of $\tau_{rz}$.

5.4. Effects of Friction at the Fiber-Matrix Interface in Ceramic Matrix Composites

Problem description:

It is known that frictional forces acting between the fiber reinforcement and its matrix have a great influence on the strength and toughness of ceramic composites. In this respect, the indentation method [47] is recently proposed to measure experimentally the average frictional traction between the fiber reinforcement and its matrix in ceramic composites.

In the present work, the indentation method is analyzed for the evaluation of the frictional traction distribution along the fiber-matrix interface with the developed solution algorithm. Detailed description and physical implication of the present work are presented in Ref. [14].

Finite element model:

Fig. 49 illustrates the finite element model for the simulation of the indentation test and the properties of the fiber and matrix materials. In the model, axisymmetric four noded isoparametric linear elements are used and

NUMNP = 170, NUMEL = 99, and NCA = 34.
The aspect ratio of the fiber is 1 to 170 in term of radius to length.

No significant difference is found between the computed results with NINC=1 and 10. however, the rate of convergence in this problem decreases significantly compared to the previous problems due to the severe aspect ratio. As a comparison with the problem in Section 5.1. (refer to Fig. 33), at the iteration counter i=200 for \( \mu = 0.2 \),

\[
\log \left( \frac{\| \epsilon Z \|^1}{\| \epsilon Z \|^0} \right) = -4.26 \text{ for the present problem, and} \]

\[
= -13.30 \text{ for the problem in Fig. 33.}
\]

For all values of \( \mu \), the error converges logarithmically as depicted in Fig. 33.

Results and discussion:

Figs. 50 and 51 show the distributions of the normal and frictional tractions along the fiber length for different friction coefficients, respectively. Fig. 52 shows the associated normalized traction ratio. The peak value of the frictional traction (the interfacial shear stress) increases drastically with increasing friction coefficients with the maximum value bounded by that for the perfectly-bonded case. Correspondingly, the interfacial shear stress concentrations for specified friction coefficients are localized in the vicinity of the applied load and exhibit a pronounced jump in the slip region due to the Poisson effect. The shear stress profile, for the
perfectly-bonded case, declines monotonically. The average interfacial shear stress, i.e., \( \tau_{av} = \int_{0}^{L} \rho_{t}(z) \, dz \), is 1.74 MPa for all friction coefficients and the perfectly-bonded case.

Of particular interest is the non-linear compliance of the composite system for different values of friction coefficient. The load-point displacement versus coefficient of friction is plotted in Fig. 53. To conform to the load-point displacement measured experimentally [47] for the SiC-LAS composite system, the present plot suggests that the coefficient of friction between the fiber and the matrix is approximately 0.1. As expected, the computed load-point displacement value \( \delta = 0.57 \mu m \) for the perfectly-bonded case is approached asymptotically.

5.5. Application to Hydraulic Fracturing

Hydraulic fracturing is an established and effective technique for the stimulation of oil and gas recovery in a variety of reservoir rocks. In this process, a pressurized fluid is injected through a wellbore to induce a fracture in the zone of interest (payzone). Among many factors which may have a great influence on the migration of the induced fractures, of major concerns are the variation of material properties in a multilayered geological medium, the in-situ stress distributions in each layer, the characteristics of
the layer interfaces, and so on.

In the past decade, Advani, Lee, and their associates have numerically simulated the hydraulic fracturing problems using the finite element methods [1,2]. Recently, they have attempted [30,45] to investigate the effects of friction at the layer interfaces on the induced fracture behavior (i.e., penetration into adjacent layers, arrest, or interfacial propagation). Of particular interest is the case when the crack tip is located at the interface (called the terminal crack). In the present work, it is assumed that the singularity field vanishes as the the crack tip slides along the interface and it is blunted. Following the previous investigations, this interfacial behavior is studied with the present solution method focused on the correlation of the effect of friction with other primary factors.

**Problem description:**

Fig. 54 depicts a representative hydraulic fracturing model. The quantities $\sigma_v$ and $\sigma_h$ represent the vertical and the horizontal in-situ stresses, respectively. The parameters $\Delta \sigma$ and $p$ then represent the differential horizontal in-situ stress between two layers and the fracture fluid pressure on the crack surface, respectively. The basic values for the selected model are:

$B = 900$ ft, $H_1 = 100$ ft, $H_2 = 800$ ft, $E_1 = 1 \times 10^7$ psi, $v_1 = v_2 = 0.2$, and
\[ \sigma_H = 2000 \text{ psi}, \]
and other parameters are then varied within the following ranges:

\[ \frac{a}{H_1} = (1. - 1.5), \quad m = \frac{E_2}{E_1} = (0.5 - 3.). \]

\[ \frac{\sigma_Y}{\sigma_H} = (1. - 4.), \quad \text{and} \quad \frac{\Delta \sigma}{\sigma_H} = (-0.1 - 0.25). \]

where \( a \) is one-half of the crack length, \( H_1 \) and \( H_2 \) are the heights of two layers, i.e., the payzone and the adjacent layers, respectively, and \( B \) is the width of the present model. Also, \( E \) and \( v \) are Young's modulus and Poisson's ratio, respectively, and \( m \) is the modulus contrast. The plane-strain condition is assumed.

For all cases, the fluid pressure \( p \) is determined so that the stress intensity factor \( K_I \) becomes 1000 psi-\( \sqrt{a} \) for the respective perfectly-bonded interface problem.

**Finite element model:**

Due to symmetry, the upper right quadrant of the model is discretized into finite elements. Fig. 55 illustrates both the overall finite element mesh of the present model and the detailed mesh around the crack tip. Both 3 and 4 noded isoparametric linear elements are used except the crack tip elements attached to the crack tip. In the present work, three different meshes are used depending on the location of the crack tip. For the case \( a/H_1 = 1 \), we take \( \text{NUMNP} = 268, \text{ NUMEL} = 270, \text{ NCA} = 38, \text{ and } \text{NCB} = 38. \)
The meshes for the other cases have almost the same patterns as above except for the location of the crack tip.

The number of load increments NINC does not have any significant influence on the global solution. In this work, however, the local field near the crack tip is of major concern. Thus, for all cases, NINC = 10 and ERON = 10^{-6} in terms of \|e\|.

**Crack tip element:**

In the present work, the crack tip element is generated by replacing the interpolation functions for a collapsed quadri-lateral isoparametric element by a singular mapping with an arbitrarily given order of singularity [80]. The subroutine for the crack tip element with an arbitrary order of singularity is also implemented into FAST. In order to check the validity of the subroutine and to examine the effect of the crack tip element size, numerical experiments for simulating the mode I type loading are carried out, and the results are plotted in Fig. 56. In this figure, the size of the crack tip element is represented by the length of its equi-lateral edge having 30 degrees of the vertex angle. The results connected with solid lines are obtained using the same mesh as in Fig. 55, but with the perfectly-bonded interface. The results for the length r = 1.25 ft (isolated data points) are obtained using the mesh having 14 nodes and 12 elements less than the above mesh near the
crack tip. Two different types of the mode I loading are considered; namely, the $\sigma_{H}$ and $p$ loadings with $\sigma_{H} = p = 1589$ psi. The stress intensity factor $K_{I}$ is then evaluated from the relation:

$$K_{I} = \frac{E}{4(1-\nu^{2})} \sqrt{\frac{2\Pi}{r}} (u_{x})_{\theta=\Pi},$$

where $(u_{x})_{\theta=\Pi}$ is the crack opening displacement at a distance $r$ from the crack tip. For both loading cases, the ratio of $K^\text{comp}_{I}$ to $K^\text{theory}_{I}$ has a transition point at $r = 0.4$, which is generally understood as a common characteristic of crack tip elements [80].

For the case of the $\sigma_{H}$ loading, without using the singular elements with the same mesh, the ratio is also compared with that for the case with the singular element. For $r = 0.4$ ft, we obtain $K^\text{theor}_{I}/K^\text{comp}_{I} = 1.006$ and 0.913 with and without the crack tip elements, respectively, and for $r = 1.25$ ft, we obtain $K^\text{theor}_{I}/K^\text{comp}_{I} = 0.990$ and 0.897, respectively.

The combined loading represents the case when both $\sigma_{H}$ and $p$ are applied simultaneously. In this case, the loading ratio $p/\sigma_{H}$ must be obtained first for the present case. Unlike the above trend for single loading cases, the combined loading case seems to
converge to the theoretical value within the range considered here.

It is also noted that the present crack tip element yields a better performance than Akin's crack tip element in the above numerical tests. In the present work, the mesh with \( r = 0.2 \) is used in the case of \( a/H_4 = 1.5 \), and \( r = 0.4 \) for all other cases.

**Pre-strained artificial spring element:**

As shown in Fig.s 54 and 55, the payzone layer is not restricted geometrically in the horizontal direction, and the adjacent layer is not restricted geometrically in the vertical direction. Due to the arbitrariness involved in the displacement solution, the stiffness matrix of each body becomes singular. To avoid difficulties in evaluating the respective compliance matrix, two methods are usually used. First, the traction boundary condition is approximated simply by the respective displacement boundary condition corresponding to the solution for the perfectly-bonded problem. Second, artificial springs are installed additionally at some appropriate positions [45]. Generally speaking, the frictional contact status is closely related to the compliances of contacting bodies as studied before. In this respect, the latter method may be better than the former method.

In the present work, pre-strained artificial spring
elements are used in order to reduce the energy stored in
the artificial spring elements, which may have an influence
on the solution. The amount of pre-strain for each spring
is then determined from the solution of the perfectly-bonded
problem. The pre-strained spring elements are installed
uniformly on the right surface for the payzone layer and on
the top surface of the adjacent layer. According to the
present numerical results, the amount of strain energy
stored in these elements becomes negligible after
deformation.

Computational procedures:

The computational procedures for the examination of the
effects of friction on the slip/stick status of the crack
tip are as follows:

(1) For given values of the material properties and loading,
the fluid pressure \( p \) is first determined so that the stress
intensity factor \( K_I \) is equal to 1000 psi-\( \sqrt{\text{in}} \) for the
respective perfectly-bonded problem. The superposition
principle and the linearity of elastic solutions can be used
conveniently for the evaluation of the fluid pressure \( p \).

(2) The displacement solution for the above perfectly-bonded
problem is obtained to determine the amount of pre-strain of
the artificial elements.

(3) With the developed solution method, the main problem is
then solved for several different values of the coefficient
of friction.

Results:

(a) Terminal crack:

Figs. 57 (a) and (b) illustrate the stick/slip status of the crack tip for different interface friction coefficients versus the stress ratio \((\sigma_y/\sigma_H)\) and modulus contrast \((m)\), respectively. For the case of the terminal crack, crack blunting results from sliding along the interface, and so the transition point separating the slip and stick regions may represent the limiting condition for the containment of the fracture in the payzone. The qualitative trend illustrated in Fig. 57 (a) agrees with experimental results [75]. The differential horizontal stress \(\Delta \sigma\) between layers has no significant influence on the transition point.

(b) Penetrated crack:

Fig. 58 shows the stress intensity factor for the interface model, normalized with respect to that for the perfectly-bonded case, versus friction coefficient for two examples representing significant fracture penetration across the interface. It is noteworthy that crack closure may result for low values of friction coefficient, influenced by \(\sigma_y/\sigma_H, m,\) and \(\Delta \sigma/\sigma_H\).
CHAPTER VI
CONCLUSIONS AND RECOMMENDATIONS

In the present work, a new computational model for two-dimensional elastic contact problems with friction has been studied. The accuracy of solution for friction problems and the convergence of the associated iterative scheme are investigated in detail. Three major components of the present work are summarized below.

First, the general structure of mathematical contact problems with Coulomb's friction law is studied within the framework of non-linear field theory in continuum mechanics. In Chapter II, the continuum concept of the pairing map is explained for the adequate description of the path-dependent nature of friction problems. Also, the contact conditions are expressed in the non-integrable rate form with use of the convective relative velocity. The consistency condition for friction problems is compared with the corresponding condition for plasticity problems.

A semi-implicit scheme for numerical time-integration of the contact conditions is then proposed. According to the proposed integration scheme, the pairing map and the
local normal direction are updated at the beginning of each incremental step with reference to an 'assumed' deformed configuration instead of taking full account of kinematic relations between configurations. Based on the integration scheme, the incremental contact boundary-value problem is described as PROBLEM I.

Also, a variational principle is proposed for elastic contact problems with friction on the basis of 'mini-max' principles in mechanics. Both displacement and contact traction vectors are taken as independent state variables. Thus, the variational equation for the equilibrium of each body changes to a bilateral form with respect to the unknown displacement field of the respective body, and that for the contact conditions between two bodies reduces to unilateral form with respect to the unknown contact traction field. The variational problem is now called as PROBLEM II, and the mathematical equivalence between two problems is proved in the context of functional analysis.

Secondly, the above variational continuum problem is reformulated into the finite dimensional problem associated with the finite element method. Instead of attempting to satisfy the conformity requirement along the contact boundary, in the present study, the ad-hoc scheme called the node-to-segment contact is employed. The governing discrete contact equations are derived in Chapter III, which is then
defined as PROBLEM III.

As a central part of the present numerical scheme, a method of embedding the contact conditions into the derived governing contact equations is investigated in Chapter IV. The associated non-linear surjective mapping is then called the embedding map.

With the help of the embedding map, PROBLEM III can be represented in the form of a fixed point problem for the nodal contact force vector. As a consequence, the existence and uniqueness of solution is studied using the Banach fixed point theorem. Finite dimensional solution(s) always exists as long as the matrices of Green's function for contacting bodies are RSPD (real, symmetric, and positive definite) and the associated iteration function is a contraction. On the other hand, the conditions for the uniqueness of solution depend not only on the system compliance (M) of contacting bodies but also on the data (W) representing all kinds of applied loading. Sufficient conditions for uniqueness independent of the data W are of considerable interest. For sufficiently small friction coefficient and also for frictionless cases, the sufficient conditions are always satisfied. The conditions for the uniqueness of solution are also used effectively for the iterative solution scheme.

An iterative scheme based on the block Gauss-Seidel method is proposed for solving the above non-linear fixed point problem. The convergence criteria for the present
iterative scheme is also studied.

Finally, the validity of the present solution method is demonstrated by several numerical examples in Chapter V. Several factors which may influence the accuracy of solution such as the number of load increments, prescribed tolerance for the error norm, finite element meshes, etc. are investigated. The convergence of the iterative scheme is also examined through these examples.

The number of load increments has a great influence on both the accuracy of solution and computing cost. Although it cannot be selected simply for a given problem, the number, in general, has to be increased as the change of contact boundary for a given finite element mesh becomes significant, the coefficient of friction becomes bigger, the change of the normal direction of contact surfaces becomes significant, and the allowable error norm becomes smaller.

For friction problems, the refinement of the finite element mesh for the contact boundary must be accompanied by the refinement of other parameters such as the number of load increments, the allowable error norm, etc. An accurate evaluation of the motion trajectories of contact points can reduce the number of load increments.

The present iterative algorithm is shown to converge for all the example problems studied. The rate of convergence depends on both the compliances of contacting
bodies and the coefficient of friction, as discussed in Section 4.3.2. For the presented computations, the rate of convergence is more significantly influenced by the compliances of contacting bodies than by the coefficient of friction.

In conclusion, the concept of pairing map and the numerical integration of the rate form of contact conditions paved a concrete way to formalize a sound variational principle. The finite element discretization of the variational problem and the embedding of contact conditions into the block Gauss-Seidel method are shown to be computationally effective. This type of unification of the continuum considerations and numerical methodology appears to be new and physically realistic.

The following recommendations are provided for future study.
(1) For large deformation problems, most kinematic and kinetic relations described in Section 2.2 can be applied directly. The invariant requirement under superposed rigid body motions should be studied carefully in describing the frictional contact conditions with objective kinematic and kinetic rate variables.
(2) For inelastic problems, the coupling effects between frictional and inelastic dissipative mechanisms should be investigated in view of the development of an effective time-integration scheme. The variational principle in the
present work should be extended for general inelastic contact problems with internal state variables.

(3) For dynamic problems, an effective method for accurate evaluation of momentum exchanges between contacting bodies through contact surfaces is recommended. The imposition of the local equilibrium condition should be reconsidered from both theoretical and numerical viewpoints.

(4) For three dimensional problems, the embedding map should be extended for direct application to the present solution algorithm.


78. Truesdell, C. and Noll, W., The Non-linear Field Theories of Mechanics, in Encyclopedia of Physics (Ed. S. Flugge), III/3, Springer-Verlag, 1975


Table 1  A comparison between Coulomb's friction model and J₂ plastic flow model with isotropic hardening

<table>
<thead>
<tr>
<th></th>
<th>Frictional slip model</th>
<th>Plastic flow model</th>
</tr>
</thead>
<tbody>
<tr>
<td>domain</td>
<td>boundary point ( x \in I_c^\alpha(\tau) )</td>
<td>interior point ( x \in \Omega )</td>
</tr>
<tr>
<td>kinematic variable</td>
<td>( \dot{z}(X,\tau) )</td>
<td>( \dot{d}_P(x,\tau) )</td>
</tr>
<tr>
<td></td>
<td>( = v^a(X,\tau) - v^b_0 \mathbf{n}(X,\tau) )</td>
<td>( = d(x,\tau) - d^e(x,\tau) )</td>
</tr>
<tr>
<td>kinetic variable</td>
<td>( p(X,\tau) )</td>
<td>( s(x,\tau) )</td>
</tr>
<tr>
<td>slip (yield) function</td>
<td>( f(p) =</td>
<td>p_t</td>
</tr>
<tr>
<td>slip (flow) rule</td>
<td>( \dot{z} = \dot{\alpha} I(p_n) + \beta I(f) \frac{\partial f}{\partial p_t} )</td>
<td>( \dot{d}_P = \lambda I(f) \frac{\partial f}{\partial s} )</td>
</tr>
<tr>
<td>evolution variable</td>
<td>none</td>
<td>( Y = Y(\bar{e}_P) )</td>
</tr>
<tr>
<td>unloading</td>
<td>none</td>
<td>( d_P = 0 )</td>
</tr>
</tbody>
</table>

Note: (a) \( d_P \) and \( s \) are deviatoric parts of the rate of plastic deformation tensor and the Cauchy stress tensor, respectively,  
(b) \( \|s\| = \sqrt{s \cdot s} \) and \( J_2 = [(3/2) s \cdot s]^{1/2} \),  
(c) \( \bar{e}_P \) is the equivalent plastic strain defined by  
\[
\bar{e}_P = \int_0^t \sqrt{2/3} \|d_P\| \, dt
\]  
(d) \( Y \) is the yield stress which is a function of \( \bar{e}_P \),  
(e) \( \dot{\alpha} \) and \( \beta \) are non-positive constants, and \( \lambda \) is a non-negative constant,  
(f) \( I(\cdot) \) denotes the indicator defined in (2.3.15).
Table 2  Admissible contact conditions for Case 2

(a) Case 2(a): $D_{nn} < \mu |D_{nt}|$ and $D_{nt} > 0$

<table>
<thead>
<tr>
<th>$\frac{z_n}{z_{nt}}$</th>
<th>$\frac{z_t}{z_{nt}}$</th>
<th>$\frac{z_t}{z_{nt}}$</th>
<th>Admissible contact conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 0$</td>
<td>$&gt; \gamma_1$</td>
<td>$-$</td>
<td>(c)</td>
</tr>
<tr>
<td>$\leq \gamma_1$</td>
<td>$-$</td>
<td>$-$</td>
<td>(b)</td>
</tr>
<tr>
<td>$&gt; \gamma_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>(a),(b),(d)</td>
</tr>
<tr>
<td>$&lt; 0$</td>
<td>$= \gamma_2$</td>
<td>$-$</td>
<td>(a),(b)</td>
</tr>
<tr>
<td>$&lt; \gamma_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>(a)</td>
</tr>
<tr>
<td>$= 0$</td>
<td>$-$</td>
<td>$&lt; 0$</td>
<td>(a),(b)</td>
</tr>
<tr>
<td></td>
<td>$-$</td>
<td>$&gt; 0$</td>
<td>(a)</td>
</tr>
</tbody>
</table>

(b) Case 2(b): $D_{nn} < \mu |D_{nt}|$ and $D_{nt} < 0$

<table>
<thead>
<tr>
<th>$\frac{z_n}{z_{nt}}$</th>
<th>$\frac{z_t}{z_{nt}}$</th>
<th>$\frac{z_t}{z_{nt}}$</th>
<th>Admissible contact conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>$&gt; 0$</td>
<td>$&lt; \gamma_2$</td>
<td>$-$</td>
<td>(d)</td>
</tr>
<tr>
<td>$\geq \gamma_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>(b)</td>
</tr>
<tr>
<td>$&lt; \gamma_1$</td>
<td>$-$</td>
<td>$-$</td>
<td>(a),(b),(c)</td>
</tr>
<tr>
<td>$&lt; 0$</td>
<td>$= \gamma_1$</td>
<td>$-$</td>
<td>(a),(b)</td>
</tr>
<tr>
<td>$&gt; \gamma_1$</td>
<td>$-$</td>
<td>$-$</td>
<td>(a)</td>
</tr>
<tr>
<td>$= 0$</td>
<td>$-$</td>
<td>$&gt; 0$</td>
<td>(a),(b)</td>
</tr>
<tr>
<td></td>
<td>$-$</td>
<td>$\leq 0$</td>
<td>(a)</td>
</tr>
</tbody>
</table>
Table 3 Admissible contact conditions for Case 3

(a) Case 3(a): $D_{nn} = \mu |D_{nt}|$ and $D_{nt} > 0$

<table>
<thead>
<tr>
<th>$\bar{z}_n$</th>
<th>$\bar{z}_t / \bar{z}_n$</th>
<th>$\bar{z}_t$</th>
<th>Admissible contact conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0</td>
<td>$&gt; \gamma_1$</td>
<td>–</td>
<td>(c)</td>
</tr>
<tr>
<td>$\leq \gamma_1$</td>
<td>–</td>
<td>(b)</td>
<td></td>
</tr>
<tr>
<td>= 0</td>
<td>–</td>
<td>$&lt; 0$</td>
<td>(a),(b),(d)†</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>–</td>
<td>–</td>
<td>(a)</td>
</tr>
</tbody>
</table>

† Any $P_n$, $0 > P_n > \frac{-\bar{z}_t}{D_{nt} - \mu D_{tt}}$, satisfies condition (d).

(b) Case 3(b): $D_{nn} = \mu |D_{nt}|$ and $D_{nt} < 0$

<table>
<thead>
<tr>
<th>$\bar{z}_n$</th>
<th>$\bar{z}_t / \bar{z}_n$</th>
<th>$\bar{z}_t$</th>
<th>Admissible contact conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt; 0</td>
<td>$&lt; \gamma_2$</td>
<td>–</td>
<td>(d)</td>
</tr>
<tr>
<td>$\geq \gamma_2$</td>
<td>–</td>
<td>(b)</td>
<td></td>
</tr>
<tr>
<td>= 0</td>
<td>–</td>
<td>$&gt; 0$</td>
<td>(a),(b),(c)¶</td>
</tr>
<tr>
<td>&lt; 0</td>
<td>–</td>
<td>$\leq 0$</td>
<td>(a)</td>
</tr>
</tbody>
</table>

¶ Any $P_n$, $0 > P_n > \frac{-\bar{z}_t}{D_{nt} + \mu D_{tt}}$, satisfies condition (c).
Table 4  Limiting values of $\|Q^i_{II}\|_2$ for all possible combinations of contact status between $z^i_I$ and $z^I_I$

<table>
<thead>
<tr>
<th></th>
<th>(a)</th>
<th>(b)</th>
<th>(c)</th>
<th>(d)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>$= 1$</td>
<td>$\leq 1$</td>
<td>$\leq \delta_1$</td>
<td>$\leq \delta_2$</td>
</tr>
<tr>
<td>b</td>
<td>$\leq 1$</td>
<td>$= 0$</td>
<td>$\leq \delta_1$</td>
<td>$\leq \delta_2$</td>
</tr>
<tr>
<td>c</td>
<td>$\leq \delta_1$</td>
<td>$\leq \delta_1$</td>
<td>$\leq \delta_1$</td>
<td>$\leq 1$</td>
</tr>
<tr>
<td>d</td>
<td>$\leq \delta_2$</td>
<td>$\leq \delta_2$</td>
<td>$\leq 1$</td>
<td>$\leq \delta_2$</td>
</tr>
</tbody>
</table>

where $\delta_1 = [1 + \gamma_i^2]^{1/2}$ and $\delta_2 = [1 + \gamma_S^2]^{1/2}$. 
Table 5  A comparison between analytic and numerical solutions (frictionless cases)

<table>
<thead>
<tr>
<th>δ</th>
<th>analytic</th>
<th>refined meshes</th>
<th>coarse mesh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>F</td>
<td>51.76</td>
<td>54.49</td>
<td>53.63</td>
</tr>
<tr>
<td>Pmax</td>
<td>47.638</td>
<td>52.928</td>
<td>52.476</td>
</tr>
<tr>
<td>b'</td>
<td>0.691</td>
<td>0.65† (0.658♯)</td>
<td>0.65 (0.653)</td>
</tr>
<tr>
<td>F</td>
<td>118.00</td>
<td>122.80</td>
<td>120.65</td>
</tr>
<tr>
<td>Pmax</td>
<td>72.058</td>
<td>78.160</td>
<td>77.320</td>
</tr>
<tr>
<td>b'</td>
<td>1.040</td>
<td>1.00 (1.004)</td>
<td>0.95 (1.001)</td>
</tr>
<tr>
<td>F</td>
<td>193.04</td>
<td>199.61</td>
<td>195.95</td>
</tr>
<tr>
<td>Pmax</td>
<td>92.349</td>
<td>99.004</td>
<td>97.744</td>
</tr>
<tr>
<td>b'</td>
<td>1.326</td>
<td>1.25 (1.291)</td>
<td>1.25 (1.283)</td>
</tr>
<tr>
<td>F</td>
<td>275.25</td>
<td>283.36</td>
<td>278.04</td>
</tr>
<tr>
<td>Pmax</td>
<td>110.52</td>
<td>117.51</td>
<td>115.82</td>
</tr>
<tr>
<td>b'</td>
<td>1.578</td>
<td>1.50 (1.543)</td>
<td>1.50 (1.536)</td>
</tr>
<tr>
<td>F</td>
<td>363.78</td>
<td>373.20</td>
<td>366.14</td>
</tr>
<tr>
<td>Pmax</td>
<td>127.35</td>
<td>134.51</td>
<td>132.37</td>
</tr>
<tr>
<td>b'</td>
<td>1.807</td>
<td>1.75 (1.773)</td>
<td>1.70 (1.774)</td>
</tr>
</tbody>
</table>

b'  half-contact-width in the undeformed configuration,
†  the position of the last node in contact
♯ the interval between nodes are 0.05 for refined meshes, 0.1 for coarse mesh,
♯ the results of data regression with reference to the elliptic distribution of normal stresses (analytic solution).
Table 6 Effects of friction (refined mesh)

<table>
<thead>
<tr>
<th>δ</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>μ</td>
<td>F_n</td>
<td>P_max</td>
<td>F_n</td>
<td>P_max</td>
<td>F_n</td>
<td>P_max</td>
<td>F_n</td>
<td>P_max</td>
<td>F_n</td>
</tr>
<tr>
<td></td>
<td>F_t</td>
<td>sk-sp</td>
<td>F_t</td>
<td>sk-sp</td>
<td>F_t</td>
<td>sk-sp</td>
<td>F_t</td>
<td>sk-sp</td>
<td>F_t</td>
</tr>
<tr>
<td>0.0</td>
<td>54.488</td>
<td>52.928</td>
<td>78.160</td>
<td>199.61</td>
<td>99.004</td>
<td>283.36</td>
<td>117.51</td>
<td>373.20</td>
<td>134.51</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0.00-0.65</td>
<td>0</td>
<td>0.00-1.00</td>
<td>0</td>
<td>0.00-1.25</td>
<td>0</td>
<td>0.00-1.50</td>
<td>0</td>
</tr>
<tr>
<td>0.05</td>
<td>54.554</td>
<td>53.884</td>
<td>79.836</td>
<td>199.82</td>
<td>101.24</td>
<td>283.62</td>
<td>120.15</td>
<td>373.46</td>
<td>137.43</td>
</tr>
<tr>
<td></td>
<td>2.593</td>
<td>0.00-0.65</td>
<td>5.947</td>
<td>0.05-1.00</td>
<td>9.671</td>
<td>0.10-1.25</td>
<td>13.562</td>
<td>0.35-1.50</td>
<td>17.356</td>
</tr>
<tr>
<td>0.10</td>
<td>54.620</td>
<td>54.720</td>
<td>123.07</td>
<td>80.876</td>
<td>200.02</td>
<td>102.30</td>
<td>283.84</td>
<td>121.21</td>
<td>373.68</td>
</tr>
<tr>
<td></td>
<td>5.039</td>
<td>0.10-0.65</td>
<td>11.034</td>
<td>0.35-0.95</td>
<td>17.072</td>
<td>0.65-1.25</td>
<td>22.816</td>
<td>1.00-1.50</td>
<td>27.932</td>
</tr>
<tr>
<td>0.15</td>
<td>54.674</td>
<td>54.904</td>
<td>123.19</td>
<td>81.088</td>
<td>200.16</td>
<td>102.51</td>
<td>283.99</td>
<td>121.41</td>
<td>373.82</td>
</tr>
<tr>
<td></td>
<td>6.603</td>
<td>0.35-0.65</td>
<td>13.965</td>
<td>0.65-0.95</td>
<td>21.028</td>
<td>1.00-1.25</td>
<td>27.486</td>
<td>1.30-1.50</td>
<td>33.072</td>
</tr>
<tr>
<td>0.20</td>
<td>54.710</td>
<td>54.968</td>
<td>123.25</td>
<td>81.160</td>
<td>200.24</td>
<td>102.58</td>
<td>284.06</td>
<td>121.48</td>
<td>373.90</td>
</tr>
<tr>
<td></td>
<td>7.403</td>
<td>0.50-0.65</td>
<td>15.347</td>
<td>0.80-0.95</td>
<td>22.818</td>
<td>1.15-1.25</td>
<td>29.544</td>
<td>1.40-1.50</td>
<td>35.322</td>
</tr>
<tr>
<td>0.30</td>
<td>54.744</td>
<td>54.988</td>
<td>123.31</td>
<td>81.180</td>
<td>200.30</td>
<td>102.60</td>
<td>284.12</td>
<td>121.50</td>
<td>373.94</td>
</tr>
<tr>
<td></td>
<td>8.032</td>
<td>0.60-0.65</td>
<td>16.395</td>
<td>0.95-0.95</td>
<td>24.154</td>
<td>1.20-1.25</td>
<td>31.096</td>
<td>1.45-1.50</td>
<td>37.022</td>
</tr>
<tr>
<td>0.50</td>
<td>54.756</td>
<td>55.008</td>
<td>123.33</td>
<td>81.200</td>
<td>200.32</td>
<td>102.62</td>
<td>284.12</td>
<td>121.52</td>
<td>373.96</td>
</tr>
<tr>
<td></td>
<td>8.226</td>
<td>0.65-0.65</td>
<td>16.725</td>
<td>0.95-0.95</td>
<td>24.564</td>
<td>1.25-1.25</td>
<td>31.548</td>
<td>1.50-1.50</td>
<td>37.498</td>
</tr>
</tbody>
</table>

1 sk: the position of the last node in the stick condition.
sp: the half contact-width
<table>
<thead>
<tr>
<th>$\delta$</th>
<th>0.1</th>
<th>0.2</th>
<th>0.3</th>
<th>0.4</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>0.1</td>
<td>0.2</td>
<td>0.3</td>
<td>0.4</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>2.7244</td>
<td>12.280</td>
<td>29.942</td>
<td>56.672</td>
<td>93.300</td>
</tr>
<tr>
<td>0.05</td>
<td>1.1347</td>
<td>4.4930</td>
<td>9.1580</td>
<td>13.918</td>
<td>17.668</td>
</tr>
<tr>
<td></td>
<td>2.7277</td>
<td>12.294</td>
<td>29.974</td>
<td>56.724</td>
<td>93.365</td>
</tr>
<tr>
<td>0.10</td>
<td>1.1429</td>
<td>3.8145</td>
<td>6.6043</td>
<td>8.7750</td>
<td>10.200</td>
</tr>
<tr>
<td></td>
<td>2.7310</td>
<td>12.307</td>
<td>30.003</td>
<td>56.768</td>
<td>93.420</td>
</tr>
<tr>
<td>0.15</td>
<td>0.61842</td>
<td>1.7494</td>
<td>2.7193</td>
<td>3.4196</td>
<td>3.9131</td>
</tr>
<tr>
<td></td>
<td>2.7337</td>
<td>12.319</td>
<td>30.024</td>
<td>56.796</td>
<td>93.455</td>
</tr>
<tr>
<td>0.20</td>
<td>0.29664</td>
<td>0.76948</td>
<td>1.1452</td>
<td>1.4504</td>
<td>1.6607</td>
</tr>
<tr>
<td></td>
<td>2.7355</td>
<td>12.325</td>
<td>30.036</td>
<td>56.812</td>
<td>93.475</td>
</tr>
<tr>
<td>0.30</td>
<td>0.06968</td>
<td>0.19246</td>
<td>0.27287</td>
<td>0.32541</td>
<td>0.36319</td>
</tr>
<tr>
<td></td>
<td>2.7372</td>
<td>12.331</td>
<td>30.045</td>
<td>56.824</td>
<td>93.485</td>
</tr>
<tr>
<td>0.50</td>
<td>0.01020</td>
<td>0.05088</td>
<td>0.06853</td>
<td>0.10977</td>
<td>0.16257</td>
</tr>
<tr>
<td></td>
<td>2.7378</td>
<td>12.333</td>
<td>30.048</td>
<td>56.824</td>
<td>93.490</td>
</tr>
</tbody>
</table>

(1) Values are 0.25 x (Frictional dissipation energy) x $10^{-3}$.
(2) Values are 0.25 x (external work-done).
Table 8 The CPU time consumed at each step of the present solution algorithm (IBM 3081-D computer system)

<table>
<thead>
<tr>
<th>Description of each step</th>
<th>CPU (sec)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preliminary stage</strong></td>
<td></td>
</tr>
<tr>
<td>Evaluation of $K^a$ and $W^a$ with the triangular decomposition of $K^a$</td>
<td>4.057</td>
</tr>
<tr>
<td>Evaluation of $G^a$</td>
<td>0.988</td>
</tr>
<tr>
<td><strong>Incremental stage</strong></td>
<td></td>
</tr>
<tr>
<td>Updating the pairing map and normal</td>
<td>NINCx0.007</td>
</tr>
<tr>
<td>Updating $t_M$ and $t_W$</td>
<td>NINCx0.036</td>
</tr>
<tr>
<td><strong>Iteration stage</strong></td>
<td></td>
</tr>
<tr>
<td>Obtain new contact force</td>
<td>NCx0.000688</td>
</tr>
<tr>
<td>Compute the error norm</td>
<td>NCx0.000688</td>
</tr>
<tr>
<td><strong>Final stage</strong></td>
<td></td>
</tr>
<tr>
<td>Obtain the displacement by back-substitution routine</td>
<td>0.157</td>
</tr>
<tr>
<td>Compute and print strains, stresses, etc.</td>
<td>1.436</td>
</tr>
</tbody>
</table>

Notes: (1) NUMNP=178, NUMEL=220, and NCA=49,  
(2) NC=No. of contact nodes  
  (e.g., NC=26 for refined mesh with $\delta=0.3$)
Fig. 1 A two-body contact problem
Fig. 2  The motion trajectory of a representative contact point $X^a$
185

outward normal to the cone surface

(a) Coulomb's friction model

\[ \phi = \tan^{-1} \mu \]

(b) Drucker-Prager plastic flow model

Direction cosines

\[ (\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}) \]

Fig. 3 A comparison between (a) Coulomb's friction model and (b) Drucker-Prager plastic flow model
Fig. 4 A comparison between two functions:
(a) a non-differentiable function $|u_T|$
(b) a regularized function $\phi_\varepsilon(u_T)$
Fig. 5 A comparison between (a) the interface model and (b) the conventional contact model.
Fig. 6 Three possible cases of transition for an incremental loading stage
Fig. 7 The consistency condition and the role of normal vector
Fig. 8 A simplified time-integration scheme
Fig. 9 Evaluation of the pairing points and the normal vector
Fig. 10 Admissible domains for $P$ and $Z$

- (a): separation
- (b): stick
- (c)(d): slip
(a) : separation
(b) : stick
(c) (d) : slip

Fig. 11 Feasible solution region
Case (1): $D_{nn} > \mu |D_{n1}|$
(a) $D_{nt} > 0$
(b) $D_{nt} < 0$

(a) separation
(b) stick
(c)(d) slip

Fig. 12 Feasible solution region
Case (2): $D_{nn} < \mu |D_{nt}|$
Fig. 13 Feasible solution region
Case (3): $D_{nt} = \mu |D_{n}|$
frictionless and rigid string

slider

weightless block

frictional surface

an unstretched spring \( (K_1) \)
a pre-strained spring \( (K_2) \)
applied force \( F = K_1 x_0 \)

Contact condition

Equilibrium and friction conditions

\[ x \leq x_0 \quad \Sigma F_x = F - K_1 x - p_T = 0 \]

where \( p_T \leq \text{sign}(x) \mu K_2 (x_0 - x) \)

<table>
<thead>
<tr>
<th>Case</th>
<th>Admissible solutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1)</td>
<td>( K_1 &gt; \mu K_2 )</td>
</tr>
<tr>
<td>(2)</td>
<td>( K_1 &lt; \mu K_2 )</td>
</tr>
<tr>
<td>(3)</td>
<td>( K_1 = \mu K_2 )</td>
</tr>
</tbody>
</table>

Fig. 14 A hypothetical spring system illustrating Cases (1), (2), and (3)
flat and rigid wall with friction

\( r_i = 1, \ r_o = 2, \ h = 2, \ E = 100,000, \ v = 0.2, \ t = 100 \)

NUMEL = 4
NUMNP = 9
NCA = 3

(a) an illustrative example

Fig. 15 The effect of the coefficient of friction on the rate of convergence
\[
\begin{bmatrix}
Z_n \\
Z_l
\end{bmatrix}_{1} = (10^{-4}) \begin{bmatrix}
.554 & .207 \\
.605 & .173 & .151 \\
-.020 & .422 & -.178 & .373 \\
\end{bmatrix}_{11} \begin{bmatrix}
.042 & .132 \\
-.178 & .373 \\
\end{bmatrix}_{13} \begin{bmatrix}
R_h \\
R_l
\end{bmatrix}_{1} + (10^{-1}) \begin{bmatrix}
-.045 \\
-.296
\end{bmatrix}_{1}
\]

\[
\begin{bmatrix}
Z_n \\
Z_l
\end{bmatrix}_{2} = (10^{-4}) \begin{bmatrix}
.257 & -.018 \\
.434 & -.182 & .383 \\
\end{bmatrix}_{22} \begin{bmatrix}
.159 & -.164 \\
-.182 & .383 \\
\end{bmatrix}_{23} \begin{bmatrix}
R_h \\
R_l
\end{bmatrix}_{2} + (10^{-1}) \begin{bmatrix}
.087 \\
-.303
\end{bmatrix}_{2}
\]

\[
\begin{bmatrix}
Z_n \\
Z_l
\end{bmatrix}_{3} = (10^{-4}) \begin{bmatrix}
\text{SYMMETRIC} \\
\end{bmatrix} \begin{bmatrix}
.417 & -.206 \\
.464 & .33 \\
\end{bmatrix}_{33} \begin{bmatrix}
R_h \\
R_l
\end{bmatrix}_{3} + (10^{-1}) \begin{bmatrix}
.141 \\
-.328
\end{bmatrix}_{3}
\]

(b) the governing contact equation

Fig. 15 The effect of the coefficient of friction on the rate of convergence
(continued)
(c) The rate of convergence in terms of $e^p$

Fig. 15 The effect of friction on the rate of convergence (continued)
Fig. 16 Representation of the relation (4.3.12)
START

Triangular decomposition of $K^a$ and $K^b$
Compute $W^a$ and $W^b$
Evaluate $G^a$ and $G^b$

LOOP FOR LOAD INCREMENTS

Update the pairing map and the normal vector
Update $M$ and $W$ in (3.4.4)

ITERATION LOOP

For each pair of contact points,

Compute $\bar{Z}_i$
Determine the contact status
Obtain the nodal contact force from the corresponding embedding map

Another pair

Yes

No

Compute the error norm

Converged

No

Yes

Another load increment

Compute stresses, etc.

STOP

Fig. 17 Solution algorithm
Fig. 18 Compression of an elastic cylinder between rigid and flat dies (Plane-strain state)
Fig. 19 FEM meshes and deformed shapes

(a) coarse mesh with displacement loading
(b) coarse mesh with traction loading
Fig. 19 FEM meshes and deformed shapes
(continued)
Fig. 20 A comparison between analytic and numerical solutions for frictionless cases (displacement loading)
Fig. 21 A comparison between analytic and numerical solutions for frictionless cases (traction loading)
(a) depth of compression vs. resultant force on the top surface of half-cylinder

(b) distribution of reaction force on the top surface of half-cylinder

**Fig. 22 A comparison between types of loading**
Fig. 23 A comparison between analytic and numerical solutions for frictionless cases (summary)
Fig. 24 Effects of friction
(refined mesh; δ=0.1; variable-increment; ERON=1.0E-15)
Fig. 25 Effects of friction
(refined mesh; $\delta=0.3$; variable-increment; $\text{ERON}=1.0\text{E}-15$)
Fig. 26 Effects of friction
(refined mesh; $\delta=0.5$; variable-increment; EROH=1.0E-15)
Fig. 27 Effects of number of load increments
(refined mesh; $\delta=0.3; \mu=0.2; \text{ERON}=1.0\times10^{-15}$)
Fig. 28 Effects of number of load increments
(coarse mesh; $\delta=0.3; \mu=0.3; \text{ERON}=1.0\text{E}-15$)
Fig. 29: Effects of number of load increments (coarse mesh: φ=0.3; ν=0.5; ERON=1.0E-15)

(a) contact traction

(b) normalized traction ratio
Fig. 30 Effects of allowable error norm
(coarse mesh; $\delta=0.3; \mu=0.5; \text{NINC}=1$)
Fig. 31 Effects of allowable error norm
(coarse mesh; δ = 0.3; λ = 0.5; NINC = 20)
Fig. 32 Effects of allowable error norm
(coarse mesh: $\delta = 0.3; \alpha = 0.5; NINC = 100$)

(a) contact traction

(b) normalized traction ratio
Fig. 33 Rate of convergence vs. friction coefficient (refined mesh; δ=0.3; variable-increment scheme)
Fig. 34 Frictional indentation problems
Fig. 35 FEM meshes and deformed shapes

(a) axisymmetric flat punch
Fig. 35 FEM meshes and deformed shapes
(continued)

(b) axisymmetric spherical punch
(c) plane-strain flat punch with transverse motion

Fig. 35  FEM meshes and deformed shapes
(continued)
Fig. 36 Distribution of normal contact traction (axisymmetric flat punch)
Fig. 37  Distribution of normal contact traction (axisymmetric spherical punch)
Fig. 38 Distribution of normal contact traction (flat punch with transverse motion $U$; plane-strain condition)
Fig. 39 Distribution of normalized traction ratio (axisymmetric flat punch)
Fig. 40 Distribution of normalized traction ratio (axisymmetric spherical punch)
Fig. 41 Distribution of normalized traction ratio (flat punch with transverse motion; plane-strain condition)
Fig. 42 Stick/slip region vs. normalized traction ratio
Fig. 43 Ring compression test
Fig. 44 Deformation modes and stick/slip regions
Fig. 44  Deformation modes and stick/slip regions (continued)
Fig. 45 Distribution of contact traction with various friction coefficients
Fig. 46 Normalized traction ratio with various friction coefficients
Fig. 47 Shear stress distribution for different coefficient of friction
Fig. 48 Shear stress distribution for different Poisson's ratio
Fig. 49 The indentation test for ceramic composites

(a) axisymmetric model
(b) FEM mesh
Fig. 50  Distribution of normal contact traction along fiber-matrix interface
Fig. 51 Distribution of frictional traction along fiber-matrix interface
Fig. 52 Distribution of normalized traction ratio along fiber-matrix interface
Fig. 53 Computed load-point displacement vs. coefficient of friction
Fig. 54  Plane-strain hydraulic fracturing model
Fig. 55  FEM mesh and details of crack tip region
Fig. 56 The effect of crack tip element size on the stress intensity factor $K$. 

The graph shows the ratio of $K_{comp}$ to $K_{theory}$ as a function of the length of the equilateral edge. Different symbols represent different loading conditions: 
- □: $\sigma_h$ loading
- △: $p$ loading
- ○: Combined loading

The data indicates a decrease in the ratio of $K_{comp}$ to $K_{theory}$ as the length of the equilateral edge increases, suggesting an improvement in the accuracy of the stress intensity factor calculation with larger element sizes.
(a) stress ratio vs. coefficient of friction

(b) modulus contrast vs. coefficient of friction

Fig. 57 Stick/slip response prediction
\[ m = 3 \]
\[ \sigma_v / \sigma_H = 1.7 \]
\[ \Delta \sigma / \sigma_H = 0.7 \]

**Kp**: mode I stress intensity factor for perfectly-bonding case

Fig. 58 Crack opening mode stress intensity factor ratio vs. friction coefficient for penetrating crack geometries
The objective of the work presented in Appendix is to study the general structure of inelastic constitutive relations with finite strain within the framework of continuum thermodynamics with internal state variables [1,2]. Of particular interest is the determination of an objective stress rate for adequately describing the constitutive equations. The convective rate of the Kirchhoff stress tensor, which is identical to the Truesdell rate of Cauchy stress, is a theoretically sound as well as computationally effective measure.

As an illustrative example, the above formalism is applied to the constitutive relations for a thermo-elasto-visco-plastic material based on the Perzyna's viscoplastic theory with isotropic hardening [11-13].
A. General Description of Response of Inelastic Materials within the framework of Continuum Thermodynamics with Internal State Variables

The approach presented here is based on continuum thermodynamics concepts with internal state variable definitions developed by Coleman and Gurtin [1,2]. Within this thermodynamics theory, the general structure of constitutive relations for inelastic materials is investigated. It is shown that there exists a unique pair of the rates of stress and strain tensors under the above formalism developed in the sequel.

A.1. Basic relations

The thermomechanical behavior of a continuous body is described below by nine functions, designated as state variables. In order to describe the internal mechanisms of energy dissipation, a set of internal state variables is also introduced. These basic state variables and a set of internal state variables are first defined as follows:

**Primary variables**

(i) \( x = x(X,t) \) denotes the spatial position of a material particle \( X \) in the motion \( x \).

(ii) \( \theta = \theta(X,t) \) designates the absolute temperature.

**Complementary variable**

(iii) \( T = T(X,t) \) is the Cauchy stress tensor.
(iv) \( e = e(X, t) \) is the specific internal energy per unit mass.

(v) \( \eta = \eta(X, t) \) is the specific entropy per unit mass.

(vi) \( q = q(X, t) \) is the heat flux vector.

**Internal state variables**

(vii) \( \xi_k = \xi_k(X, t) \) are the internal state variables,
where \( k = 1, 2, \ldots \). We note that any order of tensors can be admitted at this time.

**Supplementary variables**

(viii) \( \rho = \rho(X, t) \) is the mass density.

(ix) \( r = r(X, t) \) is the heat supply per unit mass and unit time.

(x) \( b = b(X, t) \) is the specific body force per unit mass.

In addition to the aforementioned nine basic variables, the following auxiliary variables are also defined.

**Auxiliary variables**

\[ F = \frac{\partial X}{\partial \xi} \] defines the deformation gradient with respect to the referential position vector \( X \).

\[ g = \frac{\partial \theta}{\partial x} \] is the temperature gradient with respect to the spatial position vector \( x \).

Similarly, \( G = \frac{\partial \theta}{\partial \xi} = F^T g \) is the temperature gradient with
respect to $X$, where the superposed $T$ denotes the transpose.

Various definitions of strains, stresses and their rates are listed as follows:

**Strains and the associated strain rates**

$$E = \frac{1}{2} (F^T F - I)$$ is the Green strain tensor, where $I$ is an identity tensor.

$$e = \frac{1}{2} (I - F^{-T} F^{-1})$$ is the Almansi strain tensor, where $F^{-1}$ and $F^{-T}$ are the inverse and the inverse transpose of $F$.

Therefore,

$$E = F^T e F. \quad (A.1.1)$$

Also,

$$v = \dot{x}$$ is the velocity vector, where the superposed dot denotes the material time derivative.

$$L = \frac{\partial v}{\partial x}$$ is the velocity gradient.

$$D = \frac{1}{2} (L + L^T)$$ is the rate of deformation tensor.

$$W = \frac{1}{2} (L - L^T)$$ is the spin tensor.

From the above definitions,

$$\dot{F} = L F. \quad (A.1.2)$$

$$\dot{E} = F^T D F. \quad (A.1.3)$$

$$D = \dot{e} \equiv \dot{e} + L^T e + e L. \quad (A.1.4)$$

According to the polar decomposition theory, we have
\[ F = R U = V R, \] where \( U \) and \( V \) are the right and left stretch tensor, respectively, and \( R \) is a proper orthogonal rotation tensor.

Then, a rotational velocity \( \Omega \) is defined by
\[ \Omega = \dot{R} R^T. \]
where \( R^T = R^{-1} \).  \hfill (A.1.5)

**Stresses**

The Kirchhoff stress tensor is denoted by \( \tau = J T \),
where \( J = \frac{\rho_R}{\rho} = \det |F| \) is the determinant of \( F \), and \( \rho_R \) and \( \rho \) are the mass densities in the referential and current configurations, respectively.

The quantity \( P = \tau F^{-T} \) denotes the 1st kind Piola-Kirchhoff stress tensor, and
\[ S = F^{-1} \tau F^{-T} \] defines the 2nd kind Piola-Kirchoff stress tensor.

From the above definitions,
\[ J T = \tau = P F^T = F S F^T. \] \hfill (A.1.6)

**Various definitions of energy**

In addition to the specific internal energy \( \varepsilon \), the specific Helmholtz free energy per unit mass is given by
\[ \psi = \varepsilon - \theta \eta, \quad \text{(Canonical transformation)} \] \hfill (A.1.7)
and the corresponding complementary specific free energy per unit mass is also given by

\[ \rho R \phi = S \cdot E - \rho R \psi. \]  
(Legendre transformation)  \hspace{1cm} (A.1.8)

**Stress rates**

Several objective stress rates\(^\dagger\) are listed as follows:

**Jaumann rate** (or the co-rotational rate) is defined by [7]

\[ \nabla \cdot ( ) = ( ) - W ( ) + ( ) W. \]  
(A.1.9)

**Oldroyd rate** (or the convective rate) is defined by [10]

\[ \circ ( ) = ( ) - L ( ) - ( ) L^T. \]  
(A.1.10)

**Hill rate** is defined by [6]

\[ \nabla \cdot ( )^H = ( ) - W ( ) + ( ) W + ( ) \text{tr}(D). \]  
(A.1.11)

**Truesdell rate** is defined by [17]

\[ \circ ( )^T = ( ) - L ( ) - ( ) L^T + ( ) \text{tr}(D). \]  
(A.1.12)

and **Cotter-Rivlin rate** is defined by [3]

\[ \Lambda \circ ( )^{CR} = ( ) + L^T ( ) + ( ) L. \]  
(A.1.13)

where \( \text{tr}(D) \) is the trace of \( D \).

It can be shown that

\[ \circ \tau = F \, \hat{S} \, F^T \]  
(A.1.14)

\[ = \overline{\tau} - D \, \tau - \tau \, D. \]  
(A.1.15)

\(^\dagger\) In the present work, objectivity implies invariance under superposed rigid body motion. (Refer to the general discussions in Ref.[18])
Also,

\[ \tau^0 = J \tau^T \]  \hspace{1cm} (A.1.16)

and

\[ \tau^I = J \tau^H . \]  \hspace{1cm} (A.1.17)

It is well known that use of the Jaumann rate in simple shear large deformation problems results in an unrealistic oscillatory shear stress response \[8,9\]. Recently, two modified Jaumann rates \( \tau \) have been proposed, i.e.,

\[ \nabla \tau^{GN} = (•) - \Omega (•) + (•) \Omega , \]  \hspace{1cm} (A.1.18)

and

\[ \nabla \tau^L = (•) - W'(•) + (•) W' . \]  \hspace{1cm} (A.1.19)

by Green and Naghdi [5] and Lee et al. [8], respectively. In the above, \( W' = W + Dn n^T - nn^T D \), and \( n \) is the unit vector in the direction of the largest absolute principal value of \( \tau \). Fig. A.1 illustrates a sinusoidal stress-strain relation for the case of the Jaumann rate, compared with two other stress rates i.e. the convective and the Green-Naghdi rates in simple shear for a hypoelastic material defined by,

\[ \tau^{*}_{ij} = [2\mu \delta_{ia}\delta_{ib} + \lambda \delta_{ij}\delta_{ab}] D_{ab} . \]  \hspace{1cm} (A.1.20)

where \( \mu \) and \( \lambda \) are the Lame elastic constants and superscript \( * \) denotes an appropriate stress rate. Fig. A.2 illustrates the stress-strain relations in uniaxial tension for the same material. In uniaxial tension, both the Jaumann and the
Green-Naghdi rates yield the same result.

Despite such an unrealistic property of the Jaumann rate, the Jaumann rate or the modified Jaumann rates are still predominantly employed for most plasticity problems. This preference for the Jaumann rate over other stress rates stems from the fact that the invariants of a stress tensor become stationary when this stress rate vanishes [14]. In the subsequent section, we advance a different proposition on the requirements of stress rates for large deformation problems.

A.2. **Discussion on the stress rate definition requirements**

Prager [14] has shown that only the Jaumann rate, among other stress rates, satisfies the requirement that the invariants of a stress tensor must be stationary (i.e., more specifically, the yield function must be stationary) when an objective rate of the stress vanishes. However, it has now been shown that the convective rate of the Kirchhoff stress tensor, which is the same as the Truesdell rate of the Cauchy stress tensor, also satisfies the above restriction with the inclusion of constitutive relations.

To show this, we consider the following general constitutive relations for elastoplastic materials with isotropic hardening:

\[ D_{ij} = D_{ij}^e + D_{ij}^p \]  \hspace{1cm} (A.2.1)
Equations (A.2.1) through (A.2.4) represent the additive decomposition of the rate of deformation tensor into elastic and plastic parts, the elastic constitutive relation, the plastic constitutive relation, and the loading function, respectively, with the yield criterion

\[ f = k(\tau^P), \quad \text{(A.2.5)} \]

and the associated hardening law

\[ \dot{f} = h \frac{\dot{\tau}^P}{\tau^P}. \quad \text{(A.2.6)} \]

In the above equation, \( \tau^P \) denotes the deviatoric part of the Kirchhoff stress tensor, and the effective plastic shear strain \( \tau^P \) is defined by

\[ \tau^P \equiv \int_0^t [2 D^P_{ij} D^P_{ij}]^{1/2} \, dt, \quad \text{(A.2.7)} \]

with

\[ h = \frac{\partial k}{\partial \tau^P}. \quad \text{(A.2.9)} \]

denoting the slope of the shear stress-plastic shear strain curve.

After some manipulations, (A.2.3) can be rewritten as

\[ D^P_{ij} = \frac{\tau^P_{ij} \tau^P_{kl}}{4h k^2} \dot{\tau}_{kl}. \quad \text{(A.2.10)} \]
Noting that

\[ \tau'_{kl}^2 = \tau'_{kl}^2 + 2 \tau'_{kl} \tau'_m D_{km} \]  

(A.2.11)

we get

\[ D_{ij} = \left[ \frac{1}{2\mu} \left\{ \delta_{ij} \delta_{kl} - \frac{\lambda}{2\mu + 3\lambda} \delta_{ij} \delta_{kl} \right\} \right] \tau^2_{kl} + \frac{\tau'_{ij} \tau'_{kl}}{4 h k^2} \tau'_m D_{km} \]  

(A.2.12)

for \( \tau = \tau^\circ \).

For \( \tau = 0 \),

\[ D_{ij} = \frac{\tau'_{ij} \tau'_{kl}}{4 h k^2} \tau'_m D_{km} \]  

(A.2.13)

Since (A.2.13) holds for any \( \tau \),

\[ D = 0 \quad \text{when} \quad \tau^\circ = 0. \]

This implies that \( \tau^\circ = \tau \) when \( \tau = 0 \). Thus, the invariants of \( \tau \) also become stationary when the convective stress rate of \( \tau \) vanishes.

Unlike the Jaumann stress rate, the convective stress rate and other rates do or do not satisfy the restriction, depending on the specified constitutive relations. As shown above, the convective stress rate satisfies the restriction for an elasto(inviscid)plastic material. For thermo or viscoplastic materials, the above restriction need not be satisfied because the yield function is not stationary even when the stress rate vanishes.

Besides the above restriction, in the present work, it
is shown that the convective rate of the Kirchhoff stress is thermodynamically conjugate to the rate of deformation tensor \( D \) through the Legendre transformation of energy in (A.1.8). The following relationships between inner products and their time derivatives of the pairs of stress and strain tensors can be established:

\[
S \cdot E = \tau \cdot e , \quad (A.2.14)
\]

\[
\dot{S} \cdot \dot{E} = \dot{S} \cdot E + S \cdot \dot{E} = \dot{\tau} \cdot e + \dot{\tau} \cdot \dot{e} = \dot{\tau} \cdot e , \quad (A.2.15)
\]

\[
\dot{S} \cdot E = \dot{\tau} \cdot e , \quad \text{and} \quad S \cdot \dot{E} = \tau \cdot D . \quad (A.2.16)
\]

Thus,

\[
\dot{S} \cdot E + S \cdot \dot{E} = \dot{\tau} \cdot e + \tau \cdot D . \quad (A.2.17)
\]

Eq (A.2.16) defines that the relationship between the stress power in terms of tensors invariant under superposed rigid body motion. In other words, once we select stress and strain tensors, namely, \( S, E, \tau, \) and \( e, \) the only admissible objective stress rate conjugate to \( D \) is the convective rate of the Kirchhoff stress tensor.

**A.3. Thermodynamic process**

The local balance laws can be derived from the corresponding integral forms under sufficient smoothness assumptions, i.e.,

\[
\dot{\rho} + \rho \, v \cdot v = 0 , \quad (A.3.1)
\]

\[
v \cdot T + \rho \, b = 0 , \quad (A.3.2)
\]
T = T^T, \quad (A.3.3)
\rho \dot{e} = T \cdot D + \rho r - v \cdot q, \quad (A.3.4)

representing the mass, linear and angular momentum, and energy balance laws in the spatial description, respectively, where \cdot and \nabla \cdot denote the inner product and the divergence with respect to the spatial position vector \( x \), respectively.

A set of ten state variables, introduced in the preceding section, defined for all material particle \( X \) in a body \( B \) and time \( t \) is called a thermodynamic process whenever it satisfies the balance laws in (A.2.1-4).

A set of constitutive equations is then provided for the characterization of a material as follows:

\[ T = \hat{T}(\Psi), \quad (A.3.5) \]
\[ \varepsilon = \varepsilon(\Psi), \quad (A.3.6) \]
\[ \eta = \hat{\eta}(\Psi), \quad (A.3.7) \]
\[ q = \hat{q}(\Psi), \quad (A.3.8) \]

where \( \Psi = \{ F, \theta, g, \xi_k \} \) denotes a set of arguments, and the superposed caret denotes the various quantities above as functionals of \( \Psi \). In addition, the evolution of internal state variables is described by the following set of rate equations provided for each \( \xi_k \):

\[ \dot{\xi}_k = \hat{\xi}_k(\Psi). \quad (A.3.9) \]

It is worth noting that the above constitutive assumptions
satisfy the principle of equipresence postulated by Truesdell [18]. A thermodynamic process is called as an admissible process when it is compatible with the preceding constitutive equations (A.3.5-9).

Additionally, it is required that the thermodynamic process must satisfy a thermodynamic postulate expressed in the form of the Clausius-Duhem inequality:

\[ \rho \theta \dot{\eta} - \rho \dot{\varepsilon} + T \dot{D} - \frac{1}{\theta} \mathbf{q} \cdot \mathbf{g} \geq 0. \]  \hspace{1cm} (A.3.10)

or, equivalently,

\[ \rho \theta \dot{\eta} - \rho \dot{\varepsilon} + T \dot{D} - \frac{1}{\theta} \mathbf{q} \cdot \mathbf{g} \geq 0. \]  \hspace{1cm} (A.3.11)

The constitutive equations are thus restricted to satisfy the above inequality in the subsequent section.

A.4. General restrictions on constitutive equations

Using the canonical transformation of energy in (A.1.7), Eq (A.3.11) can be rewritten as

\[ - \rho \ddot{\psi} - \rho \dot{\theta} \eta + T \dot{D} - \frac{1}{\theta} \mathbf{q} \cdot \mathbf{g} \geq 0. \]  \hspace{1cm} (A.4.1)

From the relationship \( \psi = \psi(\mathbf{u}) = \psi(F, \theta, g, \xi_k) \),

\[ \dot{\psi} = \frac{\partial \psi}{\partial F} \dot{F} + \frac{\partial \psi}{\partial \theta} \dot{\theta} + \frac{\partial \psi}{\partial g} \dot{g} + \frac{\partial \psi}{\partial \xi_k} \dot{\xi}_k \].  \hspace{1cm} (A.4.2)

Substituting (A.4.2) into (A.4.1), we get

\[ (T - \rho \frac{\partial \psi}{\partial F} \dot{F}) \cdot \mathbf{L} - \rho \left( \frac{\partial \psi}{\partial \theta} + \eta \right) \dot{\theta} - \rho \frac{\partial \psi}{\partial g} \cdot \dot{g} \]

\[ - \rho \frac{\partial \psi}{\partial \xi_k} \dot{\xi}_k - \frac{1}{\theta} \mathbf{q} \cdot \mathbf{g} \geq 0 \].  \hspace{1cm} (A.4.3)
By choosing arbitrary values for \( L, \dot{\theta} \) and \( \dot{g} \), the following relations for an admissible thermodynamic process are derived:

\[
T = \rho \frac{\partial \hat{\psi}}{\partial F} F^T, \tag{A.4.4}
\]

\[
\eta = - \frac{\partial \hat{\psi}}{\partial \theta}, \tag{A.4.5}
\]

\[
\frac{\partial \hat{\psi}}{\partial g} = 0, \quad \text{and} \tag{A.4.6}
\]

\[
\rho \frac{\partial \hat{\psi}}{\partial \xi_k} \xi_k + \frac{1}{\theta} \hat{q} \cdot \mathbf{g} \leq 0. \tag{A.4.7}
\]

From (A.4.6),

\[
\hat{\psi} = \hat{\psi}(F, \theta, \xi_k). \tag{A.4.8}
\]

For each internal state variable \( \xi_k \), the associated thermodynamically conjugate force is defined by

\[
f_k = - \frac{\partial \hat{\psi}}{\partial \xi_k}. \tag{A.4.9}
\]

Also, an internal dissipation power function \( \sigma \) is defined by

\[
\sigma = \hat{\sigma}(F, \theta, g, \xi_k) = \rho f_k \cdot \xi_k. \tag{A.4.10}
\]

From (A.4.7), we obtain the heat conduction inequality

\[
\hat{q} \cdot \mathbf{g} \leq 0 \quad \text{when } \sigma = 0, \tag{A.4.11}
\]

and the internal dissipation inequality

\[
\sigma \leq 0 \quad \text{when } g = 0. \tag{A.4.12}
\]

We now reduce the stress-strain relation by applying the invariance requirement under superposed rigid body
motion. Eq (A.4.4) can be rewritten as

$$T = \rho F \frac{\partial \psi}{\partial E} F^T,$$  \hspace{1cm} (A.4.13)

where $\psi = \tilde{\psi}(E, \theta, \xi_k)$. From (A.1.6), i.e., $J T = F S F^T$,

$$S = \rho R \frac{\partial \tilde{\psi}}{\partial E} = \frac{\partial \bar{\psi}}{\partial E}.$$  \hspace{1cm} (A.4.14)

where $\bar{\psi} = \rho R \tilde{\psi}$. Also, (A.4.14) can be transformed into the following spatial form:

$$\tau = \frac{\partial \bar{\psi}}{\partial e},$$  \hspace{1cm} (A.4.15)

where $\psi = \bar{\psi}(e, \theta, \xi_k)$ and the following relations hold

$$\frac{\partial}{\partial E_{ij}} = \frac{\partial}{\partial e_{kl}} \frac{\partial e_{kl}}{\partial E_{ij}}$$

$$= \frac{\partial}{\partial e_{kl}} \frac{\partial (F^{-1}_{mk} E_{mn} F^{-1}_{nl})}{\partial E_{ij}}$$

$$= \frac{\partial}{\partial e_{kl}} F^{-1}_{ik} \bar{F}^{-1}_{jl}.$$  \hspace{1cm} (A.4.16)

It is assumed that the internal state variables are chosen so as to meet the invariance requirement.

Now consider the Legendre transformation in (A.1.8),

$$\rho R \psi = S^E - \rho R \phi.$$  \hspace{1cm} (A.1.8)

where $\psi = \tilde{\psi}(E, \theta, \xi_k)$ and $\phi = \tilde{\phi}(S, \theta, \xi_k)$. By taking admissible variations for (A.1.8), $\psi$ and $\phi$, we have

$$\rho R \delta \psi = \delta S^E + S \delta E - \rho R \delta \phi,$$  \hspace{1cm} (A.4.17)

$$\delta \psi = \frac{\partial \tilde{\psi}}{\partial E} \delta E + \frac{\partial \tilde{\psi}}{\partial \theta} \delta \theta + \frac{\partial \tilde{\psi}}{\partial \xi_k} \delta \xi_k.$$  \hspace{1cm} (A.4.18)
respectively. From (A.4.17-19), we get

\[
(\rho_R \frac{\partial \phi}{\partial E} - S) \delta E + (\rho_R \frac{\partial \phi}{\partial S} - E) \delta S + \rho_R (\frac{\partial \phi}{\partial \theta} + \frac{\partial \phi}{\partial \xi_k}) \delta \theta
\]

\[+ \rho_R (\frac{\partial \phi}{\partial \xi_k} + \frac{\partial \phi}{\partial \xi_k}) \delta \xi_k = 0. \tag{A.4.20}
\]

Therefore,

\[
E = \frac{\partial \phi}{\partial S}, \quad \frac{\partial \psi}{\partial \theta} = - \frac{\partial \phi}{\partial S}, \quad \text{and} \quad \frac{\partial \psi}{\partial \xi_k} = - \frac{\partial \phi}{\partial \xi_k}. \tag{A.4.21}
\]

where \( \phi = \rho_R \phi \).

By introducing the relations in Eqs. (A.4.21), (A.4.4, 5, and 9) can be written as

\[
E = \frac{\partial \phi}{\partial S}, \tag{A.4.22}
\]

\[
\eta = \frac{\partial \phi}{\partial \theta}. \tag{A.4.23}
\]

and

\[
f_k = \frac{\partial \phi}{\partial \xi_k}, \tag{A.4.24}
\]

respectively. Also, the evolution equations for each \( \xi_k \) in (A.3.9) become

\[
\dot{\xi}_k = \xi_k (E(S, \theta, \xi_n), \theta, g, \xi_m)
\]

\[= \xi_k (S, \theta, g, \xi_m). \tag{A.4.25}
\]

By taking the time derivative for (A.4.22),
\[ \dot{\mathbf{E}} = \frac{\partial^2 \phi}{\partial \mathbf{S} \partial \mathbf{S}} \mathbf{\dot{S}} + \frac{\partial^2 \phi}{\partial \mathbf{S} \partial \mathbf{\theta}} \mathbf{\dot{\theta}} + \frac{\partial^2 \phi}{\partial \mathbf{S} \partial \xi_k} \mathbf{\dot{\xi}_k}. \] (A.4.26)

The above equation implies that the rate of deformation is comprised of the sum of the contributions arising from the changes in stress, temperature, and internal states.

The internal dissipation power function in (A.4.10) can now be written as,

\[ \sigma = \tilde{\sigma}(\mathbf{S}, \mathbf{\theta}, g, \xi_k) = \rho \mathbf{f}_k \cdot \mathbf{\dot{\xi}_k}. \] (A.4.27)

where \( f_k = \frac{\partial \tilde{\sigma}}{\partial \xi_k} \).

We let

\[ \rho_R \mathbf{\Sigma} = J \tilde{\sigma} = \rho_R \mathbf{f}_k \cdot \mathbf{\dot{\xi}_k}, \] (A.4.28)

and take its variational form,

\[ \delta \mathbf{\Sigma} = \mathbf{f}_k \cdot \delta \mathbf{\dot{\xi}_k} + \delta \mathbf{f}_k \cdot \mathbf{\dot{\xi}_k}. \] (A.4.29)

where

\[ \delta I \equiv \mathbf{f}_k \cdot \delta \mathbf{\dot{\xi}_k} \quad \text{and} \quad \delta \Omega \equiv \delta \mathbf{f}_k \cdot \mathbf{\dot{\xi}_k} \]

are denoted as the virtual power of the internal dissipation and the associated complementary power, respectively. From the relation \( f_k = f_k(\mathbf{S}, \mathbf{\theta}, \xi_k) \) in (A.4.27),

\[ \delta f_k = \frac{\partial f_k}{\partial \mathbf{S}} \delta \mathbf{S} + \frac{\partial f_k}{\partial \mathbf{\theta}} \delta \mathbf{\theta} + \frac{\partial f_k}{\partial \xi_m} \delta \xi_m, \] (A.4.30)

and

\[ \delta \Omega = \mathbf{\dot{\xi}_k} \cdot \delta f_k \] (A.4.31)

\[ = (\mathbf{\dot{\xi}_k} \cdot \frac{\partial f_k}{\partial \mathbf{S}}) \delta \mathbf{S} + (\mathbf{\dot{\xi}_k} \cdot \frac{\partial f_k}{\partial \mathbf{\theta}}) \delta \mathbf{\theta} + (\mathbf{\dot{\xi}_k} \cdot \frac{\partial f_k}{\partial \xi_m}) \delta \xi_m. \]
Introducing the definition of the Frechet derivative, we get

$$\frac{\partial \Omega}{\partial S} = \frac{\partial f_k}{\partial S} \dot{S}_k = \frac{\partial^2 \phi}{\partial S \partial \theta} \dot{\theta}. \quad (A.4.32)$$

Then, (A.4.26) can be written as

$$\dot{E} = \frac{\partial^2 \phi}{\partial S \partial \theta} \dot{S} + \frac{\partial^2 \phi}{\partial S \partial \theta} \dot{\theta} + \frac{\partial \Omega}{\partial S}. \quad (A.4.33)$$

where $\bar{\Omega} = \rho R \Omega$, or in indicial form,

$$\dot{E}_{ij} = \frac{\partial^2 \phi}{\partial S_{ij} \partial \theta} \dot{S}_{kl} + \frac{\partial^2 \phi}{\partial S_{ij} \partial \theta} \dot{\theta} + \frac{\partial \bar{\Omega}}{\partial S_{ij}}. \quad (A.4.34)$$

It is worthy of note that (A.4.33) has the identical form derived by Rice [15,16] starting from a different thermodynamic approach [4], where $\bar{\Omega}$ is called the flow potential.

In view of the fact that the response of a material is independent of a particular choice of reference states, it is desirable to express Eq (A.4.33) in terms of spatial variables. From the following relations,

$$D_{ij} = F^{-1}_{ji} \dot{E}_{kl} F^{-1}_{lj}, \quad (A.4.35)$$

$$\dot{S}_{ij} = F^{-1}_{ik} \dot{\tau}_{kl} F^{-1}_{jl}, \quad (A.4.36)$$

and

$$\frac{\partial}{\partial S_{ij}} = \frac{\partial}{\partial \tau_{kl}} \frac{\partial \tau_{kl}}{\partial S_{ij}} = \frac{\partial}{\partial \tau_{kl}} F_{ki} F_{lj}. \quad (A.4.37)$$

We can write (A.4.34) as

$$D_{ij} = \frac{\partial^2 \phi}{\partial \tau_{ij} \partial \tau_{kl}} \dot{\tau}_{kl} + \frac{\partial^2 \phi}{\partial \tau_{ij} \partial \theta} \dot{\theta} + \frac{\partial \bar{\Omega}}{\partial \tau_{ij}}.$$
In the above expression, \( C_{ijkl} \) is the elasticity tensor, \( B_{ij} \) is the thermal expansion tensor, and \( D_{ij}^P \) is the inelastic part of the rate of deformation.

Similarly, from (A.4.23) and (A.4.24), we obtain

\[
\dot{\eta} = \frac{\partial^2 \dot{\phi}}{\partial \eta \partial \theta} \dot{\phi} + \frac{\partial^2 \dot{\phi}}{\partial \theta \partial \phi} \dot{\phi} + \frac{\partial^2 \dot{\phi}}{\partial F_k \partial \theta} \dot{F}_k
\]

\[
= \frac{B}{\rho \theta} \dot{\phi} \dot{\phi} + \frac{c_v}{\theta} \dot{\phi} \dot{\phi} + \frac{\eta^p}{\theta}, \quad (A.4.39)
\]

where \( B \) is the thermal expansion tensor, \( c_v \) is the specific heat capacity, and \( \eta^p \) is the rate of entropy change due to the internal dissipation. The heat conduction equation (A.3.8) is usually assumed by Fourier's law,

\[
q = -K \mathbf{g}, \quad (A.4.40)
\]

which always satisfies the heat conduction inequality,

\[
q \cdot \mathbf{g} \leq 0,
\]

provided that \( K \) is a positive definite tensor. Finally, the evolution equations (A.4.25) for the internal state variables are taken so as to satisfy the internal dissipation inequality (A.4.12).

In summary, the admissible thermodynamic process for a set of functions \{x, \theta, T, \phi, \eta, q, p, r, b, F_k\} defined for every material particle \( X \) and time \( t \) should satisfy:

(i) the balance laws expressed by (A.3.1-4),
(ii) the constitutive equations (A.4.38, 39, and 40),
(iii) the heat conduction inequality (A.4.11) and the internal dissipation inequality ((A.4.12), and
(iv) the evolutionary laws (A.4.25).
B. A Special Case of Characterization of Constitutive Equations for Thermo-Elasto-Viscoplastic Materials

As an illustrative example, the developed formalism for constitutive modeling with finite strain is applied to the constitutive relations for a thermo-elasto-viscoplastic material based on the Perzyna's viscoplastic theory [11-13] with isotropic hardening.

B.1. Characterization of thermoelastic parts

From (A.4.38), for a homogeneous and isotropic material

\[ D_{ij} = C_{ijkl} \frac{\partial}{\partial t} \delta_{kl} + \frac{B_{ij}}{\theta} \delta + D^P_{ij} \]  

(B.1.1)

where the elasticity tensor is given by

\[ C_{ijkl} = \frac{1}{2\mu} (\delta_{ik} \delta_{jl} - \frac{\lambda}{3\lambda + 2\mu} \delta_{ij} \delta_{kl}) \]  

(B.1.2)

the normalized thermal expansion tensor is given by

\[ B_{ij} = \alpha \delta_{ij} \]  

(B.1.3)

\( \mu \) and \( \lambda \) are Lame constants, and \( \alpha \) is the thermal expansion coefficient. It is noted that

\[ \int_{t_R}^{t} D_{ij}^{\theta} \, dt = \alpha \delta_{ij} \ln \left( \frac{\theta}{\theta_R} \right). \]

From (A.4.39),
B.2. Characterization of viscoplastic parts

It is assumed that the dissipative behavior for the present model can be described with a scalar internal state variable \( \kappa \). The thermodynamically conjugate force \( f \) is defined by

\[
f = f(\tau) = \sqrt{J_2},
\]

where

\[
J_2 = \frac{1}{2} \tau_{ij} \tau_{ij},
\]

and

\[
\tau_{ij} = \tau_{ij} - \frac{1}{3} \delta_{ij} \tau_{kk}.
\]

Also, the evolution equation for \( \kappa \) is defined by

\[
\dot{\kappa} = \kappa(\tau, \theta, \kappa) = \gamma \langle \phi(F) \rangle,
\]

where

\[
\gamma = \gamma(\theta) \text{ is a viscosity constant which strongly depends on the temperature,}
\]

\[
F = F(\tau, \kappa) = \frac{f(\tau)}{\kappa} - 1, \text{ is a static yield function following the Perzyna's viscoplastic theory,}
\]

and

\[
\langle \phi(F) \rangle = \begin{cases} 
0 & \text{when } F \leq 0 \\
\phi(F) & \text{when } F > 0
\end{cases}
\]

The above choice of the static yield function coincides with von Mises yield criterion and \( \phi(F) \) is defined from the results of dynamic loading tests.
Thus,
\[ D_{ij}^P = \gamma \langle \Phi(F) \rangle \frac{\partial f}{\partial \tau_{ij}} \]
\[ = \gamma \langle \Phi(F) \rangle \frac{\tau_{ij}}{2\sqrt{J_2}}. \]  \hspace{1cm} (B.2.3)

and
\[ D_{ii}^P = 0 \] (the plastic incompressibility condition).

From (B.2.3), it can be shown that
\[ D_{ij}^P D_{ij}^P = \frac{1}{2} k^2, \]
or
\[ \dot{k} = \sqrt{I_2}, \]
where
\[ I_2 = 2 D_{ij}^P D_{ij}^P. \]
Thus, \[ f \dot{k} = \sqrt{J_2} \sqrt{I_2}. \] \hspace{1cm} (B.2.4)

If we consider \( k \) as an integrity of normalized dislocation movement, \( f \) behaves as a thermodynamic force acting on the dislocation which depends largely on the stress field, but \( \dot{k} \) depends on temperature (concerning short-range obstacles), stress (long-range obstacles), and \( k \) (an accumulated pattern of dislocation density).

Since \( f \) is assumed to be a function of stress only, the constitutive equations for the present model can be characterized as follows:
\[ D_{ij} = D_{ij}^e + D_{ij}^0 + D_{ij}^P, \] \hspace{1cm} (B.2.5)

where
\[ D_{ij}^e = \frac{1}{2\mu} \left( \delta_{ik}\delta_{jl} - \frac{\lambda}{3\lambda+2\mu} \delta_{ij}\delta_{kl} \right) \tilde{\tau}_{kl}, \]

\[ D_{ij}^\theta = \alpha \delta_{ij} \dot{\theta}, \]

and

\[ D_{ij}^p = \gamma \langle \phi(F) \rangle \frac{\tilde{\tau}_{ij}}{2\sqrt{J_2}}. \]

Also,

\[ \dot{\eta} = \frac{\alpha}{\rho_R} \frac{\phi}{\theta} + \frac{c_v}{\theta} \dot{\theta}, \quad \text{(B.2.6)} \]

and

\[ \dot{\bar{f}} = \frac{\tilde{\tau}_{ij} \tilde{\tau}_{ij}}{2\sqrt{J_2}}. \quad \text{(B.2.7)} \]

The energy balance equation can then be written in the form

\[ \rho_R \dot{\eta} \frac{\phi}{\theta} = \rho_R \frac{r - \text{Div} q_R + f \kappa}{\theta}, \quad \text{in } \Omega_R. \quad \text{(B.2.8)} \]

or

\[ \rho_R c_v \dot{\theta} + \alpha \frac{\phi}{\theta} \tilde{\tau}_{kk} = \text{Div} (K \text{ grad } \theta) + \rho_R \frac{r + \sqrt{J_2} \sqrt{T_2}}{\theta}. \quad \text{(B.2.9)} \]

where the last term \( \sqrt{J_2 T_2} \) represents the heat generation due to the internal dissipation mechanism (here, the rate of plastic work-done).
REFERENCES for APPENDIX


Fig. A.1 A comparison of stress rates in simple shear deformation
Fig. A.2 A comparison of stress rates in uniaxial tension