# CHAPTER 10 CONTACT-IMPACT 

by Ted Belytschko<br>Northwestern University<br>Copyright 1996<br>\section*{X. 1 INTRODUCTION}

This Chapter introduces the treatment of problems with contact and impact. Many problems in the simulation of prototype tests and manufacturing processes involve contact and impact. For example, in the simulation of a drop test on a product, the various parts must be separated by so-called sliding interfaces which can model contact, sliding and separation. In the simulation of manufacturing processes, sliding interfaces are also important: the modeling of the surfaces between the die and workpiece in sheet metal forming, the modeling of the tool-workpiece interface in machining, the modeling of extrusion are some examples of where sliding interfaces are needed. In crashworthiness simulation of automobiles, many components, including the engine, wheels, radiator, etc. can contact during the crash and their surfaces automatically must be treated as sliding interfaces. The treatment of impact always requires a subsequent treatment of contact, since bodies which impact will stay in contact until rarefaction waves result in release.

In this Chapter, the governing equations and finite element procedures for problems with contact-impact are introduced for Lagrangian meshes; the modeling of contact with Eulerian meshes introduces difficulties which have not been resolved yet. The governing equations for bodies in contact are identical to the equations introduced previously, except that it is necessary to add the kinetic and kinematic conditions on the contact interface. The key condition is the condition of impenetrability: namely, the condition that two bodies cannot interpenetrate. The general condition of impenetrability cannot be expressed as a useful equation, so several approaches to developing specialized forms of these conditions have evolved. We will consider two of these forms: a rate form which is useful for explicit dynamics methods and a form based on closest point projection; the latter is primarily useful for implicit methods. Friction is treated by both the classical Coulomb friction models and by interface constitutive models wherein the tangential tractions are developed through constitutive laws in terms of relative normal and tangential velocities of the interface.

Next, the weak forms of the governing equations are developed. Four approaches to treating the contact surface conditions are considered:

1. the Lagrange multiplier method;
2. the penalty method;
3. the augmented Lagrangian method;
4. the perturbed Lagrangian method.

The weak form for contact-impact for the Lagrange multiplier methods differs from the weak form for single bodies in that they are inequalities; they are often called weak
inequalities or variational inequalities. In penalty methods, these inequalities are incorporated by means of the Heaviside step function. It will be shown that the weak forms are equivalent to the strong forms.

The discretization of contact problems is similar to problems without contact except that in Lagrange multiplier methods, the Lagrange multiplier fields must be approximated. The Lagrange multiplier fields are constrained fields which must observe the inequality that they be nonnegative across the contact interface (the sign of the inequality depends on the structure of the weak form; the Lagrange multipliers may also be contrained to be nonpositive). These constraints on the Lagrange multiplier ultimately imply the constraint that the normal tractions be compressive. In penalty methods, the traction inequalities emerge from the Heaviside step function which is embedded in the penalty force.

Contact-impact problems are among the most difficult nonlinear problems because the response in contact-impact problems is not smooth. The velocities normal to the contact interface are discontinuous in time when impact occurs. When Coulomb friction models are used, the tangential velocities along the interface are discontinuous when stick-slip behavior is encountered. These characteristics of contact-impact problems introduce significant difficulties in the time integration of the governing equations and impair the performance of numerical algorithms. Therefore, the appropriate choice of methodologies and algorithms is crucial in the successful treatment of these problems. Techniques such as regularization are highly useful in obtaining robust solution procedures, but the analyst must understand their effect so that important aspects of the response are not eliminated.

The implementation of contact-impact for general models is quite difficult. In our discussion of implementation, we will begin with the simplest examples, one dimensional problems, which illustrate how the contact inequalities are imposed. We will then sketch some of the difficulties that arise in large-scale multidimensional problems, but we will not dwell on these since many of the approaches are based more on heuristics and computer science than computational mechanics.

## X. 2 CONTACT INTERFACE EQUATIONS

X.2.1. Notation and Preliminaries. Contact-impact algorithms in general purpose software can treat the interaction of many bodies, but for purposes of simplicity, we limit ourselves to two bodies as illustrated in Fig. 1. The treatment of multi-body contact is identical: the interaction of any pair of bodies is exactly like the two body problem. We have denoted the configurations of the two bodies by $\Omega^{A}$ and $\Omega^{B}$ and denote the union of the two bodies by $\Omega$. The boundaries of the bodies are denoted by $\Gamma^{A}$ and $\Gamma^{B}$. Although the two bodies are interchangeable with respect to their mechanics, it is sometimes useful to express the equations in term of one of the bodies, which is called the master; body A is designated as the master, body B as the slave. When we wish to distinguish field variables that are associated with a particular body, we append a superscript $A$ or $B$; when neither of these superscripts appears, the field variable applies to the union of the two bodies. Thus the velocity field $\mathbf{v}(\mathbf{X}, t)$ refers to the velocity field in both bodies, whereas $\mathbf{v}^{A}(\mathbf{X}, t)$ refers to the velocity in body $A$.

The contact interface consists of the intersection of the surfaces of the two bodies and is denoted by $\Gamma^{c}$.

$$
\begin{equation*}
\Gamma^{c}=\Gamma^{A} \cap \Gamma^{B} \tag{X.2.1}
\end{equation*}
$$

This contact interface consists of the two physical surfaces of the two bodies which are in contact, but since they are theoretically coincident we refer to a single interface $\Gamma^{c}$. In numerical solutions, the two surfaces will usually not be coincident. In those cases, $\Gamma^{c}$ refers to the master surface. Moreover, although the two bodies may be in contact on several disjoint interfaces, we designate their union by a single symbol $\Gamma^{c}$. The contact interface is a function of time, and its determination is an important part of the solution of the contact-impact problem.


Figure

1. Model problem for contact-impact showing notation.

In constructing the equations, it is convenient to express vectors in terms of local components of the contact surface. A local coordinate system is set up at each point of the master contact surface as shown in Fig. 2. At each point, we can construct unit vectors tangent to the surface of the master body $\hat{\mathbf{e}}_{1}^{A} \equiv \hat{\mathbf{e}}_{x}^{A}$ and $\hat{\mathbf{e}}_{2}^{A} \equiv \hat{\mathbf{e}}_{y}^{A}$. The procedure for obtaining these unit vectors is identical to that used in shell elements, see Chapter 8. The normal for body $A$ is given by

$$
\begin{equation*}
\mathbf{n}^{A}=\hat{\mathbf{e}}_{1}^{A} \times \hat{\mathbf{e}}_{2}^{A} \tag{X.2.2}
\end{equation*}
$$

On the contact surface

$$
\begin{equation*}
\mathbf{n}^{A}=-\mathbf{n}^{B} \tag{X.2.3}
\end{equation*}
$$

that is, the normals of the two bodies are in opposite directions.


Figure 2. Contact interface showing local unit vectors referred to master surface $A$.

The velocity fields can be expressed in the local coordinates of the contact surface by

$$
\begin{align*}
& \mathbf{v}^{A}=v_{N}^{A} \mathbf{n}^{A}+\hat{v}_{\alpha}^{A} \hat{\mathbf{e}}_{\alpha}^{A}=v_{N}^{A} \mathbf{n}^{A}+\mathbf{v}_{T}^{A}  \tag{X.2.4a}\\
& \mathbf{v}^{B}=v_{N}^{B} \mathbf{n}^{A}+\hat{v}_{\alpha}^{B} \hat{\mathbf{e}}_{\alpha}^{A}=v_{N}^{A} \mathbf{n}^{B}+\mathbf{v}_{T}^{A} \tag{X.2.4b}
\end{align*}
$$

where the range of Greek subscripts is 2 in three dimensional problems. When the problem is two dimensional, the contact surface becomes a line, so we have a single unit vector $\hat{\mathbf{e}}_{1} \equiv \hat{\mathbf{e}}_{x}$ tangent to this line; the range of the Greek subscripts in (4) is then one and the tangential component is a scalar. As can be seen in the above, the components are
expressed in terms of the local coordinate system of the master surface. The normal velocities are given by

$$
\begin{equation*}
v_{N}^{A}=\mathbf{v}^{A} \cdot \mathbf{n}^{A} \quad v_{N}^{B}=\mathbf{v}^{B} \cdot \mathbf{n}^{A} \tag{X.2.5}
\end{equation*}
$$

which can easily be seen by taking the dot product of the expressions in (4) with $\mathbf{n}^{A}$ and using the fact that the normal is orthogonal to the unit vectors tangent to the plane $\hat{\mathbf{e}}_{i}^{A}$.

The bodies are governed by the standard field equations given in Boxes 4.1 and 5.1: conservation of mass, momentum and energy, a strain measure, and the constitutive equations. Contact adds the following conditions: the bodies can not interpenetrate and the tractions must satisfy momentum conservation on the interface. Furthermore, the normal traction across the contact interface cannot be tensile. We classify the requirements on the displacements and velocities as kinematic conditions and the requirements on the tractions as kinetic conditions.
X.2.2. Impenetrability Condition. In a multi-body problem, the bodies must observe the impenetrability condition. The impenetrability condition for a pair of bodies can be stated as

$$
\begin{equation*}
\Omega^{A} \cap \Omega^{B}=0 \tag{X.2.6}
\end{equation*}
$$

that is, the intersection of the two bodies is the null set. In other words, the two bodies are not allowed to overlap, which can also be viewed as a compatibility condition. The impenetrability condition is highly nonlinear for large displacement problems, and in general cannot be expressed as an algebraic or differential equation in terms of the displacements. The difficulty arises because in an arbitrary motion it is impossible to anticipate which points of the two bodies will contact. For example, in Fig. 1, if the bodies are spinning, it is possible for point $P$ to contact point $Q$, whereas a different relative motion can result in contact of point $P$ with point $S$. Consequently, an equation which expresses the fact that point $P$ does not penetrate body $A$ cannot be written except in general terms such as (6).

Because it is not feasible to express Eq. (6) in terms of the displacements, it is convenient to express the impenetrability equations in rate form or incremental form in each stage of the process. The rate form of the impenetrability condition is applied to those portions of bodies $A$ and $B$ which are already in contact, i.e. to those points which are on the contact surface $\Gamma^{c}$. It can be written as

$$
\begin{equation*}
\gamma_{N}=\left(\mathbf{v}^{A}-\mathbf{v}^{B}\right) \cdot \mathbf{n}^{A} \equiv v_{N}^{A}-v_{N}^{B} \leq 0 \text { on } \Gamma^{c} \tag{X.2.7}
\end{equation*}
$$

where $v_{N}^{A}$ and $v_{N}^{B}$ are defined in Eq. (5). Here $\gamma_{N}(\mathbf{X}, t)$ is the rate of interpenetration of the two bodies; see Fig. 3. The impenetrability condition (7) restricts the interpenetration rate for any points on the contact surface to be negative, i.e. Eq. (7) expresses the fact that when the two bodies are in contact, then they must either remain in contact $\left(\gamma_{N}=0\right)$ or they must separate $\left(\gamma_{N}<0\right)$. When (7) is met for all points which are in contact, the impenetrability condition is met exactly. However, the equivalence between (7) and (6) does not hold when (7) is only observed at discrete points in time as in most numerical
methods, since interpenetration is then possible for points which are close but not on the contact surface during the intervening time intervals.


Figure 3. Nomenclature for velocities on contact surface; the same nomenclature and relations hold for incremental displacements $\Delta \mathbf{u}$ or variations $\delta \mathbf{u}$ or $\delta \mathbf{v}$; the contacting surfaces are shown separated for clarity.

Equation (7) can introduce discontinuities in the velocity time histories. Prior to contact, the normal velocities are not equal whereas subsequent to impact, the normal velocity components must observe (7). These discontinuities in time complicate the time integration of the discrete equations.

Equation (7) is useful only for point-pairs that are in contact or separated by small distances, since it defines the interpenetration rate exactly only when the two surfaces are coincident. However, it gives the correct sign on the interpenetration and is representative of the speed of relative surface motion when the gap between the two surfaces is small. When the interpentration is moderately large or used as the basis of the contact traction calculation, Eq. (7) is not recommended because the rate $\gamma_{N}$ is not integrable and therefore depends on the path of interpenetration. Later in this Section, formulas are discussed which are applicable for moderate amounts of interpenetration .

Many authors use the quantity $-\gamma_{N}$ to characterize the interaction of the two bodies and call it the gap rate. The gap rate is the negative of the interpenetration rate; we prefer to use the term interpenetration rate. Some authors define an interpenetration but call it a gap. It may appear inconsistent to speak of an interpenetration rate when impenetrability is a fundamental condition on the solution. However, in many numerical methods, a small amount of interpenetration is allowed, and inequality (7) will not be observed exactly.

The relative tangential velocity is given by

$$
\begin{equation*}
\gamma_{T}=\hat{\gamma}_{T x} \hat{\mathbf{e}}_{x}-\hat{\gamma}_{T y} \hat{\mathbf{e}}_{y}=\mathbf{v}_{T}^{A}-\mathbf{v}_{T}^{B} \tag{X.2.8}
\end{equation*}
$$

The middle term is included to illustrate that the relative tangential velocity in three dimensions is a two-component vector which can be expressed in terms of the local
coordinates of each point on the contact surface. As can be seen from (8), the expression for the relative tangential velocity is similar to the expression for the normal relative velocities, Eq. (2).
X.2.3. Traction Conditions. The tractions must observe the balance of momemtum across the contact interface. Since the interface has no mass, this requires that the sum of the tractions on the two bodies vanish

$$
\begin{equation*}
\mathbf{t}^{A}+\mathbf{t}^{B}=0 \tag{X.2.9a}
\end{equation*}
$$

The tractions on the surfaces of the two bodies are defined by Cauchy's law

$$
\begin{align*}
\mathbf{t}^{A} & =\sigma^{A} \cdot \mathbf{n}^{A} \quad \text { or } t_{i}^{A}=\sigma_{i j}^{A} n_{j}^{A}  \tag{X.2.9b}\\
\mathbf{t}^{B} & =\sigma^{B} \cdot \mathbf{n}^{B}  \tag{X.2.9c}\\
\text { or } t_{i}^{B} & =\sigma_{i j}^{B} n_{j}^{B}
\end{align*}
$$

The normal tractions are defined by

$$
\begin{align*}
& t_{N}^{A}=\mathbf{t}^{A} \cdot \mathbf{n}^{A} \text { or } t_{N}^{A}=t_{j}^{A} n_{j}^{A}  \tag{X.2.9d}\\
& t_{N}^{B}=\mathbf{t}^{B} \cdot \mathbf{n}^{A} \text { or } t_{N}^{B}=t_{j}^{B} n_{j}^{A} \tag{X.2.9e}
\end{align*}
$$

Note that the normal components, like all local components on the contact surface, refer to the master body. The momentum balance condition on the normal tractions can be obtained by taking a dot product of Eq. (8a) with the normal vector $\mathbf{n}^{A}$, which gives

$$
\begin{equation*}
t_{N}^{A}+t_{N}^{B}=0 \tag{X.2.9f}
\end{equation*}
$$

We do not consider any adhesion between the contact surfaces in the normal direction, so the normal tractions cannot be tensile. We will subsequently often use the phrase that the normal tractions must be compressive, although the normal tractions can also vanish. The condition that the normal tractions cannot be tensile can be stated as

$$
\begin{equation*}
t_{N} \equiv t_{N}^{A}(\mathbf{x}, t)=-t_{N}^{B}(\mathbf{x}, t) \leq 0 \tag{X.2.9g}
\end{equation*}
$$

The condition that the normal tractions be compressive requires $t_{N}^{B}$ to be positive since $t_{N}^{B}$ is the projection of the traction on body $B$ onto the unit normal of $A$, which points into body $B$.

The tangential tractions are defined by

$$
\begin{equation*}
\mathbf{t}_{T}^{A}=\mathbf{t}^{A}-t_{N}^{A} \mathbf{n}^{A} \quad \mathbf{t}_{T}^{B}=\mathbf{t}^{B}-t_{N}^{B} \mathbf{n}^{A} \tag{X.2.10a}
\end{equation*}
$$

so the tangential tractions are the total tractions projected on the master contact surface. Momentum balance requires that

$$
\begin{equation*}
\mathbf{t}_{T}^{A}+\mathbf{t}_{T}^{B}=0 \tag{X.2.10b}
\end{equation*}
$$

The above equation can be obtained by substituting (10a) into (9a) and using (9f).
When a frictionless model of contact is used, the tangential tractions vanish:

$$
\begin{equation*}
\mathbf{t}_{T}^{A}=\mathbf{t}_{T}^{B}=0 \tag{X.2.10c}
\end{equation*}
$$

We have used the phrase "frictionless model of contact" to emphasize that it is not implied that friction is absent, but rather that friction is neglected in the model because it is deemed unimportant. Subsequently we shall just say frictionless contact, but it should be understood that friction never vanishes in reality.

Although one of the bodies has been chosen as the master body in developing the preceding contact interface equations, these equations are symmetrical with respect to the bodies when the two contact surfaces are coincident and Eq. (3) is observed. Thus it does not matter which body is chosen as the master body. However, when the two surfaces are not coincident, as in most numerical solutions, then the choice of the master body changes the equations somewhat.
X.2.4. Unitary Contact Condition. Conditions (7) and (9g) can be combined into a single equation

$$
\begin{equation*}
t_{N} \gamma_{N}=0 \tag{X.2.10d}
\end{equation*}
$$

which is called the unitary contact condition. This equation also expresses the fact that the contact forces do no work. That this condition must hold on the contact surface can be seen as follows: when the bodies are in contact and remain in contact, $\gamma_{N}=0$, whereas when contact ceases, $\gamma_{N} \leq 0$ but the normal traction must vanish, so the product always vanishes. It will also be seen that this is a Kuhn-Tucker condition when a Lagrange multiplier approach is used, for the normal traction is then equivalent to a Lagrange multiplier, and the unitary condition states that the product of the Lagrange multiplier and the constraint on the velocities vanishes.
X.2.5. Surface Description*. In penalty treatments of the contact conditions and for some interface constitutive equations, it is useful to allow a certain amount of interpenetration on the contact interface and to compute it precisely. To develop such expressions for the interpenetration, a referential description of the contact surface is used. If the reference coordinates in a three dimensional problem are $\xi \equiv\left(\xi_{1}, \xi_{2}, \xi_{3}\right)$,i.e. $\xi \in R^{3}$, then the contact surface can be described by a manifold $\zeta=\left(\zeta_{1}, \zeta_{2}\right)$ i.e. $\zeta \in R^{2}$. These referential coordinates will be usually element reference coordinates in FE discretizations; an example is given later. In two dimensions, $\xi \in R^{2}$ and $\zeta \in R^{1}$, so the contact surface is a curve.

The contact surface can be described by the reference coordinates of either body, but it is conventional to choose one body as the master and use the reference coordinates of the master body for the contact interface. Body $A$ is chosen to be the master and the contact interface is described by $\mathbf{x}^{A}(\zeta, t)=\phi^{A}(\zeta, t)$. The covariant base vectors are given by

$$
\begin{equation*}
\mathbf{a}_{\alpha}=\frac{\partial \phi^{A}}{\partial \zeta^{\alpha}} \equiv \phi_{, \alpha}^{A} \equiv \mathbf{x}_{, \alpha}^{A} \tag{X.2.11a}
\end{equation*}
$$

In the above, the second through fourth term in (11) are different expressions for the same equation. The normal vector is given by

$$
\begin{equation*}
\mathbf{n}=\mathbf{a}_{1} \times \mathbf{a}_{2} / / \mathbf{a}_{1} \times \mathbf{a}_{2} \| \tag{X.2.11b}
\end{equation*}
$$

The covariant base vectors $\mathbf{a}_{\alpha}$ are useful primarily in derivations; they are tangent to the surface but not necessarily orthogonal nor of unit length. The variable Cartesian base vectors $\hat{\mathbf{e}}_{\alpha}$ are orthonormal and can be constructed from $\mathbf{a}_{\alpha}$ by $\mathbf{e}_{1}=\mathbf{a}_{1} /\left\|\mathbf{a}_{1}\right\|, \mathbf{e}_{2}=\mathbf{n} \times \mathbf{e}_{1}$, where $\mathbf{n}$ is given above; a better choice of $\mathbf{e}_{\alpha}$ is given in Chapter (Shells).
X.2.6. Interpenetration Measure. In many implementations of contact, the impenetrability condition is relaxed, i.e. a certain amount of interpenetration is permitted. When the points of two contacting areas have interpenetrated, it is useful to write the interpenetration $g_{N}\left(\zeta^{\alpha}, t\right)$ in the form of an explicit equation. We follow here the work of Wriggers(1995) and Wriggers and Miehe(1992). Consider a situation such as shown in Fig. 4, where point $P$ has penetrated body $A$. The objective is to find the penetration, which is denoted by $g_{N}\left(\zeta^{\alpha}, t\right)$


Figure 4. Interpenetration of point $P$ on body slave $B$ defined as orthonormal projection from master body $A$..

The interpenetration is defined as the minimum distance from point $P$ on body $B$ to a point on body $A$. The distance between $P$ and any point on $A$ is given by

$$
\begin{aligned}
& l_{A B}=\left\|\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)\right\| \\
& \quad \equiv\left[\left(x^{B}-x^{A}\right)^{2}+\left(y^{B}-y^{A}\right)^{2}+\left(z^{B}-z^{A}\right)^{2}\right]^{\frac{1}{2}}
\end{aligned}
$$

The referential coordinates $\bar{\zeta}$ and $\zeta$ pertain to bodies $A$ and $B$, respectively. The interpenetration $g_{N}(\zeta, t)$ is then defined as the minimum distance of point $P$ to the surface of $A$ when point $P$ has penetrated body $A$ :

$$
\begin{align*}
g_{N}(\zeta, t)= & \min \| \mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t) \mid \\
& \text { if }\left[\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)\right] \cdot \mathbf{n}^{A} \leq 0 \tag{X.2.12}
\end{align*}
$$

otherwise $g_{N}(\zeta, t)=0$
According to this definition, $g_{N}(\zeta, t)$ is positive when interpenetration occurs and vanishes when the bodies have not interpenetrated.

To evaluate the $g_{N}(\zeta, t)$, the referential coordinate $\bar{\zeta}$ which minimizes the interpenetration must be found, i.e. we must find the location of the point $\mathbf{x}^{A}(\bar{\zeta}, t)$ on the master body which corresponds to the stationary point of the distance, so we take the derivative of $l_{A B}$ with respect to $\bar{\zeta}$ and set the result to zero. This yields

$$
\begin{equation*}
\frac{\partial l_{A B}}{\partial \bar{\zeta}^{\alpha}}=\frac{\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)}{\left|\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)\right|} \cdot \frac{\partial \mathbf{x}^{A}(\bar{\zeta}, t)}{\partial \bar{\zeta}} \equiv \mathbf{e} \cdot \mathbf{a}_{\alpha}=0 \tag{X.2.13}
\end{equation*}
$$

where $\mathbf{a}_{\alpha}$ is given by Eq.(11) and $\mathbf{e}=\left(\mathbf{x}^{B}-\mathbf{x}^{A}\right) / \mid \mathbf{x}^{B}-\mathbf{x}^{A} \|$, so $\mathbf{e}$ is a unit vector from body $A$ to body $B$. The last term in the above shows that the distance is minimum, i.e. the derivative vanishes, when $\mathbf{e}$ is orthogonal to the two tangent vectors $\mathbf{a}_{\alpha}$. This implies that $\mathbf{e}$ is normal to the surface of $A$. Thus $\mathbf{x}^{A}(\bar{\zeta}, t)$ is the orthogonal projection of the point $P$ with coordinates $\mathbf{x}^{B}$ onto the master surface. This is a result that permeates mathematics: the shortest distance is always the orthogonal projection. The result is illustrated in Fig 4 in two dimensions. Note that when the bodies have interpenetrated, $\mathbf{e}$ is opposite to the direction of the outward pointing normal, so $\mathbf{e}=-\mathbf{n}^{A}$. Therefore the interpenetration $g_{N}(\zeta, t)$ is the distance from $P$ to the surface $A$ along the direction opposite to the normal of $A$. As a matter of fact, the result developed in this section is obvious from the definition: since the point corresponding to $\bar{\zeta}$ is the minimizer of the distance, it must be the orthogonal projection.

The minimizer $\bar{\zeta}$ is determined by solving the nonlinear algebraic equations (13). In three-dimensional problems, (13) involves two equation in two unknowns, in two dimensions a single equation. Once $\bar{\zeta}$ is determined, the interpenetration $g_{N}$ can be found by Eq. (12).

This approach to defining the interpenetration poses difficulties when the two bodies are not smooth and locally convex. For example, in the situation shown in Fig. 5 the minimizer of $l_{A B}$ is not unique: there are two points which are orthogonal projections of the point $P$. In these situations, it is difficult to develop schemes which lead to a uniquely defined measure of the interpenetration. Furthermore, if the discontinuous surface is the slave, the point of maximum interpenetration is not reflected in the interpenetration measure $g_{N}(\zeta, t)$ because the point of maximum interpenetration is not the orthogonal projection of any point on the master surface.


Figure 5. Penetration by a surface with a kink showing the resulting nonuniqueness of the point of orthogonal projection.
X.2.7. Path Independent Form of Interpentration Rate. In this Section, the rate of interpenetration will be developed from the interpenetration formula (12) and compared to the rate formula developed previously, Eq. (7). The rate of $g_{N}(\zeta, t)$ provides a path-independent measure of the interpentration rate so its derivative is integrable, in contrast to $\gamma_{N}$, which is not integrable. The rate $\dot{g}_{N}(\zeta, t)$ can be found by taking the derivative of $g_{N}(\zeta, t)$ in Eq. (12):

$$
\begin{equation*}
\dot{g}_{N}=\frac{d}{d t}\left(\min l_{A B}\right)=\frac{\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)}{\left\|\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)\right\|} \cdot\left(\frac{\partial \mathbf{x}^{B}(\zeta, t)}{\partial t}-\frac{\partial \mathbf{x}^{A}(\bar{\zeta}, t)}{\partial t}\right) \tag{X.2.14}
\end{equation*}
$$

Based on the discussion following Eq. (13), we know that the minimimum is attained when $\mathbf{x}^{B}-\mathbf{x}^{A} / \mid \mathbf{x}^{B}-\mathbf{x}^{A} \|$ corresponds to the normal to body $B$. Using this fact and that $\mathbf{v}^{B}=\partial \mathbf{x}^{B}(\zeta, t) / \partial t$, the above can be rewritten as

$$
\begin{equation*}
\dot{g}_{N}=\mathbf{n}^{B} \cdot\left(\mathbf{v}^{B}-\frac{\partial \mathbf{x}^{A}(\bar{\zeta}, t)}{\partial t}\right) \tag{X.2.15}
\end{equation*}
$$

It is important to observe that $\bar{\zeta}$ is not a material coordinate, because in order to remain the closest point projection, this point moves independently of the material. Thus the second term in the parenthesis of the RHS of (15) is not a material velocity. The point can be considered an ALE point: it is neither fixed in space nor coincident with a material point. Using the concept of ALE derivatives from Section X, which is based on the chain rule, it follows that

$$
\begin{equation*}
\frac{\partial \mathbf{x}^{A}\left(\bar{\zeta}_{\alpha}, t\right)}{\partial t}=\mathbf{v}^{A}-\frac{\partial \mathbf{x}^{A}}{\partial \bar{\zeta}^{\alpha}} \frac{d \bar{\zeta}^{\alpha}}{d t} \equiv \mathbf{v}^{A}-\mathbf{x}_{, \alpha}^{A} \bar{\zeta}_{, t}^{\alpha} \tag{X.2.16}
\end{equation*}
$$

Substituting (16) into (15), and using Eq. (2) it follows that

$$
\begin{equation*}
\dot{g}_{N}=\mathbf{n}^{B} \cdot\left(\mathbf{v}^{B}-\mathbf{v}^{A}-\mathbf{x}_{, \alpha}^{A} \bar{\zeta}_{, t}^{\alpha}\right)=\mathbf{n}^{A} \cdot \mathbf{v}^{A}-\mathbf{n}^{A} \cdot \mathbf{v}^{B}+\mathbf{n}^{A} \cdot \mathbf{x}_{, \alpha}^{A} \bar{\zeta}_{, t}^{\alpha} \tag{X.2.17a}
\end{equation*}
$$

Comparing Eqs. (7) and (17a), it can be seen that the normal interpenetration rate differs from the normal projection of the relative velocities $\gamma_{N}$ unless $\bar{\zeta}_{, t}^{\alpha}=0$. Whenever the two surfaces of the contacting bodies are coincident $\bar{\zeta}_{, t}=0$, so

$$
\begin{equation*}
\gamma_{N}=\dot{g}_{N} \quad \text { when } \quad g_{N} \ll 1 \tag{X.2.17b}
\end{equation*}
$$

X.2.8. Tangential Relative Velocity for Interpentrated Bodies. If the bodies have interpenetrated, Eq. (8) does not give the relative tangential velocities of two points on the contact surfaces; Eq. (8) is exact only when the two bodies are in contact but have not interpenetrated. To obtain a relation for the tangential velocities which holds for interpenetrated bodies, we follow Wriggers(1995) and Wriggers. In this approach, the relative tangential velocity is defined in terms of the velocities of a point $P$ on body $B$ and its closest point projection. The relative velocity is then projected onto the master surface. So the relative tangential velocity is defined by

$$
\begin{equation*}
\dot{\mathbf{g}}_{T}=\bar{\zeta}_{, t}^{\alpha} \mathbf{a}_{\alpha} \tag{X.2.17c}
\end{equation*}
$$

which involves the rate $\bar{\zeta}, t$ which appears in Eq. (16). This rate $\bar{\zeta}_{, t}^{\alpha}$ can be obtained from Eq. (13) as follows. Since Eq. (13) always holds for the point which is the closest point projection, the derivative of the LHS must vanish, i.e. multiplying Eq.(13) by $\mid \mathbf{x}^{B}-\mathbf{x}^{A} \|$ and using Eq. (11), $\mathbf{a}_{\alpha}=\partial \mathbf{x}^{A} / \partial \zeta^{\alpha}$, we have

$$
\begin{equation*}
\frac{d}{d t}\left[\left(\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)\right) \cdot \mathbf{a}_{\alpha}\right]=0 \tag{X.2.18}
\end{equation*}
$$

To expand the time derivative of the covariant base vector $\mathbf{a}_{\alpha}$ use (see Section X)

$$
\frac{d \mathbf{a}_{\alpha}}{d t}=\frac{d}{d t}\left(\frac{\partial \mathbf{x}^{A}}{\partial \zeta^{\alpha}}\right)=\frac{\partial}{\partial \zeta_{\alpha}}\left(\frac{d \mathbf{x}^{A}}{d t}+\frac{\partial \mathbf{x}^{A}}{\partial \zeta^{\beta}} \frac{d \zeta^{\beta}}{d t}\right)
$$

$$
\begin{equation*}
=\frac{\partial}{\partial \zeta_{\alpha}}\left(\mathbf{v}^{A}+\mathbf{x}_{, \beta}^{A} \zeta_{, t}^{\beta}\right)=\mathbf{v}_{, \alpha}^{A}+\mathbf{x}_{, \alpha \beta}^{A} \zeta_{, t}^{\beta} \tag{X.2.19}
\end{equation*}
$$

The remaining step are as follows (the independent variables are suppressed when convenient):
(derivative of product in (18))

$$
\begin{equation*}
\left(\mathbf{x}_{, t}^{B}(\zeta, t)-\mathbf{x}_{, t}^{A}(\bar{\zeta}, t)\right) \cdot \mathbf{a}_{\alpha}+\left(\mathbf{x}^{B}-\mathbf{x}^{A}\right) \cdot \mathbf{a}_{\alpha, t}=0 \tag{X.2.20}
\end{equation*}
$$

(using $\mathbf{v}^{B A} \equiv \mathbf{v}^{B}-\mathbf{v}^{A}, \mathbf{x}^{B A} \equiv \mathbf{x}^{B}-\mathbf{x}^{A}$, Eq. (19) for $\mathbf{a}_{\alpha, t}$ )

$$
\begin{equation*}
\left(\mathbf{v}^{B A}-\mathbf{x}_{, \beta}^{A} \bar{\zeta}_{, t}^{\beta}\right) \cdot \mathbf{a}_{\alpha}+\mathbf{x}^{B A} \cdot\left(\mathbf{v}_{, \alpha}^{A}+\mathbf{x}_{, \alpha \beta}^{A} \bar{\zeta}_{, t}^{\beta}\right)=0 \tag{X.2.21}
\end{equation*}
$$

(using $g_{N} \mathbf{n}^{A}=\mathbf{x}^{B}-\mathbf{x}^{A} \equiv \mathbf{x}^{B A}, \mathbf{x}_{, \beta}^{A}=\mathbf{a}_{\beta}$ Eq. (11) )

$$
\begin{equation*}
\left(\mathbf{a}_{\alpha} \cdot \mathbf{a}_{\beta}-g_{N} \mathbf{n}^{A} \cdot \mathbf{x}_{, \alpha \beta}^{A}\right) \bar{\zeta}_{, t}^{\beta}=g_{N} \mathbf{n}^{A} \cdot \mathbf{v}_{, \alpha}^{A}+\mathbf{v}^{B A} \cdot \mathbf{a}_{\alpha} \tag{X.2.22}
\end{equation*}
$$

The above is a system of two linear algebraic equations in the two unknowns $\bar{\zeta}_{, t}^{\beta}$; all terms on the right hand side are known. Once the time derivatives $\bar{\zeta}_{, t}^{\beta}$ are known, $\dot{\mathbf{g}}_{T}$ can be determined from (17c). The first terms on the LHS and RHS of the above equations are of fundamental importance in the theory of surfaces: they are the first and second fundamental forms of the surface.

When $g_{N}=0$ (or when $g_{N}$ is sufficiently small), Eq. (22) can be simplified to

$$
\begin{equation*}
\mathbf{a}_{\alpha} \cdot \mathbf{a}_{\beta} \bar{\zeta}_{, t}^{\beta}=\left(\mathbf{v}^{B}-\mathbf{v}^{A}\right) \cdot \mathbf{a}_{\alpha} \tag{X.2.23}
\end{equation*}
$$

Taking the tensor product of the above with $\mathbf{a}^{\alpha}$ we obtain

$$
\begin{equation*}
\dot{\mathbf{g}}_{T}=\mathbf{a}_{\beta} \bar{\zeta}_{, t}^{\beta}=\left(\mathbf{v}^{B}-\mathbf{v}^{A}\right) \mathbf{a}_{\alpha} \otimes \mathbf{a}^{\alpha}=\mathbf{v}_{T}^{B}-\mathbf{v}_{T}^{A} \tag{X.2.24}
\end{equation*}
$$

where the second line follows from the fact that the projection of any vector on the surface is the tangential component. Since the RHS by Eq. (8) is $-\gamma_{T}$, we can see that when that when the surfaces are coincident, i.e. when $g_{N}=0$, then

$$
\begin{equation*}
\dot{\mathbf{g}}_{T}=-\gamma_{T} \tag{X.2.25}
\end{equation*}
$$

Thus the displacement-based definition of relative tangential velocity, Eqs. (17c) and (22), is consistent with the tangential velocity defined in Eq. (8) in the absence of interpenetration (except for the sign, which is irrelevant)

The kinetic and kinematic contact interface equations are summarized in Box X.1.

Example 5.2.1. Consider the two surfaces shown in Fig. 6, which have partially interpenetrated as shown. The master body is a 9 -node isoparametric element, so the 3 nodes of surface $A$ are defined by a quadratic mapping:

$$
\begin{gathered}
\left\{\begin{array}{l}
x \\
y
\end{array}\right\}^{A}=\left(1-r^{2},\left\{\begin{array}{l}
2 \\
1
\end{array}\right\}+\frac{1}{2} r(1+r)\left\{\begin{array}{l}
3 \\
3
\end{array}\right\}\right. \\
r \equiv \zeta^{A},-1 \leq r \leq 1
\end{gathered}
$$

The surface of the slave body $B$ is a horizontal line given by

$$
\left\{\begin{array}{l}
x \\
y
\end{array}\right\}^{B}=\left\{\begin{array}{l}
4 s \\
1.5
\end{array}\right\}, \quad s \equiv \zeta^{B}, 0 \leq s \leq 1
$$

The interpenetration in the example has been exaggerated. Note that $\mathbf{n}^{B} \neq-\mathbf{n}^{A}$ along the interface.

Part A. For the point Pon slave surface $B$ with coordinates (1,1.5), find the interpenetration.

The first step is to find the orthogonal projection point $Q$ which minimizes $l_{P Q}$ :

## BOX X.1. Contact Interface Conditions

kinetic conditions

$$
\mathbf{t}^{A}+\mathbf{t}^{B}=0
$$

$$
\text { normal: } t_{N}^{A}+t_{N}^{B}=0, \quad t_{N}^{A} \equiv \mathbf{t}^{A} \cdot \mathbf{n}^{A}, \quad t_{N}^{B} \equiv \mathbf{t}^{B} \cdot \mathbf{n}^{A}, \quad t_{N} \equiv t_{N}^{A} \leq 0
$$

$$
\text { tangential: } \mathbf{t}_{T}^{A}+\mathbf{t}_{T}^{B}=0, \quad \mathbf{t}_{T}^{A} \equiv \mathbf{t}^{A}-t_{N}^{A} \mathbf{n}^{A}, \quad \mathbf{t}_{T}^{B} \equiv \mathbf{t}^{B}-t_{N}^{B} \mathbf{n}^{A}
$$

kinematic conditions in velocity form

$$
\begin{aligned}
& \gamma \equiv \gamma_{N}=\left(\mathbf{v}^{A}-\mathbf{v}^{B}\right) \cdot \mathbf{n}^{B} \equiv v_{N}^{A}-v_{N}^{B} \leq 0 \\
& \gamma_{T}=\mathbf{v}_{T}^{A}-\mathbf{v}_{T}^{B}=\mathbf{v}^{A}-\mathbf{v}^{B}-\left(\mathbf{v}^{A}-\mathbf{v}^{B}\right) \cdot \mathbf{n}^{A}
\end{aligned}
$$

unitary contact condition

$$
t_{N} \gamma_{N}=0
$$

kinematic conditions and definitions in displacement form

$$
\begin{aligned}
& g \equiv g_{N}=\min _{\bar{\zeta}} \mid \mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t) \| \quad \text { if }\left[\mathbf{x}^{B}(\zeta, t)-\mathbf{x}^{A}(\bar{\zeta}, t)\right] \cdot \mathbf{n}^{A} \leq 0 \\
& \dot{g}_{N}=\mathbf{n}^{B} \cdot\left(\mathbf{v}^{B}-\mathbf{v}^{A}-\mathbf{x}_{, \alpha}^{A} \bar{\zeta}_{, t}^{\alpha}\right)=\mathbf{n}^{A} \cdot \mathbf{v}^{A}-\mathbf{n}^{A} \cdot \mathbf{v}^{B}+\mathbf{n}^{A} \cdot \mathbf{x}_{, \alpha}^{A} \bar{\zeta}_{, t}^{\alpha} \\
& \dot{\mathbf{g}}_{T}=\bar{\zeta}_{, t}^{\alpha} \mathbf{a}_{\alpha} \text { where }\left(\mathbf{a}_{\alpha} \cdot \mathbf{a}_{\beta}-g \mathbf{n}^{A} \mathbf{v}_{, \alpha \beta}\right) \bar{\zeta}_{\zeta, t}^{\beta}=g \mathbf{n}^{A} \cdot \mathbf{v}_{, \alpha}^{A}+\left(\mathbf{v}^{B}-\mathbf{v}^{A}\right) \cdot \mathbf{a}_{\alpha}
\end{aligned}
$$

$$
\begin{aligned}
l_{P Q} & =\left\|\mathbf{x}^{B}\left(\zeta^{B}\right)-\mathbf{x}^{A}\left(\zeta^{A}\right)\right\|=\left(\left(x^{B}-x^{A}\right)^{2}+\left(y^{B}-y^{A}\right)^{2}\right)^{1 / 2} \\
& =\left\{\left[1-\left(2\left(1-r^{2}\right)+\frac{3}{2} r(1+r)\right)\right]^{2}+\left[\frac{3}{2}-\left(\left(1-r^{2}\right)+\frac{3}{2} r(1+r)\right)\right]^{2}\right\}^{1 / 2}
\end{aligned}
$$

The minimizer satisfies

$$
0=\frac{d l_{P Q}}{d r}=\frac{1}{l_{P Q}}\left(r^{3}+3 r+\frac{3}{4}\right)
$$

The root is found numerically to be $r=-0.2451$, so $\left(x_{Q}, y_{Q}\right)=(1.6023,0.6624)$.

## X. 3 FRICTION MODELS

X.3.1. Classification. The models used for the computation of the tangential tractions are collectively called friction models. There are basically three types of friction models:

1. Coulomb friction models, which are based on the classical theories of friction commonly taught in undergraduate mechanics and physics courses;
2. Interface constitutive equations, which approximate the behavior of the tangential forces by equations similar to constitutive equations used for materials;
3. Asperity-lubricant models, which model the behavior of the physical characteristics of the interface, often on a microscale.

The demarcations between these classes are not sharp; some models adopt features of more than one of the above classes, but the above roughly describes the current state of affairs.
X.3.2. Coulomb Friction. Coulomb friction models originate from classical friction, which is used for the total frictional forces between rigid bodies. In the application of classical Coulomb friction models to continua, they are applied at each point of the contact interface. A direct translation of the Coulomb friction law to a pointwise law gives
if $A$ and $B$ are in contact at $\mathbf{x}$, then

$$
\begin{align*}
& \text { a) } i f\left\|\mathbf{t}_{T}(\mathbf{x}, t)\right\|<-\mu_{F} t_{N}(\mathbf{x}, t), \gamma_{T}(\mathbf{x}, t)=0  \tag{X.3.1a}\\
& \text { b) } i f\left\|\mathbf{t}_{T}(\mathbf{x}, t)\right\|=-\mu_{F} t_{N}(\mathbf{x}, t), \gamma_{T}(\mathbf{x}, t)=-\alpha(\mathbf{x}, t) \mathbf{t}_{T}(\mathbf{x}, t), \quad \alpha \geq 0 \tag{X.3.1b}
\end{align*}
$$

where $\alpha$ is a variable which is determined from the solution of the complete problem. The condition that the two bodies are in contact at a point implies that the normal traction $t_{N} \leq 0$, so the RHS of the two expressions, $-\mu_{F} t_{N}$, is always positive. Condition (a) is known as the stick condition, for when the tangential traction at a point is less than the critical value, no relative tangential motion is permitted according to this condition, i.e. the two bodies stick. Condition (b) corresponds to frictional sliding, and the second part of that equation expresses the fact that the tangential traction arising from friction must be in the direction opposite to the direction of the relative tangential velocity.

The classical Coulomb friction law closely resembles a rigid-plastic constitutive equation. If the tangential velocity $\gamma_{T}$ is interpreted as a strain and the tangential traction components are interpreted as stresses, the first relation in Eq. (1a) can be interpreted as a yield function. According to (1), when the yield criterion is not met, the tangential velocity vanishes. Once the yield function is satisfied, the tangential velocity is in the
direction of the tangential traction but its magnitude is unspecified. These attributes of the response parallel the rigid plastic model desribed in Section 6.?.

There are several alternative ways of stating Coulomb's law which are equivalent to the above. For example, Demkowicz and Oden(1981) state Coulomb's law as (the spatial dependence of the variables has been dropped for simplicity):

$$
\begin{align*}
& \text { if Aand Bareincontact at } x \text {, then } \\
& \qquad \mid \mathbf{t}_{T} \| \leq-\mu_{F} t_{N} \quad \text { and } \quad \mathbf{t}_{T} \cdot \mathbf{v}_{T}+\mu_{F}\left|t_{N}\right|\left|\gamma_{T}\right|=0 \tag{X.3.2}
\end{align*}
$$

The stick condition of Coulomb friction is its most troublesome characteristic, since it introduces discontinuities in the time history of the relative tangential velocity. When the motion of point changes from relative slipping to sticking, the relative tangential velocity $\gamma_{T}$ discontinuously jumps to zero. Thus the tangential velocities at that point are not smooth, but exhibit the same discontinous character as the normal velocities at the time of impact. Furthermore, the inequalities result in the Coulomb friction law result in weak forms which involve inequalities. Therefore, Coulomb friction is difficult to handle in numerical solutions and we consider it only for some special cases.
X.3.3. Interface Constitutive Equations. A different approach to defining interface laws has been pioneered by Michalowski and Mroz (1978) and Curnier(1984). This approach is motivated by the theory of plasticity and the analogy between Coulomb friction and rigid-perfect plasticity we alluded to above. Interface constitutive equations can model behavior similar to Coulomb friction by means of the Mohr-Coulomb criterion (see Section XX). Plastic models of interface behavior are motivated by the fact that microscopic examination of even the smoothest surfaces reveals surface roughness due to asperities, such as shown in Fig. X.3.1. Even when the surfaces appear smooth, friction is generated by the interaction of these asperities during sliding. Sliding initially causes elastic deformations of these asperities, so a true stick condition cannot exist in actual siding, i.e. the stick condition is an idealization of observed behavior. The elastic deformation of the asperities is followed by "grinding" down of the asperities as the sliding proceeds. The elastic deformations of the asperities are reversible, whereas the grinding down is irreversible, so ascribing an elastic character to the initial sliding and a plastic character to subsequent sliding is natural.


Figure X.3.1. Asperities on contacting surface.
As an example of an interface constitutive law we describe an adaptation of Curnier's plasticity theory for friction. This model contains all of the ingredients of a plasticity theory for continua: a decomposition of deformation into reversible and irreversible components, a yield function and a flow law. In this description of the model Curnier(1984), we have replaced displacements by rates, which appears appropriate for problems involving arbitrary time histories and large relative sliding.

In this theory, the rate of relative velocities $\gamma$ is subdivided into that ascribed to adherence, which is the elastic deformation of the asperities, and that ascribed to slip, the grinding down of the asperities:

$$
\begin{equation*}
\gamma=\gamma^{a d h}+\gamma^{s l i p} \equiv \gamma^{a}+\gamma^{s} \tag{X.3.3}
\end{equation*}
$$

Here $\gamma^{a d h}$ is the reversible part, $\gamma^{s l i p}$ is the irreversible part. A wear function is defined by

$$
\begin{equation*}
D^{c}=\int_{0}^{t}\left(\gamma_{T}^{s} \cdot \gamma_{T}^{s}\right)^{\frac{1}{2}} d t \tag{X.3.4}
\end{equation*}
$$

which is reminiscent of the definition of effective plastic strain.
Two functions are defined to construct the plastic interface law:

1. a yield function, $f(\mathbf{t})$
2. a potential function for the flow law, $h(\mathbf{t})$

The yield function determines the onset of plastic response, the potential function the relationship between the slip (plastic strain rate) and the tangential tractions.

The theory is similar to the nonassociative plasticity theories given in Section?. Therefore, we will only sketch the steps so that the equations are available and to enable us to point out the need for nonassociative plasticity in a model of frictional sliding.

The yield function for Coulomb type behavior is obtained from Eq. (1):

$$
\begin{equation*}
f\left(t_{N}, \mathbf{t}_{T}\right)=\left\|\mathbf{t}_{T}\right\|+\mu_{F} t_{N}=0 \tag{X.3.5}
\end{equation*}
$$



Figure X.3.2. Coulomb yield surface in two dimensions.
In two dimensions this yield function takes the form shown in Fig. X.3.2: $\mathbf{t}_{T}=t_{T} \hat{\mathbf{e}}_{x}$ in that case, so the yield function consists of two lines with slopes $\pm \mu_{F}$ as shown. For the three-dimensional case, $\mathbf{t}_{T}=\hat{t}_{\alpha} \hat{\mathbf{e}}_{\alpha}=\hat{t}_{x} \hat{\mathbf{e}}_{x}+\hat{t}_{y} \hat{\mathbf{e}}_{y}$, we can write Eq. (5) as

$$
\begin{equation*}
f\left(t_{N}, \mathbf{t}_{T}\right)=\left(\hat{t}_{x}^{2}+\hat{t}_{y}^{2}\right)^{1 / 2}+\mu_{F} t_{N}=0 \tag{X.3.6}
\end{equation*}
$$

so the yield function is a cone as shown in Fig. X.3.3.


Figure X.3.3. Coulomb surface for contact in 3D.
In a nonassociative theory, the potential function for the slip differs from the yield function. One possible potential function for a nonassociative theory is

$$
\begin{equation*}
h\left(t_{N}, \mathbf{t}_{T}\right)=\left\|\mathbf{t}_{T}\right\|-\beta=0 \tag{X.3.7}
\end{equation*}
$$

where $\beta$ is a constant whose magnitude is irrelevant. This potential function is also shown in Fig. X.3.4.


Figure X.3.4. Non-associated flow law.
To write the complete relations for a plasticity theory of friction in two and three dimensions, it is convenient to define

$$
\begin{align*}
& \mathbf{g}=\left\{\begin{array}{l}
\gamma_{N} \\
\mathbf{g}_{T}
\end{array}\right\} \text { in 2D, } \mathbf{g}=\left\{\begin{array}{l}
\gamma_{N} \\
\mathbf{g}_{T}
\end{array}\right\}=\left\{\begin{array}{l}
\gamma_{N} \\
\hat{\gamma}_{x} \\
\hat{\gamma}_{y}
\end{array}\right\} \text { in 3D }  \tag{X.3.10}\\
& \mathbf{Q}=\left\{\begin{array}{l}
t_{N} \\
t_{T}
\end{array}\right\} \text { in 2D, } \mathbf{Q}=\left\{\begin{array}{l}
t_{N} \\
\mathbf{t}_{T}
\end{array}\right\}=\left\{\begin{array}{l}
t_{N} \\
\hat{t}_{x} \\
\hat{t}_{y}
\end{array}\right\} \text { in 3D } \tag{X.3.11}
\end{align*}
$$

The adhesive strains are then related to the stresses by

$$
\begin{equation*}
\dot{\mathbf{Q}}=\mathbf{C}_{F} \mathbf{g}^{a d h} \text { or } \quad \dot{\mathbf{Q}}_{i}=C_{i j}^{F} \gamma_{j}^{a d h} \tag{X.3.12}
\end{equation*}
$$

which is the counterpart of the linear elastic law for continua. Usually $\mathbf{C}_{F}$ is diagonal since little experimental information is available on coupling between different components of the frictional traction and the relative motion.

The adhesive slip rates are given by the nonassociative flow law. Perfectly-plastic sliding, in which there is no increase in the tractions with the accumulation of slip, closely resembles Coulomb friction and is given by

$$
\begin{equation*}
\mathbf{g}^{s l i p}=\alpha \frac{\partial h}{\partial \mathbf{Q}} \text { or } \gamma_{i}^{s l i p}=\alpha \frac{\partial h}{\partial Q_{i}} \tag{X.3.13}
\end{equation*}
$$

We define

$$
\begin{equation*}
\mathbf{f}_{Q}=\frac{\partial f}{\partial Q} \quad \mathbf{h}_{Q}=\frac{\partial h}{\partial Q} \tag{X.3.14}
\end{equation*}
$$

The steps for developing the constitutive equation for the frictional surface are then:

$$
\begin{array}{ll}
\mathbf{f}_{Q}^{T} \dot{\mathbf{Q}}=0 & \text { consistency } \\
\dot{\mathbf{Q}}=\mathbf{C}\left(\mathbf{g}-\mathbf{g}^{s l i p}\right) & \text { (12) and (3) } \\
\mathbf{f}_{Q}^{T} \mathbf{C}\left(\mathbf{g}-\alpha \mathbf{h}_{Q}\right)=0 & \text { (13) and (15) intc } \\
\alpha=\frac{\mathbf{f}_{Q}^{T} \mathbf{C} \mathbf{g}}{\mathbf{f}_{Q}^{T} \mathbf{C h}} & \text { solve (17) for } \alpha  \tag{X.3.18}\\
\dot{\mathbf{Q}}=\mathbf{C}\left(\mathbf{g}-\frac{\mathbf{f}_{Q}^{T} \mathbf{C g}}{\mathbf{f}_{Q}^{T} \mathbf{C} \mathbf{h}_{Q}} \mathbf{h}_{Q}\right) & \text { (18) and (??) into }
\end{array}
$$

Since the above is a traction rate, it is not objective (i.e. frame invariant) so a frameinvariant rate must be used for integration.

$$
\begin{equation*}
\mathbf{Q}^{\nabla}(\zeta, t)=\frac{\partial \mathbf{Q}(\zeta, t)}{\partial t}-\mathbf{Q} \cdot \mathbf{W} \tag{X.3.20}
\end{equation*}
$$

where $\mathbf{Q}^{\nabla}$ is a frame invariant rate and $\mathbf{W}$ is the projection of the spin given by Eq. (3.X.X) onto the surface. In the above, $\partial \mathbf{Q} / \partial t$ is the rate of the tractions due to the slip rates. The update proccedures are analogous to that in elasto-plasticity and are discussed in Section X.

The reason for choosing a nonassociative flow law can be clarified by considering sliding in a two dimensional problem. If we were to use an associated flow law, the irreversible slips are given by $\gamma_{N}^{s l i p}=\alpha \frac{\partial f}{\partial t_{N}}=-\alpha \mu_{F} \gamma_{T}^{s l i p}=\alpha \frac{\partial f}{\partial t_{T}}=-\alpha \operatorname{sign}\left(t_{T}\right)$. Since $\alpha \geq 0$, this implies that, $\gamma_{N}^{s l i p}<0$ so the bodies would separate after the onset of slip (recall $\gamma_{N}$ is positive in interpenetration). If the slips are then given by the potential flow laws using the nonassociated potential (7), the slips in two dimensions can be written as

$$
\begin{align*}
& \gamma_{N}^{s l i p}=\alpha \frac{\partial h}{\partial t_{N}}=0  \tag{X.3.21}\\
& \gamma_{T}^{s l i p}=\alpha \frac{\partial h}{\partial t_{T}}=\alpha \tag{X.3.22}
\end{align*}
$$

Thus the normal slip vanishes, i.e. that no irreversible normal interpenetration occurs during to sliding.

Hardening can also be incorporated in a manner analogous to the procedure in elasto-plasticity. The constitutive equation for the interface is then developed as in plasticity with hardening, see Section ?. Under large pressures, the asperities are often significantly ground down, and some degree of permanent change occurs in the normal interpentration. This can be modeled by a cap model such as described in Section?.

## X. 4 WEAK FORMS

X.4.1. Notation and Preliminaries. The weak form of the momentum equation and the contact interface conditions will be developed for a Lagrangian mesh. This development is also applicable to an ALE mesh when the contact surface is treated as Lagrangian. For simplicity, we start with frictionless contact and defer the treatment of tangential tractions to the last part of this Section. We restrict the following developments to the case where all traction or velocity components are prescribed on a traction or dispacement boundary, respectively.

The contact surface is neither a traction nor a displacement boundary. Thus the total boundary of body $A$ is given by

$$
\begin{equation*}
\Gamma^{A}=\Gamma_{t}^{A} \cup \Gamma_{u}^{A} \cup \Gamma^{c} \tag{X.4.2a}
\end{equation*}
$$

$$
\begin{equation*}
\Gamma_{t}^{A} \cap \Gamma_{u}^{A}=0 \quad \Gamma_{t}^{A} \cap \Gamma^{c}=0 \quad \Gamma_{u}^{A} \cap \Gamma^{c}=0 \tag{X.4.2b}
\end{equation*}
$$

Similar relations hold for body $B$.
The trial solutions are in the space of kinematically admissible velocities, and as in Chapter 4 we choose the velocities to be the cardinal dependent variable. The trial solution is $\mathbf{v}(\mathbf{X}, t) \in \mathcal{U}$ where the space of trial functions is defined by

$$
\begin{equation*}
\mathcal{U}=\left\{\mathbf{v}(\mathbf{X}, t) \mid \mathbf{v} \in C^{0}\left(\Omega^{A}\right), \mathbf{v} \in C^{0}\left(\Omega^{B}\right), \mathbf{v}=\overline{\mathbf{v}} \text { on } \Gamma_{u}\right\} \tag{X.4.3}
\end{equation*}
$$

The space is similar to that for the single body problem, but the velocities are separately approximated in the two bodies; the velocity fields in $\mathcal{U}$ are not required to be continuous across the contact interface. (*while the admissible velocity fields are here given as $C^{0}$, i.e. in $\mathcal{H}^{1}$, for purposes of convergence analysis in linear elastostatics the displacements for the contact problem are defined in the space $\mathcal{H}^{1 / 2}$, see Kikuchi and Oden(1988). This is the same space that is used in fracture mechanics problems to handle the singular stresses at the crack tip. In contact problems, singularities occur at the edge of the contact zone so the same space must be used in convergence analysis. However, unlike in fracture mechanics, these singularities do not appear to be of any engineering significance, since the roughness of surfaces appears to eliminate the appearance of even near singular behavior in the stresses.)

The space of test functions is defined by

$$
\begin{equation*}
\mathcal{U}_{0}=\mathcal{U} \cap\left\{\delta \mathbf{v}(\mathbf{X}) \mid \quad \delta \mathbf{v}=0 \text { on } \Gamma_{u}\right\} \tag{X.4.4}
\end{equation*}
$$

which parallels the definition in Section 4.3: The test functions are identical to the trial functions except that they vanish on prescribed displacement boundaries.
X.4.2. Lagrange Multiplier Weak Form. A common approach to imposing the contact constraints is by means of Lagrange multipliers. We will follow the description given by Belytschko and Neal(1991). Let the Lagrange multiplier trial functions $\lambda\left(\zeta^{\alpha}, t\right)$ and the corresponding test functions be in the following spaces

$$
\begin{align*}
& \lambda\left(\zeta^{\alpha}, t\right) \in g^{+}, g^{+}=\left\{\lambda\left(\zeta^{\alpha}, t\right) \mid \lambda \in C^{-1}, \lambda \geq 0 \text { on } \Gamma^{c}\right\}  \tag{X.4.5}\\
& \delta \lambda\left(\zeta^{\alpha}\right) \in \mathcal{I}^{-}, g^{-}=\left\{\delta \lambda\left(\zeta^{\alpha}\right) \mid \delta \lambda \in C^{-1}, \delta \lambda \leq 0 \text { on } \Gamma^{c}\right\} \tag{X.4.6}
\end{align*}
$$

The weak form is:

$$
\begin{equation*}
\text { if } \delta \mathcal{P}_{L}(\mathbf{v}, \delta \mathbf{v}, \lambda, \delta \lambda) \equiv \delta \mathcal{P}+\delta \mathcal{G}_{L} \geq 0 \quad \forall \delta \mathbf{v} \in \mathcal{U}_{0}, \forall \delta \lambda \in \mathcal{I}^{-} \tag{X.4.7}
\end{equation*}
$$

$$
\begin{equation*}
\delta \mathcal{G}_{L}=\int_{\Gamma^{c}} \delta\left(\lambda \gamma_{N}\right) d \Gamma \tag{X.4.8}
\end{equation*}
$$

where $\delta \mathscr{P}$ is defined in Table B4.2 and $v \in \mathcal{U}, \lambda \in \mathcal{I}^{+}$. This weak form is equivalent to the momentum equation, the traction boundary conditions and the following contact interface conditions: impenetrability (2.7), momentum balance on normal tractions (2.9f) and the frictionless condition $(2.10 \mathrm{c})$. The restriction of the normal interface traction to be compressive will result from the constraints on the trial set of Lagrange multipliers. Note that the above weak form is an inequality.

The above is a standard way of appending a constraint to a weak form by means of a Lagrange multipliers: compare to the Hu-Washizu variational principle. The only difference from the Hu-Washizu form is that the constraint is an inequality.

The equivalence of the weak form to the momentum equation, the traction boundary conditions and the contact conditions is shown by a procedure that parallels that given in Section 4.2. Recall that $\delta \mathscr{P}$ is given in Box 4.1 as

$$
\begin{equation*}
\delta \mathcal{P}=\int_{\Omega}\left[\delta v_{i, j} \sigma_{j i}-\delta v_{i}\left(\rho b_{i}-\rho \dot{v}_{i}\right)\right] d \Omega-\int_{\Gamma_{t}} \delta v_{i} \bar{t}_{i} d \Gamma \tag{X.4.9}
\end{equation*}
$$

where we have used commas to denote derivatives with respect to the spatial variables and a superposed dot to denote the material time derivative. All integrals in the above apply to the union of both bodies, i.e. $\Omega=\Omega^{A} \cup \Omega^{B}, \Gamma_{t}=\Gamma_{t}^{A} \cup \Gamma_{t}^{B}$. The first step is to integrate the internal virtual power by parts and apply Gauss's theorem:

$$
\begin{equation*}
\int_{\Omega}\left(\delta v_{i} \sigma_{j i}\right)_{, j} d \Omega=\int_{\Gamma_{t}} \delta v_{i} \sigma_{j i} n_{j} d \Gamma+\int_{\Gamma^{c}}\left(\delta v_{i}^{A} t_{i}^{A}+\delta v_{i}^{B} t_{i}^{B}\right) d \Gamma \tag{X.4.12}
\end{equation*}
$$

We have used the fact that the integral over the displacement boundary $\Gamma_{u}$ vanishes because $\delta v_{i}=0$ on $\Gamma_{u}$ and Cauchy's law ( $9 \mathrm{~b}-\mathrm{c}$ ) has been applied to obtain the expressions in the last integral. The first integral on the right hand side of the above applies to both bodies, as can be seen from the definition (2c). The contact surface integral appears for each body when Gauss's theorem is applied, so to express the result as a single integral, the field variables associated with the two bodies have been specifically indicated the superscripts $A$ and $B$.

The integrand of the second integral on the RHS of the above is now broken up into components normal and tangential to the contact surface. In indicial notation this gives

$$
\begin{equation*}
\delta v_{i}^{A} t_{i}^{A}=\delta v_{N}^{A} t_{N}^{A}+\delta \hat{v}_{\alpha}^{A} t_{\alpha}^{A} \tag{X.4.13}
\end{equation*}
$$

where, as usual in this book, the range of alpha is 1 for two dimesional problems and 2 for three dimensional problems. A similar relationship can be written for body $B$. The above is clearer to some people in vector notation, where using (2.10a) we can write

$$
\begin{align*}
\delta \mathbf{v}^{A} \cdot \mathbf{t}^{A} & =\left(\delta v_{N}^{A} \mathbf{n}^{A}+\delta \mathbf{v}_{T}^{A}\right) \cdot\left(t_{N}^{A} \mathbf{n}^{A}+\mathbf{t}_{T}^{A}\right) \\
& =\delta v_{N}^{A} t_{N}^{A}+\delta \mathbf{v}_{T}^{A} \cdot \mathbf{t}_{T}^{A} \tag{X.4.14}
\end{align*}
$$

The simplification to the second line is obtained by noting that $\mathbf{n}$ is normal to the tangent vectors $\mathbf{t}_{N}$ and $\mathbf{v}_{N}$. The second term in (14) is an alternative expression for $\hat{t}_{\alpha} \hat{v}_{\alpha}$.

Substituting Eqs. (12) and (13) into (9) gives

$$
\begin{align*}
& \delta \mathcal{P}=\int_{\Gamma^{C}} \delta v_{i}\left(\rho v_{i}-b_{i}-\sigma_{i j, j}\right) d \Omega+\int_{\Gamma_{t}} \delta v_{i}\left(\sigma_{j i} n_{j}-\bar{t}_{i}\right) d \Gamma  \tag{X.4.16}\\
& \quad+\int_{\Gamma^{c}}\left(\delta v_{N}^{A} t_{N}^{A}+\delta v_{N}^{B} t_{N}^{B}+\delta \hat{v}_{\alpha}^{A} \hat{t}_{\alpha}^{A}+\delta \hat{v}_{\alpha}^{B} \hat{t}_{\alpha}^{B}\right) d \Gamma
\end{align*}
$$

Now consider Eq. (8):

$$
\begin{equation*}
\delta \mathcal{G}_{L}=\int_{\Gamma^{C}} \delta\left(\lambda \gamma_{N}\right) d \Gamma=\int_{\Gamma^{C}}\left(\delta \lambda \gamma_{N}+\delta \gamma_{N} \lambda\right) d \Gamma \tag{X.4.17}
\end{equation*}
$$

Substituting Eq.(2.2) into the above gives

$$
\begin{equation*}
\delta \mathcal{G}_{L}=\int_{\Gamma^{c}}\left(\delta \lambda \gamma_{N}+\lambda\left(\delta v_{N}^{A}-\delta v_{N}^{B}\right)\right) d \Gamma \tag{X.4.18}
\end{equation*}
$$

Combining Eqs. (16) and (18) yields

$$
\begin{align*}
0 \leq & \delta \mathcal{P}_{L}=\int_{\Omega} \delta v_{i}\left(\sigma_{j i, j}-\rho b_{i}-\rho \dot{v}_{i}\right) d \Omega+\int_{\Gamma_{t}} \delta v_{i}\left(\sigma_{j i} n_{j}-\bar{t}_{i}\right) d \Gamma \\
& +\int_{\Gamma^{c}}\left[\delta v_{N}^{A}\left(t_{N}^{A}+\lambda\right)+\delta v_{N}^{B}\left(t_{N}^{B}-\lambda\right)+\left(\delta \hat{v}_{\alpha}^{A} \hat{t}_{\alpha}^{A}+\delta \hat{v}_{\alpha}^{B} t_{\alpha}^{B}\right)+\delta \lambda \gamma_{N}\right] d \Gamma \tag{X.4.19}
\end{align*}
$$

Extracting the strong form from the weak inequality is similar to the procedure described in Section 4.2. Whenever the test function is unconstrained, there is then no restriction on the sign of the term which multiplies the test function and the term must vanish by the density theorem. Thus it follows from the first two integrals of the above that

$$
\begin{align*}
& \sigma_{j i, j}-\rho b_{i}=\rho \dot{v}_{i} \text { in } \Omega \\
& \sigma_{j i} n_{j}=\dot{t}_{i} \text { on } \Gamma_{t} \tag{X.4.21}
\end{align*}
$$

i.e. that the momentum equation and the natural boundary conditions are satisfied in bodies $A$ and $B$. In all terms of the integrand on the contact surface except the last, the test function is also unconstrained, and we obtain the equalities

$$
\begin{align*}
& \hat{t}_{\alpha}^{A}=0 \text { and } \hat{t}_{\alpha}^{B}=0 \text { on } \Gamma^{c}, \quad \text { or } \mathbf{t}_{T}^{A}=\mathbf{t}_{T}^{A}=0 \text { on } \Gamma^{c}  \tag{X.4.22}\\
& \lambda=-t_{N}^{A} \text { and } \lambda=t_{N}^{B} \text { on } \Gamma^{c} \tag{X.4.23}
\end{align*}
$$

By eliminating $\lambda$ from (23) we obtain the momentum balance condition on the normal tractions

$$
\begin{equation*}
t_{N}^{A}+t_{N}^{B}=0 \text { on } \Gamma^{c} \tag{X.4.25a}
\end{equation*}
$$

Since the space of trial functions for $\lambda$ is constrained to be positive, see Eq. (5), it follows from (23) that the normal traction on the contact interface is compressive. Thus we can write

$$
\begin{equation*}
t_{N}^{A}+t_{N}^{B}=0 \text { on } \Gamma^{c} \tag{X.4.25b}
\end{equation*}
$$

In the last term of the integrand of Eq.(18), the variation $\delta \lambda$ is constrained to be negative. Therefore, it cannot be deduced that its coefficient $\gamma_{N}$ vanishes. However it can be deduced that the coefficient must be nonpositive, i.e. that the weak inequality is equivalent to

$$
\begin{equation*}
\gamma_{N} \leq 0 \text { on } \Gamma^{c} \tag{X.4.26}
\end{equation*}
$$

which is the interpenetration inequality (2.2).
Equations (20-22) and (25-26) constitute the strong form corresponding to the weak form given in Eq. (7). This set of equations includes the momentum equation and the traction (natural) boundary conditions on both bodies. On the contact surface, the strong form enforces the momentum balance of the normal tractions and the inequality on the interpenetration rate. The compressive character of the normal tractions follows from the restriction on the Lagrange multiplier field (5).
X.4.3. Contribution of Virtual Power to Contact Surface. At this point, for the purpose of simplifying subsequent proofs, we observe that the only contribution of $\delta \mathscr{P}$ to the conditions on the contact interface is the term in Eq. (12). We call this term $\delta \mathscr{P}_{1}$ and from (12) it can be seen that it is given by

$$
\begin{align*}
\delta \mathcal{P}_{1}\left(\Gamma^{c}\right) & =\int_{\Gamma^{c}}\left(\delta v_{i}^{A} t_{i}^{A}+\delta v_{i}^{B} t_{i}^{B}\right) d \Gamma  \tag{X.4.27}\\
& =\int_{\Gamma^{c}}\left(\delta v_{N}^{A} t_{N}^{A}+\delta v_{N}^{B} t_{N}^{B}+\delta \mathbf{v}_{T}^{A} \cdot \mathbf{t}_{T}^{A}+\delta \mathbf{v}_{T}^{B} \cdot \mathbf{t}_{T}^{B}\right) d \Gamma
\end{align*}
$$

The remaining terms in $\delta \mathcal{P}$ are equivalent to the momentum equation and traction boundary conditions, so by replacing $\delta \mathscr{P}$ by $\delta \mathscr{P}_{1}$ the momentum equation and traction boundary conditions are observed.

If the contact surface is frictionless, then the last two terms in the integrand of (27a) vanish, so the contribution of $\delta \mathcal{P}$ to the contact interace is

$$
\begin{equation*}
\delta \mathcal{P}_{2}\left(\Gamma^{c}\right) \equiv \int_{\Gamma^{c}}\left(\delta v_{N}^{A} t_{N}^{A}+\delta v_{N}^{B} t_{N}^{B}\right) d \Gamma \tag{X.4.28}
\end{equation*}
$$

Replacing $\delta \mathcal{P}$ by $\delta \mathscr{P}_{2}$ implies the momentum equation, the traction boundary conditions, and the frictionless condition (10c). These results will be used in the proofs which follow.
X.4.4. Penalty Method with Rate-dependent Penalty. In the penalty method, the impenetrabilty constraint is imposed as a penalty normal traction along the contact surface. In contrast to the Lagrange multiplier method, the penalty method allows some interpenetration. However, it is easier to implement and is quite widely used. We consider two forms of the penalty method:

1. a penalty which is proportional to the square of the normal interpentration rate $\gamma_{N}$;
2. a penalty which can be an arbitrary function of the interpenetration and its rate.

The second is more useful for applications in nonlinear problems, because a strictly velocity-dependent penalty allows too much interpenetration. However, the interpenetration-rate dependent penalty leads to a form which is of interest in the elastostatic problem when we replace velocities by displacements, so it is included.

In the penalty methods, we use the same test and trial functions for the velocities as in the Lagrange multiplier method, Eqs. (3) and (4), respectively. The equivalence of the weak form to the strong form for the penalty method can be stated as follows:

$$
\begin{equation*}
\text { if } \mathbf{v} \in \mathcal{U} \text { and } \delta \mathcal{P}_{p}(\mathbf{v}, \delta \mathbf{v})=\delta \mathcal{P}+\delta \mathcal{G}_{p}=0 \quad \forall \delta \mathbf{v} \in \mathcal{U}_{0} \tag{X.4.29a}
\end{equation*}
$$

where $\delta \mathcal{G}_{p}=\int_{\Gamma^{C}} \frac{\beta}{2} \delta\left(\gamma_{N}^{2}\right) \mathcal{H}\left(\gamma_{N}\right) d \Gamma$
then the momentum equation and natural boundary conditions are satisfied in the two bodies and the normal tractions on $\Gamma^{c}$ satisfy momentum balance and are compressive, and vice versa

In the above $\mathcal{H}\left(\gamma_{N}\right)$ is the Heaviside step function,

$$
\mathcal{H}\left(\gamma_{N}\right)= \begin{cases}1 & \text { if } \gamma_{N} \geq 0  \tag{X.4.30}\\ 0 & \text { if } \gamma_{N}=0\end{cases}
$$

The functional $\delta \mathcal{P}$ is defined in Eq. (9) and $\beta$ is an arbitrary parameter known as the penalty parameter. The penalty parameter can be a function of the spatial coordinates. The weak form associated with the penalty method is not an inequality; the discontinuous nature of the contact-impact problem is introduced by the Heaviside step function in Eq. (29b). This weak form does not include the impenetrability condition, which is satisfied only approximately in the penalty method.

To show that the weak form implies the strong form, we begin by taking the variation of $\delta \mathcal{G}_{P}$, which gives

$$
\begin{equation*}
\delta \mathcal{G}_{P}=\int_{\Gamma^{C}} \beta \gamma_{N} \delta \gamma_{N} \mathcal{H}\left(\gamma_{N}\right) d \Gamma \tag{X.4.31}
\end{equation*}
$$

Using Eq. (2.3) in the above gives

$$
\delta \mathcal{G}_{P}=\int_{\Gamma^{C}} \beta \gamma_{N}^{+}\left(\delta v_{N}^{A}-\delta v_{N}^{B}\right) d \Gamma
$$

where

$$
\begin{equation*}
\gamma_{N}^{+}=\gamma_{N} \mathcal{H}\left(\gamma_{N}\right) \tag{X.4.34}
\end{equation*}
$$

We then combine the above term with the contact term, $\delta P_{2}\left(\Gamma_{C}\right)$ in Eq. (28), i.e. with what remains from $\delta \mathcal{P}$ after extracting the momentum equation and natural boundary conditions (which means that these strong forms are already implied). This yields

$$
\begin{equation*}
\delta \mathscr{P}_{P}=\int_{\Gamma^{C}}\left[\delta v_{N}^{A}\left(t_{N}^{A}+\beta \gamma_{N}^{+}\right)+\delta v_{N}^{B}\left(t_{N}^{B}-\beta \gamma_{N}^{+}\right)\right] d \Gamma=0 \tag{X.4.33}
\end{equation*}
$$

The arbitrariness of $\delta v_{N}^{A}$ and $\delta v_{N}^{B}$ on $\Gamma^{c}$ then yields

$$
\begin{align*}
& t_{N}^{A}+\beta \gamma_{N}^{+}=0 \quad \text { on } \Gamma^{c}  \tag{X.4.35}\\
& t_{N}^{B}-\beta \gamma_{N}^{+}=0 \tag{X.4.36}
\end{align*} \text { on } \Gamma^{c}
$$

Combining the two above equations gives

$$
\begin{equation*}
t_{N}^{A}=-t_{N}^{B}=-\beta \gamma_{N}^{+} \leq 0 \tag{X.4.37}
\end{equation*}
$$

where the inequality follows from Eq. (34). Thus the weak form implies that the normal tractions satisfy momentum balance and are compressive. The weak form, unlike the Lagrange multiplier technique, does not enforce the continuity of the velocities of the two bodies across the contact interface; in fact, the velocities will be discontinuous across the interface. The magnitude of the discontinuity can be obtained from (37), which gives

$$
\gamma_{N}^{+}=v_{N}^{A}-v_{N}^{B}=t_{N}^{A} / \beta
$$

Thus the discontinuity in the relative normal velocity component is inversely proportional to the penalty parameter $\beta$; as $\beta$ is increased, the diccontinuity will decrease.
X.4.5. Interpenetration-dependent Penalty. The above form of the penalty method often performs quite poorly since it may allow excessive interpenetration. The normal traction is applied only when the relative velocities lead to continued interpenetration. As soon as the relative velocities of contiguous points of the two
surfaces become equal or negative, the normal traction vanishes. Substantial interpenetration may consequently persist in the solution. Therefore, in penalty methods, it is recommended that the normal traction also be a function of the interpenetration as defined in (2.12). For this purpose, we define the following relation for the normal traction :

$$
\begin{equation*}
p=p\left(g_{N}, \gamma_{N}\right) \mathcal{H}\left(g_{N}\right) \tag{X.4.38}
\end{equation*}
$$

where $g_{N}$ is defined in Eq. (2.12). The weak form is then given by Eq. (28) with

$$
\begin{equation*}
\delta \mathcal{G}_{p}=\int_{\Gamma^{C}} \delta \gamma_{N} p d \Gamma \tag{X.4.39}
\end{equation*}
$$

The same procedure as before then gives

$$
\begin{align*}
& t_{N}^{A}+p=0 \text { on } \Gamma^{c}  \tag{X.4.40}\\
& t_{N}^{B}-p=0 \text { on } \Gamma^{c} \tag{X.4.41}
\end{align*}
$$

Combining the two above equations gives

$$
\begin{equation*}
t_{N}^{A}=-t_{N}^{B}=-p\left(g_{N}, \gamma_{N}\right) \mathcal{H}\left(g_{N}\right) \tag{X.4.42}
\end{equation*}
$$

Thus the tractions are always compressive and satisfy momentum balance. The tractions are functions of the interpenetration and rate of interpenetration. An example of a suitable penalty function is

$$
\begin{equation*}
p=\left(\beta_{1} g_{N}+\beta_{2} \gamma_{N}\right) \mathcal{H}\left(\beta_{1} g_{N}+\beta_{2} \gamma_{N}\right) \tag{X.4.43a}
\end{equation*}
$$

where $\beta_{1}, \beta_{2}$ are penalty parameters whose selection is discussed in Section ?. The step function in this expression avoids tensile normal tractions across the interface.
X.4.6. Perturbed Lagrangian Weak Form. The perturbed Lagrangian method is primarily of interest in small displacement elastostatics. In the perturbed Lagrangian method, the weak form is

$$
\begin{equation*}
\text { if } v \in \mathcal{U}, \lambda \in C^{-1} \text { and } \delta \mathcal{P}_{P L}=\delta \mathcal{P}+\delta \mathcal{G}_{P L} \geq 0 \quad \forall \delta v \in \mathcal{U}_{0}, \delta \lambda \in C^{-1} \tag{X.4.44}
\end{equation*}
$$

In the above

$$
\begin{equation*}
\delta \mathcal{G}_{P L}=\int_{\Gamma^{C}} \delta\left(\lambda \gamma_{N}^{+}-\frac{1}{2 \beta} \lambda^{2}\right) d \Gamma \tag{X.4.45}
\end{equation*}
$$

where $\gamma_{N}^{+}$is defined by Eqs. (34) and (2.3) and $\beta$ is a large constant, analogous to a penalty parameter. It can be seen that the second term in the above integrand is a perturbation of the Lagrangian multiplier, Eq. (8); the quadratic perturbation term is small since $\beta$ is large.

In this weak form, the test and trial functions for the Lagrange multiplier are unconstrained. This weak form is equivalent to the momentum equation, the traction boundary condition and the momentum balance and traction inequalities ( 2.9 g ) on the contact interface. It will be shown that as in the penalty method, the impenetrability condition (2.7) is only met approximately.

The equivalence to the strong form is shown as follows. From (45),

$$
\begin{equation*}
\delta \mathcal{G}_{P L}=\int_{\Gamma^{C}}\left(\delta \lambda \gamma_{N}^{+}+\lambda \delta \gamma_{N}-\frac{1}{\beta} \lambda \delta \lambda\right) d \Gamma \tag{X.4.46}
\end{equation*}
$$

Combining $\delta \mathcal{G}_{P L}$ with the terms that emerge from $\delta P$ once the momentum equation, traction boundary conditions and frictionless interface conditions are met, $\delta \mathscr{P}_{2}\left(\Gamma^{c}\right)$ in Eq. (27), yields

$$
\begin{align*}
0=\delta \mathcal{G}_{P L}+\delta \mathcal{P}_{2}= & \int_{\Gamma^{c}} \delta \lambda\left(\gamma_{N}^{+}-\frac{\lambda}{\beta}\right) d \Gamma  \tag{X.4.47}\\
& +\int_{\Gamma^{c}} \delta v_{N}^{A}\left(t_{N}^{A}+\lambda\right)+\delta v_{N}^{B}\left(t_{N}^{B}-\lambda\right) d \Gamma
\end{align*}
$$

Since the test functions $\delta v_{N}^{A}$ and $\delta v_{N}^{B}$ are arbitrary, it follows that

$$
\begin{align*}
t_{N}^{A} & =-\lambda \text { on } \Gamma^{c}  \tag{X.4.48a}\\
t_{N}^{B} & =\lambda \text { on } \Gamma^{c} \tag{X.4.48b}
\end{align*}
$$

The test function $\delta \lambda$ is constrained to be negative, so the variational inequality yields

$$
\begin{equation*}
\lambda=\beta \gamma_{N}^{+} \text {on } \Gamma^{c} \tag{X.4.48c}
\end{equation*}
$$

Combining the above yields

$$
\begin{equation*}
t_{N}^{A}=-t_{N}^{B}=-\beta \gamma_{N}^{+}=-\beta\left(v_{N}^{A}-v_{N}^{B}\right) \mathcal{H}\left(\gamma_{N}\right) \text { on } \Gamma^{c} \tag{X.4.49}
\end{equation*}
$$

So the tractions satisfy momentum balance and are compressive on the contact interface.
The above strong form of the contact surface conditions are almost identical to those which emanate from the penalty method. This similarity is also found in the discrete equations, so the perturbed Lagrangian is a penalty method in disguise.
X.4.7. Augmented Lagrangian. The augmented Lagrangian formulation has been developed to exploit improved methods for solving the Lagrange multiplier problem, c.f. Bertsekas (1984). The weak form is given by

$$
\delta P_{A L}(\mathbf{v}, \delta \mathbf{v}, \lambda, \delta \lambda)=\delta P+\delta G_{A L} \geq 0 \quad \forall \delta \mathbf{v} \in \mathcal{U}_{0}, \delta \lambda \in \mathcal{I}^{-}
$$

$$
\begin{equation*}
\left.\delta G^{A L}=\int_{\Gamma^{C}} \delta^{[ } \lambda \gamma_{N}(\mathbf{v})+\frac{\alpha}{2} \gamma_{N}^{2}(\mathbf{v})\right] d \Gamma \tag{X.4.52}
\end{equation*}
$$

where $v \in \mathcal{U}, \lambda \in \mathcal{I}^{+}\left(\Gamma^{C}\right) ; \gamma_{N}(\mathbf{v})$ is defined by Eq. (2.3) and $\alpha$ is a positive parameter determined as part of the solution process.

The equivalence of this weak form to the strong form is shown in the following. Expanding the integrand in (52) gives

$$
\begin{equation*}
\delta G_{A L}=\int_{\Gamma^{C}}\left[\delta \lambda \gamma_{N}+\lambda\left(\delta v_{N}^{A}+\delta v_{N}^{B}\right)+\alpha \gamma_{N}\left(\delta v_{N}^{A}-\delta v_{N}^{B}\right)\right] d \Gamma \tag{X.4.53}
\end{equation*}
$$

where Eq. (2.3) has been used for $\delta \gamma$. Combining the above with the terms associated with $\delta \mathcal{G}_{P L}$ from Eq. (28) gives

$$
\begin{equation*}
\int_{\Gamma^{C}}\left[\delta \lambda \gamma+\delta v_{N}^{A}\left(\lambda+\alpha \gamma+t_{N}^{A}\right)+\delta v_{N}^{B}\left(\lambda+\alpha \gamma-t_{N}^{B}\right)\right] d \Gamma \geq 0 \tag{X.4.54}
\end{equation*}
$$

Since all of the variations are arbitrary, we obtain that on $\Gamma^{c}$

$$
\begin{array}{ll}
\delta \lambda: & \gamma_{N}=v_{N}^{A}-v_{N}^{B} \leq 0 \\
\delta v_{N}^{A}: & \lambda=-\alpha \gamma-t_{N}^{A} \\
\delta v_{N}^{B}: & \lambda=-\alpha \gamma+t_{N}^{B} \tag{X.4.57}
\end{array}
$$

Eqs. (56) and (57) can be combined to yield

$$
\begin{equation*}
t_{A}^{N}=-t_{B}^{N}=-\lambda-\alpha \gamma \leq 0 \tag{X.4.58}
\end{equation*}
$$

where the inequality follows because $\lambda \geq 0$ and $\gamma \geq 0$ when interpenetration occurs. Thus the normal interface traction is compressive and satisfies momentum balance.
X.4.8. Tangential Tractions by Lagrange Multipliers. All of the above formulations can be modified to handle interface friction laws by adding a term to the weak form which enforces continuity of the tangential tractions. We simply let

$$
\begin{equation*}
\delta \mathcal{P}_{C}=\delta \mathcal{P}+\delta \mathcal{G}_{N}+\delta \mathcal{G}_{T} \tag{X.4.59}
\end{equation*}
$$

where

$$
\delta \mathscr{P}_{C} \geq 0 \text { if } \delta \mathcal{G}_{N}=\delta \mathcal{G}_{L} \operatorname{or} \delta \mathcal{G}_{A L}
$$

and

$$
\begin{equation*}
\delta \mathcal{P}_{C}=0 \text { if } \delta \mathcal{G}_{N}=\delta \mathcal{G}_{P} \operatorname{or} \delta \mathcal{G}_{P L} \tag{X.4.60b}
\end{equation*}
$$

The tangential weak form is given by

$$
\begin{equation*}
\delta \mathcal{G}_{T}=\int_{\Gamma_{C}} \delta \gamma_{T} \cdot \mathbf{t}_{T} d \Gamma \equiv \int_{\Gamma_{C}} \delta \hat{\gamma}_{\alpha} \hat{t}_{\alpha} d \Gamma \tag{X.4.61}
\end{equation*}
$$

where $\mathbf{t}_{T}$ is a traction which is computed by a friction model. We have put hats on the expressions which are expressed in indicial notation to indicate that these components are in the local coordinates of the tangent plane of the contact interface.

To obtain the equations, we take what remains from $\delta \mathcal{P}$ after extracting the momentum equation and traction boundary conditions, Eq. (27a) . The normal kinetic and possibly kinematic conditions are then extracted as indicated in the preceding sections. What remains is combined with $\delta G_{T}$, giving

$$
\begin{gather*}
0=\delta \mathcal{P}_{1}\left(\Gamma_{C}\right)+\delta \mathcal{G}+\delta \mathcal{G}_{T}=\int_{\Gamma^{c}}\left(\delta v_{\alpha}^{A} t_{\alpha}^{A}+\delta v_{\alpha}^{B} t_{\alpha}^{B}+\delta \hat{\gamma}_{\alpha} \hat{t}_{\alpha}\right) d \Gamma  \tag{X.4.62}\\
\equiv \int_{\Gamma^{c}}\left(\delta \mathbf{v}_{T}^{A} \cdot \mathbf{t}_{T}^{A}+\delta \mathbf{v}_{T}^{B} \cdot t_{T}^{B}+\delta \gamma_{T} \cdot \mathbf{t}_{T}\right) d \Gamma
\end{gather*}
$$

Note that $\mathbf{t}_{T}$ differs from $\mathbf{t}_{T}^{A}$ and $\mathbf{t}_{T}^{B} ; \mathbf{t}_{T}$ is the prescribed traction, which can be computed by an interface constitutive equation, whereas $\mathbf{t}_{T}^{A}$ and $\mathbf{t}_{T}^{B}$; are the tractions on the interface which result from the interior stresses by Eqs. ( $2.9 \mathrm{~b}-\mathrm{c}$ ). Using the definition of $\gamma_{T}$, Eq. (2.8) we can write $\delta \gamma_{T}=\delta \mathbf{v}_{T}^{A}-\delta \mathbf{v}_{T}^{B}$. Substituting into the above we have, after rearranging the terms

$$
\begin{equation*}
0=\delta \mathcal{P}_{1}\left(\Gamma_{C}\right)+\delta \mathcal{G}_{T}=\int_{\Gamma^{c}}\left[\delta \mathbf{v}_{T}^{A} \cdot\left(\mathbf{t}_{T}^{A}+\mathbf{t}_{T}\right)+\delta \mathbf{v}_{T}^{B} \cdot\left(\mathbf{t}_{T}^{B}-\mathbf{t}_{T}\right)\right] d \Gamma \tag{X.4.63}
\end{equation*}
$$

From this we can extract

$$
\begin{equation*}
\mathbf{t}_{T}^{A}=-\mathbf{t}_{T} \quad \mathbf{t}_{T}^{B}=\mathbf{t}_{T} \tag{X.4.64}
\end{equation*}
$$

Eliminating $\mathbf{t}_{T}$ from the above we have

$$
\begin{equation*}
\mathbf{t}_{T}^{A}+\mathbf{t}_{T}^{B}=0 \quad \text { or } \quad \hat{t}_{\alpha}^{A}+\hat{t}_{\alpha}^{B}=0 \tag{X.4.65}
\end{equation*}
$$

Thus the additional term $\delta \mathcal{G}_{T}$ in the weak form corresponds to the momentum balance of the tangential tractions on the contact interface. Without this term in the weak form, the tangential tractions vanish, i.e. the interface is frictionless.

This approach can be viewed as considering the $\hat{x}$ and $\hat{y}$ components of the contact surfaces to be prescribed traction surfaces. The traction term in the external power would then be equivalent (61).

When the stick condition applies to a part of the contact interface, it is possible to use a Lagrange multiplier to impose the constraint of no tangential slip. The form of the term is similar to that which imposes the interpenetration condition, (8). It is given by

$$
\begin{equation*}
\delta \mathcal{G}_{T}=\int_{\Gamma_{C}} \delta\left(\gamma_{T} \cdot \mathbf{t}_{T}\right) d \Gamma \equiv \int_{\Gamma_{C}} \delta\left(\hat{\gamma}_{\alpha} \hat{t}_{\alpha}\right) d \Gamma \tag{X.4.66}
\end{equation*}
$$

This tangential weak form is associated with an equality, so if the original weak form to which it is appended is an equality, then the weak form remains an equality, whereas if the original weak is an inequality it remains an inequality. The strong forms corresponding to (66) are (65) and $\gamma_{T}=0$.

## BOX X. 2 Weak Forms

$\delta \mathcal{P}_{C}=\delta \mathscr{P}+\delta \mathcal{G}+\delta \mathcal{G}_{T} \quad$ note $\gamma \equiv \gamma_{N}$
Tangential tractions: $\delta \mathcal{G}=\int_{\Gamma^{c}} \delta \gamma_{T} \cdot \lambda_{T} d \Gamma \equiv \int_{\Gamma^{c}} \delta \hat{\gamma}_{\alpha} \hat{\lambda}_{\alpha} d \Gamma$
Lagrangian: $\quad \delta \mathcal{G}=\delta \mathcal{G}_{L}=\int_{\Gamma^{c}} \delta(\lambda \gamma) d \Gamma, \quad \delta P_{C} \geq 0$
Penalty: $\quad \delta \mathcal{G}=\delta \mathcal{G}_{P}=\int_{\Gamma^{c}} \frac{1}{2} \beta \delta\left(\gamma^{2}\right) d \Gamma, \quad \delta P_{C}=0$
Augmented Lagrangian: $\quad \delta \mathcal{G}=\delta \mathcal{G}_{A L}=\int_{\Gamma^{c}} \delta\left(\lambda \gamma+\frac{\alpha}{2} \gamma^{2}\right) d \Gamma, \quad \delta P_{C} \geq 0$
Perturbed Lagrangian: $\quad \delta \mathcal{G}_{N}=\delta \mathcal{G}_{P L}=\int_{\Gamma^{c}} \delta\left(\lambda \gamma-\frac{1}{2 \beta} \lambda^{2}\right) d \Gamma, \quad \delta \mathcal{P}_{C}=0$

## X. 5 FINITE ELEMENT DISCRETIZATION

X.5.1 Overview. In the following, the finite element equations for the various treatments of contact-impact are developed. The weak statements for all of the approaches to the contact-impact problem, (penalty, Lagrange multiplier, etc.) involve a sum of the standard virtual power and a contribution from the contact interface. The standard virtual power is discretized exactly as in the absence of contact, so we will use the results developed in Chapter 4. This Section concentrates on the discretization of the various contact interface weak forms.

The developments that follow here are applicable to both updated and total Lagrangian formulations. However in total Lagrangian formulations, the contact interface conditions must be imposed in terms of the tractions on the deformed surface areas. The following discretizations are also applicable to ALE formulations as long as the nodes on the contact surface are Lagrangian. They are not directly applicable to Eulerian formulations since we assume that we have at our disposal a referential coordinate that describes the contact surface. Such a coordinate system cannot easily be defined in an Eulerian mesh. In a Lagrangian mesh, the contact surface corresponds to a subset of the boundary of the mesh.

We will first develop the FEM discretization for the Lagrangian multiplier method in indicial notation. Indicial notation enables us to go through some subtle steps which will subsequently be glossed over in the matrix derivations; anyone who wishes to replicate these steps for other formulations can rederive these in indicial notation.
X.5.2 Lagrange Multiplier Method. For the purpose of developing a finite element discretization, the velocities and the Lagrange multipliers must be approximated as functions of space and time. The velocity $\mathbf{v}(\mathbf{X}, t)$ is approximated by $C^{0}$ interpolants in each body as in the single body problem; as can be seen from (4.3), continuity of velocities between two bodies across the contact interface is not built into the approximation, so the interpenetration condition will emanate from the discretization of the weak form. The velocity field can also be expressed in terms of the reference coordinates $\zeta$ on the contact surface when needed. As in Chapter 4, we note that the approximation of the velocity field directly defines the approximation of the displacement field.

The finite element approximation for the velocity field is expressed in terms of the material coordinates since we are dealing with a Lagrangian mesh. It can alternatively be written in terms of the element reference coordinates, since as pointed out in Chapter 4 the two sets of coordinates are equivalent. To clarify certain issues, we will initially discard the summation convention on repeated nodal indices and indicate sums explicitly. The velocity field is

$$
\begin{align*}
& v_{i}^{A}(\mathbf{X}, t)=\sum_{I \in \Omega^{A}} N_{I}(\mathbf{X}) v_{I i}^{A}(t)  \tag{X.5.1a}\\
& v_{i}^{B}(\mathbf{X}, t)=\sum_{I \in \Omega^{B}} N_{I}(\mathbf{X}) v_{I i}^{B}(t) \tag{X.5.1b}
\end{align*}
$$

If the node numbers of bodies $A$ and $B$ are treated as distinct, then the two velocity fields can be written as a single expression

$$
\begin{equation*}
v_{i}(\mathbf{X}, t)=\sum_{I \in \Omega^{A} \cup \Omega^{B}} N_{I}(\mathbf{X}) v_{I i}(t) \equiv N_{I}(\mathbf{X}) v_{I i}(t) \tag{X.5.1c}
\end{equation*}
$$

where the last expression uses the implicit summation convention on node numbers

The Lagrange multiplier field $\lambda(\zeta, t)$, as can be seen from (4.5) and (4.6), is approximated by a $C^{-1}$ field on the contact surface. The Lagrange multiplier field need only be piecewise continuous because its derivatives do not appear in any of the weak forms. We will use the element coordinates of the master body, $\zeta$, as the independent variables in approximating the Lagrange multiplier field.

$$
\begin{equation*}
\lambda(\zeta, t)=\sum_{I \in \Gamma_{\lambda}^{c}} \Lambda_{I}(\zeta) \lambda_{I}(t) \equiv \Lambda_{I}(\zeta) \lambda_{I}(t) \quad \lambda(\zeta, t) \geq 0 \tag{X.5.2}
\end{equation*}
$$

The shape functions for the Lagrange multiplier field often differ from those used for the velocities, so different symbols have been used for the two approximations. Moreover, when the nodes of bodies $A$ and $B$ are not coincident, the mesh structure differs from that for the velocity field and a subscript $\lambda$ has been added to $\Gamma^{c}$ to indicate this fact. The need for a different nodal structure for the Lagrange multipliers is discussed in more detail later.

The test functions are given by

$$
\begin{align*}
& \delta v_{i}(\mathbf{X})=N_{I}(\mathbf{X}) \delta v_{I i}  \tag{X.5.3}\\
& \delta \lambda(\zeta)=\Lambda_{I}(\zeta) \delta \lambda_{I} \quad \delta \lambda(\zeta) \leq 0 \tag{X.5.4}
\end{align*}
$$

where the implicit sums are defined in Eqs. (1) and (2).
To develop the semidiscrete equations, the above approximations for the velocity and Lagrange multiplier fields and the test functions are substituted into the weak form, Eq. (BX.2.3), which is repeated below:

$$
\begin{equation*}
\delta \mathcal{P}+\int_{\Gamma^{c}} \delta\left(\lambda \gamma_{N}\right) d \Gamma \geq 0 \tag{X.5.5}
\end{equation*}
$$

The terms emerging from $\delta \mathcal{P}$ are identical to the nodal forces developed in Chapter 4, so they will not be rederived; the results are given in Table B4.1. From Eq. (B4.1.?) it follows that

$$
\begin{equation*}
\delta \mathcal{P}=\delta v_{I i}\left(f_{I i}^{i n t}-f_{I i}^{e x t}+M_{I J i j} \dot{j}_{J j}\right) \equiv \delta \mathbf{d}^{T}\left(\mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{M} \ddot{\mathbf{d}}\right) \equiv \delta \mathbf{d}^{T} \mathbf{f}^{\text {res }} \tag{X.5.6}
\end{equation*}
$$

The interpenetration rate can be expressed in terms of the nodal velocities by using (2.7) and (8):

$$
\begin{equation*}
\gamma_{N}=\sum_{I \in \Gamma^{C} \cap \Gamma^{A}} N_{I} v_{I i}^{A} n_{i}^{A}+\sum_{I \in \Gamma^{C} \cap \Gamma^{B}} N_{I} v_{I i}^{B} n_{i}^{B} \tag{X.5.7}
\end{equation*}
$$

where the first sum, as indicated, is over the nodes of body $A$ which are on the contact interface, and the second sum is over the nodes of body $B$ which are on the contact interface. If we assign these nodes distinct node numbers, we can eliminate the distinction between nodes of body $A$ and $B$ and express the above as

$$
\begin{equation*}
\gamma_{N}=N_{I} v_{\text {In }} \tag{X.5.8}
\end{equation*}
$$

(lower case $n$ is used to indicate some of the normal components in this equation and the following). The range of the sum on the repeated index $I$ is implied and defined in (7). The normal components are defined as in (2.5) by

$$
\begin{equation*}
v_{I n}=v_{I i}^{A} n_{i}^{A} \text { if Iin } A, v_{I n}=v_{I i}^{B} n_{i}^{A} \text { if } \operatorname{Iin} B \tag{X.5.9}
\end{equation*}
$$

Then using the approximations (1-4) it follows that

$$
\begin{equation*}
\int_{\Gamma^{c}} \delta\left(\lambda \gamma_{N}\right) d \Gamma=\delta v_{I n} \hat{G}_{I J}^{T} \lambda_{J}+\delta \lambda_{I} \hat{G}_{I J} v_{J n} \tag{X.5.10}
\end{equation*}
$$

where

$$
\begin{equation*}
\hat{G}_{I J}=\int_{\Gamma^{c}} \Lambda_{I} N_{J} d \Gamma \tag{X.5.11}
\end{equation*}
$$

A superposed hat has been placed on $\hat{G}_{I J}$ to indicate that it pertains to the velocities in the local coordinate system of the contact interface. Combining Eqs. (5), (6), and (10) we can write the weak form as

$$
\begin{equation*}
\sum_{I \in \Omega} \delta v_{I I} f_{I i}^{r e s}+\sum_{I \in \Gamma_{\lambda}^{c}} \delta v_{I n} \hat{G}_{I J}^{T} \lambda_{J}+\sum_{I \in \Gamma_{\lambda}^{c}} \delta \lambda_{I} \hat{G}_{I J} v_{J n} \geq 0 \tag{X.5.12}
\end{equation*}
$$

where the implicit sum on the index $J$ holds, but the sums on the index $I$ are explicitly stated to indicate the relevant nodes.

The governing equations must be extracted carefully because of the inequalities and the different roles different velocity components play in this equation. The equations for nodes which are not on the contact interface can be directly extracted from the first sum since the nodal velocities are arbitrary, which yields the standard nodal equations of motion at the nodes which are not on the contact interface

$$
\begin{equation*}
f_{I i}^{\text {res }}=0 \text { or } M_{I J} \dot{\mathbf{v}}_{J j}=f_{I i}^{e x t}-f_{I i}^{\text {int }} \quad \text { for } I \in \Omega-\Gamma^{c} \tag{X.5.13}
\end{equation*}
$$

To obtain the equations on the contact interface, what remains of the first sum after extracting Eq. (13) is rewritten in the local coordinate systems of the contact interface and combined with the second sum, giving

$$
\begin{equation*}
\sum_{I \in \Gamma^{c}}\left(\delta v_{I n} f_{I n}^{r e s}+\delta \hat{v}_{I \alpha} \hat{f}_{I \alpha}^{r e s}+\delta v_{I n} \hat{G}_{I J}^{T} \lambda_{J}\right)+\sum_{I \in \Gamma_{\lambda}^{c}} \delta \lambda_{I} \hat{G}_{I J} v_{J n} \geq 0 \tag{X.5.14}
\end{equation*}
$$

Since the tangential nodal velocities are unconstrained, the weak inequality yields an equality for the coefficients of the nodal velocities. First we set the coefficient of $\delta \hat{v}_{I \alpha}$ to zero, which gives

$$
\begin{equation*}
\hat{f}_{I \alpha}^{\text {res }}=0 \text { or } M_{I J} \dot{\hat{v}}_{J \alpha}=\hat{f}_{I \alpha}^{e x t}-\hat{f}_{I \alpha}^{\text {int }} \text { for } I \in \Gamma^{c} \tag{X.5.15}
\end{equation*}
$$

The equation for the normal component at the contact interface nodes involves the first and third terms of the first sum in (13) and gives

$$
\begin{equation*}
f_{I n}^{r e s}+\hat{G}_{I J}^{T} \lambda_{J}=0 \text { or } M_{I J} \dot{v}_{J n}+f_{I n}^{e x t}-f_{I n}^{i n t}+\hat{G}_{I J}^{T} \lambda_{J}=0 \quad \text { for } I \in \Gamma^{c} \tag{X.5.16}
\end{equation*}
$$

To extract the equations associated with the Lagrange multipliers, we note that the variations of the nodal Lagrange multipliers must be negative. Therefore the inequality (5) implies

$$
\begin{equation*}
\hat{G}_{I J} v_{J n} \leq 0 \tag{X.5.17}
\end{equation*}
$$

In addition, we have from Eq. (4.6) the requirement that the test function for the Lagrange multiplier field must be positive

$$
\begin{equation*}
\lambda(\zeta, t) \geq 0 \tag{X.5.18}
\end{equation*}
$$

The above inequality is difficult to enforce. For elements with piecewise linear displacements along the edges, this condition is often enforced only at the nodes by $\lambda_{I} \geq 0$. This simplification is only appropriate with piecewise linear approximations since the local minima of the Lagrange multipliers then occur at the nodes.

The above equations, in conjunction with the strain-displacement equations and the constitutive equation, comprise the complete system of equations for the semidiscrete model. The semidiscrete equations consist of the equations of motion and the contact interface conditions. The equations of motion for nodes not on the contact interface are unchanged from the unconstrained case. On the contact interface, additional forces $\hat{G}_{I J} \lambda_{J}$ which represent the normal contact tractions appear. In addition, the impenetrability constraint in weak form (17) must be imposed. Like the equations without contact, the semidiscrete equations are ordinary differential equations, but the variables are subject to algebraic inequality constraints on the velocities and the Lagrange multipliers. These inequality constraints substantially complicate the time integration, since the smoothness which is implicitly assumed by most time integration procedures is lost.

For purposes of implementation, it is convenient to write the above equations in matrix form in global components. Let the interpenetration rate be defined in terms of the nodal velocities by

$$
\begin{equation*}
\gamma=\Phi_{I i}(X) v_{I i}(t) \tag{X.5.19}
\end{equation*}
$$

where

$$
\Phi_{I i}(X)=\left\{\begin{array}{l}
N_{I} n_{i}^{A} \text { if IonA }  \tag{X.5.20}\\
N_{I} n_{i}^{B} \text { if Ion } B
\end{array}\right.
$$

The contact weak term is then given by

$$
\begin{equation*}
\delta \mathcal{G}_{L}=\int_{\Gamma^{c}} \delta\left(\lambda_{I} \Lambda_{I} \Phi_{J j} v_{J j}\right) d \Gamma=\lambda^{T} \mathbf{G} \mathbf{v} \tag{X.5.21a}
\end{equation*}
$$

where

$$
\begin{equation*}
G_{J j I}=\int_{\Gamma^{c}} \Lambda_{I} \Phi_{J j} d \Gamma \quad \mathbf{G}=\int_{\Gamma^{c}} \Lambda^{\mathrm{T}} \Phi d \Gamma \tag{X.5.21b}
\end{equation*}
$$

The equations of motion can be written in matrix form by combining this form with matrix forms of the internal, external and inertial power, which gives

$$
\begin{equation*}
\delta \mathbf{v}^{T}\left(\mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{M} \ddot{\mathbf{d}}\right)+\delta\left(\mathbf{v}^{T} \mathbf{G}^{T} \lambda\right)=0 \quad \forall \delta \mathbf{v} \in \mathcal{U}^{h} \quad \forall \delta \lambda \in g^{h-} \tag{X.5.22}
\end{equation*}
$$

We will skip the steps represented by Eqs. (7-17) and invoke the arbitrariness of $\delta \mathbf{v}$ and $\delta \lambda$. The matrix forms of the equations of motion and the interpenetration condition are

$$
\begin{equation*}
\mathbf{M} \ddot{\mathbf{d}}+\mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{G}^{T} \lambda=\mathbf{0} \tag{X.5.23a}
\end{equation*}
$$

$\mathbf{G v} \leq 0$
The construction of the interpolation, and hence the nodal arrangement, for the Lagrange multipliers poses some difficulties. In general, the nodes of the two contacting bodies are not coincident, as shown in Fig. 5.1. Therefore it is necessary to develop a scheme to deal with noncontiguous nodes. One possibility is indicated in Fig. 5.1, where the nodes for the Lagrange multiplier field are chosen to be the nodes of the master body which are in contact. This is a simple


Figure. X.5.1. Nodal arrangements for two contacting bodies with noncontiguous nodes showing (a) a Lagrange multiplier mesh based on the master body and (b) an independent Lagrange multiplier mesh.
scheme, but when the nodes of body B are much more finely spaced a coarse nodal structure for the Lagrange multipliers will lead to interpenetration. An alternative is to place Lagrange multiplier nodes wherever a node appears in either body $A$ or $B$, as shown in Fig. 5.1b. The disadvantage of that scheme is that when nodes of A and B are closely spaced, the Lagrange multiplier element is then very small. This can lead to illconditioning of the equations.
X.5.3. Assembly of Interface Matrix. The G matrix can be assembled from "element" matrices like any other global finite element matrix. To illustrate the assembly procedure, let the nodal velocities and Lagrange multipliers of element $e$ be expressed in terms of the global matrices by

$$
\begin{equation*}
\mathbf{v}_{e}=\mathbf{L}_{e} \mathbf{v} \quad \lambda_{e}=\mathbf{L}_{e}^{\lambda} \lambda \tag{X.5.24a}
\end{equation*}
$$

with identical relations for the test functions

$$
\delta \mathbf{v}_{e}=\mathbf{L}_{e} \mathbf{v} \quad \delta \lambda_{e}=\mathbf{L}_{e}^{\lambda} \delta \lambda
$$

Substituting into (18) gives

$$
\lambda^{\mathbf{T}} \mathbf{G} \mathbf{v}=\int_{\Gamma^{c}} \lambda \gamma d \Gamma=\sum_{e} \int_{\Gamma_{e}^{c}} \lambda \gamma d \Gamma=\lambda^{T}\left(\mathbf{L}_{e}^{\lambda}\right)^{T} \int_{\Gamma_{e}^{c}} \Phi^{T} \Lambda d \Gamma \mathbf{L}_{e} \mathbf{v}
$$

Since (18) must hold for arbitrary $\dot{\mathbf{d}}$ and $\lambda$ it can be seen by comparing the first and last term of the above that

$$
\begin{equation*}
\mathbf{G}=\sum_{e}\left(\mathbf{L}_{e}^{\lambda}\right)^{T} \mathbf{G}_{e} \mathbf{L}_{e}, \quad \mathbf{G}_{e}=\int_{\Gamma_{e}^{c}} \Lambda^{T} \phi d \Gamma \tag{X.5.25}
\end{equation*}
$$

Thus the assembly of $\mathbf{G}$ from $\mathbf{G}_{e}$ is identical to assembly of global matrices such as the stiffness matrix.
X.5.4. Lagrange Multipliers for Small-Displacement Elastostatics. We will call the analysis of small-displacement problems with linear, elastic materials smalldisplacement elastostatics. We have used the nomenclature of small-displacement, elastostatics rather than linear elasticity because these problems are not linear due to the inequality constraint on the displacements which arises from the contact condition. For small-displacement elastostatics, the governing relations for the impenetrability constraint can be obtained from the preceding by replacing the velocities by the displacements. Thus Eq. (2.7) and (19) are replaced by

$$
\begin{equation*}
g_{N}=\left(\mathbf{u}^{A}-\mathbf{u}^{B}\right) \cdot \mathbf{n}^{A} \leq 0 \text { on } \Gamma^{c} \quad g_{N}=\Phi \mathbf{d} \tag{X.5.26}
\end{equation*}
$$

The discretization procedure is then identical to the above except for substituting velocities by displacements and omitting the inertia, giving

$$
\begin{equation*}
\delta \mathbf{d}^{T}\left(\mathbf{f}^{\text {int }}-\mathbf{f}^{e x t}\right)+\delta\left(\mathbf{d}^{T} \mathbf{G} \lambda\right)=0 \quad \forall \delta \mathbf{d} \in \mathcal{U} \quad \forall \delta \lambda \in \mathcal{I}^{-} \tag{X.5.27}
\end{equation*}
$$

Since the internal nodal forces are not effected by contact, for the small displacement elastostatic problem they can be expressed in terms of the stiffness matrix by

$$
\begin{equation*}
\mathbf{f}^{i n t}=\mathbf{K d} \tag{X.5.26a}
\end{equation*}
$$

Taking the variation of the second term and using the arbitrariness of $\delta \mathbf{d}$ and the arbitrary but negative character of $\delta \lambda$ gives

$$
\left[\begin{array}{cc}
\mathbf{K} & \mathbf{G}^{T}  \tag{X.5.27}\\
\mathbf{G} & \mathbf{0}
\end{array}\right\}\left[\begin{array}{l}
\mathbf{d} \\
\lambda
\end{array}\right\} \leq\left\{\begin{array}{c}
\mathbf{f}^{e x t} \\
\mathbf{0}
\end{array}\right\}
$$

This is the standard form for Lagrange multiplier problems except that an equality has been replaced by an inequality in the second matrix equation.

If we recall other Lagrange multiplier problems, two properties of this system come to mind:

1. the system of linear algebraic equations is no longer positive definite;
2. the equations as given above are not banded and it is difficult to find an arrangement of unknowns so that they are banded;
3. the number of unknowns is increased as compared to the system without the contact constraints.

In addition, for the contact problem, the solution of the equations is complicated by the presence of the inequalities. These are very difficult to deal with and often the small-displacement, elastostatic problem is posed as a quadratic programming problem, see Section ?. These difficulties also arise in the nonlinear implicit solution of contact problems.

A major disadvantage of the Lagrange multiplier method is the need to set up a nodal and element topology for the Lagrange multipliers. As we have seen in the simple two dimensional example, this can introduce complications even in two dimensions. In three dimensions, this task is far more complicated. In penalty methods we see there is no need to set up an additional mesh.

In comparison to the penalty method, the advantage of the Lagrange multiplier method is that there are no user-set parameters and the contact constraint can be met almost exactly when the nodes are contiguous. When the nodes are not contiguous, impenetrability can be violated slightly, but not as much as in penalty methods. However, for high velocity impact, Lagrange multipliers often result in very noisy solutions. Therefore, Lagrange multiplier methods are most suited for static and low velocity problems.
X.5.5. Penalty Method for Nonlinear Frictionless Contact. The nonlinear discretization is developed only for the second form of the penalty method, (X.4.47). In the penalty method only the velocity field needs to be approximated. Again, the velocity field is $C^{0}$ within each body, but no stipulation of continuity between bodies need be made. Continuity between bodies on the contact interface is enforced by the penalty method. We only develop the weak penalty term

$$
\begin{equation*}
\delta \mathcal{G}_{p}=\int_{\Gamma^{c}} \delta \gamma p(g, \gamma) d \Gamma \tag{X.5.28}
\end{equation*}
$$

since the other weak terms are unchanged from the unconstrained problem. Substituting

$$
\begin{equation*}
\delta \mathcal{G}_{P}=\delta \mathbf{v}^{T} \int_{\Gamma^{c}} \phi^{T} p d \Gamma \equiv \delta \mathbf{v}^{T} \mathbf{f}^{c} \tag{X.5.29}
\end{equation*}
$$

where $\phi$ the second equality defines $\mathbf{f}^{c}$ by

$$
\begin{equation*}
\mathbf{f}^{c}=\int_{\Gamma^{c}} \phi^{T} p d \Gamma \tag{X.5.30}
\end{equation*}
$$

Note the similarity of this formula to that for the internal forces; they express the same thing, the relation between discrete forces and continuous tractions. Using (29) and (6) in the weak form (4.28) with (4.39) the above definition of $\mathbf{f}^{c}$ gives

$$
\begin{equation*}
\delta \mathscr{P}=\delta \mathbf{v}^{T} \mathbf{f}^{r e s}+\delta \mathbf{v}^{T} \mathbf{f}^{c} \tag{X.5.31}
\end{equation*}
$$

So using the arbitrariness of $\delta \mathbf{v}$ and (5.6) gives

$$
\begin{equation*}
\mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{M a}+\mathbf{f}^{c}=0 \tag{X.5.32}
\end{equation*}
$$

Thus in the penalty method the number of equations is unchanged from the unconstrained problem. The inequalities (B1.3) do not appear explicitly among the discrete equations but are enforced by appearance of the step function in the calculation of the contact penalty forces by (30) and (4.38).
X.5.6. Penalty for Small-Displacement Elastostatics. For smalldisplacement elastostatics, we replace velocities by displacements as previously. Equation (4.43a) with $\beta_{2}=0$ and (26b) give

$$
\begin{equation*}
p=\beta_{1} g_{N}=\beta_{1} \phi \mathbf{d} \tag{X.5.33}
\end{equation*}
$$

Substituting the above into (30) gives

$$
\mathbf{f}^{c}=\int_{\Gamma^{c}} \phi^{T} p\left(g_{N}\right) \mathcal{H}(\gamma) d \Gamma=\int_{\Gamma} \beta_{1} \phi^{T} \phi \mathcal{H}(\gamma) d \Gamma \mathbf{d}
$$

or

$$
\begin{equation*}
\mathbf{f}^{c}=\mathbf{P}_{c} \mathbf{d}, \quad \mathbf{P}_{c}=\int_{\Gamma} \beta_{1} \phi^{T} \phi \mathcal{H}(\gamma) d \Gamma \tag{X.5.34}
\end{equation*}
$$

Substituting (34) and (26a) into (32) after dropping the inertial term, gives,

$$
\begin{equation*}
\left(\mathbf{K}+\mathbf{P}_{c}\right) \mathbf{d}=\mathbf{f}^{e x t} \tag{X.5.35}
\end{equation*}
$$

This is a system of algebraic equations of the same order as the problem without contact impact. The contact interface constraints appear strictly through the penalty forces $\mathbf{P}_{c} \mathbf{d}$. The algebraic equations are not linear because as can be seen from (34), the matrix $\mathbf{P}_{c}$ involves the Heaviside step function of the gap, which depends on the displacements.

In contrast to the Lagrange multiplier methods it can be seen that:

1. the number of unknowns does not increase due to the enforcement of the contact constraints.
2. the system equations remain positive definite since $\mathbf{K}$ is positive definite and $\mathbf{G}$ is positive definite.

The disadvantage of the penalty approach is that the enforcement of the impenetrability condition is only approximate and its effectiveness depends on the appropriateness of the penalty parameters. If the penalty parameters is too small, excessive interpenetration occurs causing errors in the solution. In impact problems, small penalty parameters reduce the maximum computed stresses. We have seen some shenanigans in calculations where analysts met stress criteria by reducing the penalty parameters. Picking the correct
penalty parameter is a challenging problem. Some guidelines are given in Section ?, where we discuss implementation of various solution procedures with penalty methods.
X.5.7. Augmented Lagrangian. In the augmented Lagrangian method, the weak contact term is

$$
\begin{equation*}
\delta \mathcal{G}_{A L}=\int_{\Gamma^{c}} \delta\left(\lambda \gamma+\frac{\alpha}{2} \gamma^{2}\right) d \Gamma \tag{X.5.36}
\end{equation*}
$$

Using the approximation for the velocity $v(\mathbf{X}, t)$ and the Lagrange multiplier $\lambda(\xi, t)$ gives

$$
\begin{equation*}
\delta \mathcal{G}_{A L}=\int_{\Gamma_{C}} \delta\left(\lambda^{T} \Lambda^{T} \phi \mathbf{v}+\frac{\alpha}{2} \mathbf{v}^{T} \phi^{T} \phi \mathbf{v}\right) d \Gamma \tag{X.5.37}
\end{equation*}
$$

Taking the variations gives

$$
\begin{equation*}
\delta \mathcal{G}_{A L}=\delta \lambda^{T} \mathbf{G} \mathbf{v}+\delta \mathbf{v}^{T} \mathbf{G}^{T} \lambda+\delta \mathbf{v}^{T} \mathbf{P}_{c}(\alpha) \mathbf{v} \tag{X.5.38}
\end{equation*}
$$

where $\mathbf{P}_{c}(\alpha)$ is defined by (34). Writing out the weak form $\delta P_{A L}=\delta \mathcal{P}+\delta \mathcal{G}_{A L} \geq 0$ using Eqs. (36-38) then gives

$$
\begin{align*}
& \mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{M a}+\mathbf{G}^{T} \lambda+\mathbf{P}_{c} \mathbf{v}=0  \tag{X.5.40a}\\
& \mathbf{G v} \leq 0 \tag{X.5.40b}
\end{align*}
$$

Comparing Eqs. (40) with (23) and (35), we can see that the augmented Lagrangian method gives contact forces which are a sum of those in the Lagrangian method and the penalty method. The impenetrability constraint (40b), is the same as in the Lagrange multiplier method.

For small-displacement elastostatics, we use the same procedure as before. We change the dependent variables to displacements so we replace the nodal velocities by nodal displacements, and using( ??) and (27a), the counterpart of Eqs. (39) and (40)

$$
\left[\begin{array}{cc}
\mathbf{K}+\mathbf{P}_{c} & \left.\mathbf{G}^{T}\right\rceil\left\{\begin{array}{l}
\mathbf{d} \\
\mathbf{G}
\end{array} \mathbf{O}\right.
\end{array}\right]=\left\{\begin{array}{c}
\mathbf{f}^{\text {ext }}  \tag{X.5.41}\\
\lambda
\end{array}\right\} \leq\left\{\begin{array}{c}
\mathbf{O}
\end{array}\right\}
$$

which further illustrates that the augmented Lagrangian method is a synthesis of penalty and Lagrange multiplier methods, Eqs. (27) and (35).
X.5.8. Perturbed Lagrangian. The semidiscretization of the perturbed Lagrangian formulation is obtained by using (4.45) with velocity and Lagrange multiplier approximations are given by Eqs. (1) and (2), respectively. We won't go through the steps, since they are identical to the previous discretizations. The discrete equations are

$$
\begin{equation*}
\mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{M a}+\mathbf{G}^{T} \lambda=\mathbf{O} \tag{X.5.42}
\end{equation*}
$$

$$
\begin{equation*}
\mathbf{G v}-\mathbf{H} \lambda=\mathbf{O} \tag{X.5.43}
\end{equation*}
$$

Equation (42) corresponds to the momentum equation, Eq. (43) to the impenetrability condition. The matrix $\mathbf{G}$ is defined by Eq. (21b) and

$$
\begin{equation*}
\mathbf{H}=\int_{\Gamma^{c}} \frac{1}{\beta} \Lambda^{T} \Lambda d \Gamma \tag{X.5.44}
\end{equation*}
$$

The constraint equations (43) can be eliminated to yield a single system of equations. Solving Eq.(43) for $\lambda$ and substituting into (42) gives

$$
\begin{equation*}
\mathbf{f}^{i n t}-\mathbf{f}^{e x t}+\mathbf{M a}+\mathbf{G}^{T} \mathbf{H}^{-1} \mathbf{G}=0 \tag{X.5.45}
\end{equation*}
$$

The above is similar to the discrete penalty equation (35) with the penalty parameter $\beta$ appearing through $\mathbf{H}$ in (44). The last term in the above equations represents the contact forces.

The semidiscrete equations for small-displacement elastostatics for the perturbed Lagrangian methods are

$$
\left[\begin{array}{cc}
\mathbf{K} & \left.\mathbf{G}^{T}\right\rceil  \tag{X.5.46}\\
\mathbf{G} & -\mathbf{H}
\end{array}\right]\left\{\begin{array}{l}
\mathbf{d} \\
\lambda
\end{array}\right\}=\left\{\begin{array}{c}
\mathbf{f}^{e x t} \\
\mathbf{O}
\end{array}\right\}
$$

Comparing the above to the Lagrangian method, Eq. (27), we can see that it differs only in the lower left submatrix, which is $\mathbf{0}$ in the Lagrangian method but consists of the matrix $\mathbf{H}$ in the perturbed Lagrangian method.

BOX X. 3 Semidiscrete Equations for Nonlinear Contact
$\mathbf{f}=\mathbf{f}^{e x t}-\mathbf{f}^{i n t}$
Lagrange Multiplier

$$
\mathbf{M a}-\mathbf{f}+\mathbf{G}^{T} \lambda=0, \quad \mathbf{G v} \leq 0, \quad \lambda(x) \geq 0
$$

Penalty
$\mathbf{M a}-\mathbf{f}+\mathbf{f}^{c}=0, \quad \mathbf{f}^{c}=\int_{\Gamma^{c}} \Phi^{T} p\left(g_{N}\right) \mathcal{H}\left(g_{N}\right) d \Gamma$
Augmented Lagrangian

$$
\mathbf{M a}-\mathbf{f}+\mathbf{G}^{T} \lambda+\mathbf{P}_{c} \mathbf{v}=0, \quad \mathbf{G v} \leq 0
$$

Perturbed Lagrangian
$\mathbf{M a}-\mathbf{f}+\mathbf{G}^{T} \lambda=0, \quad \mathbf{G v}-\mathbf{H} \lambda=0$

$$
\mathbf{G}=\int_{\Gamma^{c}} \Lambda^{T} \phi d \Gamma \quad \mathbf{H}=\int_{\Gamma^{c}} \Lambda^{T} \Lambda d \Gamma \quad \mathbf{P}_{c}=\int_{\Gamma^{c}} \alpha \phi^{T} \phi d \Gamma
$$



Figure X.5.1. One dimensional example of contact; example 1.

Example X.5.1. Finite Element Equations for One Dimensional Contact-Impact. Consider the two rods shown in Fig. X.5.1. We consider a rod of unit cross-sectional area. The contact interface consists of the nodes at the ends of the rods, which are numbered 1 and 2. The unit normals, as shown in Fig. X.5.1, are $n_{x}^{A}=1, n_{x}^{B}=-1$. The contact interface in one-dimensional problems is rather odd since it consists of a single point. The velocity fields in the two elements which border the contact interface are given by

$$
\begin{equation*}
v(\xi, t)=\mathbf{N}(\xi, t) \dot{\mathbf{d}}=\left[\xi^{A}, 1-\xi^{B}, \xi^{B}\right] \dot{\mathbf{d}} \tag{X.5.47}
\end{equation*}
$$

where the column matrix of nodal velocities is

$$
\dot{\mathbf{d}}^{T}=\left[\begin{array}{lll}
v_{1} & v_{2} & v_{3} \tag{X.5.48}
\end{array}\right]
$$

The G matrix is given by Eqs. (20) and (21); in a one-dimensional problem, the integral is replaced by a single function value, with the function evaluated at the contact point:

$$
\begin{align*}
\mathbf{G}^{T} & =\left.\left[\xi^{A} \cdot n^{A},\left(1-\xi^{B}\right) n^{B}, \xi^{B}\right]\right|_{\xi^{A}=1, \xi^{B}=0} \\
& =[(1)(+1), 1(-1), 0]  \tag{X.5.49}\\
& =\left[\begin{array}{lll}
1, & -1, & 0
\end{array}\right]
\end{align*}
$$

The impenetrability condition in rate form, (23b), is given by

$$
\mathbf{G}^{T} \dot{\mathbf{d}} \leq 0 \text { or }\left[\begin{array}{lll}
1 & -1 & 0 \tag{X.5.50}
\end{array}\right] \dot{\mathbf{d}}=v_{1}-v_{2} \leq 0
$$

The last equation can easily be obtained by inspection: when the two nodes are in contact, the velocity of node 1 must be less or equal than the velocity of node 2 to preclude overlap. If they are equal, they remain in contact, whereas when the inequality holds, they release. These conditions are not sufficient to check for initial contact, which should be checked in terms of the nodal displacements: $x_{1}-x_{2} \geq 0$ indicates contact has occurred during the previous time step.

Since there is only one point of contact, only a single Lagrange multiplier appears in the equations of motion. The equations of motion, Eqs. (BX.3.2) are then

$$
\begin{align*}
& \left\lceil\begin{array}{lll}
M_{11} & M_{12} & \left.M_{13}\right\rceil \\
M_{21} & M_{22} & M_{23} \\
M_{31} & M_{32} & M_{33}
\end{array} \left\lvert\,\left\{\begin{array}{l}
\ddot{d}_{1} \\
\ddot{d}_{2} \\
\ddot{d}_{3}
\end{array}\right\}-\left\{\begin{array}{l}
f_{1} \\
f_{2} \\
f_{3}
\end{array}\right\}+\left\{\begin{array}{c}
1 \\
-1 \\
0
\end{array}\right\} \lambda_{1}=0\right.\right. \tag{X.5.51}
\end{align*}
$$

and

$$
\begin{equation*}
\lambda_{1} \geq 0 \tag{X.5.52}
\end{equation*}
$$

The last terms in (51) are the nodal forces resulting from contact between nodes 1 and 2 . The forces on the nodes are equal and opposite and vanish when the Lagrange multiplier vanishes. The equations of motion are identical to the equations for an unconstrained finite element mesh except at the nodes which are in contact. The equations for a diagonal mass matrix with unit area can be written as

$$
\begin{align*}
& M_{1} a_{1}-f_{1}+\lambda_{1}=0 \\
& M_{2} a_{2}-f_{2}-\lambda_{1}=0  \tag{X.5.53}\\
& M_{3} a_{3}-f_{3}=0
\end{align*}
$$

where $a_{I}=\ddot{d}_{I}$.
The equations for small-displacement elastostatics, Eq. (27) can be written by combining the G matrix, Eq. (49), with the assembled stiffness as in (27c) giving
where $k_{I}$ is the stiffness of element $I$. The assembled stiffness matrix in the absence of contact, i.e. the upper left hand $3 x 3$ matrix, is singular, but with the addition of the contact interface conditions, the complete $4 \times 4$ matrix becomes regular.

Penalty Method. To write the equation for the penalty method, we will use the penalty law $p=\beta g=\beta\left(x_{1}-x_{2}\right) \mathcal{H}(g)=\beta\left(X_{1}-X_{2}+u_{1}-u_{2}\right) \mathcal{H}(g)$. Then evaluating Eq. (30) gives

$$
\mathbf{f}^{c}=\int_{\Gamma^{c}} \phi^{T} p d \Gamma=\left[\begin{array}{c}
1\rceil  \tag{X.5.55}\\
-1 \\
0
\end{array}\right] \beta g
$$

The above integral consists of the integrand evaluated at the interface point since $\Gamma^{c}$ is a point. Equations (32) for a diagonal mass are then

$$
\begin{align*}
& M_{1} a_{1}-f_{1}+\beta g=0 \\
& M_{2} a_{2}-f_{2}-\beta g=0  \tag{X.5.56}\\
& M_{3} a_{3}-f_{3}=0
\end{align*}
$$

The equations are identical to that for the Lagrange multiplier method, (53) except that the Lagrange multiplier is replaced by the penalty force.

To construct the small displacement, elastostatic equations for the penalty method, we first evaluate $\mathbf{P}_{\mathrm{c}}$ by Eq. (34):

$$
\begin{align*}
& \left.\mathbf{P}_{c}=\int_{\Gamma^{c}} \beta_{1} \phi^{T} \phi \mathcal{H}(g) d \Gamma=\beta_{1} \mathcal{H}(\gamma) \left\lvert\, \begin{array}{c}
\lceil+1\rceil \\
-1 \\
0
\end{array}\right.\right]\left[\begin{array}{lll}
+1 & -1 & 0
\end{array}\right] \\
& =\beta_{1} \mathcal{H}(g)\left[\begin{array}{ccc}
+1 & -1 & 0 \\
-1 & +1 & 0 \\
0 & 0 & 0
\end{array}\right] \tag{X.5.57}
\end{align*}
$$

If we define $\bar{\beta}=\beta_{1} \mathcal{H}(g)$, and add $\mathbf{P}_{\mathrm{c}}$ to the linear stifness, then the resulting equations are

$$
\left\{\begin{array}{ccc}
k_{1}+\bar{\beta} & -\bar{\beta} &  \tag{X.5.58}\\
-\bar{\beta} & k_{2}+\bar{\beta} & -k_{2} \\
& -k_{2} & k_{2}
\end{array}\right\}\left\{\begin{array}{l}
d_{1} \\
d_{2} \\
d_{3}
\end{array}\right\}=\left\{\begin{array}{l}
f_{1} \\
f_{2} \\
f_{3}
\end{array}\right\}
$$

It can be seen from the above equation that the penalty method simply adds a spring with a spring constant $\bar{\beta}$ between nodes 1 and 2. The above equation is nonlinear since $\bar{\beta}$ is a nonlinear function of $g=u_{1}-u_{2}$.


Figure X.5.2
Example X2. Two Dimensional Example. Figure 2 shows two dimensional bodies modeled by 4 -node quadrilaterals which are in contact along a line parallel to the $x$-axis. The approximations along the contact surface are written in terms of the element coordinates of one of the master body $A$., which in this case is the identical to that of body $B$. The velocity field along the contact interface is given by

$$
\left\{\begin{array}{l}
v_{x}(\xi, t)  \tag{X.5.59}\\
v_{y}(\xi, t)
\end{array}\right\}=\left[\begin{array}{cccccccc}
N_{1} & 0 & N_{2} & 0 & N_{3} & 0 & N_{4} & 0 \\
0 & N_{1} & 0 & N_{2} & 0 & N_{3} & 0 & N_{4}
\end{array}\right]^{\mathbf{v}}
$$

where

$$
\mathbf{v}^{T}=\left[\begin{array}{llllllll}
v_{1 x} & v_{1 y} & v_{2 x} & v_{2 y} & v_{3 x} & v_{3 y} & v_{4 x} & v_{4 y} \tag{X.5.60}
\end{array}\right]^{T}
$$

$$
\begin{equation*}
N_{1}=N_{3}=1-\xi, N_{2}=N_{4}=\xi, \xi=x / l \tag{X.5.61}
\end{equation*}
$$

The unit normals are given by $\mathbf{n}^{A}=\left[\begin{array}{ll}0 & -1\end{array}\right]^{T}, \mathbf{n}^{B}=\left[\begin{array}{ll}0 & 1\end{array}\right]^{T}$ so the $\Phi$ matrix is given by Eq. (20):

$$
\begin{align*}
& \Phi=\left[\begin{array}{llllllll}
N_{1} n_{1}^{A} & N_{1} n_{2}^{A} & N_{2} n_{1}^{A} & N_{2} n_{2}^{A} & N_{3} n_{1}^{B} & N_{3} n_{2}^{B} & N_{4} n_{1}^{B} & N_{4} n_{2}^{B}
\end{array}\right] \\
& =\left[\begin{array}{llllllll}
-N_{1} & 0 & -N_{2} & 0 & N_{3} & 0 & N_{4} & 0
\end{array}\right] \tag{X.5.62}
\end{align*}
$$

The Lagrange multiplier field is approximated by the same linear field (we will discuss appropriate fields later)

$$
\lambda(\xi, t)=\Lambda \lambda=\left[\begin{array}{ll}
N_{1} & N_{2}
\end{array}\right]\left[\begin{array}{l}
\lambda_{1}  \tag{X.5.63}\\
\lambda_{2}
\end{array}\right\}
$$

where the same shape functions as in (61) are used. The $\mathbf{G}$ matrix is given by

$$
\mathbf{G}=\int_{\Gamma^{c}} \Lambda^{T} \Phi d \Gamma=\frac{l}{6}\left[\begin{array}{llllllll}
0 & -2 & 0 & -1 & 0 & 2 & 0 & 1  \tag{X.5.64}\\
0 & -1 & 0 & -2 & 0 & 1 & 0 & 2
\end{array}\right]
$$

The terms of the rows resemble the terms of the consistent mass for a rod, and the behavior for this Lagrange multiplier field is similar: a contact at node 1 results in forces at node 2, and vice versa. Nodal forces due to contact are strictly in the $y$ direction; all $x$ components of forces from contact in this example will vanish since the odd rows of the G matrix vanish. This is consistent with what is expected physically, since the contact surface is along the x -direction and the contact interface is frictionless.

## MISCELLANEOUS TOPICS

Regularization. The penalty approach may be thought of as a regularization of the interface conditions; the exact solution of the impact of two rods leads to solutions discontinuous in time, cf. Fig. . A regularization procedure in mathematics is a procedure which by an artifact replaces a problem whose solutions are difficult to deal with because of warts such as discontinuities or singularities by one with smoother, more regular solutions. The classic example of regularization is von Neumann's addition of artificial viscosity to the Euler fluid equations to smooth shocks. Without this artificial viscosity, solutions of the Euler equations in the vicinity of shocks by the central difference method are so oscillatory that they look like lash. Von Neumann showed that his regularization conserves momentum, so only part of the system is modified by regularization.

The penalty method plays the same role as artificial viscosity in impact. With the Lagrange multiplier method, the velocities are discontinuous in time at the point of impact, and these discontinuities propagate through the body as waves and result in considerable noise. The penalty regularization preserves momentum conservation, and the other conservation equations are also observed exactly. It only relaxes one condition,
the impenetrability condition, by allowing some overlap of the two bodies. It is a small price to pay for smoother solutions if the interpenetration is small.

The Curnier-Mroz plasticity models of friction can also be considered regularization, in this case, of the discontinuous character of the friction laws. The discontinuous nature of Coulomb friction can be gleamed from a simple illustration. Consider an element on a rigid surface with interface tractions modeled by Coulomb friction. A vertical force is applied to the top nodes, a horizontal force on the two lefthand nodes as shown, and we neglect the deformability of the element. If the vertical force is kept constant while the horizontal force has the time history shown, the velocity will have the time history shown in Fig. Xd. The discontinuity in time arises because the inequalities in the Coulomb friction law embody Heaviside step functions exactly as they embodied in the interpenetration inequalities.

The Curnier-Mroz friction model eliminates the discontinuity as shown in Fig. X. Regularization of Coulomb friction differs from regularization of interpenetration in that, superficially at least, it smoothes the response by introducing additional mechanics to the model, namely the asperities, whereas the relaxation of the interpenetrability condition appears to be quite $a d$ hoc and not motivated by physical arguments. In fact, one can also attribute some interpenetration of the idealized bodies which comprise the models in contact-impact problems to compression of asperities. Usually, however, the penalty parameters are not chosen by such physical characteristics, but instead by the desirability of eliminating frequencies above a certain threshold.

## REFERENCES

T. Belytschko and M.O. Neal (1991), "Contact-Impact by the Pinball Algorithm with Penalty and Lagrangian Methods," Int. J. for Numerical Methods in Engineering, 31. 547-572.
D.P. Bertsekas (1984), Constrained Optimization and Lagrange Multiplier Methods, Academic Press, New York.
A. Curnier (1984), "A Theory of Friction," Int. J. Solids and Structures, 20, 637-647.

Demkowicz and J. T. Oden (1981), On some existence and uniqueness results in contact problems with nonlocal friction, Texas Institute of Computational Mechanics (TICOM) Report 81-13, University of Texas at Austin.
N. Kikuchi and J.T. Oden, Contact Problems in Elasticity: A Study of Variational Inequalities and Finite Element Methods, SIAM ????
R. Michalowski and Z. Mroz (1978), "Associated and non-associated sliding rules in contact friction problems," Archives of Mechanics, 30, 259-276.
P. Wriggers (1995), "Finite Element Algorithms for Contact Problems," Archives of Computational Methods in Engineering, 2,4, pp. 1-49.
P. Wriggers and C. Miehe (1994), "Contact Constraints within Coupled Thermomechanical Analysis - a Finite Element Model," Comp. Meth. in Appl. Mech. and Eng., 113, 301-319.

## ERRATA

1. p. 38 should read

$$
\begin{equation*}
G_{J j I}=\int_{\Gamma^{c}} \Lambda_{I} \Phi_{J j} d \Gamma \quad \mathbf{G}=\int_{\Gamma^{c}} \Lambda^{\mathrm{T}} \Phi d \Gamma \tag{X.5.21b}
\end{equation*}
$$

2. in Box X. 3 last equation should read

$$
\mathbf{G}=\int_{\Gamma^{c}} \Lambda^{T} \phi d \Gamma
$$

3. equation before (X.5.55) should read

$$
p=\beta g=\beta\left(x_{1}-x_{2}\right) \mathcal{H}(g)=\beta\left(X_{1}-X_{2}+u_{1}-u_{2}\right) \mathcal{H}(g)
$$

## X.6. EXPLICIT METHODS OF TIME INTEGRATION

In this Section we describe the procedures for treating contact impact with explicit time integration. Explicit time integration is well suited to contact-impact problems because the small time steps imposed by numerical stability can treat the discontinuities in contact-impact. The large time steps made possible by unconditionally stable implicit methods are not effective for discontinous response. Furthermore, contactimpact also introduces discontinuities in the Jacobian, which impedes the convergence of Newton methods.

Another advantage of explicit algorithms is that the bodies can first be integrated completely independently, as if they were not in contact. This uncoupled solution correctly indicates which parts of the body are in contact. The contact conditions are imposed after the two bodies have been updated in an uncoupled manner; no iterations are needed to establish the contact interface. An explicit algorithm with contact-impact is almost identical to the algorithm described in Chapter X except that the bodies are checked for interpenetration. In each time step, the displacements and velocities of those nodes which have penetrated into another body are modified to reflect momentum balance and impenetrability on the interface.

We will here describe several implementations of contact-impact algorithms in explicit methods. Only the Lagrange multiplier and the penalty methods will be considered. the issues to be discussed include: 1. the approximations for the Lagrange multiplier fields; 2. structure of the algorithm; 3. effects of contact-impact methods on numerical stability. We will also describe certain characteristics of explicit solutions which arise from the physics and numerical characteristics of the contact-impact problem. In order to illustrate the characteristics of contact-impact in a simple setting, we first consider a one dimensional problem.

Example of Contact in One Dimension. The one-dimensional example is shown in Fig. ??. We first consider the premise that uncoupled updates of bodies A and B followed by modifications of the interpenetrating nodes for contact-impact lead to consistent solutions. For the two points R and S , which correspond to nodes 1 and 2,
respectively, of bodies A and B , there are four possibilities during a contact-impact problem

1. R and S are not in contact and do not contact during the time step;
2. R and S are not in contact but impact during the time step;
3. $R$ and $S$ are in contact and remain in contact;
4. $R$ and $S$ are in contact and separate during a time step, often known as release.

For case 3, the statement "remain in contact" does not imply that if two points must remain contiguous, because relative tangential motion, or sliding, which separates contiguous points is always possible. When two bodies remain in contact, they are assumed not to separate.

All of these possibilities can be correctly accounted for by integrating the two bodies independently as if they were not in contact and subsequently adjusting the velocities and the displacements. The possibilities which need to be explained are cases 2,3 and 4.

The governing equations for the nodes 1 and 2 have been given in Example Eq. (53); although the problem shown in Fig. ?? is somewhat different, the equations for the contact nodes are unchanged. We will show that when the velocities from the uncoupled update predict initial or continuing contact, then the Lagrange multiplier $\lambda \geq 0$. The accelerations of nodes 1 and 2 when the two bodies are updated as uncoupled are

$$
M_{1} \bar{a}_{1}-f_{1}=0, \quad M_{2} \bar{a}_{2}-f_{2}=0
$$

where bars have been superposed on the accelerations to indicate that these are trial accelerations computed with the uncoupled bodies, as can be seen from the absence of the Lagrange multipliers. The central difference form of the update of Eqs. (53)

$$
\begin{aligned}
& M_{1} v_{1}^{+}-M_{1} v_{1}^{-}-\Delta t f_{1}+\Delta t \lambda=0 \\
& M_{1} v_{2}^{+}-M_{1} v_{2}^{-}-\Delta t f_{2}-\Delta t \lambda=0
\end{aligned}
$$

When the bodies contact during the time step, these equations must be solved with the subsidiary condition $v_{1}^{+}=v_{2}^{+}$. Eliminating $\lambda$ from the above equations by adding them and using the equality $v_{1}^{+}=v_{2}^{+}$gives

$$
v_{1}^{+}=v_{2}^{+}=\frac{M_{2} v_{1}^{-}+M_{2} v_{1}^{-}+\Delta t\left(f_{1}+f_{2}\right)}{M_{1}+M_{2}}
$$

where all unmarked variables are a time step $n$. By means of the above equations, the updated velocities can be updated whenever impact occurs or the nodes were in contact in the previous time step. The above equations can be recognized to be the well known equations of conservation of mass for plastic impact of rigid bodies; more will be said on this later.

We will now show that whenever the updated velocities of any nodes which interpenetrate are computed by (), then the Lagrange multiplier will be positive, i.e. the interface force will be compressive. In other words, if the two nodes are updated as if the bodies were uncoupled and if the velocities are subsequently modified by (), then the Lagrange multipliers will have the correct sign, This amounts to showing that

$$
\text { if } \bar{v}_{1}^{+} \geq \bar{v}_{2}^{+} \text {, then } \lambda \geq 0 .
$$

Multiplying Eq top by M2 and Eq. bot by M1 and subtract bot form top; this gives

$$
M_{1} M_{2}\left(v_{1}^{-}-v_{2}^{-}\right)+\Delta t\left(M_{2} f_{1}-M_{1} f_{2}\right)=\lambda \Delta t\left(M_{1}+M_{2}\right)
$$

Substituting the expressions for f 1 and $\mathfrak{f} 2$ from () into the above and rearranging gives

$$
\frac{\Delta t\left(M_{1}+M_{2}\right)}{M_{1} M_{2}} \lambda=\left(v_{1}^{-}-v_{2}^{-}\right)+\Delta t\left(\bar{a}_{1}-\bar{a}_{2}\right)=\bar{v}_{1}^{+}-\bar{v}_{2}^{+}
$$

where the last equality is obtained by using the central difference formulas for the uncoupled integration of the two bodies: $\bar{v}_{I}^{+}=v_{I}^{-}+\Delta t a_{I}$. The coefficient of $\lambda$ is positive, so the sign of the RHS gives the sign of $\lambda$. Thus Eq () has been demonstrated.

To examine this finding in more detail, we now consider the three of the cases listed above (case 1 is trivial since it requires no modification of the nodal velocities since there is no contact):
case 2 (not in contact /contacts during $\Delta t$ ): then $\bar{v}_{1}^{+} \succ \bar{v}_{2}^{+}$and $\lambda \geq 0$ by Eq ()
case 3 (in contact/remains in contact): then $\bar{v}_{1}^{+} \succ \bar{v}_{2}^{+}$and $\lambda \geq 0$ by Eq ()
case 4 (in contact/release during $\Delta t$ ): then $\bar{v}_{1}^{+} \prec \bar{v}_{2}^{+}$and $\lambda \prec 0$ by Eq ()
Thus the velocities obtained by uncoupled integration correctly predict the sign of the Lagrange multiplier $\lambda$.

Two other interesting properties of explicit integration that can be learned from this example are:

1. initial contact, i.e. impact cannot occur in the same time step as release;
2. energy is dissipated during impact;

The first statement rests on the fact that the Lagrange multiplier at time step $n$ is computed so that the velocities at time step $n+1 / 2$ match. Hence there is no mechanism in an explicit method for forcing release during the time step in which impact occurs. This property is consistent with the mechanics of wave propagation. In the mechanics of impacting bodies, release is caused by rarefaction waves which are generated when the compressive waves due to impact reflect from a free surface and reach the point of contact. When the magnitude of these rarefaction is sufficient to cause tension across the contact interface, release occurs. Therefore the minimum time required for release
subsequent to impact is two traversals of the distance to the nearest free surface. The stable time step, you may recall, allows the any wave generated by impact to move at most to the node nearest to the contact nodes. Therefore, in explicit time integration, there is insufficient time in a stable time step for the waves to traverse twice the distance to the nearest free surface.

The second statement can be explained by Eq. () which shows that the postimpact velocities are obtained by the plastic impact conditions, for rigid bodies, which always dissipate energy. The energy dissipated when two rods as shown in Fig.() are at constant but equal velocities is given by

As can be seen from the above, the amount of dissipation decreases with the refinement of the mesh. In the continuos impact problem, no energy is dissipated because the condition of equal velocities after impact is limited to the impact surfaces. A surface is a set of measure zero in three dimensions, so a change of energy over the surface has no effect on the total energy. (For one-dimensional problems the impact surface is a point, which is also a set of measure zero.) In a discrete model, the impacting nodes represent the material layer of thickness $h / 2$ adjacent to the contact surface. Therefore, the dissipation in a discrete model is always finite. The correspondence between the continuous model and the discrete model also substantiates the correctness of the plastic impact condition. Since release cannot occur until the rarefaction waves reach the contact interface, the velocities of the two contacting bodies must be equal until that time. Thus it is inappropriate to use impact conditions other than perfectly plastic impact for discrete models of continuous systems. It should be stressed that such arguments do not apply to strictly multi-body models, where each node represents a body whose stiffness is not modeled, or to structural models, where the thickness direction has no deformability. The release and impact conditions are then more complex.

Penalty Method. The discrete equation at the impacting nodes for the two body problem can be taken directly from those given in Eq. ():

$$
\begin{aligned}
& M_{1} a_{1}-f_{1}+f_{1}^{c}=0 \\
& M_{2} a_{2}-f_{2}-f_{2}^{c}=0
\end{aligned}
$$

where the contact forces $f_{I}^{c}$ replace the Lagrange multiplier replace the Lagrange multiplier in (). When the nodes are initially almost coincident, then $X_{1}=X_{2}$ and the interface normal traction can be written as

$$
f^{c}=p=\beta_{1} g+\beta_{1} g=\beta_{1}\left(u_{1}-u_{2}\right) \mathcal{H}(g)+\beta_{2}\left(v_{1}-v_{2}\right) \mathcal{H}(\dot{g})
$$

The unitary condition is now approximately enforced by the step functions in the normal contact force; it is violated since the normal traction is positive while the interpenetration rate is positive, so its product no longer vanishes. The post-impact velocities are now given by

The velocities of the two nodes are not equal since the penalty method only enforces the impenetrability constraint approximately. As the penalty parameter is increased, the condition of impenetrability is observed more closely. However, as indicated in the next paragraph, in dynamics the penalty parameter cannot be made arbitrarily large.

The condition that release not occur in the same time step as impact, which has been described to be a natural consequence of the physics of contact and numerical stability conditions of explicit intergrators, is not automatically satisifed by the penalty method. If the penalty force is very large, it is possible for the relative nodal velocity to reverse, so that decreasing gap rate is is followed in the same time step by an increasing gap rate. In view of the behavior of the continuous model described previously, which does not permit release until rarefaction waves reflected from the free surfaces reach the contact interface, this possibility in penalty methods does not appear physically correct. This anomaly can be eliminated by placing an upper bound on the penalty force, so that the impact is at most perfectly plastic. In other words, the penalty force should be bounded so that the velocities at the end of the impact time step are given by (). This yields the following upper bound on the contact force:

This bopund can be very useful since it provides a
In contrast to the Lagrange multiplier method, the penalty method usually decreases the stable time step. An estimate of the stable time step can be made by using the linear stiffness for the penalty method given in Eq. () in conjunction with the eigenvalue element inequality. In using the element eigenvalue inequlitu, a group of elements consisting of the penalty spring and the two surrounding elements should be used, since the penalty element has no mass by itself and therefore has an infinite frequency. This analysis shows that the stable time step for an interpenetration dependent penalty is given by
where $\alpha$ is given by

The decrease in the time step depends on the stiffness of the penalty spring. As the interpenetration stifness $b$ is increased, the stable time step decreases. As in the case of the Lagrange multiplier method, this estimate of the stable time step is not a conservative estimate, even though it is based on the element eigenvalue inequaality. The analysis presumes linear behavior, whereas contact-impact is a very non-linear process.

## EXPLICIT ALGORITHM

A flowchart for explicit time integration with contact-impact is shown in Box ??. As can be seen from the flowchart, the contact impact conditions are enforced immediately after the boundary conditions. Prior to the contact-impact step, all nodes in the model have been updated as if they were not in contact, including the nodes which were in contact in the previous time step. The nodes which are in contact are not treated differently in the rest of the algorithm. Some difficulties may occur dus to making the contact-impact
modifications after the boundary condition enforcement. For example, for a pair of contacting nodes on a plane of symmetry, it is possible for the contact-impact modifications to result in violation of the condition that the velocities normal to the plane of symmetry vanish. This can occur when the normal of the element adjacent to the plane of symmetry does not lie in that plane. Therefore, boundary conditions sometimes have to be imposed at contact nodes after the modifications.

The CONTACT module is limited to low-order elements in which the maximum interpenetration always occurs of the nodes of the master or slave body. It is then only necessary to check all nodes for interpenetration into elements of another body. The second statement in the CONTACT nevertheless conceals many challenging tasks. In a large model, on the order of $10^{5}$ nodes may have to be checked against penetration into a similar number of elements. Obviously a brute force approach to this task is not going to work. Some of the strategies for dealing with this task are described in Section ?.

The

Box X.?
Flowchart for Explicit Integration with Contact-Impact

## Main Program

1. initial conditions and setup: $t=0, n=0$, set $\mathbf{v}^{0}, \sigma^{0}$
2. get nodal forces $\mathbf{f}^{t}=\left(\mathbf{f}^{e x t}-\mathbf{f}^{\text {int }}\right)^{t}$
3. velocity update: if $n \succ 0, \mathbf{v}^{\Delta t / 2}=\mathbf{v}^{0}+\mathbf{M}^{-1} \mathbf{f}^{0}$;

$$
\text { otherwise } \mathbf{v}^{t+\Delta t / 2}=\mathbf{v}^{t-\Delta t / 2}+\mathbf{M}^{-1} \mathbf{f}^{t}
$$

4. displacement update: $\mathbf{d}^{t+\Delta t}=\mathbf{d}^{t}+\Delta t \mathbf{v}^{t+\Delta t / 2}$
5. modify velocities and displacements for velocity boundary conditions
6. go to CONTACT
7. get $\mathbf{f}^{t+\Delta t}=\left(\mathbf{f}^{e x t}-\mathbf{f}^{i n t}\right)^{t+\Delta t}$
8. accelerations: $\mathbf{a}^{t}=\mathbf{M}^{-1} \mathbf{f}^{t}$
9. if if $t \prec E N D$, go to 3

## CONTACT

1. find node-element pairs which are in contact;
2. 

Lagrange multiplier method. The discrete equations for the system are obtained by combining the semidiscrete equations with an explicit integration formula. For simplicity, we consider here only the central difference method. Substituting an expression for the accelerations at time step $n$, Eq.(), we obtain from () that

$$
\mathbf{M}\left(\mathbf{v}^{n+1 / 2}-\mathbf{v}^{n-1 / 2}\right)-\mathbf{f}^{n}+\mathbf{G}^{T} \lambda^{n}=0
$$

Referring to the flowchart in Box $X$, it can be seen that when the contact conditions are enforced, the nodal forces at time step $n$ are already known. However the Lagrange
multiplier are unknown. If we combine the above with the velocity constraint, Eq. () in Box ?, we obtain

$$
\left[\begin{array}{cc}
\mathbf{M} & \Delta t \mathbf{G}^{T} \\
\mathbf{G} & \mathbf{0}
\end{array}\right]\left\{\begin{array}{c}
\mathbf{v}^{n+1 / 2} \\
\lambda^{n}
\end{array}\right\} \leq\left\{\begin{array}{c}
\Delta t \mathbf{f}^{n}+\mathbf{M} \mathbf{v}^{n-1 / 2} \\
\mathbf{0}
\end{array}\right\}
$$

If a consistent mass matrix is used, solving for these variables appears to involve a system of equations which is larger than the unconstrained system, since the Lagrange multipliers have been added. In fact, for most systems, the size of the matrix can be reduced substantially, since trial values of $\mathbf{v}^{n+1 / 2}$ are already known and only the velocities of nodes on the contact interface are modified by contact .

In the above, everything on the right hand side is known at time step n when the modifications for the contact are made. The unknowns are $\lambda^{n}$ and $\mathbf{v}^{n+y 2}$, although trial values for the nodal velocities have already been obtained by the uncoupled update. The solution for the Lagrange multipliers is obtained by first solving the top of the above equation for $\mathbf{v}^{n+y 2}$ and then solving for $\lambda^{n}$, which gives

$$
\mathbf{G} \mathbf{M}^{-1} \mathbf{G}^{T} \lambda^{n} \leq-\left(\mathbf{M}^{-1} \mathbf{f}^{n}+\frac{1}{\Delta t} \mathbf{v}^{n+y / 2}\right) \equiv \mathbf{r} w
$$

When the mass matrix is diagonal, the solution for the Lagrange multipliers can be streamlined by taking advantage of the fact that the inverse of the mass matrix consists of the reciprocals of the diagonal terms. To preserve the symmetric structure of the equations we take the square root of the mass matrix and multiply $\mathbf{G}$, and define the resulting matrix as $\overline{\mathbf{G}}$ :

$$
\overline{\mathbf{G}}=\mathbf{M}^{-1 / 2} \mathbf{G} \quad \bar{G}_{a b}=M_{a d}^{-1 / 2} G_{d b}
$$

Equations () can then be written as

$$
\overline{\mathbf{G}} \overline{\mathbf{G}}^{T} \lambda^{n}=\mathbf{r}
$$

An interesting characteristic of these equations is that they are already in the form of a triangulation. It is only necessary to eliminate all terms of the $\overline{\mathbf{G}}$ matrix to obtain a matrix from which the solution can easily be found. Moreover, the above equations involve only the nodes on the contact interfaces. Thus the system of equations to be solved is usually much smaller than the complete model. Nevertheless, for large-scale explicit solutions, the burden of solving these equations is too great, so simplifications are usually made to avoid solving these equations; these are discussed later.

Lagrange multiplier interpolation. In order to develop explicit forms of Eq.(), the interpolation for the Lagrange multipliers must be defined. We have already mentioned in Section X.5.2 that the construction of these interpolants can be complicated when the nodes of the bodies are noncontiguous. As indicated there, two possibilities are: 1. the master body mesh is chosen to be the $\mathbf{l}$ mesh; 2. construct a new mesh. Examples of the G matrix for noncontiguous nodes have already been described in Example XX. We now
explore the consequences of various I approx imations and their effects on computational efficency.

The implementation of implicit time integration and statics will be combined because the procedure are almost identical. The reader is urged to consult Sections X, where these topics are discussed for problems without contact. As in the aforementioned, both classes of problems are treated by the Newton method.

In the Newton method, the solution to the discrete equations is found by using a local linear model for the nonlinear equations. The linear model is based on a linearization of the governing discrete equations. We will consider the Lagrange multiplier methods and the penalty methods. In both cases, as before, we write the nonlinear equations in the form

$$
\overline{\mathbf{f}}(\mathbf{d}, \dot{\mathbf{d}}, \lambda)=0
$$

where $\mathbf{d}, \mathbf{d}$, and $\lambda$ are, respectively, the nodal displacements, nodal velocities, and discrete Lagrange multipliers at time $t+\Delta t ; \lambda$ appears only in the Lagrange multiplier method. The internal force is only a function of $\mathbf{d}$, i.e. the material is rate-independent. The extension to rate dependent materials involves a combination of the techniques described here and in Section X, but they obscure the effects of contact-impact, so are omitted in this exposition.

In the Lagrange multiplier method the governing equations are

$$
0=\dot{f}(\mathbf{d}, \dot{\mathbf{d}}, \lambda, t)=\mathbf{M} \ddot{\mathbf{d}}(t)+\mathbf{f}^{i n t}(\mathbf{d})-f^{e x t}(\mathbf{d}, t)-\mathbf{G}(\mathbf{d}) \lambda(t)
$$

where the independent variables are indicated in the above. All of the above terms are functions of time since $\mathbf{d}=\mathbf{d} / t$. The development is restricted to rate-independent materials, so the internal nodal forces are only functions of the nodal displacements, see Section X.

We now expand the nodal forces by the chain rule, giving


Figure 4. Illustration of nomenclature in a two dimensional contact problem.

An interesting simplification of the above example is shown in Fig. . The equations for this system can be obtained by just eliminating rows 1 and 3 and columns 1 and 3 , giving

$$
\left[\begin{array}{cc}
k_{2} & -1 \\
-1 & 0
\end{array}\right]\left\{\begin{array}{l}
d_{1} \\
\lambda_{1}
\end{array}\right\}=\left\{\begin{array}{c}
f_{1} \\
0
\end{array}\right\}
$$

The potential energy $\Pi(d, \lambda)=\frac{1}{2} k_{2} d^{2}-f_{1} d$ is plotted for $f_{1}=$ in Fig. as a function of $d$ and $\lambda$; to obtain the plot, Eqs. ( ) have been solved.
$y$


Figure 9.
Example X3. Figure 9 shows the two two-dimensional bodies of unit thickness which are in contact along a line; a total of 5 nodes are in contact, 3 nodes from body A, 2 nodes from body B.. The nodes on the endpoints are coincident, but the center node of body A is not coincident with a node on body B . This example introduces some of the difficulties arising from noncoincident nodes. We will restrict our attention to the nodes which are in contact, since the equation at the other nodes are unchanged. The nodal velocities of the contact nodes are denoted by $\mathbf{d}_{c}$ where $\dot{\mathbf{d}}_{c}^{T}=\left[\begin{array}{lllll}v_{1} & v_{2} & v_{3} & v_{4} & v_{5}\end{array}\right]$. The elements in the two bodies are bilinear 4-node quadrilaterals, so the displacement and velocity fields along the contact lines are linear, and will be represented by

$$
v(\xi, t)=\sum_{I=1}^{5} N_{I}(\xi) v_{I}(t)
$$

where $\xi=\left(s-s_{I}\right) / l_{I}$ along each of the segments.

There are many choices for the approximation of the Lagrange multiplier field, but a good choice is usually a field of the same order as that of the field being constrained. In this case the tractions are constrained and the tractions are piecewise constant, so we let

$$
\begin{aligned}
& \lambda(x, t)=\lambda_{1}(t) \text { for } 0 \leq s \leq \frac{l}{2} \\
& \lambda(s, t)=\lambda_{2}(t) \text { for } l / 2 \leq s \leq l
\end{aligned}
$$

where s parametrizes the contact line.
The G matrix here will be assembled from the segment, or element, matrices. The $\mathbf{T}$ matrix is constant along the line joining nodes 1 and 3 and given by

$$
T=\left[\begin{array}{cc}
s & -c \\
0 & 0
\end{array}\right], c=\cos \theta, s=\sin \theta
$$

The element $G$ matrices are then given by

$$
\begin{aligned}
& \left.\mathbf{G}_{e=1}=\mathbf{G}_{e=2}=\int_{0}^{1 \mid} \begin{array}{c}
s(1-\xi)] \\
|-c(1-\xi)| \\
s \xi \\
-c \xi
\end{array}\right][1] l d \xi=\frac{l}{2}\left|\begin{array}{c} 
\\
|-c| \\
s \\
-c\rfloor
\end{array}\right| \\
& \left.\left.\left.\mathbf{G}_{e=3}=\int_{0}^{1 / 2 \mid} \left\lvert\, \begin{array}{c}
-s(1-\xi)\rceil \\
c(1-\xi) \\
-s \xi \\
c \xi
\end{array}\right.\right] \quad\left[\begin{array}{ll}
1 & 0
\end{array}\right] 2 l d \xi+\int_{1 / 2}^{1} \left\lvert\, \begin{array}{c}
-s(1-\xi)\rceil \\
c(1-\xi) \\
-s \xi \\
c \xi
\end{array}\right.\right] \quad \left\lvert\, \begin{array}{ll}
0 & 1
\end{array}\right.\right] 2 l d \xi=\frac{l}{4}\left|\begin{array}{cc}
-3 s & -s \\
3 c & c
\end{array}\right|
\end{aligned}
$$

Assembling the three matrices gives

$$
\mathbf{G}^{T}=\frac{l}{4}\left[\begin{array}{cccccccccc}
2 s & -2 c & 2 s & -2 c & 0 & 0 & -3 s & 3 c & -s & c \\
0 & 0 & 2 s & -2 c & 2 s & -2 c & -s & c & -3 s & -3 c
\end{array}\right]
$$

