Chapter 5
Finite Element Analysis for Contact Problems

5.1 Introduction

When two or more bodies collide, contact occurs between two surfaces of the bodies so that they cannot overlap in space. Metal formation, vehicle crash, projectile penetration, various seal designs, and bushing and gear systems are only a few examples of contact phenomena. During sheet-metal formation, for example, a simple-shaped blank is formed into a desired shape through contact against a punch and die. In such a case, it is important to determine contact locations between a deformable blank and a rigid or deformable punch and die. In a broader sense, contact is a common and important aspect of mechanical systems, where multiple parts are assembled to compose the system. In fact, contact is the main tool to join multiple parts together, which includes screws, bolts, welds, etc.

The objective of contact analysis is to answer the following questions: (a) whether two or more bodies are in contact, (b) if they are, where the location or region of contact is, (c) how much contact force or pressure occurs in the contact interface, and (d) if there is a relative motion after contact in the interface. In this chapter, these questions will be addressed in the continuum and finite element domains.

Contact is categorized as boundary nonlinearity, in contrast to both geometric nonlinearity, which emerges from finite deformation problems, and material nonlinearity, which is a product of nonlinear constitutive relations. The nonlinearity of contact can be explained in two aspects. Firstly, if two separate bodies come into contact, the graph of the contact force vs. displacement looks like a cliff because the contact force stays at zero when two bodies are separate and increases vertically after the bodies come into contact. In such a case, a functional relationship is not available because there is no one-to-one relationship between contact force and displacement. A similar phenomenon happens in the tangential direction under friction where two bodies are stuck together until the tangential force reaches a threshold, after which continuous sliding occurs without further increasing the

N.-H. Kim, *Introduction to Nonlinear Finite Element Analysis*,
DOI 10.1007/978-1-4419-1746-1_5

tangential force. Such an abrupt change in contact force and slip makes the problem highly nonlinear. Secondly, in order to be a well-posed problem in mechanics, either displacement (kinematics) or force (kinetics), but not both, must be given for every material point. Then, the finite element equation solves for unknown information with given information. On the displacement boundary, for example, if displacement is given, reaction force should be calculated. On the other hand, on the traction boundary, if the applied force is given, the corresponding displacement is to be calculated. Note that these two boundaries are clearly identified in the problem definition stage. In the case of contact, however, both displacement and contact force are unknown, except for very limited cases; that is, the contact boundary is a part of the solution. The user can only identify a candidate of contact boundary before solving the problem. Therefore, the finite element analysis procedure must find (a) whether a material point in the boundary of a body is in contact with the other body, and if it is in contact, (b) the corresponding contact force must be calculated. Since the contact force at a material point can affect the deformation of neighboring points, this process needs to be repeated until finding right states for all points that are possible in contact. Because of this procedural nature, contact nonlinearity is often addressed algorithmically (Fig. 5.1).

For the case of an elastic system, equilibrium can be described as finding a displacement field that minimizes the potential energy. Contact can then be considered as a constraint of the optimization formulation, such that the potential energy is minimized while satisfying the contact constraint; that is, a body cannot penetrate the other body.[1] The constrained optimization problem can be converted into an unconstrained one by using the penalty regularization or Lagrange multiplier methods. Therefore, most contact algorithms are derived based on these two methods. Once understanding that contact can be considered as a constraint to the structural equilibrium, it can be applicable to nonelastic materials, such as elastoplastic material, as it is basically independent of material models used. Therefore, it is possible to treat the contact formulation independent of constitutive models.

Although contact problems can be formulated in a variety of ways, the slave–master concept is commonly used in finite element-based applications. In the slave–master concept, one body is called a slave body, and the other is called a master body. Although the selection of slave and master bodies is arbitrary, some guidelines will be given later in the chapter. The contact constraint is then imposed in such a way that the slave surface cannot penetrate the master surface. Or, in finite elements, the nodes on the slave boundary cannot penetrate the surface elements on the master boundary. It is also possible that the role of slave and master can be changed so that the master surface cannot penetrate the slave surface.

[1] Rigorous discussions on this topic with variational inequality and its equivalence to the constrained optimization can be found in J. Sokolowski and J. P. Zolesio, Introduction to Shape Optimization, Springer-Verlag, Berlin, 1991. A brief summary will be presented in Sect. 5.3.

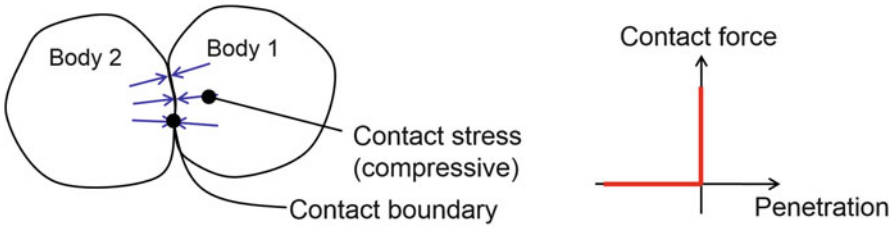

Fig. 5.1 Contact boundary and contact force

This chapter is organized as follows. In Sect. 5.2, simple one-point contact examples are presented in order to show the characteristics of contact phenomena and possible solution strategies. In Sect. 5.3, a general formulation of contact is presented based on the variational formulation similar to previous chapters. To facilitate comprehension, the complexity of formulation is gradually increased by moving from flexible-to-rigid contact to flexible-to-flexible contact, from line-to-line to surface-to-surface contact, and including friction. Section 5.4 focuses on finite element discretization and numerical integration of the contact variational form. Three-dimensional contact formulation is presented in Sect. 5.5. From the finite element point of view, all formulations involve the use of some form of constraint equation. Because of the highly nonlinear and discontinuous nature of contact problems, great care and trial and error are necessary to obtain solutions to practical problems. Section 5.6 presents modeling issues related to contact analysis, such as selecting slave and master bodies, removing rigid-body motions, etc.

5.2 Examples of Simple One-Point Contact

In order to illustrate key features of a contact problem, simple one-point contact examples are presented in this section. The concepts in this section will be generalized to curve or surface contact problems in Sect. 5.3.

5.2.1 Contact of a Cantilever Beam with a Rigid Block

Consider a cantilever beam subjected to a distributed load. The deflection of the free end of the cantilever is limited by a rigid block. There is a small gap between the end of the beam and the rigid block as shown in Fig. 5.2. The following numerical data are assumed: distributed load $q = 1$ kN/m, length of the beam $L = 1$ m, flexural rigidity $EI = 10^5$ N m^2, and initial gap $\delta = 1$ mm.

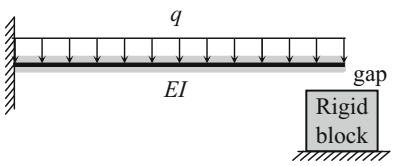

Fig. 5.2 Cantilever beam supported by a rigid block with a gap

5.2.1.1 Solution Using Trial and Error

In such a simple case, there are two possibilities. If the initial gap is larger than the deflection of the beam, then there will be no contact. On the other hand, if the deflection is larger than the gap, then the gap is closed under the given load. One solution strategy would be first to assume that the gap is large enough so that it will not close under the given load. In that case, the rigid block has no influence on the deformations. From the Euler beam model, the deflection curve and the tip deflection can be given as

$$v_{\mathrm{N}}(x) = \frac{qx^2}{24EI}\left(x^2 + 6L^2 - 4Lx\right), \quad v_{\mathrm{N}}(L) = \frac{qL^4}{8EI} = 0.00125\,\mathrm{m}.$$

The solution shows that the tip deflection is larger than the gap, and therefore, the assumption of gap not closing is wrong.

Now, when the gap is closed, contact occurs between the beam and the rigid block. Even if the rigid block has a finite width, it is assumed that the contact only occurs at the tip of the beam, i.e., one-point contact. Since the rigid block prevents the deflection of the beam, its effect can be modeled by applying a force, i.e., a contact force, such that the beam cannot penetrate the rigid block. Since the beam deflection is small, the rule of superposition is used for the effect of the two loads. The deflection curve and the tip deflection of the beam under the force at the tip can be given as

$$v_{\mathrm{c}}(x) = \frac{-\lambda x^2}{6EI}(3L - x), \quad v_{\mathrm{c}}(L) = \frac{-\lambda}{3 \times 10^5}.$$

Here, a negative sign is used for the force because the direction of contact force is opposite to the applied distributed load. At this point, the contact force, λ, is unknown, which can be calculated from the condition that the beam cannot penetrate the rigid block; that is, the deflection of the combined loads is the same with the gap, as

$$v_{\mathrm{tip}} = v_{\mathrm{N}}(L) + v_{\mathrm{c}}(L) = 0.00125 - \frac{\lambda}{3 \times 10^5} = 0.001 = \delta.$$

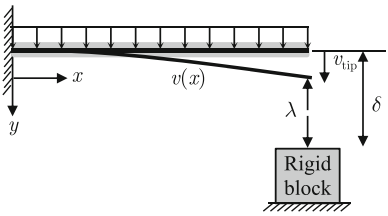

Fig. 5.3 Deflection of cantilever beam with gap and contact force

From the above relation, the contact force can be calculated to be $\lambda = 75$ N. The deflection curve can then be obtained by combining the two deflection curves, as $v(x) = v_N(x) + v_c(x)$.

5.2.1.2 Solution Using Contact Constraint

The issue in the previous trial-and-error approach is that the solution (deflection of the beam) has to be calculated first in order to determine the status of contact. When contact occurs at multiple points, the procedure can be quite complicated to check all possible combinations of the contact points. A more systematic contact formulation can be developed by considering both the contact force and the gap between the beam and the rigid block as unknowns and adding an additional constraint. The unknown contact force is denoted by λ that acts on the beam and the rigid block in the opposite directions.[2]

In order to assign consistent directions, one of the two contact points is considered a master and the other a slave. The master is assumed to be fixed while the slave moves to initiate the contact. For a general situation, when both bodies in potential contact are loaded, the choice between a master and a slave may be arbitrary. More details will be discussed in Sect. 5.5 for selecting the master and slave. In this problem, the beam obviously is the slave and the rigid block is the master.

With the downward deflection being positive, Fig. 5.3 shows the positive directions for this contact force on the beam and the rigid block. The contact force is treated as an externally applied load, even if it is unknown. Because of Newton's third law, the contact force acts in an equal and opposite direction to the beam and the block. In this particular example, since the rigid block is fixed, it is unnecessary to consider the equilibrium of the rigid block.

Treating the contact forces as externally applied loads and using the superposition rule of two independent loads, the beam deflection curve can be obtained by

[2] It will be clear later that the contact force is equivalent to the Lagrange multiplier in the constrained optimization, which is the reason to use the Greek symbol λ.

$$v(x) = \frac{qx^2}{24EI}\left(x^2 + 6L^2 - 4Lx\right) - \frac{\lambda x^2}{6EI}(3L - x). \tag{5.1}$$

This deflection curve must be supplemented by a contact constraint, which is defined using the following gap function:

$$g = v_{\text{tip}} - \delta \leq 0. \tag{5.2}$$

The physical requirements of contact are that there should be no penetration, the contact force should be positive, and when the gap is greater than zero, the contact force should be zero and vice versa. These requirements dictate that the solution satisfies the following three conditions:

$$\begin{aligned} &\text{No penetration} : g \leq 0, \\ &\text{Positive contact force} : \lambda \geq 0, \\ &\text{Consistency condition} : \lambda g = 0. \end{aligned} \tag{5.3}$$

The above requirements are exactly the same as those of the Lagrange multiplier in a constrained optimization problem. The consistency condition in the above equation can be used to find the correct contact status as well as the contact force. Since the gap is also a function of contact force, using Eqs. (5.1) and (5.2), the above consistency condition can be written as

$$\lambda g = \lambda\left(0.00025 - \frac{\lambda}{3 \times 10^5}\right) = 0.$$

The above quadratic equation has two solutions: $\lambda = 0\,\text{N}$ or $\lambda = 75\,\text{N}$. When $\lambda = 0\,\text{N}$, the gap becomes $g = 0.00025 > 0$, which violates the condition of no penetration. Therefore, this cannot be a possible configuration. On the other hand, when $\lambda = 75\,\text{N}$, the gap becomes $g = 0$. Since this solution satisfies all requirements, this is the solution. In fact, the solution is consistent with the solution from the direct method.

In the above example, the additional unknown (contact force) is added as a Lagrange multiplier, and the consistency condition is used to determine contact status and contact force. In the penalty method, it is also possible to impose the contact constraint without introducing additional unknowns. In the penalty method, a small amount of penetration is allowed, and the contact force is applied proportional to the amount of penetration. Since the gap in Eq. (5.2) can be both positive and negative, the following penetration function is defined:

$$\phi_N = \frac{1}{2}(|g| + g), \tag{5.4}$$

which is zero when $g \leq 0$ and has the same value with g when $g > 0$. Then, the contact force is defined using the penetration function

Table 5.1 Penetrations and contact forces for different penalty parameters

Penalty parameter	Penetration (m)	Contact force (N)
3×10^5	1.25×10^{-4}	37.50
3×10^6	2.27×10^{-5}	68.18
3×10^7	2.48×10^{-6}	74.26
3×10^8	2.50×10^{-7}	74.92
3×10^9	2.50×10^{-8}	75.00

$$\lambda = K_N \phi_N, \tag{5.5}$$

where K_{N} is a penalty parameter. The contact force will be zero when the gap is open and proportionally increase with the penetration. The basic concept is that this method allows a small amount of penetration and then penalizes it by applying a large force. A benefit is that the contact force is now related to the gap, albeit the relationship is nonlinear.

The definition of the gap in Eq. (5.2) can be used to calculate the contact status and contact force, as

$$g = 0.00025 - \frac{K_N}{3 \times 10^5} \frac{1}{2}(|g| + g).$$

When $g \leq 0$ is assumed, the above equation is self-conflicting, which means that penetration occurs. When $g > 0$, the above equation can be solved for the gap with a given penalty parameter K_{N}. Table 5.1 shows the amount of penetration and contact forces for different values of the penalty parameter. It can be observed that as the penalty parameter increases, the penetration decreases and the contact force converges to the accurate value.

Example 5.1. Lagrange multiplier when no contact When the distributed load is 500 N/m, calculate the tip deflection of the beam and determine if contact occurs or not using the Lagrange multiplier method.

Solution From Eq. (5.1), the tip deflection can be written in terms of the Lagrange multiplier as

$$v_{\mathrm{tip}} = 0.625 \times 10^{-3} - \frac{\lambda}{3 \times 10^5}.$$

Using the gap function in Eq. (5.2), the contact consistency condition in Eq. (5.3) can be written as

$$\lambda g = \lambda \left(-0.375 \times 10^{-3} - \frac{\lambda}{3 \times 10^5} \right) = 0.$$

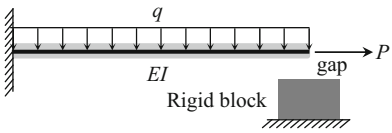

Fig. 5.4 Cantilever beam supported with a potential frictional contact at the tip with a rigid block

The above equation has two solutions: $\lambda = -112.5$ N and $\lambda = 0$. The former violates the inequality condition of the Lagrange multiplier; therefore, it is an invalid solution.[3] The latter yields the gap function of $g = -0.375 \times 10^{-3} < 0$, which satisfies the inequality condition; therefore, it is a valid solution. In fact, it physically means that the gap is not closed and contact does not occur. This can be confirmed by the tip deflection of $v_{\text{tip}} = 0.625$ mm, which is smaller than the initial gap. ■

5.2.2 *Contact of a Cantilever Beam with Friction*

Consider a slightly more complicated problem that involves both a normal contact and friction. A cantilever beam is subjected to a distributed load and an axial load. The free end of the cantilever could potentially contact the block, as shown in Fig. 5.4. Again, it is assumed that the contact can occur only at the tip of the beam. The block surface has a known coefficient of friction μ. The load sequence is such that the transversely distributed load q is applied first, followed by the axial load P. The same numerical data as in Sect. 5.2.1 are used for the beam deflection. For the axial direction, the following data are used: axial load $P = 100$ N, axial rigidity $EA = 10^5$ N, and friction coefficient $\mu = 0.5$.

5.2.2.1 Solution with No Frictional Resistance

From the assumption of infinitesimal deformation, the transverse behavior of the beam can be decoupled with the axial behavior. Therefore, the beam deflection will be identical to the previous section, and the beam will be in contact with the rigid block with a contact force of 75 N. In the axial direction, the displacement can be modeled using an axially loaded bar. Therefore, the tip displacement due to the axial load becomes

$$u_{\text{tip}}^{\text{no-friction}} = \frac{PL}{EA} = 1.0\,\text{mm}. \tag{5.6}$$

This tip displacement will be compared with the case when friction exists at the contact point.

[3] It is interesting to note that the physical interpretation of the negative Lagrange multiplier is the force that is required to apply at the tip of the beam in order to close the gap.

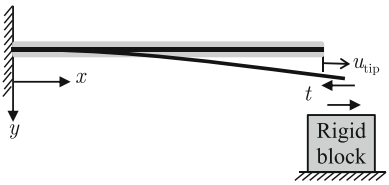

Fig. 5.5 Tangential slip of cantilever beam with friction force

5.2.2.2 Frictional Constraint Function

When friction exists on the interface, the contact point may or may not slip, depending on interface conditions, such as friction coefficient, contact force, and tangential force. In Coulomb's friction model, sliding along the contact surface will take place when the tangential component of the contact force is greater than the frictional resistance. Similar to the normal contact case, the tangential friction force also occurs both in the beam and the rigid block in equal and opposite directions. The positive directions of contact and friction forces are shown in Fig. 5.5. Denoting the normal force at the contact surface as λ and tangential force t, the physical requirements for a frictional constraint are as follows.

$$\begin{aligned} &\text{Stick condition}: \quad t - \mu\lambda < 0, \quad u_{\text{tip}} = 0, \\ &\text{Slip condition}: \quad t - \mu\lambda = 0, \quad u_{\text{tip}} > 0, \\ &\text{Consistency condition}: \quad u_{\text{tip}}(t - \mu\lambda) = 0. \end{aligned} \tag{5.7}$$

When the stick condition occurs, the contact point will not move in the tangential direction, and the tangential force will be determined based on the equilibrium with the externally applied loads, whose magnitude should be less than $\mu\lambda$. When the slip condition occurs, the tangential force will be the same as $\mu\lambda$ and the contact point will continuously move tangentially until the system finds an equilibrium.

The above frictional constraint is similar to the one in Eq. (5.3). Therefore, either the penalty method or Lagrange multiplier method can be applied. The only difference is that now the slip displacement u_{tip} is considered as a Lagrange multiplier, while the friction force is considered as a constraint.

5.2.2.3 Solution Using Trial and Error

In the trial-and-error approach, one condition is assumed first, and then after solving the problem, the other requirements are checked. If all requirements are satisfied,

then the initial assumption is correct and the state is determined. Otherwise, other conditions are assumed until all possible conditions are exhausted.

If the stick condition is assumed first, it means that

$$u_{\text{tip}} = \frac{PL}{EA} - \frac{tL}{EA} = 0 \quad \Rightarrow \quad t = P = 100\,\text{N}.$$

However, this tangential friction force violates the requirement $t - \mu\lambda < 0$. Therefore, the stick condition is not valid.

In order to check with the slip condition, the friction force is first calculated from $t = \mu\lambda = 37.5\,\text{N}$. The friction force will generate the following displacement:

$$u_{\text{tip}} = \frac{PL}{EA} - \frac{tL}{EA} = 0.625\,\text{mm}, \tag{5.8}$$

which satisfies the requirement. Therefore, the slip condition is valid. Note that the slip is less than that of the frictionless assumption in Eq. (5.6).

5.2.2.4 Solution Using Frictional Constraint

In the Lagrange multiplier method, the consistency condition in Eq. (5.7) is used to impose the constraint condition. Compared to the case of normal contact, the choice of the Lagrange multiplier is not obvious in this case. Between the tip displacement and frictional force, the tip displacement is chosen as a Lagrange multiplier and the frictional forcing term, $t - \mu\lambda$, is chosen as a constraint. For the case when the Lagrange multiplier and constraint are switched, the readers are referred to Exercise Problem P5.2. Using the tip displacement formula in Eq. (5.8), the tangential force can be written in terms of the tip displacement as

$$t = P - \frac{EA}{L} u_{\text{tip}}.$$

Therefore, the consistency condition can be written as

$$u_{\text{tip}} \left(P - \frac{EA}{L} u_{\text{tip}} - \mu\lambda \right) = 0. \tag{5.9}$$

The above consistency condition has two solutions: $u_{\text{tip}} = 0$ and

$$u_{\text{tip}} = \frac{(P - \mu\lambda)L}{EA}.$$

The first solution, $u_{\text{tip}} = 0$, corresponds to the stick condition and yields $t = P$. However, since $t - \mu\lambda = 62.5\,\text{N} > 0$, it violates the stick condition. The second solution,

$$u_{\text{tip}} = \frac{(P - \mu\lambda)L}{EA} = 0.625\,\text{mm} > 0$$

corresponds to the slip condition and yields $t = \mu\lambda$, which satisfies the slip condition. Therefore, this is the valid state.

In the penalty method, the constraints on the frictional force are penalized when it violates the condition, that is, when $t - \mu\lambda > 0$. In the same way with the normal contact case, the following penalty function is defined for the frictional force:

$$\phi_T = \frac{1}{2}(|t - \mu\lambda| + t - \mu\lambda). \tag{5.10}$$

Note that $\phi_T = 0$ when $t - \mu\lambda \leq 0$ and $\phi_T = t - \mu\lambda > 0$ when the constraint is violated. In the penalty method, the relationship between the slip displacement and the frictional force can be established by

$$u_{\text{tip}} = K_{\text{T}}\phi_T, \tag{5.11}$$

where K_{T} is the penalty parameter for the tangential slip. When $t - \mu\lambda \leq 0$, the above equation represents a stick condition exactly; i.e., $u_{\text{tip}} = 0$. Therefore, no approximation is involved in the case of a stick condition. On the other hand, the above equation shows a slip condition when $t - \mu\lambda > 0$, i.e., when the constraint is violated. However, the slip condition is penalized with a large value of penalty parameter K_{T} so that the violation remains small.

In order to find the tip displacement using the penalty method, the frictional force is expressed in terms of the tip displacement, as

$$t = P - \frac{EA}{L}u_{\text{tip}}. \tag{5.12}$$

By substituting the above equation into Eq. (5.11), the following tip displacement can be obtained:

$$u_{\text{tip}} = \frac{K_{\text{T}}L(P - \mu\lambda)}{L + K_{\text{T}}EA}.$$

For a large value of K_{T}, the above equation can be approximated by

$$u_{\text{tip}} \approx \frac{(P - \mu\lambda)L}{EA},$$

and the frictional force becomes

Table 5.2 Tip displacement and frictional forces for different penalty parameters

Penalty parameter	Tip displacement (m)	Frictional force (N)
1×10^{-4}	5.68×10^{-4}	43.18
1×10^{-3}	6.19×10^{-4}	38.12
1×10^{-2}	6.24×10^{-4}	37.56
1×10^{-1}	6.25×10^{-4}	37.50
1×10^{0}	6.25×10^{-4}	37.50

$$t = P - \frac{EA}{L} u_{\text{tip}} \approx \mu\lambda,$$

which is nothing but the slip condition. Table 5.2 shows the tip displacements and frictional forces for different values of penalty parameter. It can be observed that as the penalty parameter increases, the tip displacement and the frictional force converges to the accurate value of 0.625 mm and 37.5 N, respectively. It is noted that the penalty parameter is relatively small compared to the case of normal contact because the penalty parameter relates the frictional force to the slip displacement.

Example 5.2. Lagrange multiplier for friction When the force at the tip is $P = 25$ N, calculate the tip displacement of the beam and determine if a stick of slip occurs using the Lagrange multiplier method.

Solution From Eq. (5.9), the consistency condition has two solutions: $u_{\text{tip}} = 0$ and $u_{\text{tip}} = (P - \mu\lambda)L/EA$. The first solution, $u_{\text{tip}} = 0$, corresponds to the stick condition and yields $t = P = 25$ N. Since $t - \mu\lambda = -7.5\,\text{N} < 0$, it satisfies the stick condition. Therefore, the beam is in the stick condition. On the other hand, if the slip condition is checked, the tip displacement

$$u_{\text{tip}} = \frac{(P - \mu\lambda)L}{EA} = -0.075\,\text{mm} < 0$$

becomes negative, which violates the stick condition. Therefore, the stick condition is not a valid state. ■

5.3 General Formulation for Contact Problems

The one-point contact examples in the previous section are limited to a practical point of view, as most contact in engineering applications occurs along a line (one or two dimensional) or a surface (three dimensional). In this section, the basic concepts of one-point contact are extended to two or three dimensions. In order to simplify the presentation, only the penalty method will be discussed.

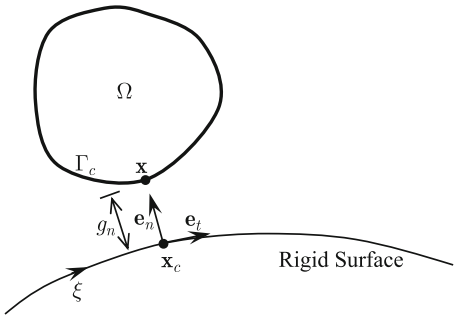

Fig. 5.6 Contact condition in two dimensions

5.3.1 Contact Condition with Rigid Surface

The general formulation is illustrated with reference to contact between two bodies, as shown in Fig. 5.6. The concepts can easily be generalized to contact involving more than two bodies. Note that each body is assumed to be properly supported such that no rigid-body motion is possible even without the contact. In the case of contact with a rigid body, it is natural that the flexible body is selected as a slave body and the rigid body as a master body.

Contact conditions can be divided into normal impenetrability and tangential slip. The impenetrability condition prevents the slave body from penetrating into the master body, while the tangential slip represents the frictional behavior on the contact surface. Figure 5.6 illustrates a general contact condition with a rigid surface in two dimensions. A part of the slave boundary is denoted by contact boundary, Γ_c. Although the actual contact region is unknown and is a part of the solution, the user specifies the contact boundary such that all possible contacts can only occur within this boundary. It is assumed that a point $\mathbf{x}$ on the contact boundary will be in contact with a point $\mathbf{x}_c$ on the master surface if the contact actually occurs. In the following, the contact condition will only be discussed with respect to a single slave point $\mathbf{x}$. Since the motion of the rigid surface is prescribed throughout the analysis, a natural coordinate ξ is used to represent the location on a rigid surface. Thus, the coordinates of contact point $\mathbf{x}_c$ on the master surface can also be represented using a natural coordinate at the contact point ξ_c by

$$\mathbf{x}_c = \mathbf{x}_c(\xi_c). \tag{5.13}$$

However, the contact point $\mathbf{x}_c$, or equivalently ξ_c, is yet unknown. In the three-dimensional space case, two natural coordinates are required to describe the master surface.

In general, contact analysis is to find the contact point and the contact force at the contact point, including contact pressure and frictional force. In finite element analysis, either displacement or force is known at the boundary and the other

unknown variable is solved through the equilibrium requirements. In contact analysis, however, both the contact point $\mathbf{x}_c$ and the contact force at that point are unknown, which makes the contact problem challenging. Usually, a trial-and-error approach is taken in which the contact point is searched from the current geometry, and the contact constraint is imposed once the point is in contact.

The first step of contact analysis is to find the contact point $\mathbf{x}_c(\xi_c)$ on the master surface corresponding to a slave point $\mathbf{x}$. It is necessary to identify this point in order to determine if the two points are in contact or not. Mathematically, this is called the orthogonal projection, or the closest point from the slave point $\mathbf{x}$. When the master boundary is a straight line, the closest point can explicitly be found. For a general nonlinear curve, however, the following nonlinear equation is solved to find the contact point:

$$\varphi(\xi_c) = (\mathbf{x} - \mathbf{x}_c(\xi_c))^T \mathbf{e}_t(\xi_c) = 0, \tag{5.14}$$

where $\mathbf{e}_t = \mathbf{t}/\|\mathbf{t}\|$ is the unit tangential vector and $\mathbf{t} = \mathbf{x}_{c,\xi}$ is the tangential vector at the contact point. The subscribed comma represents differentiation with respect to the following variable; i.e., $\mathbf{x}_{c,\xi} = \partial \mathbf{x}_c / \partial \xi$. Equation (5.14) is called the contact consistency condition, and $\mathbf{x}_c(\xi_c)$ is the closest projection point of $\mathbf{x} \in \Gamma_c$ onto the rigid surface that satisfies Eq. (5.14).

Once the contact point is found, it is necessary to determine if the contact actually occurs, which can be done by measuring the distance between the two points. At the same time, the impenetrability condition can be imposed by using the same distance, as shown in Fig. 5.5. The impenetrability condition can be defined by using the normal gap function g_n, which measures the normal distance, as

$$g_n \equiv (\mathbf{x} - \mathbf{x}_c(\xi_c))^T \mathbf{e}_n(\xi_c) \geq 0, \quad \mathbf{x} \in \Gamma_c, \tag{5.15}$$

where $\mathbf{e}_n(\xi_c)$ is the unit outward normal vector of the master surface at the contact point.

As the contact point moves along the master boundary, a frictional force in a tangential direction to the master boundary resists the tangential relative movement. The tangential slip function g_t is the measure of the relative movement of the contact point along the rigid surface and is defined as

$$g_t \equiv \left\|\mathbf{t}^0\right\| \left(\xi_c - \xi_c^0\right), \tag{5.16}$$

where both the tangential vector $\mathbf{t}^0$ and the natural coordinate ξ_c^0 are the values at the previously converged time increment or load increment. The superscript "0" will denote the previous configuration time in the following derivations.

Example 5.3. Projection to a parabola When a rigid boundary is given as $y = x^2$, find a projection point from $\mathbf{x} = \{3,\ 1\}^T$ using Eq. (5.14) and distance using Eq. (5.15) when $x > 0$.

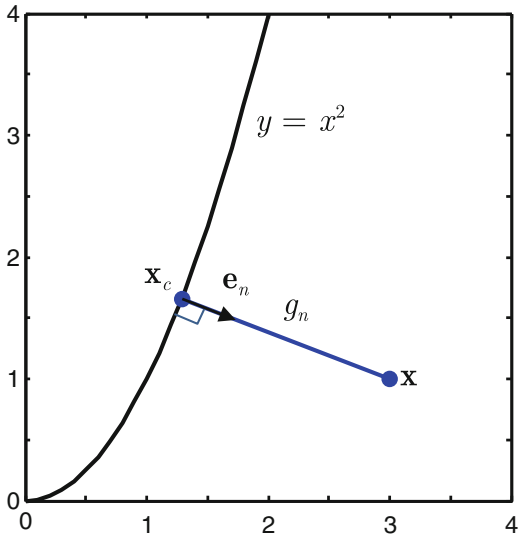

Fig. 5.7 Projection to a parabola

Solution In order to use the contact consistency condition to find a project point, it is necessary to represent the rigid boundary using a parameter. In the case of the given parabola, the parametric relation can be written as $\mathbf{x}_c = \{\xi, \xi^2\}^T$. The unit tangential vector can be calculated by

$$\mathbf{e}_t = \frac{\mathbf{x}_{c,\xi}}{\|\mathbf{x}_{c,\xi}\|} = \frac{1}{\sqrt{1+4\xi^2}} \begin{Bmatrix} 1 \\ 2\xi \end{Bmatrix}.$$

By defining the unit normal vector in the z-coordinate as $\mathbf{k} = \{0 \quad 0 \quad 1\}^T$, the unit normal vector to the rigid boundary can be defined as

$$\mathbf{e}_n = \mathbf{e}_t \times \mathbf{k} = \frac{1}{\sqrt{1+4\xi^2}} \begin{Bmatrix} 2\xi \\ -1 \end{Bmatrix}.$$

Then, the closest project point from $\mathbf{x} = \{3, 1\}^T$ can be found by

$$\varphi(\xi) = (\mathbf{x} - \mathbf{x}_c(\xi))^T \mathbf{e}_t(\xi) = \frac{3 + \xi - 2\xi^3}{\sqrt{1+4\xi^2}} = 0.$$

Since the solution is in the first quadrant, $\xi > 0$, the numerator of the above equation can be solved for $\xi_c = 1.29$. Therefore, the contact point on the rigid boundary becomes $\mathbf{x}_c = \{1.29, \ 1.66\}^T$, as shown in Fig. 5.7.

The distance between the slave point x and contact point $\mathbf{x}_c$ can be obtained using the gap function, as

$$g_{\mathrm{n}} = (\mathbf{x} - \mathbf{x}_{\mathrm{c}})^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} = \frac{-\xi_{\mathrm{c}}^2 + 6\xi_{\mathrm{c}} - 1}{\sqrt{1 + 4\xi_{\mathrm{c}}^2}} = 1.83.$$

The above result can be verified by using the distance formula between two points as

$$\sqrt{(3 - 1.29)^2 + (1 - 1.66)^2} = 1.83.$$

■

5.3.2 *Variational Inequality in Contact Problems*

Before deriving a contact variational formula, it is beneficial to discuss the fundamental properties of the contact problem. Although only a linear elastic problem will be considered for simplicity, due to the inequality constraint on the deformation field, the contact problem is nonlinear even in a linear elastic case. The differential equation of the contact problem can be written as follows:

Governing equilibrium equation:

$$\begin{aligned} &\sigma_{ij,j} + f_i^B = 0, \quad \mathbf{x} \in \Omega, \\ &u_i(\mathbf{x}) = 0, \quad \mathbf{x} \in \Gamma^g, \\ &\sigma_{ij} n_j = f_i^S, \quad \mathbf{x} \in \Gamma^S. \end{aligned} \tag{5.17}$$

Contact conditions:

$$\begin{aligned} &\mathbf{u}^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} + g_{\mathrm{n}} \geq 0, \\ &\sigma_{\mathrm{n}} \geq 0, \quad \mathbf{x} \in \Gamma_{\mathrm{c}}, \\ &\sigma_{\mathrm{n}} (\mathbf{u}^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} + g_{\mathrm{n}}) = 0. \end{aligned} \tag{5.18}$$

The first inequality in Eq. (5.18) can be obtained from the incremental form of the impenetrability condition in Eq. (5.15), since a small deformation linear problem is assumed. Note that the expression of contact conditions in Eq. (5.18) is similar to that of Eq. (5.3). Therefore, either the Lagrange multiplier method or the penalty method can be used to impose the contact condition. The inequality contact constraint in Eq. (5.18) can be considered by constructing a closed convex set $\mathbb{K}$, defined as

$$\mathbb{K} = \left\{ \mathbf{w} \in \left[H^1(\Omega) \right]^N | \mathbf{w}|_{\Gamma^g} = 0 \quad \text{and} \quad \mathbf{w}^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} + g_{\mathrm{n}} \geq 0 \ \text{ on } \ \Gamma_{\mathrm{c}} \right\}. \tag{5.19}$$

The convex set $\mathbb{K}$ satisfies all kinematic constraints (displacement conditions).

If $\mathbf{u}$ is the solution to Eqs. (5.17) and (5.18), then $\mathbf{u} \in \mathbb{K}$. The variational inequality can be derived from the weak formulation of the differential Eq. (5.17). In previous chapters, the weak form is obtained by multiplying the governing differential equation with a virtual displacement $\bar{\mathbf{u}}$, which belongs to the space of kinematically admissible displacements. In order to make the elements in the convex set $\mathbb{K}$ kinematically admissible, the virtual displacement $\bar{\mathbf{u}}$ is substituted by $\mathbf{w} - \mathbf{u}$ for all $\mathbf{w} \in \mathbb{K}$. Therefore, after multiplying $\mathbf{w} - \mathbf{u}$ and integrating by parts, the weak form becomes

$$\begin{aligned} &\int_{\Omega} \sigma_{ij}(\mathbf{u}) \varepsilon_{ij}(\mathbf{w} - \mathbf{u}) \ \mathrm{d}\Omega \\ &\quad = -\int_{\Omega} \sigma_{ij,j}(w_i - u_i) \ \mathrm{d}\Omega + \int_{\Gamma^{\mathrm{S}} \cup \Gamma_{\mathrm{c}}} \sigma_{ij} n_j (w_i - u_i) \ \mathrm{d}\Gamma \\ &\quad = \ell(\mathbf{w} - \mathbf{u}) + \int_{\Gamma_{\mathrm{c}}} \sigma_{ij} n_j (w_i - u_i) \ \mathrm{d}\Gamma, \end{aligned} \tag{5.20}$$

where the last term in Eq. (5.20), which is not known until the solution is obtained, is always nonnegative, as shown below:

$$\begin{aligned} &\int_{\Gamma_{\mathrm{c}}} \sigma_{ij} n_j (w_i - u_i) \ \mathrm{d}\Gamma \\ &\quad = \int_{\Gamma_{\mathrm{c}}} \sigma_{\mathrm{n}} (w_{\mathrm{n}} - u_{\mathrm{n}}) \ \mathrm{d}\Gamma \\ &\quad = \int_{\Gamma_{\mathrm{c}}} \sigma_{\mathrm{n}} (w_{\mathrm{n}} + g_{\mathrm{n}} - u_{\mathrm{n}} - g_{\mathrm{n}}) \ \mathrm{d}\Gamma \\ &\quad = \int_{\Gamma_{\mathrm{c}}} \sigma_{\mathrm{n}} (w_{\mathrm{n}} + g_{\mathrm{n}}) \ \mathrm{d}\Gamma \geq 0, \quad \forall \mathbf{w} \in \mathbb{K}. \end{aligned} \tag{5.21}$$

Thus, variational equation (5.20) becomes a variational inequality as

$$a(\mathbf{u}, \mathbf{w} - \mathbf{u}) \geq \ell(\mathbf{w} - \mathbf{u}), \quad \forall \mathbf{w} \in \mathbb{K}, \tag{5.22}$$

where $\mathbf{u} \in \mathbb{K}$ is the solution.

Figure 5.8 shows the relationship between the solution without contact, $\mathbf{u}' \in \mathbb{Z}$, and the solution with contact, $\mathbf{u}$. If the solution $\mathbf{u}'$ belongs to the convex set $\mathbb{K}$, it satisfies the contact condition and is the solution. However, if $\mathbf{u}'$ is out of the convex set, that is, $\mathbf{u}'$ violates the contact condition, then it has to move to $\mathbf{u}'$ through the orthogonal project, which belongs to the convex set. This conceptual explanation can be illustrated using the beam deflection problem in Sect. 5.2.1. As shown in Fig. 5.9, v' is the deflection curve when there is no rigid block or contact constraint. Since the distributed load is large enough so that the tip deflection is larger than the initial gap, the contact constraint is violated. That is, v' belongs to the space of kinematically admissible displacements, but not in the convex set. Therefore, v' is projected to v on the boundary of the convex set by applying the contact force. The

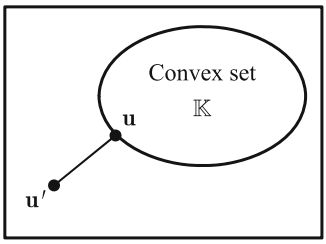

Fig. 5.8 Projection of a solution on to a convex set

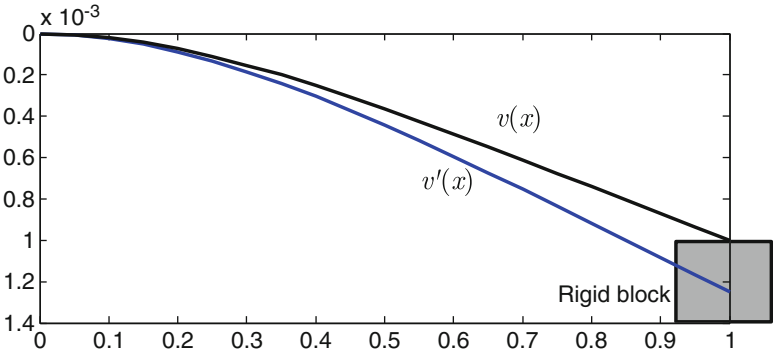

Fig. 5.9 Beam deflection with and without rigid block

physical meaning of the contact force is the force required to project v' onto the convex set, i.e., to satisfy the contact constraint.

The existence and uniqueness of the solution to the variational inequality has been extensively studied for linear elastic material by Duvaut and Lions [1] and Kikuchi and Oden [2]. The existence of a solution to Eq. (5.22) for the nonlinear elastic problem has been proved by Ciarlet [3] for a polyconvex strain energy function.

The same variational inequality in Eq. (5.22) can be used for the nonlinear elastic contact problem with the appropriate structural energy form, as seen in previous chapters. The constraint set of a large deformation problem contains the impenetrability condition in Eq. (5.15) as

$$\mathbb{K} = \left\{ \mathbf{w} \in \left[H^1(\Omega)\right]^N \middle| \mathbf{w}\big|_{\Gamma^g} = 0 \quad \text{and} \quad (\mathbf{x} - \mathbf{x}_c(\xi_c))^T \mathbf{e}_n \geq 0 \ \text{ on } \ \Gamma_c \right\}. \tag{5.23}$$

From an engineering point of view, however, it is not convenient to solve the variational inequality directly without mentioning the construction of a test function on constraint set $\mathbb{K}$. The good news is that it is possible to show that the variational inequality is equivalent to the constrained optimization problem of the total potential energy. If the total potential energy is

$$\Pi(\mathbf{u}) = \frac{1}{2}a(\mathbf{u}, \mathbf{u}) - \ell(\mathbf{u}), \tag{5.24}$$

where $a(\mathbf{u}, \mathbf{u})$ is positive definite, then the directional derivative of $\Pi(\mathbf{u})$ in the direction of $\mathbf{v}$ is defined as

$$\langle D\Pi(\mathbf{u}), \mathbf{v}\rangle = a(\mathbf{u}, \mathbf{v}) - \ell(\mathbf{v}), \tag{5.25}$$

where the bilinear property of $a(\cdot,\cdot)$ and the linear property of $\ell(\cdot)$ are used. Using the directional derivative, the variational inequality $a(\mathbf{u},\mathbf{w}-\mathbf{u}) \geq \ell(\mathbf{w}-\mathbf{u})$ can then be rewritten as

$$\langle D\Pi(\mathbf{u}), \mathbf{w} - \mathbf{u}\rangle \geq 0. \tag{5.26}$$

To show that Eq. (5.26) is equivalent to the constrained minimization problem, let us consider the following relation. For an arbitrary $\mathbf{w} \in \mathbb{K}$,

$$\Pi(\mathbf{w}) - \Pi(\mathbf{u}) = \langle D\Pi(\mathbf{u}), \mathbf{w} - \mathbf{u}\rangle + \frac{1}{2}a(\mathbf{w} - \mathbf{u}, \mathbf{w} - \mathbf{u}). \tag{5.27}$$

Since $a(\cdot,\cdot)$ is positive definite, the last term in Eq. (5.27) is always nonnegative; thus,

$$\Pi(\mathbf{w}) \geq \Pi(\mathbf{u}) + \langle D\Pi(\mathbf{u}), \mathbf{w} - \mathbf{u}\rangle, \quad \forall \mathbf{w} \in \mathbb{K}, \tag{5.28}$$

which means

$$\Pi(\mathbf{u}) = \min_{\mathbf{w}\in\mathbb{K}} \Pi(\mathbf{w}) = \min_{\mathbf{w}\in\mathbb{K}} \left[\frac{1}{2}a(\mathbf{w}, \mathbf{w}) - \ell(\mathbf{w})\right]. \tag{5.29}$$

If $\Pi(\mathbf{w})$ is convex, and set $\mathbb{K}$ is closed and convex, then both the constrained minimization problem in Eq. (5.29) and the variational inequality have a unique solution $\mathbf{u}$. The variational inequality in Eq. (5.22) can be solved using the constrained minimization problem in Eq. (5.29). Many optimization theories can be used, including mathematical programming, sequential quadratic programming, and active set strategies. For further information on the numerical treatment of contact constraints, the mathematical programming method [4, 5], active set strategies [6], and the sequential quadratic programming method [7] are available.

5.3.3 *Penalty Regularization*

In the viewpoint of finite element analysis, the constrained optimization problem in Eq. (5.29) is not trivial to solve, because the fundamental idea of finite element analysis is to build test functions that satisfy zero displacement (kinematic)

boundary conditions, which is nothing but the space of kinematically admissible displacements. Using nodal interpolation of finite elements, this condition can easily be obtained by setting the nodal displacement to zero for those nodes that belong to the displacement boundary. However, it is not trivial to build test functions that satisfy the contact constraint because the contact boundary is unknown until the problem is solved.

Instead of solving the constrained optimization, it is easier to convert the constrained optimization problem into an unconstrained optimization problem by using either the penalty method or the Lagrange multiplier method. The former penalizes the potential energy proportional to the amount of constraint violation such that the minimum of the penalized potential energy approximately satisfies the contact constraint. The latter augments the potential energy by a product of the contact constraint and a Lagrange multiplier, which corresponds to the force to impose the contact constraint, such that the minimum of augmented potential energy can satisfy the contact constraint as well as identify the Lagrange multiplier or the contact force. The advantages and disadvantages of the two methods can be found in traditional optimization textbooks [8]. In this section, only the penalty method will be discussed, but a similar approach can be developed for the Lagrange multiplier method.

In order to penalize when $\mathbf{w} \notin \mathbb{K}$ in Eq. (5.29), if a region Γ_c exists that violates the impenetrability condition in Eq. (5.15), then the potential energy is penalized using a penalty function. That is, the potential energy is penalized when penetration occurs. Similarly, the tangential movement of Eq. (5.16) can also be penalized under the stick condition. The contact penalty function must first be defined for the penetrated region by

$$P = \frac{1}{2}\omega_n \int_{\Gamma_c} {g_n}^2 \, d\Gamma + \frac{1}{2}\omega_t \int_{\Gamma_c} {g_t}^2 \, d\Gamma, \tag{5.30}$$

where ω_n and ω_t are the penalty parameters for normal contact and tangential slip, respectively. The penalty function defined in Eq. (5.30) leads to an exterior penalty method whereby the solution approaches from the infeasible region. This means that the impenetrability condition will be violated, but the amount of violation decreases as the penalty parameter is increased.

The constrained minimization problem in Eq. (5.29) is converted to an unconstrained minimization problem by adding a penalty function to the total potential energy. Thus,

$$\Pi(\mathbf{u}) = \min_{\mathbf{w} \in \mathbb{K}} \Pi(\mathbf{w}) \approx \min_{\mathbf{w} \in \mathbb{Z}} [\Pi(\mathbf{w}) + P(\mathbf{w})]. \tag{5.31}$$

Note that the solution space is changed to $\mathbb{Z}$ from $\mathbb{K}$ because of the penalty function. Therefore, it is much more convenient to build test functions $\mathbf{w} \in \mathbb{Z}$. The variation of Eq. (5.31) contains two contributions that will be examined in this section: one

from the structural potential and the other from the penalty function. The variation of P yields the contact variational form, which is defined by

$$b(\mathbf{u},\overline{\mathbf{u}}) \equiv \omega_{\mathrm{n}} \int_{\Gamma_{\mathrm{c}}} g_{\mathrm{n}} \overline{g}_{\mathrm{n}} \, \mathrm{d}\Gamma + \omega_{\mathrm{t}} \int_{\Gamma_{\mathrm{C}}} g_{\mathrm{t}} \overline{g}_{\mathrm{t}} \, \mathrm{d}\Gamma = b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}}) + b_{\mathrm{T}}(\mathbf{u},\overline{\mathbf{u}}), \tag{5.32}$$

where $b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}})$ and $b_{\mathrm{T}}(\mathbf{u},\overline{\mathbf{u}})$ are the normal and tangential contact variational forms, respectively. The variable with an over-bar represents the variation of the variable. $b_{\mathrm{T}}(\mathbf{u},\overline{\mathbf{u}})$ appears only when there is friction in the contact interface. In Eq. (5.32), $\omega_{\mathrm{n}} g_{\mathrm{n}}$ corresponds to the compressive normal contact force, and $\omega_{\mathrm{t}} g_{\mathrm{t}}$ corresponds to the tangential traction force. The latter increases linearly with the tangential slip g_{t} until it reaches a normal force multiplied by the friction coefficient. The contact variational form in Eq. (5.32) can be expressed in terms of the displacement variation. To make subsequent derivations easier to follow, it is necessary to define several scalar symbols, as follows:

$$\begin{gathered} \alpha \equiv \mathbf{e}_{\mathrm{n}}^{\mathrm{T}} \mathbf{x}_{\mathrm{c},\xi\xi}, \quad \beta \equiv \mathbf{e}_{\mathrm{t}}^{\mathrm{T}} \mathbf{x}_{\mathrm{c},\xi\xi}, \quad \gamma \equiv \mathbf{e}_{\mathrm{n}}^{\mathrm{T}} \mathbf{x}_{\mathrm{c},\xi\xi\xi} \\ c \equiv \|\mathbf{t}\|^2 - g_{\mathrm{n}}\alpha, \quad \nu \equiv \|\mathbf{t}\| \left\|\mathbf{t}^0\right\| / c. \end{gathered} \tag{5.33}$$

Note that α, β, and γ are related to the higher-order derivatives of the master boundary. If the rigid boundary is approximated by a piecewise linear function, then $\alpha = \beta = \gamma = 0$ and $\nu = 1$.

Example 5.4. Penalty method for beam contact Using the potential energy and the penalty method, calculate the deflection curve for the cantilever beam example in Sect. 5.2.1 with different values of penalty parameter. Assume the following form of beam deflection curve $v(x) = a_2 x^2 + a_3 x^3 + a_4 x^4$ and calculate unknown coefficients.

Solution The potential energy of a cantilever beam under a distributed load can be written as

$$\Pi = \frac{1}{2} \int_0^L EI (v_{,xx})^2 \, \mathrm{d}x - \int_0^L qv \, \mathrm{d}x. \tag{5.34}$$

Since the given form of deflection curve satisfies the essential boundary conditions at $x = 0$, i.e., $v(0) = v_{,x}(0) = 0$, it already belongs to the space of kinematically admissible displacements.

In order to apply the penalty constraint to the region where the impenetrability constraint is violated, the deflection curve of the beam should be calculated first by minimizing the potential energy in Eq. (5.34). If the impenetrability constraint is violated, then the penalty constraint is applied to the violated region. This process makes the problem nonlinear, and the solution can be found through an iterative

procedure. However, to simplify the presentation, it is assumed that the impenetrability constraint is violated only at the tip.

In this particular problem, the contact boundary becomes a point at the tip of the beam. In order to define the penalty function, the following form of gap function is defined first:

$$g_{\mathrm{n}} = \delta - v_{\mathrm{tip}} = \delta - a_2 - a_3 - a_4.$$

The integral form of the penalty function in Eq. (5.30) is defined at a point, $x = L$, as

$$P = \frac{1}{2}\omega_{\mathrm{n}} g_{\mathrm{n}}^2.$$

Therefore, the penalized potential energy becomes

$$\Pi + P = \frac{1}{2}\int_0^L EI(v_{,xx})^2\,\mathrm{d}x - \int_0^L qv\,\mathrm{d}x + \frac{1}{2}\omega_{\mathrm{n}} g_{\mathrm{n}}^2.$$

Note that the above penalized potential function is a function of unknown coefficients, a_2, a_3, and a_4. The requirement of its minimum is that the potential energy is stationary with respect to these unknown coefficients.[4] By substituting the expression of v and $v_{,xx}$ into the penalized potential energy, and differentiating with respect to a_2, a_3, and a_4, the following linear system of equations can be obtained:

$$\begin{bmatrix} 4EI + \omega_{\mathrm{n}} & 6EI + \omega_{\mathrm{n}} & 8EI + \omega_{\mathrm{n}} \\ 6EI + \omega_{\mathrm{n}} & 12EI + \omega_{\mathrm{n}} & 18EI + \omega_{\mathrm{n}} \\ 8EI + \omega_{\mathrm{n}} & 18EI + \omega_{\mathrm{n}} & \frac{144}{5}EI + \omega_{\mathrm{n}} \end{bmatrix} \begin{Bmatrix} a_2 \\ a_3 \\ a_4 \end{Bmatrix} = \begin{Bmatrix} \frac{1}{3}q + \omega_{\mathrm{n}}\delta \\ \frac{1}{4}q + \omega_{\mathrm{n}}\delta \\ \frac{1}{5}q + \omega_{\mathrm{n}}\delta \end{Bmatrix}.$$

For the given material, geometric, and load parameters, the three unknown coefficients can be calculated by solving the above matrix equations. For a positive penalty parameter, the coefficient matrix is positive definite. Therefore, a unique solution is expected. Table 5.3 shows the three unknown coefficients, penetration ($a_2 + a_3 + a_4 - \delta$), and the contact force ($-\omega_{\mathrm{n}} g_{\mathrm{n}}$) for different values of the penalty parameter. Similar to the results in Table 5.1, the tip displacement converges to the accurate value as the penalty parameter increases. ■

[4] This is the Rayleigh-Ritz method. For details, readers are referred to N. H. Kim and B. V. Sankar, Introduction to Finite Element Analysis and Design, Wiley & Sons, NY, 2008.

Table 5.3 Coefficients of deflection curve, penetrations, and contact forces for different penalty parameters

Penalty parameter	a_1	a_2	a_3	Penetration (m)	Contact force (N)
3×10^5	2.31×10^{-3}	-1.60×10^{-3}	4.17×10^{-4}	1.25×10^{-4}	37.50
3×10^6	2.16×10^{-3}	-1.55×10^{-3}	4.17×10^{-4}	2.27×10^{-5}	68.18
3×10^7	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	2.48×10^{-6}	74.26
3×10^8	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	2.50×10^{-7}	74.92
3×10^9	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	2.50×10^{-8}	75.00
True value	2.13×10^{-3}	-1.54×10^{-3}	4.17×10^{-4}	0.0	75.00

5.3.4 *Frictionless Contact Formulation*

As an ideal case, the contact formulation when there is no friction in the contact interface is addressed first. Computationally, the frictionless contact problem with elastic material is path independent; that is, the equilibrium state is independent of the load history. From a mechanics point of view, a potential energy (or augmented potential energy with contact penalty function) exists, and all field variables are functions of the current configuration.

The first step is to express the normal contact variational form in terms of displacement variation. By taking the first variation of the normal gap function in Eq. (5.15) and using the variation of the contact consistency condition in Eq. (5.14), the first variation of the normal gap function can be obtained as

$$\overline{g}_{\mathrm{n}}(\mathbf{u};\overline{\mathbf{u}}) = \overline{\mathbf{u}}^{\mathrm{T}}\mathbf{e}_{\mathrm{n}}, \tag{5.35}$$

where the variation of the natural coordinate at the contact point is canceled by an orthogonal condition. The normal gap function can vary only in a normal direction to the rigid surface, which is physically plausible. By using Eq. (5.35), the normal contact form is expressed in terms of displacement variation as

$$b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}}) = \omega_{\mathrm{n}} \int_{\Gamma_{\mathrm{c}}} g_{\mathrm{n}} \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} \, \mathrm{d}\Gamma. \tag{5.36}$$

This contact form originates in the impenetrability condition and the fact that the magnitude of the impenetrability force is proportional to the violation of the impenetrability condition.

Note that $b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}})$ is linear with respect to $\overline{\mathbf{u}}$ and implicit with respect to $\mathbf{u}$ through g_{n} and $\mathbf{e}_{\mathrm{n}}$. Since $b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}})$ is nonlinear in displacement, the same linearization

procedure is required that was used for the structural energy form in Chaps. 3 and 4. The increment of the normal gap function can be obtained in a similar procedure to Eq. (5.35) as

$$\Delta g_{\mathrm{n}}(\mathbf{u};\Delta\mathbf{u}) = \mathbf{e}_{\mathrm{n}}^{\mathrm{T}}\Delta\mathbf{u}. \tag{5.37}$$

To obtain the increment of the unit normal vector, it is necessary to compute the increment of natural coordinate ξ_{c} at the contact point using Eq. (5.14), since the normal vector changes along ξ_{c}. The increment of Eq. (5.14) solves $\Delta\xi_{\mathrm{c}}$ in terms of $\Delta\mathbf{u}$ as

$$\begin{aligned}
&\Delta\left[(\mathbf{x}-\mathbf{x}_{\mathrm{c}})^{\mathrm{T}}\mathbf{e}_{\mathrm{t}}\right] \\
&\quad = (\Delta\mathbf{u}-\mathbf{t}\Delta\xi_{\mathrm{c}})^{\mathrm{T}}\mathbf{e}_{\mathrm{t}} + (\mathbf{x}-\mathbf{x}_{\mathrm{c}})^{\mathrm{T}}\Delta\mathbf{e}_{\mathrm{t}} \\
&\quad = \Delta\mathbf{u}^{\mathrm{T}}\mathbf{e}_{\mathrm{t}} - \|\mathbf{t}\|\Delta\xi_{\mathrm{c}} + (\mathbf{x}-\mathbf{x}_{\mathrm{c}})^{\mathrm{T}}\mathbf{e}_{\mathrm{n}}\left(\frac{1}{\|\mathbf{t}\|}\mathbf{e}_{\mathrm{n}}^{\mathrm{T}}\mathbf{x}_{\mathrm{c},\xi\xi}\right)\Delta\xi_{\mathrm{c}} = 0.
\end{aligned} \tag{5.38}$$

Thus, using the definition in Eq. (5.33), we can calculate the increment of the natural coordinates in terms of increment of displacement, as

$$\Delta\xi_{\mathrm{c}} = \frac{\|\mathbf{t}\|}{c}\mathbf{e}_{\mathrm{t}}^{\mathrm{T}}\Delta\mathbf{u}. \tag{5.39}$$

If $\mathbf{e}_3$ is the fixed unit vector in the out-of-plane direction, then the increment of the unit normal vector can be obtained from the relation $\mathbf{e}_{\mathrm{n}} = \mathbf{e}_3 \times \mathbf{e}_{\mathrm{t}}$ as

$$\begin{aligned}
\Delta\mathbf{e}_{\mathrm{n}} &= \mathbf{e}_3 \times \Delta\mathbf{e}_{\mathrm{t}} \\
&= \mathbf{e}_3 \times \Delta\left[\frac{\mathbf{t}}{\|\mathbf{t}\|}\right] \\
&= \mathbf{e}_3 \times \frac{1}{\|\mathbf{t}\|}\left[\Delta\mathbf{t} - \mathbf{e}_{\mathrm{t}}\left(\mathbf{e}_{\mathrm{t}}^{\mathrm{T}}\Delta\mathbf{t}\right)\right] \\
&= -\frac{1}{\|\mathbf{t}\|}\left[\mathbf{e}_{\mathrm{t}}\left(\mathbf{e}_{\mathrm{n}}^{\mathrm{T}}\Delta\mathbf{t}\right)\right] \\
&= -\frac{\alpha\Delta\xi_{\mathrm{c}}}{\|\mathbf{t}\|}\mathbf{e}_{\mathrm{t}} \\
&= -\frac{\alpha}{c}\mathbf{e}_{\mathrm{t}}\left(\mathbf{e}_{\mathrm{t}}^{\mathrm{T}}\Delta\mathbf{u}\right).
\end{aligned} \tag{5.40}$$

Thus, from Eqs. (5.37) and (5.40), the linearization of the normal contact form is obtained as

$$b_{\mathrm{N}}^{*}(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) = \omega_{\mathrm{n}}\int_{\Gamma_{\mathrm{c}}}\overline{\mathbf{u}}^{\mathrm{T}}\mathbf{e}_{\mathrm{n}}\mathbf{e}_{\mathrm{n}}^{\mathrm{T}}\Delta\mathbf{u}\,\mathrm{d}\Gamma - \omega_{\mathrm{n}}\int_{\Gamma_{\mathrm{c}}}\frac{\alpha g_{\mathrm{n}}}{c}\overline{\mathbf{u}}^{\mathrm{T}}\mathbf{e}_{\mathrm{t}}\mathbf{e}_{\mathrm{t}}^{\mathrm{T}}\Delta\mathbf{u}\,\mathrm{d}\Gamma. \tag{5.41}$$

Note that there is a component in the tangential direction because of the effect of curvature. The first term is the conventional contact tangent term for linear

kinematics. The contribution of the second term is usually small, as the contact violation is reduced. If the contact boundary is linear, the second term disappears as $\alpha = 0$.

In the case of a general nonlinear material with a frictionless contact problem, the principle for virtual work can be written as

$$a(\mathbf{u}, \overline{\mathbf{u}}) + b_{\mathrm{N}}(\mathbf{u}, \overline{\mathbf{u}}) = \ell(\overline{\mathbf{u}}), \quad \forall \overline{\mathbf{u}} \in \mathbb{Z}. \tag{5.42}$$

The above equation is obtained from the first variation of the penalized potential energy function in Eq. (5.31), which is equated to zero to satisfy the Kuhn-Tucker condition. Suppose the current time is t_n and the current iteration counter is $k+1$. Assuming that the external force is independent of displacement, the linearized incremental equation of Eq. (5.42) is obtained as

$$\begin{aligned} &a^*\left({}^n\mathbf{u}^k; \Delta\mathbf{u}^{k+1}, \overline{\mathbf{u}}\right) + b_{\mathrm{N}}^*\left({}^n\mathbf{u}^k; \Delta\mathbf{u}^{k+1}, \overline{\mathbf{u}}\right) \\ &\quad = \ell(\overline{\mathbf{u}}) - a\left({}^n\mathbf{u}^k, \overline{\mathbf{u}}\right) - b_{\mathrm{N}}\left({}^n\mathbf{u}^k, \overline{\mathbf{u}}\right), \quad \forall \overline{\mathbf{u}} \in \mathbb{Z}. \end{aligned} \tag{5.43}$$

Equation (5.43) is linear in incremental displacement for a given displacement variation. The linearized system of Eq. (5.43) is solved iteratively with respect to incremental displacement until the residual forces on the right side of the equation vanish at each time step.

Example 5.5. Frictionless contact of a block A unit square block is under a uniformly distributed load at the top surface and a frictionless contact condition with a rigid body at the bottom surface, as shown in Fig. 5.10. Using the penalty method, calculate the displacement field, penetration, and contact force at the contact interface. Use $EA = 10^5$ N and $q = 1.0$ kN/m and vary the penalty parameter from 10^5 to 10^8. Assume plane strain with zero Poisson's ratio.

Solution In a two-dimensional problem, the displacement and its variation can be written as $\mathbf{u} = \{u_x,\ u_y\}^{\mathrm{T}}$ and $\overline{\mathbf{u}} = \left\{\overline{u}_x,\ \overline{u}_y\right\}^{\mathrm{T}}$, respectively. Since the contact surface is flat and parallel to the x-coordinate, the unit normal vector is constant as $\mathbf{e}_{\mathrm{n}} = \{0,\ 1\}^{\mathrm{T}}$. In this simple problem, the contact boundary can be parameterized by $\xi = x$. Accordingly, the slave contact point and corresponding master point can be written as $\mathbf{x} = \{\xi_{\mathrm{c}},\ 0\}^{\mathrm{T}}$ and $\mathbf{x}_{\mathrm{c}} = \{\xi_{\mathrm{c}},\ u_y\}^{\mathrm{T}}$. Therefore, the gap function can be defined as

$$g_{\mathrm{n}} = (\mathbf{x} - \mathbf{x}_{\mathrm{c}})^{\mathrm{T}}\mathbf{e}_{\mathrm{n}} = u_y.$$

Therefore, the contact form in Eq. (5.36) can be written in terms of displacement as

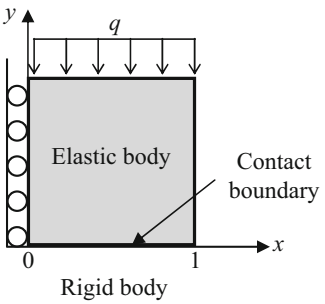

Fig. 5.10 Frictionless contact of an elastic block

$$b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}}) = \omega_{\mathrm{n}} \int_0^1 u_y \overline{u}_y \big|_{y=0} \,\mathrm{d}x.$$

The penalized potential energy for a two-dimensional plane strain problem can be written as

$$\Pi + P = \frac{1}{2}\iint_A \boldsymbol{\varepsilon}^{\mathrm{T}} \mathbf{D} \boldsymbol{\varepsilon} \,\mathrm{d}A - \int_0^1 (-q) u_y \big|_{y=1} \,\mathrm{d}x + \frac{1}{2}\omega_{\mathrm{n}} \int_0^1 g_{\mathrm{n}}^2 \big|_{y=1} \,\mathrm{d}x. \tag{5.44}$$

From the assumption of zero Poisson's ratio, the stress–strain matrix $\mathbf{D}$ becomes a diagonal matrix, and all stress–strain relations are decoupled. In addition, since the load is only applied to y-direction, it can be concluded that $\varepsilon_{xx} = \gamma_{xy} = 0$. Therefore, the only nonzero displacement component will be u_y. Based on the Rayleigh–Ritz method, the following forms of u_y and its variation are assumed:

$$u_y = a_0 + a_1 y, \quad \overline{u}_y = \overline{a}_0 + \overline{a}_1 y.$$

After substituting these approximations into Eq. (5.44) and taking the variation, we have

$$\overline{\Pi} + \overline{P} = \iint_A E a_1 \overline{a}_1 \,\mathrm{d}A - \int_0^1 (-q)(\overline{a}_0 + \overline{a}_1) \,\mathrm{d}x + \omega_{\mathrm{n}} \int_0^1 a_0 \overline{a}_0 \,\mathrm{d}x = 0.$$

Since $\overline{a}_0$ and $\overline{a}_1$ are arbitrary, their coefficients must be zero in order to satisfy the above equation, from which the two coefficients can be determined by

$$a_0 = -\frac{q}{\omega_{\mathrm{n}}}, \quad a_1 = -\frac{q}{EA}.$$

Therefore, the displacement u_y can be determined by

$$u_y = -\frac{q}{\omega_n} - \frac{q}{EA}y, \quad 0 \le y \le 1.$$

The first term on the right-hand side is the contact constraint violation due to the penalty method, while the second term represents the constant strain due to the distributed load. The constraint violation will be reduced as the penalty parameter increases. On the other hand, the contact force remains the same as $-\omega_n g_n = -\omega_n u_y|_{y=0} = q$; that is, the product of penetration and penalty parameter remains constant. Note that the contact force is equal and opposite in direction to the distributed load in order to create equilibrium for the block. ■

5.3.5 *Frictional Contact Formulation*

As mentioned before, frictionless contact is independent of load history. When friction exists at the contact interface, the solution depends on the history of the load applied to the structure. The sequence of the load needs to be considered, and the friction force is determined using not only the current but also the previous location of the contact point. Therefore, it is natural to discuss frictional behavior in the framework of load increment. The current load increment is t_n, and the previous load increment t_{n-1} is converged. For the notational convenience, all variables at load increment t_{n-1} are denoted by a right superscript "0," and all variables at the current load increment are denoted without any superscript.

The classical Coulomb friction law is commonly used in computational mechanics. However, as mentioned in Sect. 5.2.2, due to the discontinuity in the relationship between slip and friction force, it is difficult to handle in the framework of iterative solution procedures based on the Newton–Raphson method, which assumes that the solution is continuous and smooth. As an alternative, the frictional interface law of Wriggers et al. [9] is employed here. This friction law is a regularized version of Coulomb law, such that the vertical portion of the Coulomb model is changed to an inclined line, as shown in Fig. 5.11. The slope of the regularized line can be related to experimental observation.

The tangent slip form $b_T(\mathbf{u}, \overline{\mathbf{u}})$ in Eq. (5.32) can be expressed in terms of a displacement variation. The first variation of the tangential slip function, presented in Eq. (5.16), becomes

$$\overline{g}_t = \|\mathbf{t}^0\| \overline{\xi}_c = \nu \overline{\mathbf{u}}^T \mathbf{e}_t, \tag{5.45}$$

where a procedure similar to Eq. (5.39) is used. Note that the first variations of $\|\mathbf{t}^0\|$ and ξ_c^0 are zero, since they are the solutions to the previous time increment and fixed at the current time. By using Eq. (5.45), the tangential slip variational form in Eq. (5.32) can be rewritten in terms of the displacement variation as

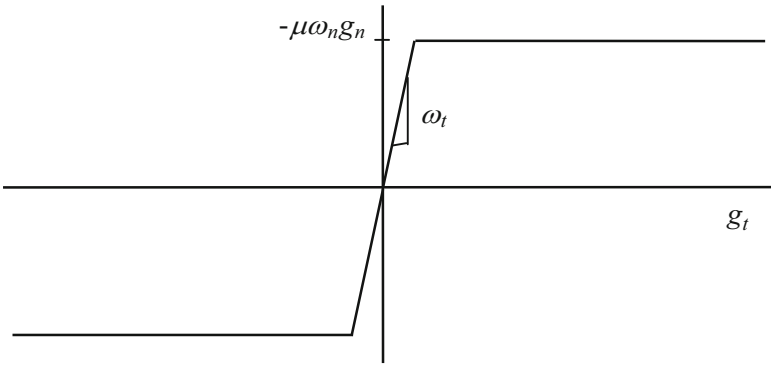

Fig. 5.11 Frictional interface model

$$b_{\mathrm{T}}(\mathbf{u}, \overline{\mathbf{u}}) = \omega_{\mathrm{t}} \int_{\Gamma_{\mathrm{c}}} \nu g_{\mathrm{t}} \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{t}} \, \mathrm{d}\Gamma. \tag{5.46}$$

The frictional traction force $\omega_{\mathrm{t}} g_{\mathrm{t}}$ works in the tangential direction, is proportional to the tangential slip, and is scaled by curvature through ν. As discussed in Eq. (5.33), the variable $v = 1$ when the contact boundary is straight.

The frictional force is bounded above by a compressive normal force multiplied by the friction coefficient in the Coulomb friction law. In the case of a small slip (micro-displacement), however, traction force is proportional to the tangential slip. The penalty parameter ω_{t} is the proportional constant for this case. An exact stick condition represented by a step function in the classical Coulomb friction law is now regularized by a piecewise linear function, with the penalty parameter ω_{t} serving as a regularization parameter. As shown in Fig. 5.11, this regularized friction law is reduced to the classical law as $\omega_{\mathrm{t}} \to \infty$. The regularized stick condition occurs when

$$\omega_{\mathrm{t}} g_{\mathrm{t}} \leq |\mu \omega_{\mathrm{n}} g_{\mathrm{n}}|. \tag{5.47}$$

Otherwise, it becomes a slip condition and $\omega_{\mathrm{t}} g_{\mathrm{t}} = -\mu \omega_{\mathrm{n}} g_{\mathrm{n}}$. In Eq. (5.47), μ is the Coulomb friction coefficient. In the case of a slip condition, the contact variational form has to be modified. Thus, Eq. (5.46) must be divided into two cases as

$$b_{\mathrm{T}}(\mathbf{u}, \overline{\mathbf{u}}) = \begin{cases} \omega_{\mathrm{t}} \displaystyle\int_{\Gamma_{\mathrm{c}}} \nu g_{\mathrm{t}} \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{t}} \, \mathrm{d}\Gamma & \text{if } |\omega_{\mathrm{t}} g_{\mathrm{t}}| \leq |\mu \omega_{\mathrm{n}} g_{\mathrm{n}}| \\ -\mu \omega_{\mathrm{n}} \mathrm{sgn}(g_{\mathrm{t}}) \displaystyle\int_{\Gamma_{\mathrm{c}}} \nu g_{\mathrm{n}} \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{t}} \, \mathrm{d}\Gamma & \text{otherwise} \end{cases}. \tag{5.48}$$

Thus, linearization of the tangential slip variational form has to be separated into stick and slip conditions.

5.3.5.1 Linearization of Stick Condition

The first equation in Eq. (5.48) implicitly depends on displacement through v, g_t, and $\mathbf{e}_t$. The incremental form of g_t can be obtained using the relation in Eq. (5.39) as

$$\begin{aligned}\Delta g_t(\mathbf{u};\Delta\mathbf{u}) &= \left\|\mathbf{t}^0\right\|\Delta\xi_c \\ &= \nu\,\mathbf{e}_t^T\Delta\mathbf{u}.\end{aligned} \tag{5.49}$$

The incremental form of the unit tangential vector can be derived using a procedure similar to that used in Eq. (5.40) with $\mathbf{e}_t = \mathbf{e}_3 \times \mathbf{e}_n$

$$\begin{aligned}\Delta\mathbf{e}_t &= -\mathbf{e}_3 \times \Delta\mathbf{e}_n \\ &= \frac{\alpha}{c}\mathbf{e}_n\left(\mathbf{e}_t^T\Delta\mathbf{u}\right).\end{aligned} \tag{5.50}$$

In addition, the increment of v can be obtained from its definition in Eq. (5.33). After some algebraic calculation, the linearization of Eq. (5.46) leads to the tangential stick bilinear form

$$\begin{aligned}b_T^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) = {}& \omega_t\int_{\Gamma_c}\nu^2\overline{\mathbf{u}}^T\mathbf{e}_t\mathbf{e}_t^T\Delta\mathbf{u}\,d\Gamma \\ &+ \omega_t\int_{\Gamma_c}\frac{\alpha\nu g_t}{c}\overline{\mathbf{u}}^T\left(\mathbf{e}_n\mathbf{e}_t^T+\mathbf{e}_t\mathbf{e}_n^T\right)\Delta\mathbf{u}\,d\Gamma \\ &+ \omega_t\int_{\Gamma_c}\frac{\nu g_t}{c^2}\left((\gamma\|\mathbf{t}\|-2\alpha\beta)g_n-\beta\|\mathbf{t}\|^2\right)\ \overline{\mathbf{u}}^T\mathbf{e}_t\mathbf{e}_t^T\Delta\mathbf{u}\,d\Gamma.\end{aligned} \tag{5.51}$$

Again, for the case of a straight contact boundary, only the first terms on the right-hand side of the above equation survives.

The contact bilinear form is the sum of Eqs. (5.41) and (5.51) as

$$b^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) = b_N^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) + b_T^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}). \tag{5.52}$$

In the case of a stick condition, the contact bilinear form in Eq. (5.52) is symmetric with respect to the incremental displacement and variation of displacement. It is noted that the elastic stick contact condition is a conservative system.

5.3.5.2 Linearization of Slip Condition

As the contact point is forced to move along the contact surface, leading to a violation of Eq. (5.47), the slip contact condition is applied and the second equation from Eq. (5.48) is used. In the case of a slip contact condition, the tangential penalty parameter ω_t is related to the impenetrability penalty parameter ω_n according to the following relation:

$$\omega_{\mathrm{t}} = -\mu\omega_{\mathrm{n}}\mathrm{sgn}(g_{\mathrm{t}}). \tag{5.53}$$

The tangential slip form for the slip condition is

$$b_{\mathrm{T}}(\mathbf{u}, \overline{\mathbf{u}}) = \omega_{\mathrm{t}} \int_{\Gamma_{\mathrm{c}}} \nu\, g_{\mathrm{n}} \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{t}}\, \mathrm{d}\Gamma. \tag{5.54}$$

The linearization of Eq. (5.54) leads to the tangential slip bilinear form as

$$\begin{aligned} b^{*}_{\mathrm{T}}(\mathbf{u}; \Delta\mathbf{u}, \overline{\mathbf{u}}) = {} & \omega_{\mathrm{t}} \int_{\Gamma_{\mathrm{c}}} \nu \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{t}} \mathbf{e}_{\mathrm{n}}^{\mathrm{T}} \Delta\mathbf{u}\, \mathrm{d}\Gamma \\ & + \omega_{\mathrm{t}} \int_{\Gamma_{\mathrm{c}}} \frac{\alpha\nu g_{\mathrm{n}}}{c} \overline{\mathbf{u}}^{\mathrm{T}} \left(\mathbf{e}_{\mathrm{n}} \mathbf{e}_{\mathrm{t}}^{\mathrm{T}} + \mathbf{e}_{\mathrm{t}} \mathbf{e}_{\mathrm{n}}^{\mathrm{T}}\right) \Delta\mathbf{u}\, \mathrm{d}\Gamma \\ & + \omega_{\mathrm{t}} \int_{\Gamma_{\mathrm{c}}} \frac{\nu g_{\mathrm{n}}}{c^2} \left((\gamma \|\mathbf{t}\| - 2\alpha\beta) g_{\mathrm{n}} - \beta \|\mathbf{t}\|^2 \right) \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{t}} \mathbf{e}_{\mathrm{t}}^{\mathrm{T}} \Delta\mathbf{u}\, \mathrm{d}\Gamma. \end{aligned} \tag{5.55}$$

In the case of a slip condition, the contact bilinear form in Eq. (5.55) is not symmetric with respect to the incremental displacement and variation of the displacement. The system is no longer conservative because frictional slip dissipates energy.

In the case of a general nonlinear material with a frictional contact problem, the principle for virtual work can be written as

$$a(\mathbf{u}, \overline{\mathbf{u}}) + b(\mathbf{u}, \overline{\mathbf{u}}) = \ell(\overline{\mathbf{u}}), \quad \forall \overline{\mathbf{u}} \in \mathbb{Z}. \tag{5.56}$$

The current time is t_n and the current iteration counter is $k+1$. Assuming that the external force is independent of displacement, the linearized incremental equation of Eq. (5.56) is obtained as

$$\begin{aligned} & a^{*}\left({}^{n}\mathbf{u}^{k}; \Delta\mathbf{u}^{k+1}, \overline{\mathbf{u}}\right) + b^{*}\left({}^{n}\mathbf{u}^{k}; \Delta\mathbf{u}^{k+1}, \overline{\mathbf{u}}\right) \\ & \quad = \ell(\overline{\mathbf{u}}) - a\left({}^{n}\mathbf{u}^{k}, \overline{\mathbf{u}}\right) - b\left({}^{n}\mathbf{u}^{k}, \overline{\mathbf{u}}\right), \quad \forall \overline{\mathbf{u}} \in \mathbb{Z}. \end{aligned} \tag{5.57}$$

Equation (5.57) is linear in incremental displacement for a given displacement variation. This linearized equation is solved iteratively with respect to incremental displacement until the residual forces (the right side of the equation) vanish at each time step.

Example 5.6. Frictional slip of a cantilever beam The cantilever beam in Example 5.4 is now under additional axial load $P = 100$ N at the tip, after the distributed load q is applied. Using the variation of the penalized potential energy, determine the stick or slip condition and calculate the tip displacement. Use friction penalty parameter $\omega_{\mathrm{t}} = 10^6$, axial rigidity $EA = 10^5$ N, and friction coefficient $\mu = 0.5$. Assume the axial displacement in the form of $u(x) = a_0 + a_1 x$.

Solution From Example 5.4, the cantilever beam is in contact with the rigid block with the contact force of $F_c = -\omega_n g_n = 75$ N. From the infinitesimal deformation assumption, the bending behavior of the beam can be decoupled (or sequential) with the axial behavior. Therefore, it is possible to write the penalized potential energy of the axial behavior and take a variation to find an equilibrium. The penalized potential energy becomes

$$\Pi_a = \int_0^L EA(u_{,x})^2 \, \mathrm{d}x - Pu(L) + \frac{1}{2}\omega_t g_t^2 \Big|_{x=L}.$$

The variation of the penalized potential energy becomes

$$\overline{\Pi}_a = \int_0^L EAu_{,x}\overline{u}_{,x} \, \mathrm{d}x - P\overline{u}(L) + \omega_t g_t \overline{g}_t|_{x=L} = 0, \quad \forall \overline{u} \in \mathbb{Z}.$$

The assumed axial displacement must satisfy the essential boundary condition, which is $u(0) = 0$ in this case. Therefore, the first coefficient should be zero, and $u(x) = a_1 x$; only one coefficient needs to be identified. The gradient and its variation of displacement can be written in terms of the unknown coefficient as $u_{,x} = a_1$ and $\overline{u}_{,x} = \overline{a}_1$.

The tangential slip function needs to be expressed in terms of displacement using the definition in Eq. (5.16). In order to simplify the calculation, it can be assumed that the parametric coordinate x has an origin at $x = L$, and it has the same length as the x-coordinate. Based on this setting, it can be derived that $\|\mathbf{x}_{c,\xi}\| = \|\mathbf{t}\| = \|\mathbf{t}^0\| = 1$ and $\xi_c^0 = 0$. In addition, the tangential slip becomes $g_t = \xi_c = u(L) = a_1$.

First, the stick condition is assumed; that is, $\omega_t g_t \leq |\mu\omega_n g_n|$ must be satisfied once the solution is obtained. After substituting the above variables, the variation of the penalized potential energy becomes

$$\overline{a}_1(EAa_1 + \omega_t a_1 - P) = 0, \quad \forall \overline{a}_1 \in \mathbb{R},$$

where $\mathbb{R}$ is the space of real number. In order to satisfy the above equation for all $\overline{a}_1$, the terms in the parenthesis must vanish, which can be solved for the unknown coefficient a_1. Therefore, the axial displacement becomes

$$u(x) = \frac{Px}{EA + \omega_t} = 9.09 \times 10^{-5} x.$$

Using the tip displacement, the stick condition should be checked, as

$$\omega_t g_t = 90.9 > 37.5 = |\mu\omega_n g_n|.$$

Since the assumption of the stick condition is violated, the beam is under the slip condition, where the variation of the penalized potential energy can be written as

$$\overline{\Pi}_a = \int_0^L EAu_{,x}\overline{u}_{,x}\,\mathrm{d}x - P\overline{u}(L) - \mu\omega_{\mathrm{n}}\mathrm{sgn}(g_{\mathrm{t}})g_{\mathrm{n}}\overline{g}_{\mathrm{t}}|_{x=L} = 0, \quad \forall\overline{u} \in \mathbb{Z}.$$

From the normal contact result in Example 5.4, it can be concluded that $-\mu\omega_{\mathrm{n}}g_{\mathrm{n}} = 37.5\,\mathrm{N}$. Therefore, the above penalized potential energy becomes

$$\overline{a}_1(EAa_1 - P + 37.5) = 0, \quad \forall\overline{a}_1 \in \mathbb{R}$$

which yields $a_1 = 62.5 \times 10^{-5}$ and the tip displacement $u_{\mathrm{tip}} = 0.625\,\mathrm{mm}$, which is consistent with the result in Eq. (5.8). ■

5.4 Finite Element Formulation of Contact Problems

As mentioned before, since the contact formulation is independent of constitutive models, it is enough to discuss finite element formulation of the contact variational form in Eq. (5.30). Then, it can be added to the matrix equation of different materials, for example, elastic material models in Chap. 3 and elastoplastic material models in Chap. 4. Therefore, in the following, only the discretization of the contact variational form will be discussed.

Since the contact problem is solved as a part of finite element analysis, it makes sense to formulate the contact problem in the same context. For that purpose, the contact variational form in Eq. (5.30) is calculated on the boundary of the discretized finite element domain. If the structural domain is discretized by two-dimensional finite elements, then the contact problem is defined on the boundary of two-dimensional finite elements, that is, along a boundary curve. In the case of three dimensions, the contact problem is defined on the boundary surface. In this section, contact conditions in two dimensions are discussed. In order to make the presentation simple, the contact between a flexible body and a rigid body will be discussed first in Sect. 5.4.1, followed by contact between two flexible bodies in Sect. 5.4.2.

5.4.1 Contact Between a Flexible Body and a Rigid Body

The simplest formulation of a contact problem can be obtained when a flexible body is in contact with a rigid body, which is the main topic of this section. In general, it is possible that the rigid body can move to satisfy the equilibrium; but in this text, it is assumed that either the rigid body is fixed or its motion is prescribed. In such a case, it is obvious to choose the flexible body as a slave body and the rigid body as a master body so that the flexible body cannot penetrate the rigid body. In fact, it is sufficient to define the master boundary, not the entire master body. In addition,

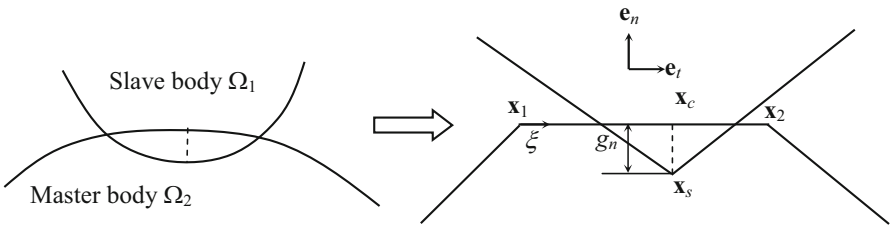

Fig. 5.12 Continuum vs. discrete contact conditions

since the master body is not governed by equilibrium, the contact variational form is only calculated on the slave boundary.

Let Γ_c be a portion of the slave boundary where the slave body penetrates the master body as shown in Fig. 5.12. This boundary is represented by a set of slave nodes that penetrate the master boundary. In the following, a single slave node is considered. Although different ways of defining contact constraints exist in finite elements, in this section it is assumed that the contact constraint is defined using a pair that includes a slave node and a master segment. In addition, only a straight master segment that is defined by two nodes is considered. Therefore, a contact pair can be defined using a slave node and two master nodes, as $\mathbf{X} = \{\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}_2\}^{\mathrm{T}}$. It is possible that one slave node can be associated with different master segments that have a possibility of making contact with the slave node. The two master nodes are ordered in such a way that the master body is located on the right-hand side of the directional line segment from node $\mathbf{x}_1$ to node $\mathbf{x}_2$. The natural coordinate ξ on the master boundary is defined such that it is zero at $\mathbf{x}_1$ and one at $\mathbf{x}_2$.

For a given contact pair $\mathbf{X}$, the objectives are (1) to find if the contact pair is in contact or in separation and (2) to calculate the contact force and penetration if it is in contact. The first objective is called "contact search." In a large-scale model, many slave nodes have a possibility of making contact with many master segments. Therefore, the number of contact pairs is huge and a lot of computational time is often consumed in search of contact pairs that are actually in contact. Once these actual contact pairs are identified, the contact force is calculated for these pairs in the second step.

5.4.1.1 Normal Contact

For a given contact pair, the unit normal and tangential vectors can be defined as

$$\mathbf{t} = \mathbf{x}_2 - \mathbf{x}_1, \quad \mathbf{e}_{\mathrm{t}} = \frac{\mathbf{t}}{\|\mathbf{t}\|}, \quad \mathbf{e}_{\mathrm{n}} = \mathbf{e}_3 \times \mathbf{e}_{\mathrm{t}}.$$

Since the master segment is straight and fixed, the above vectors are also fixed. Because of the linear master segment, the contact consistency condition in

Eq. (5.14) and the gap in Eq. (5.15) can explicitly be calculated. First, the gap can be calculated by

$$g_{\mathrm{n}} = (\mathbf{x}_s - \mathbf{x}_1)^{\mathrm{T}} \mathbf{e}_{\mathrm{n}}. \tag{5.58}$$

If $g_{\mathrm{n}} > 0$, then contact does not occur in this pair and no further calculation is required. If $g_{\mathrm{n}} \leq 0$, then one more check is required. That is, the natural coordinate at the contact point must be within $0 \leq \xi_{\mathrm{c}} \leq 1$ so that the contact occurs within this master segment. The natural coordinate at the contact point can be calculated by

$$\xi_{\mathrm{c}} = \frac{1}{\|\mathbf{t}\|}(\mathbf{x}_s - \mathbf{x}_1)^{\mathrm{T}} \mathbf{e}_{\mathrm{t}}.$$

If $\xi_{\mathrm{c}} < 0$ or $\xi_{\mathrm{c}} > 1$, then contact does not occur in this segment and no further calculation is required. If $0 \leq \xi_{\mathrm{c}} \leq 1$, then contact occurs in the segment and needs to calculate the contact force, which acts in the direction to the normal vector and proportional to penetration, as

$$\mathbf{f}_{\mathrm{n}}^{\mathrm{c}} = -\omega_{\mathrm{n}} g_{\mathrm{n}} \mathbf{e}_{\mathrm{n}}.$$

In Newton–Raphson iteration during nonlinear analysis, the tangent stiffness of the above contact force is required. Since $\mathbf{e}_{\mathrm{n}}$ is fixed, no linearization is required. Therefore, only the gap needs to be linearized, which is similar to the variation of the gap in Eq. (5.35), by replacing the displacement variation with the displacement increment. Therefore, the contact stiffness can be obtained by

$$\mathbf{k}_{\mathrm{n}}^{\mathrm{c}} = \omega_{\mathrm{n}} \mathbf{e}_{\mathrm{n}} \mathbf{e}_{\mathrm{n}}^{\mathrm{T}}.$$

In the continuum formulation, $b_{\mathrm{N}}(\mathbf{u}, \overline{\mathbf{u}})$ is expressed as an integral along the boundary Γ_{c}. However, using the slave–master pair and collocation integration, the boundary integral along the Γ_{c} is approximated by the summation for those violated slave nodes. Let the contact boundary of body Ω_2 in Fig. 5.12 be represented by piecewise linear master segments, with a slave node on the contact boundary of Ω_1. Since Γ_{c} is not known in priori, contact search has to be carried out first to find those violated nodes. Let *NC* be the number of slave nodes that penetrate the master segment. Then, the discretized contact variational form becomes

$$b_{\mathrm{N}}(\mathbf{u}, \overline{\mathbf{u}}) \approx \sum_{I=1}^{NC} \left[\overline{\mathbf{u}}^{\mathrm{T}} (\omega_{\mathrm{n}} g_{\mathrm{n}} \mathbf{e}_{\mathrm{n}})\right]_I = \sum_{I=1}^{NC} \left[\overline{\mathbf{u}}^{\mathrm{T}} \left(-\mathbf{f}_{\mathrm{n}}^{\mathrm{c}}\right)\right]_I \equiv \overline{\mathbf{u}}_{\mathrm{g}}^{\mathrm{T}} \left(-\mathbf{F}_{\mathrm{n}}^{\mathrm{c}}\right), \tag{5.59}$$

where $\mathbf{F}_{\mathrm{n}}^{\mathrm{c}}$ is the contact force in the global coordinate, which is constructed by adding contact forces at each slave node to the corresponding global degrees of freedom. Since the contact variational form occurs on the left-hand side of Eq. (5.56), it will be moved to the right-hand side as a residual force during a

Newton–Raphson iteration as in Eq. (5.57). Including a negative sign in front of the contact force in Eq. (5.59) is equivalent to adding the contact force to the global residual force.

At this point, it is a good idea to discuss collocation integrals. In general, numerical integration of a function approximates the integral by function values and associated weights at selected integration points. A collocation integral simply chooses the integration points at the node. This choice is a matter of convenience and accuracy. Since most field variables are calculated at nodes in the finite element method, it is convenient to use the nodal values in integration. This is why many finite element programs use collocation integrals for contact analysis. Since the accuracy of numerical integration depends on the number of integration points, a single point integration at a node is less accurate. The weight represents the domain that an integration point covers. For example, if a constant function is integrated over an area with a single integration point, then the weight is the same as the area. However, it is not commonly known that the weight of integration is implicitly included in the function in a collocation integral. For the case of the contact variational form in Eq. (5.59), the weight is included in the gap, g_{n}. This concept is further explained in Example 5.7.

In a Newton–Raphson iteration during nonlinear analysis, linearization of the contact variational form needs to be calculated, which yields the tangent stiffness matrix. Linearization of $b_{\mathrm{N}}(\mathbf{u}, \overline{\mathbf{u}})$ becomes

$$b_{\mathrm{N}}^{*}(\mathbf{u}; \Delta\mathbf{u}, \overline{\mathbf{u}}) = \sum_{I=1}^{NC} \left(\omega_{\mathrm{n}} \overline{\mathbf{u}}^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} \mathbf{e}_{\mathrm{n}}^{\mathrm{T}} \Delta\mathbf{u}\right)_I = \sum_{I=1}^{NC} \left(\overline{\mathbf{u}}^{\mathrm{T}} \mathbf{k}_{\mathrm{n}}^{\mathrm{c}} \Delta\mathbf{u}\right)_I \equiv \overline{\mathbf{u}}_g^{\mathrm{T}} \mathbf{K}_{\mathrm{n}}^{\mathrm{c}} \Delta\mathbf{u}_g, \qquad (5.60)$$

where $\mathbf{K}_{\mathrm{n}}^{\mathrm{c}}$ is the contact stiffness in the global coordinate, which is constructed by adding contact stiffness matrices at each slave node to the corresponding global degrees of freedom. Using a too-large value for the penalty parameter can cause a numerical difficulty because it makes the matrix ill-conditioned.

Example 5.7. Contact force and gap of a block A unit square block is under a uniformly distributed load $q = 1.0$ kN/m on the top surface. The bottom surface is under contact constraint against the rigid floor. When the block is modeled by one and four finite elements, as shown in Fig. 5.13, calculate contact forces and gaps at contact nodes. Assume isotropic material and no friction in the contact interface. Use the penalty parameter $\omega_{\mathrm{n}} = 10^5$.

Solution Since a uniformly distributed load is applied, the finite elements are under constant stress. Therefore, without detailed calculation, the contact forces at the bottom two nodes should be $f_1^{\mathrm{c}} = f_2^{\mathrm{c}} = 500$ N. Since the contact force is generated by the gap multiplied by the penalty parameter, the gap at two nodes should be $g_{\mathrm{n}1} = g_{\mathrm{n}2} = -10^5 \times 500 = -0.005$.

In the case of four elements, the two bottom elements are in contact with the rigid floor. Since each element is under constant stress and distributes the equal contact force to the two nodes, the contact forces at the three nodes become

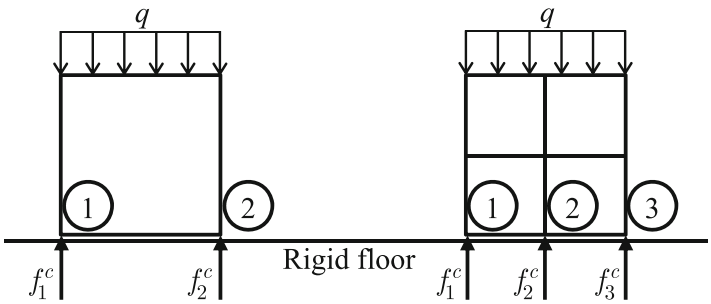

Fig. 5.13 Contact forces of square block

$f_1^c = f_3^c = 250\,\mathrm{N}$ and $f_2^c = 500\,\mathrm{N}$. Again, using the penalty method, the gap at each node can be calculated using the contact force by $g_{n1} = g_{n3} = -0.0025$ and $g_{n2} = -0.005$. As more elements are used along the contact boundary, the gap will become smaller. It is clear that the normal gap changes as the element size changes, which means that the discretized normal gap includes the integration weight implicitly. ■

5.4.1.2 Frictional Slip

Different from the normal contact, the tangential slip under friction requires information from the current as well as the reference configuration. The reference configuration can be the initial state or the previous time increment, but for a large deformation problem, the previous time increment can be more accurate. For the straight master segment, the tangential slip is first defined by

$$g_t = l^0\left(\xi_c - \xi_c^0\right), \tag{5.61}$$

where the right superscript "0" denotes the value evaluated at the previous time, $\xi_c \in [0, 1]$ is the natural coordinate corresponding to the contact point on the master segment, and l^0 is the length of the master segment. Since the master segment is rigid, its length does not change, but the above definition is used in order to be compatible with the case of two flexible bodies in contact.

In the penalty method, the friction force is generated if the tangential slip is not zero, proportional to the tangential penalty parameter ω_t, as

$$\mathbf{f}_t^c = -\omega_t g_t \mathbf{e}_t. \tag{5.62}$$

The above linear relationship is called the stick condition because the tangential slip disappears if the tangential force vanishes. Therefore, the two bodies stick together and behave similar to an elastic material. Also, in the stick condition, the friction force can be understood as a recovery force against tangential deformation. Therefore, the stick condition represents an elastic deformation before slip occurs.

However, the tangential force cannot indefinitely increase in the Coulomb friction model. The magnitude of the tangential force is limited by the friction coefficient multiplied by the normal contact force as in Eq. (5.47). Once the slip is greater than the limit, then the tangential force is limited by

$$\mathbf{f}_{\mathrm{t}}^{\mathrm{c}} = \mu\omega_{\mathrm{n}}\mathrm{sgn}(g_{\mathrm{t}})g_{\mathrm{n}}\mathbf{e}_{\mathrm{t}}, \quad if \quad |\omega_{\mathrm{t}}g_{\mathrm{t}}| \geq |\mu\omega_{\mathrm{n}}g_{\mathrm{n}}|, \tag{5.63}$$

where μ is the frictional coefficient. Thus, the tangential contribution is separated into two cases: the stick and slip conditions. The stick condition is applied when the tangential force is small such that only a microscopic relative movement is observed and the frictional force is proportional to the relative deformation. The slip condition is applied when macroscopic movement is occurred with the critical force. In this case, Eq. (5.63) is used to calculate tangential friction force.

For the stick condition, the tangent stiffness of the above frictional force becomes

$$\mathbf{k}_{\mathrm{t}}^{\mathrm{c}} = \omega_{\mathrm{t}}\mathbf{e}_{\mathrm{t}}\mathbf{e}_{\mathrm{t}}^{\mathrm{T}}$$

while the tangent stiffness of the slip condition becomes

$$\mathbf{k}_{\mathrm{t}}^{\mathrm{c}} = \mu\omega_{\mathrm{n}}\mathrm{sgn}(g_{\mathrm{t}})\mathbf{e}_{\mathrm{t}}\mathbf{e}_{\mathrm{n}}^{\mathrm{T}}.$$

Note that the tangent stiffness matrix is unsymmetric for the slip condition.

Now, the tangential slip form representing the frictional behavior of the contact interface can be discretized by

$$b_{\mathrm{T}}(\mathbf{u}, \overline{\mathbf{u}}) \approx \sum_{I=1}^{NC} \left[\overline{\mathbf{u}}^{\mathrm{T}}\left(-\mathbf{f}_{\mathrm{t}}^{\mathrm{c}}\right)\right]_I \equiv \overline{\mathbf{u}}_g^{\mathrm{T}}\left(-\mathbf{F}_{\mathrm{t}}^{\mathrm{c}}\right), \tag{5.64}$$

where $\mathbf{F}_{\mathrm{t}}^{\mathrm{c}}$ is the frictional force in the global coordinate, which is constructed by adding frictional forces at each slave node to the corresponding global degrees of freedom. And the linearization of the tangential slip form yields

$$b_{\mathrm{T}}^{*}(\mathbf{u}; \Delta\mathbf{u}, \overline{\mathbf{u}}) \approx \sum_{I=1}^{NC} \left(\overline{\mathbf{u}}^{\mathrm{T}}\mathbf{k}_{\mathrm{t}}^{\mathrm{c}}\Delta\mathbf{u}\right)_I \equiv \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{t}}^{\mathrm{c}}\Delta\mathbf{u}_g,$$

where $\mathbf{K}_{\mathrm{t}}^{\mathrm{c}}$ is the contact stiffness in the global coordinate, which is constructed by adding contact stiffness matrices at each slave node to the corresponding global degrees of freedom.

5.4.2 Contact Between Two Flexible Bodies

When both the slave and master bodies are flexible, the finite element discretization becomes more complicated as the contact force is applied to both bodies. In this case, the slave nodes are the boundary nodes of the slave finite elements, while the master segments are the edges of master finite elements. In the case of self-contact, the slave finite elements are the same with the master finite elements.

The contact force is directly calculated at the slave node. However, in the case of the master segment, the contact force is applied at the ξ_c location of the segment. Therefore, the contact force is distributed to the two master nodes proportional to the distance from the contact point. For example, when the contact force $\mathbf{f}_n^c$ occurs at ξ_c, this force is distributed to two master nodes by $[\mathbf{f}_{n1}^c,\ \mathbf{f}_{n2}^c] = [-(1-\xi_c)\mathbf{f}_n^c,\ -\xi_c\mathbf{f}_n^c]$. The negative sign is added in the contact force because the direction of contact force is reversed at the master segment. Therefore, in the contact pair, $\hat{\mathbf{x}} = \{\mathbf{x}_s, \mathbf{x}_1, \mathbf{x}_2\}^T$, the contact forces can also be written in the same format as $\hat{\mathbf{f}}_n^c = \{\mathbf{f}_n^c, -(1-\xi_c)\mathbf{f}_n^c, -\xi_c\mathbf{f}_n^c\}^T$. Considering the contact force as an internal force, the sum of contact forces in a contact pair vanishes. In the following, the superposed "hat" symbol will be used to represent the nodal values in a contact pair, for example, $\hat{\mathbf{u}} = \{\mathbf{u}_s,\ \mathbf{u}_1,\ \mathbf{u}_2\}^T$.

In the multi-body contact, the contact variational forms in Eq. (5.32) need to be modified to include the effect of master surface, as

$$b_N(\mathbf{u}, \overline{\mathbf{u}}) = \omega_n \int_{\Gamma_c} g_n \mathbf{e}_n^T (\overline{\mathbf{u}}_s - \overline{\mathbf{u}}_c)\, d\Gamma,$$

$$b_T(\mathbf{u}, \overline{\mathbf{u}}) = \omega_t \int_{\Gamma_c} g_t \left(\nu \mathbf{e}_t^T (\overline{\mathbf{u}}_s - \overline{\mathbf{u}}_c) + \frac{g_n \left\| \mathbf{t}^0 \right\|}{c} \mathbf{e}_n^T \overline{\mathbf{u}}_{c,\xi} \right) d\Gamma.$$

Note that the tangential slip form has a normal component; the contact point can move in both normal and tangential directions due to the movement of master surface. For detailed derivations of the above equations, readers are referred to Kim et al. [10].

For notational convenience of derivations, the following sets of vectors (6×1) are defined, which has become quite standard [11]:

$$\hat{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_s \\ \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}, \quad \mathbf{N} = \begin{bmatrix} \mathbf{e}_n \\ -(1-\xi_c)\mathbf{e}_n \\ -\xi_c \mathbf{e}_n \end{bmatrix}, \quad \mathbf{T} = \begin{bmatrix} \mathbf{e}_t \\ -(1-\xi_c)\mathbf{e}_t \\ -\xi_c \mathbf{e}_t \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{e}_n \\ \mathbf{e}_n \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} \mathbf{0} \\ -\mathbf{e}_t \\ \mathbf{e}_t \end{bmatrix}$$

$$\mathbf{C}_n = \mathbf{N} - \frac{g_n}{l}\mathbf{Q}, \quad \mathbf{C}_t = \mathbf{T} + \frac{g_n}{l}\mathbf{P}. \tag{5.65}$$

Then, the contact variational form $b(\mathbf{u}, \overline{\mathbf{u}})$ can be discretized by

$$
\begin{aligned}
b(\mathbf{u},\overline{\mathbf{u}}) &= b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}}) + b_{\mathrm{T}}(\mathbf{u},\overline{\mathbf{u}}) \\
&\approx \sum_{I=1}^{NC}\left[\hat{\overline{\mathbf{u}}}^{\mathrm{T}}(\omega_{\mathrm{n}} g_{\mathrm{n}}\mathbf{N})\right]_I + \sum_{I=1}^{NC}\left[\hat{\overline{\mathbf{u}}}^{\mathrm{T}}(\omega_{\mathrm{t}} g_{\mathrm{t}}\mathbf{C}_{\mathrm{t}})\right]_I \\
&\equiv \overline{\mathbf{u}}_g^T(-\mathbf{F}_{\mathrm{C}}),
\end{aligned}
\tag{5.66}
$$

where $\mathbf{F}_{\mathrm{C}}$ is the contact force in the global coordinate, which is constructed by adding contact forces at slave and master nodes to the corresponding global degrees of freedom.

The linearization of the above contact variational form can be obtained by following a similar procedure as with the previous section. First, linearization of $b_{\mathrm{N}}(\mathbf{u},\overline{\mathbf{u}})$ becomes

$$
b_{\mathrm{N}}^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) = \sum_{I=1}^{NC}\left(\omega_{\mathrm{n}}\hat{\overline{\mathbf{u}}}^{\mathrm{T}}\left[\mathbf{C}_{\mathrm{n}}\mathbf{C}_{\mathrm{n}}^{\mathrm{T}}\right]\Delta\hat{\mathbf{u}}\right)_I \equiv \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{N}}\Delta\mathbf{u}_g. \tag{5.67}
$$

Linearization of $b_{\mathrm{T}}(\mathbf{u},\overline{\mathbf{u}})$ should be considered in two different cases.

For the stick condition,

$$
\begin{aligned}
b_{\mathrm{T}}^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) &= \sum_{I=1}^{NC}\left(\omega_{\mathrm{t}}\hat{\overline{\mathbf{u}}}^{\mathrm{T}}\left[\mathbf{C}_{\mathrm{t}}\mathbf{C}_{\mathrm{t}}^{\mathrm{T}} + \frac{2g_{\mathrm{t}}}{l}\left(\mathrm{sym}\left(\mathbf{C}_{\mathrm{n}}\mathbf{P}^{\mathrm{T}}\right) - \mathrm{sym}\left(\mathbf{C}_{\mathrm{t}}\mathbf{Q}^{\mathrm{T}}\right)\right)\right]\Delta\hat{\mathbf{u}}\right)_I \\
&\equiv \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{T}}\Delta\mathbf{u}_g.
\end{aligned}
\tag{5.68}
$$

For the slip condition,

$$
\begin{aligned}
b_{\mathrm{T}}^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) &= \sum_{I=1}^{NC}\left(\omega_{\mathrm{t}}\hat{\overline{\mathbf{u}}}^{\mathrm{T}}\left[\mathbf{C}_{\mathrm{t}}\mathbf{N}^{\mathrm{T}} + \frac{2g_N}{l}\left(\mathrm{sym}\left(\mathbf{C}_{\mathrm{n}}\mathbf{P}^{\mathrm{T}}\right) - \mathrm{sym}\left(\mathbf{C}_{\mathrm{t}}\mathbf{Q}^{\mathrm{T}}\right)\right)\right]\Delta\hat{\mathbf{u}}\right)_I, \\
&\equiv \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{T}}\Delta\mathbf{u}_g,
\end{aligned}
\tag{5.69}
$$

where $\omega_{\mathrm{t}} = \mu\omega_{\mathrm{n}}\mathrm{sgn}(g_{\mathrm{t}})$ is used for the slip condition in Eq. (5.69) and $\mathrm{sym}(\cdot)$ is the symmetric part of the matrix. Note that the matrix $\mathbf{K}_{\mathrm{T}}$ in the slip condition is not symmetric. Thus, the linearization of the contact variational form is obtained as

$$
\begin{aligned}
b^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) &= b_{\mathrm{N}}^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) + b_{\mathrm{T}}^*(\mathbf{u};\Delta\mathbf{u},\overline{\mathbf{u}}) \\
&\approx \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{N}}\Delta\mathbf{u}_g + \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{T}}\Delta\mathbf{u}_g \\
&= \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{C}}\Delta\mathbf{u}_g,
\end{aligned}
\tag{5.70}
$$

where $\mathbf{K}_{\mathrm{C}}$ is the contact tangent stiffness matrix in the global coordinate. After combining with the structural matrix equation and transforming to the physical coordinate, the incremental variational equation and corresponding matrix equation are obtained as

$$\overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}\Delta\mathbf{u}_g + \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{K}_{\mathrm{C}}\Delta\mathbf{u}_g = \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{F}_{\mathrm{res}} + \overline{\mathbf{u}}_g^{\mathrm{T}}\mathbf{F}_{\mathrm{C}}. \tag{5.71}$$

The incremental discrete variational equation (5.71) must satisfy for all $\overline{\mathbf{u}}_g$ that satisfy the homogeneous essential boundary conditions. One of the common methods in imposing this condition is to delete those rows that correspond to the essential boundary from the above matrix equation. After performing this removal process, we can obtain the reduced form of the incremental matrix equation:

$$(\mathbf{K} + \mathbf{K}_{\mathrm{C}})\Delta\mathbf{u}_g = \mathbf{F}_{\mathrm{res}} + \mathbf{F}_{\mathrm{C}}. \tag{5.72}$$

As can be seen in the above equation, the contribution from the contact constraints is separated from the contribution from the structural problem. Thus, the contact constraints can be implemented independently of the structural constitutive model. For both the elastic problem in Chap. 3 and elastoplastic problem in Chap. 4, the contact stiffness matrix $\mathbf{K}_{\mathrm{C}}$ and the contact residual force $\mathbf{F}_{\mathrm{C}}$ need to be added to the system matrix equation.

5.4.3 MATLAB Code for Contact Analysis

The MATLAB program **cntelm2d** calculates the contact force and contact tangent stiffness for a contact pair, whose current coordinates are defined in the **ELXY** array. The format of the **ELXY** array is

$$\mathrm{ELXY} = \begin{bmatrix} x_s & x_1 & x_2 \\ y_s & y_1 & y_2 \end{bmatrix}.$$

ELXYP is the same as **ELXY**, except that the array stores the coordinates of the contact pair at the previous time increment. **OMEGAN, OMEGAT**, and **CFRI** are, respectively, the two penalty parameters and the coefficient of friction. If **LTAN** is not zero, then **cntelm2d** calculates the contact tangent stiffness matrix **STIFF**. The contact force, **FORCE**, will always be calculated.

The program first checks if contact occurs in the given contact pair. This check is performed in two ways: (1) the gap must be negative and (2) the natural coordinate at the contact point must be between zero and one. Also, the program checks if the contact interface has friction, based on the value of the coefficient of friction. Once these two conditions are satisfied, then the contact force vector, whose dimension is (6×1), is calculated. If **LTAN** is not zero, then the contact tangent stiffness matrix, whose dimension is (6×6), is also returned.

PROGRAM cntelm2d

```
function [FORCE, STIFF] = cntelm2d(OMEGAN, OMEGAT, CFRI, ELXY, ELXYP, LTAN)
%*********************************************************************
% SEARCH CONTACT POINT AND RETURN STIFFNESS AND RESIDUAL FORCE
% IF CONTACTED FOR NORMAL CONTACT
%*********************************************************************
%
  ZERO = 0.D0; ONE = 1.D0; EPS = 1.E-6; P05 = 0.05; FORCE=[]; STIFF=[];
  XT = ELXY(:,3)-ELXY(:,2); XLEN = norm(XT);
  if XLEN < EPS, return; end
  XTP = ELXYP(:,3)-ELXYP(:,2); XLENP = norm(XTP);
%
% UNIT NORMAL AND TANGENTIAL VECTOR
  XT = XT/XLEN;
  XTP = XTP/XLENP;
  XN = [-XT(2); XT(1)];
%
% NORMAL GAP FUNCTION Gn = (X_s - X_1).N
  GAPN = (ELXY(:,1)-ELXY(:,2))'*XN;
%
% CHECK IMPENETRATION CONDITION
  if (GAPN >= ZERO) || (GAPN <= -XLEN), return; end
%
% NATURAL COORDINATE AT CONTACT POINT
  ALPHA = (ELXY(:,1) - ELXY(:,2))'*XT/XLEN;
  ALPHA0 = ((ELXYP(:,1)-ELXYP(:,2))'*XTP)/XLENP;
%
% OUT OF SEGMENT
  if (ALPHA > ONE+P05) || (ALPHA < -P05), return; end
%
% CONTACT OCCURS IN THIS SEGMENT
  XLAMBN = -OMEGAN*GAPN;
  XLAMBT = 0;
  LFRIC = 1; if CFRI == 0, LFRIC = 0; end
 if LFRIC
   GAPT = (ALPHA - ALPHA0)*XLENP;
   XLAMBT = -OMEGAT*GAPT;
   FRTOL  = XLAMBN*CFRI;
   LSLIDE = 0;
   if abs(XLAMBT) > FRTOL
    LSLIDE = 1;
    XLAMBT = -FRTOL*SIGN(ONE,GAPT);
   end
 end
%
% DEFINE VECTORS
  NN = [XN; -(ONE-ALPHA)*XN; -ALPHA*XN];
  TT = [XT; -(ONE-ALPHA)*XT; -ALPHA*XT];
  PP = [ZERO; ZERO; -XN; XN];
  QQ = [ZERO; ZERO; -XT; XT];
```

```
  CN = NN - GAPN*QQ/XLEN;
  CT = TT + GAPN*PP/XLEN;
%
% CONTACT FORCE
  FORCE = XLAMBN*CN + XLAMBT*CT;
%
% FORM STIFFNESS
  if LTAN
   STIFF = OMEGAN*(CN*CN');
   if LFRIC
    TMP1 = -CFRI*OMEGAN*SIGN(ONE,GAPT);
    TMP2 = -XLAMBT/XLEN;
    if LSLIDE
     STIFF = STIFF + TMP1*(CT*CN') + TMP2*(CN*PP'+PP*CN'-CT*QQ'-QQ*CT');
    else
     STIFF = STIFF + OMEGAT*(CT*CT') + TMP2*(CN*PP'+PP*CN'-CT*QQ'-QQ*CT');
    end
   end
  end
end
```

5.5 Three-Dimensional Contact Analysis

The two-dimensional contact formulation in Sect. 5.3 and its finite element discretization in Sect. 5.4 can be extended to three dimensions. However, three-dimensional contact formulations are quite complicated without providing much insight in physical understandings. In this section, a finite element formulation of three-dimensional contact is introduced without considering continuum variational formulation. In order to simplify the presentation, it is assumed that the master body is a rigid body and the master surface is discretized by four-node quadrilateral elements. Only frictionless contact between a slave node and a master element is considered. An extended formulation of three-dimensional contact formulations can be bound in the work by Laursen and Simo [12] or Kim et al. [13].

Figure 5.14 shows the contact situation between a flexible slave body and a rigid master body. In the discretized domain, the contact condition between a slave node and a master element is considered. The reference coordinates in finite elements can be used as natural coordinates in contact formulation. Therefore, the master element is be represented by the two parameters ξ_1 and ξ_2 such that a point on the element can be expressed as $\mathbf{x}_c(\xi_1, \xi_2)$.

Two tangential vectors in the parametric direction on the master element are defined as

$$\mathbf{t}_\alpha = \mathbf{x}_{c,\alpha}, \quad \alpha = 1, 2, \tag{5.73}$$

where the subscribed comma denotes a partial derivative with respect to the parametric coordinate, i.e., $\mathbf{x}_{c,\alpha} = \partial \mathbf{x}/\partial \xi_\alpha$, $\alpha = 1, 2$. In this section, Greek letters

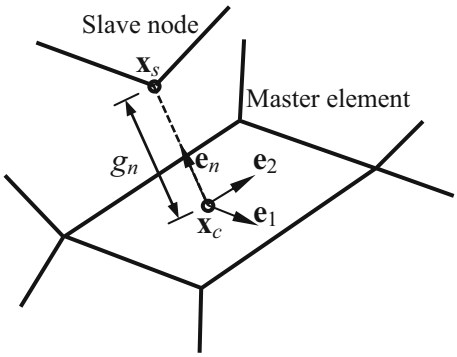

Fig. 5.14 Contact kinematics and design velocities of two bodies

are used for the index in the direction of the parametric coordinates. Note that $\mathbf{t}_1$ and $\mathbf{t}_2$ are not necessarily orthogonal to each other, but are parallel to the contact surface. In the quadrilateral master element, the two tangent vectors can be calculated by differentiating shape functions

$$\mathbf{t}_\alpha = \sum_{I=1}^{4} \frac{\partial N_I(\xi_\alpha)}{\partial \xi_\alpha} \mathbf{x}_I, \quad \alpha = 1, 2.$$

The unit outward normal vector on the master surface can be obtained using Eq. (5.73) as

$$\mathbf{e}_\mathrm{n} = \frac{\mathbf{t}_1 \times \mathbf{t}_2}{\|\mathbf{t}_1 \times \mathbf{t}_2\|}. \tag{5.74}$$

One of the most important steps in contact analysis process is locating the contact point in an accurate and efficient way. The contact point on the master element corresponding to the slave point can be found from the following consistency condition:

$$\varphi_\alpha = (\mathbf{x}_s - \mathbf{x}_c(\xi_1, \xi_2))^\mathrm{T} \mathbf{t}_\alpha(\xi_1, \xi_2) = 0, \quad \alpha = 1, 2, \tag{5.75}$$

which provides the closest projection point $\mathbf{x}_\mathrm{c}$ of $\mathbf{x}_s$, and the corresponding parametric coordinates at the contact point are denoted by $(\xi_1^\mathrm{c}, \xi_2^\mathrm{c})$. For a general master surface, no explicit form of the solution to Eq. (5.75) is available. Finding contact point $\mathbf{x}_\mathrm{c}$ efficiently is very important for a large deformation problem. A local Newton–Raphson method can be used to solve nonlinear Eq. (5.75) with a close initial estimate. The contact consistency Eq. (5.75) is two equations with two

unknowns. Using the first-order Taylor series expansion, the following equation for Newton–Raphson iteration can be obtained:

$$\begin{bmatrix} \dfrac{\partial \varphi_1}{\partial \xi_1} & \dfrac{\partial \varphi_1}{\partial \xi_2} \\ \dfrac{\partial \varphi_2}{\partial \xi_1} & \dfrac{\partial \varphi_2}{\partial \xi_2} \end{bmatrix} \begin{Bmatrix} \Delta\xi_1 \\ \Delta\xi_2 \end{Bmatrix} = \begin{Bmatrix} -\varphi_1 \\ -\varphi_2 \end{Bmatrix}.$$

Once the contact point is found, it is necessary to check if the contact point is within the master element or not. If $\xi_\alpha < -1$ or $\xi_\alpha > 1$, then contact does not occur in this element and no further calculation is required. If $-1 \le \xi_\alpha \le 1$, the gap function is defined by the distance between the slave node and the contact point on the master element as

$$g_{\mathrm{n}} = (\mathbf{x}_s - \mathbf{x}_{\mathrm{c}})^{\mathrm{T}} \mathbf{e}_{\mathrm{n}} \ge 0, \tag{5.76}$$

where the inequality constraint represents the impenetrability condition: the slave point cannot penetrate the master surface. If the gap at the contact point is greater than zero, contact does not occur and no further calculation is required. The violated region of constraint Eq. (5.76) is penalized by applying the contact force, which acts in the direction of the normal vector and proportional to penetration, as

$$\mathbf{f}_{\mathrm{n}}^{\mathrm{c}} = -\omega_{\mathrm{n}} g_{\mathrm{n}} \mathbf{e}_{\mathrm{n}}.$$

In a Newton–Raphson iteration during nonlinear analysis, the tangent stiffness of the above contact force is required. Since $\mathbf{e}_{\mathrm{n}}$ is fixed, no linearization is required. Therefore, only the gap needs to be linearized, which is similar to the variation of the gap in Eq. (5.35) by replacing the displacement variation with the displacement increment. Therefore, the contact stiffness can be obtained by

$$\mathbf{k}_{\mathrm{n}}^{\mathrm{c}} = \omega_{\mathrm{n}} \mathbf{e}_{\mathrm{n}} \mathbf{e}_{\mathrm{n}}^{\mathrm{T}}.$$

The assembly process is identical to two-dimensional contact.

The MATLAB program **cntelm3d** calculates the contact force and contact tangent stiffness for a contact pair, whose current coordinates are defined in the **ELXY** array. Since the master segment is discretized by four-node quadrilateral elements, the format of the **ELXY** array is

$$\mathrm{ELXY} = \begin{bmatrix} x_s & x_1 & x_2 & x_3 & x_4 \\ y_s & y_1 & y_2 & y_3 & y_4 \\ z_s & z_1 & z_2 & z_3 & z_4 \end{bmatrix}.$$

OMEGAN and **LTAN** are the same with **cntelm2d**.

The master element is parameterized by two natural coordinates, **XI** and **ETA**, and the contact point XC is found by determining these two natural coordinates at the contact point. Different from a two-dimensional contact problem, these natural coordinates cannot be determined explicitly; that is, the contact consistency condition must be solved iteratively using the Newton–Raphson-type method. Since the convergence of the Newton–Raphson method strongly depends on the initial estimate, cntelm3d projects the slave node to the master element and approximately estimates the natural coordinates by projecting the projected slave node to the two tangent vectors.

```
% INITIAL CONTACT POINT ESTIMATE.
  [T1, T2, XS]=CUTL(0,0,ELXY);
  XN=cross(T1, T2); XN=XN/norm(XN);
  XI=(XS'*T1)/(2*norm(T1)^2);
  ETA=(XS'*T2)/(2*norm(T2)^2);
  GN=XN'*XS;
```

If the estimated **XI** and **ETA** are out of their ranges with a safety margin [–2, 2], then it is clear that the current contact pair is not in contact and no further calculation is performed. Also, if the estimated gap is positive with a safety margin, it is also concluded that no contact occurs.

Once the initial estimate is within the thresholds, then Newton–Raphson iteration is performed to find the accurate contact point. Since the master element is linear quadrilateral, the iteration should converge within two or three iterations. Once the accurate contact point is determined and if the gap is negative, then the contact force is calculated proportional to the amount of penetration. Also, if **LTAN** is not zero, then the contact tangent stiffness matrix, whose dimension is (6×6), is also returned.

PROGRAM cntelm3d

```
function [FORCE, STIFF] = cntelm3d(OMEGAN, ELXY, LTAN)
%*********************************************************************
% CALCULATE CONTACT FORCE AND STIFFNESS FOR NORMAL CONTACT FOR 3D
%*********************************************************************
%
 EPS=1.E-6; TL1=2; TL2=0.1; TL3=1.01; FORCE=[]; STIFF=[];
%
% INITIAL CONTACT POINT ESTIMATE.
  [T1, T2, XS] = CUTL(0,0,ELXY);
  XN = cross(T1, T2); XN=XN/norm(XN);
  XI = (XS'*T1)/(2*norm(T1)^2);
  ETA = (XS'*T2)/(2*norm(T2)^2);
  GN = XN'*XS;
  XX=(ELXY(:,2)-ELXY(:,3)+ELXY(:,4)-ELXY(:,5))/4;
%
% INITIAL SCREENING OF OUT OF BOUNDS
  if((XI<-TL1)||(XI>TL1)||(ETA<-TL1)||(ETA>TL1)||GN>TL2), return; end
%
```

```
% FIND EXACT CONTACT POINT THROUGH NEWTON-RAPHSON METHOD
  for ICOUNT=1:20
   [T1, T2, XS] = CUTL(ETA,XI,ELXY);
   A=[-T1'*T1, XS'*XX-T2'*T1; XS'*XX-T2'*T1, -T2'*T2];
   B=[-XS'*T1; -XS'*T2];
   DXI=A\B;
   XI=XI+DXI(1); ETA=ETA+DXI(2);
   if(norm(DXI)<EPS), break; end
  end
%
% CHECK THE RANGE OF NATURAL COORD.
  if((XI<-TL3)||(XI>TL3)||(ETA<-TL3)||(ETA>TL3)), return; end
%
% NORMAL GAP FUNCTION AND CONTACT FORCE
  XN = cross(T1, T2); XN=XN/norm(XN);
  GN = XN'*XS;
  if GN>0, return; end
  FORCE = -OMEGAN*GN*XN;
%
% FORM STIFFNESS (NONFRICTION)
  if LTAN, STIFF = OMEGAN*(XN*XN'); end
end
function [T1, T2, XS] = CUTL(ETA,XI,ELXY)
%*********************************************************************
% COMPUTE COORD. OF CENTEROID AND TWO TANGENT VECTORS
%*********************************************************************
 XNODE=[0 -1 1 1 -1; 0 -1 -1 1 1];
 T1 = zeros(3,1); T2 = zeros(3,1); XC = zeros(3,1); XS = zeros(3,1);
 for J = 1:3
  T1(J) = sum(XNODE(1,2:5).*(1+ETA*XNODE(2,2:5)).*ELXY(J,2:5)./4);
  T2(J) = sum(XNODE(2,2:5).*(1+XI *XNODE(1,2:5)).*ELXY(J,2:5)./4);
  XC(J) = sum((1+XI*XNODE(1,2:5)).*(1+ETA*XNODE(2,2:5)).*ELXY(J,2:5)./4);
  XS(J) = ELXY(J,1) - XC(J);
 end
end
```

5.6 Contact Analysis Procedure and Modeling Issues

The contact formulations in the previous sections are relatively straightforward compared to nonlinear constitutive models in the previous chapters. However, in practice, users often experience difficulty of solving nonlinear problems due to contact. The lack of convergence and significant amount of calculation error can be caused by poorly modeled contact conditions. Therefore, it is important to understand the modeling characteristics of contact problems, which is the objective of this section.

5.6.1 Contact Analysis Procedure

In general, contact analysis requires three steps: (1) defining contact pairs and types, (2) searching for the contact point, and (3) calculating contact force and tangent stiffness.

5.6.1.1 Definition of Contact Pairs and Types

Since the user does not know the location of the contact boundary, it is necessary to define contact pairs that are already in contact or have a possibility of contact. This is especially important for a large deformation problem where the structural boundary can change its shape significantly during the analysis. Many commercial programs provide a tool to generate all contact pairs automatically or with minimum user actions. In addition to contact pairs, it is also necessary to define the properties of contact interface, including (a) weld contact, (b) rough contact, (3) stick contact, and (4) slip contact.

In weld contact, the slave node is bonded to the master element and there is no relative motion in the interface. There will be no contact search, and all contact pairs are assumed already in contact, which makes this formulation fastest. Conceptually, this is equivalent to the rigid link element or multipoint constraint; the only difference is that the force in the interface is decomposed into normal and tangential components. However, the interface is still under infinitesimal elastic deformation, as it is a part of an elastic body.

Rough contact is similar to weld contact, except that the contact interface may not be initially in contact or the initial contact point can be separated. But once it is in contact, it behaves similar to a weld contact. Therefore, its behavior is similar to the case where the contact interface is rough such that there is no relative motion in the interface, independent of the magnitude of normal contact force.

Stick contact is similar to rough contact in the sense that the contact interface can be closed or separated. The difference from rough contact is that the interface can have a relative motion similar to elastic deformation. When a tangential force acts on the contact interface, rough contact behaves like a rigid link, while stick contact shows a small elastic deformation in the interface. The user needs to specify the tangential stiffness. It is noted that the tangential stiffness matrix is symmetric.

Slip contact is the most general contact formation, where the contact point can be closed or separated. In addition, the contact interface can have a relative motion governed by the Coulomb friction model. In practice, the stick condition is first applied before the slip condition is used in many contact algorithms. Different from stick contact, the tangent stiffness of slip contact is unsymmetric.

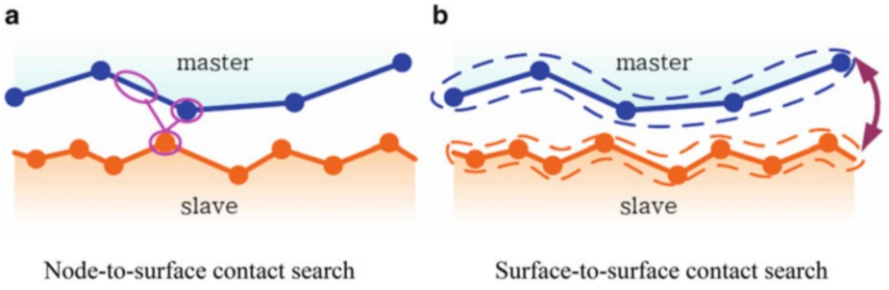

Fig. 5.15 Contact search methods. (**a**) Node-to-surface contact search. (**b**) Surface-to-surface contact search

5.6.1.2 Contact Search

The easiest way of performing a contact search is for the user to specify the master element with which that a slave node will contact. This is only possible when deformation is small and no relative motion exists in the contact interface. Slave and master nodes are often located at the same position and connected by a compression-only spring (node-to-node contact). This type of contact pairing works for very limited cases where the user makes slave elements and master elements such that both elements coincide at the contact interface. Also, the contact surface must be simple enough so that the user knows the exact contact region in advance.

In general cases, however, the user does not know about the contact pairs that are actually in contact. Instead, the user specifies all possible candidates. During the contact analysis, the program searches for all contact pairs and determines those pairs that are actually in contact, that is, the violated pairs of the impenetrability condition. Since contact pairs include all possible pairs, the number of pairs is significantly large. For example, if 1,000 slave nodes have a possibility of contacting 1,000 master elements, then theoretically it is necessary to check one million contact pairs. Considering this is required during a single iteration of nonlinear analysis, the program will repeat this search numerous times in order to finish iterative nonlinear analysis. Therefore, it is important to effectively search for contact pairs. Sometimes, it is useful to store the currently contacting master element information for a given slave node, such that in the following iteration, the contact search is performed for only neighboring master elements of the previous one.

In general, contact search is categorized by a node-to-surface and surface-to-surface search (Fig. 5.15). The former is to search if a slave node penetrates the master surface, which is often used when the master surface is rigid. The latter is to search for an impenetrability condition between a slave surface and a master surface. This is useful when the two flexible bodies are under a large slip such that the distinction of slave and master is unclear. Although the latter represents the impenetrability condition more accurately, it takes more computational time due to bidirectional contact (Fig. 5.16).

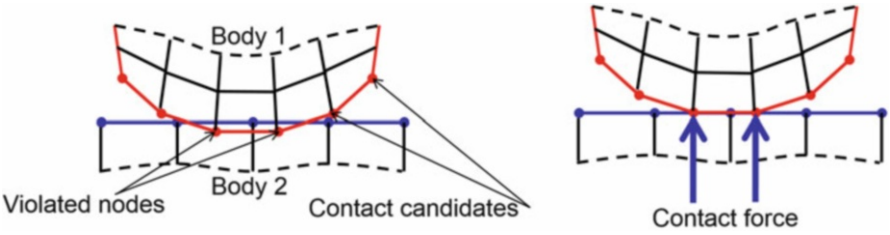

Fig. 5.16 Contact search and contact force

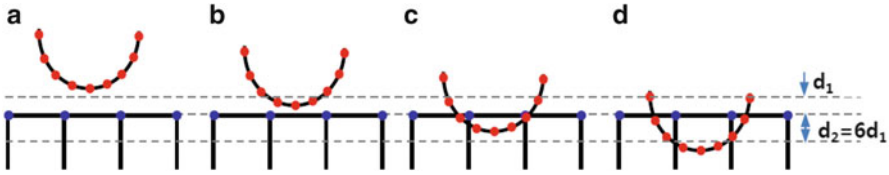

Fig. 5.17 Contact tolerance and detecting contact

Contact tolerance: Since searching for all possible contact pairs is very expensive, commercial programs often use the concept of contact tolerance, which is the minimum distance to search for contact. The default value can be 1 % of the contact element length. This can be used for detecting bodies about to make contact as well as excluding bodies that are on opposite sides. The contact tolerance can be used for compensating for geometric tolerance in the case of weld contact. If two contact surfaces are within the contact tolerance, they are considered in contact and the contact force is calculated. In the case of rough and general contact, contact pair is established when two surfaces are within the contact tolerance.

For example, the two separate bodies in Fig. 5.17 are going to be in contact. Contact tolerance is set in two ways: d_1 for separation and d_2 for penetration. In the cases of (b) and (c), the initial separation or penetration is within the tolerance, the contact pair is established, and the convergence analysis is performed by generating appropriate contact force. In the case of (c), since penetration is relatively large, the load increment is bisected to reduce the amount of penetration. However, in the cases of (a) and (d), the initial separation or penetration is larger than the contact tolerance, the search algorithm fails to detect contact, and, as a result, the two surfaces will be penetrated without contact.

Therefore, an appropriate load increment, as in Fig. 5.18a, should be used in order to make initial contact detection. If the load increment is too large, as in Fig. 5.18c, then the contact search algorithm fails to detect contact because the movement is larger than the contact tolerance. In this case, the contact condition is not established, and, as a result, a too-large penetration occurs in the next load increment, which may completely miss contact detection. In the case of Fig. 5.18b, the contact surface is within the contact tolerance; even if a large penetration may occur, the contact pair will be generated, and the impenetrability condition will be satisfied through bisecting the load increment.

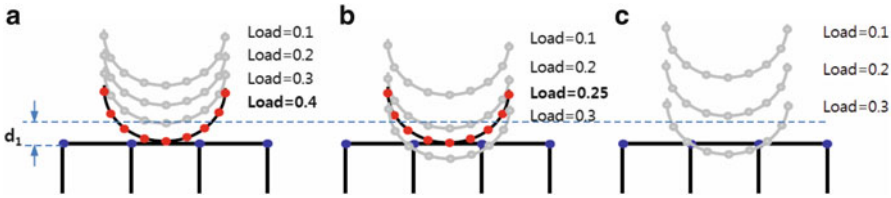

Fig. 5.18 The effect of load increment in contact detection

5.6.1.3 Contact Force and Tangent Stiffness

Once the contact pairs are actually in contact (or violated the impenetrability condition), either the penalty method or the Lagrange multiplier method can be applied to satisfy contact constraint. The penalty method is simple and intuitive but allows a small amount of constraint violation. That is, the impenetrability condition will be slightly violated. The amount of violation can be controlled by the penalty parameter. A large penalty parameter allows only a small amount of violation, but a too-large penalty parameter can cause numerical instability because it makes the stiffness matrix ill-conditioned.

Contact stiffness: In practice, the penalty parameter is better selected based on material stiffness, element size, and element height normal to the contact interface. Therefore, it is often called the contact stiffness. If two contacting bodies have different material stiffness, it is calculated based on the softer material. A large value of contact stiffness can reduce penetration, but can also cause a problem in convergence. Therefore, a proper value of contact stiffness must be determined based on allowable penetration, which requires experience. Normally many programs suggest the contact stiffness based on the elastic modulus of contacting bodies and allow users to change it by multiplying a scale factor with a default of one. The user can start with a small initial scale factor and gradually increase it until a reasonable penetration can be achieved.

Tangential stiffness: If the contact stiffness is for the normal contact, tangential stiffness is for the frictional force in the contact interface. Since the frictional force is generated through normal contact force, it depends on the contact stiffness, and its behavior is more complicated because of friction. In the penalty formulation, an elastic stick condition applies before slip occurs under a tangential load. If the tangential load is removed, then the body returns to its original state. The tangential stiffness controls this stick condition. If the tangential stiffness is too large, then the contact interface shows slip without stick. If too small, then the stick condition will be overextended.

Contact force: When two bodies are in contact, the contact force in the interface can be considered as either an internal or external force, depending on how the system is defined. If a free-body diagram is constructed of each body separately,

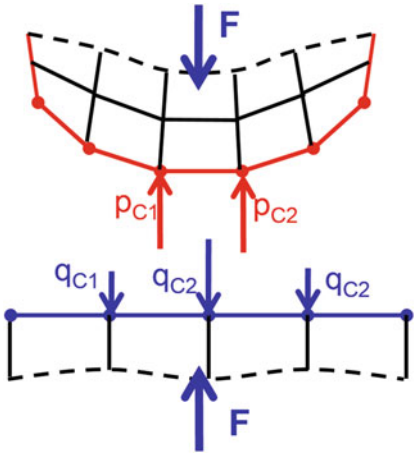

Fig. 5.19 Contact force on slave nodes and master elements

then the contact force is the externally applied force on the boundary. From this viewpoint, the contact problem is called boundary nonlinearity, because both the boundary and force are unknown. However, if the free-body diagram includes both contacting bodies, then the contact force can be viewed as an internal force. If the entire system is in equilibrium, then all internal forces must vanish. Therefore, the contact force on the slave nodes must be equal and opposite in direction to the contact force on the master elements. This can also be viewed from Newton's third law: equal and opposite forces act on interface. Figure 5.19 shows two contacting bodies in equilibrium. Because the individual bodies as well as both bodies together are in equilibrium, the following relation should be satisfied:

$$\mathbf{F} = \sum_{i=1}^{N_p} p_{ci} = \sum_{i=1}^{N_q} q_{ci}. \tag{5.77}$$

It is noted that in Eq. (5.77), individual p_{ci} and q_{ci} are different in magnitudes because of discretization. The force distribution can be different. However, the resultants should be the same, as the two bodies are in equilibrium.

5.6.2 Contact Modeling Issues

In this section, several modeling issues in contact analysis are summarized. The contents that are covered in this section are by no means complete. However, users should be familiar to these issues in order to solve convergence problems as well as accuracy of analysis results.

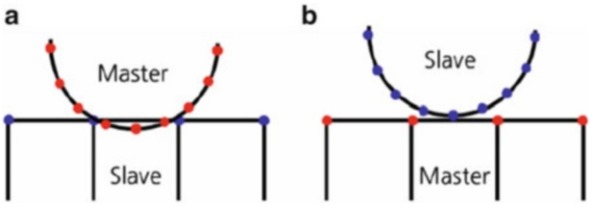

Fig. 5.20 Definition of slave and master

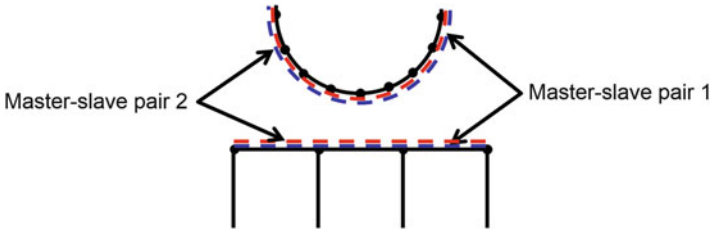

Fig. 5.21 Alternating definition of slave–master pairs in order to prevent penetration from either body

5.6.2.1 Definition of Slave and Master

When two bodies are in contact, the slave–master concept distinguishes body 1 from body 2. Although there is no theoretical reason to distinguish body 1 from body 2, the distinction is often made for numerical convenience. One body is called a slave body, while the other is called a master body. Then, the contact condition is imposed such that the slave body cannot penetrate into the master body. This means that hypothetically the master body can penetrate into the slave body, which is not physically possible but numerically possible because it is not checked. There is not much difference in a fine mesh, but the results can be quite different in a coarse mesh, as shown in Fig. 5.20. When a curved boundary with a fine mesh is selected as a master body, a straight slave boundary with a coarse mesh shows a significant amount of penetration, even if none of slave nodes penetrate into the master body. Therefore, it is important to select the slave and master body in order to minimize this type of numerical error. In general, in order to minimize penetration, a flat and stiff body is selected as a master body, while a concave and soft body is selected as a slave body. Also, it is suggested that a body with a fine mesh be a slave and a body with a coarse mesh be a master. In the case of flexible–rigid body contact, the rigid body is selected as a master body and the flexible one as a slave body.

No matter how the slave and master are selected, it is possible that a master node can penetrate into the slave element. In order to prevent penetration from either body, it is necessary to define the slave–master pair twice by changing their role, as shown in Fig. 5.21. Some surface-to-surface algorithms use this technology to prevent penetration from either body.

5.6.2.2 Flexible Contact vs. Flexible–Rigid Contact

Since all bodies are flexible in the viewpoint of mechanics, it seems natural to model all contacting bodies as flexible and apply contact conditions between flexible bodies. However, since modeling is an abstraction of physical phenomena, it is possible to consider one body as a rigid body, even if in reality it is flexible. Therefore, in such a case, a flexible–rigid body contact condition can be applied. The question is why we want to use flexible–rigid body contact and when we can apply that condition.

The flexible–flexible contact can be applied when two bodies have a similar stiffness and both can deform. For example, metal-on-metal contact can be modeled as flexible–flexible contact. However, when the stiffness of two bodies are significantly different, such as contact between rubber and metal, the behavior of metal can be approximated as a rigid body, because the deformation of metal can be negligible compared to that of rubber. However, this can also depend on physical behavior of the system. For example, if a rubber ball impacts on a thin metal plate, then the plate needs to be modeled as a flexible body because the deformation of the plate can be large.

There are obvious advantages of using flexible–rigid body contact over two flexible–body contact. When two bodies have a large difference in stiffness, the stiffness matrix becomes ill-conditioned and the matrix equation loses many significant digits. Therefore, accurate calculation becomes difficult. In addition, as shown in previous section, the numerical implementation of flexible–rigid body contact formulation is much easier than multi-body contact formulation.

5.6.2.3 Sensitivity of Mesh Discretization

At the continuum level, it is assumed that the contact boundary varies smoothly and the boundary is differentiated two or three times in deriving contact force and tangent stiffness. In the numerical model, however, the contact boundary is approximated by piecewise continuous curves (or straight lines), and only C^0 continuity is guaranteed across the element boundary. Therefore, the slope of the contact boundary is not continuous. Unfortunately, the contact force is very sensitive to the boundary discretization and strongly depends on this slope: contact force acts in the normal direction of the contact boundary. Therefore, if the actual contact point is near the boundary of two elements with a large slope change, it is possible that the Newton–Raphson iteration may have difficulty in convergence.

Another important aspect related to mesh is the distribution of contact stress/contact pressure. As shown in Fig. 5.22, if a uniform pressure is applied on top of a slave body, it is natural to think that the contact pressure on the bottom surface will also be uniform. However, due to the effect of a large master surface at the bottom, the contact pressure is high on the edge of the contacting region. Therefore, the

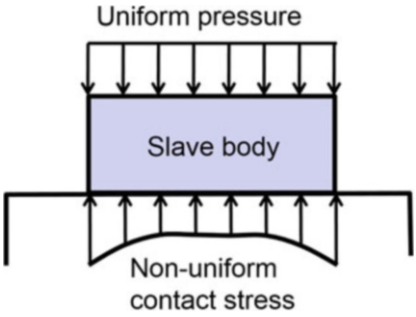

Fig. 5.22 Contact stress distribution under uniform pressure load

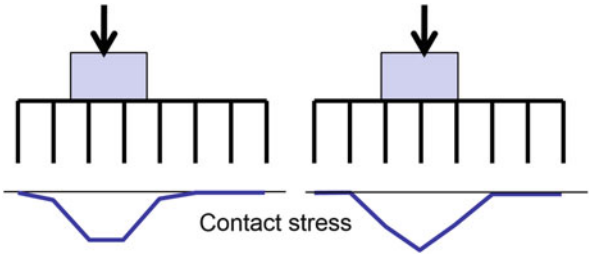

Fig. 5.23 Variation of contact stress distribution as a function of block location

contact stress/pressure is not uniform. Theoretically, the contact stress on the edge can be twice the inside contact stress.

Another important observation on contact stress distribution is that it is sensitive to mesh discretization. As shown in Fig. 5.23, the contact stress distribution is different for different locations of the block. Therefore, it is dangerous to determine the maximum contact stress using a single coarse mesh. It is always recommended to perform mesh sensitivity study to show convergence of contract stress.

5.6.2.4 Rigid-Body Motion

Rigid-body motion in contact is one of the most commonly confused concepts to users. This is also a good example of contact boundary conditions that are different from the displacement boundary conditions. Figure 5.24 shows a cylindrical slave body between two rigid masters. It is assumed that the slave body slightly penetrates into the lower master body, while it has a slight gap with the upper master body. Since the contact force is generated proportional to penetration, the upward contact force will be applied at the lower part of the cylinder, which will move the cylinder upward, as in Fig. 5.24a. Next, the body now penetrates the upper master body because of the previous upward motion. Then, the contact force is now applied from the upper master body and it is not in contact with the lower master

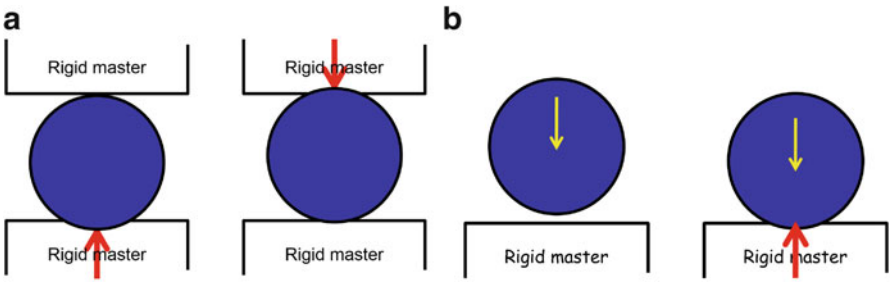

Fig. 5.24 The effect of rigid-body motion in contact

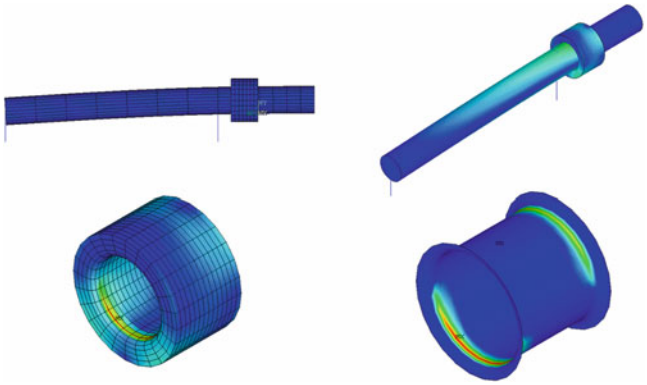

Fig. 5.25 Contact stress at bushing due to shaft bending

body, which will cause a downward contact force. Under this situation, the slave body can either oscillate between the two master surfaces (Fig. 5.24a) or fly out if it is in contact with a single master body (Fig. 5.24b). In fact, without contact, the cylinder is not well constrained. Even if in real physics a body can be stable between two contacting bodies, in numerical analysis, it is better to constrain the flexible body without contact, so that the rigid-body motion can be removed. When a body has rigid-body motion, an initial gap can cause a singular matrix (infinite/very large displacements). The same is true when there is an initial overlap. In order to remove rigid-body motion, it is possible to add a small, artificial bar element so that the body is well constrained while minimally affecting analysis results, as shown in Fig. 5.25, where the shaft is constrained by two bar elements.

5.6.2.5 Convergence Difficulty

Common difficulties in contact analysis are (a) the contact condition does not work, i.e., penetration occurs, and (b) the Newton–Raphson iteration does not converge. The former is related to the contact definition or a too-large load increment.

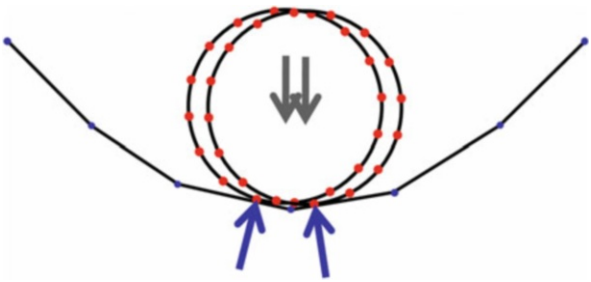

Fig. 5.26 Discontinuity of contact force by nonsmooth contact boundary

Therefore, this type of problem can be solved relatively easily. On the other hand, the lack of convergence is the most common difficulty in nonlinear analysis, and it is not trivial to find the cause because they can be caused by different reasons.

As the convergence of Newton–Raphson method depends on the initial estimate, it is possible that the method can improve the convergence by starting with the initial estimate that is close to the solution. In the increment force method in Chap. 2, the solution is a function of load increment. A small increment means that the solution from a previous increment is close to the solution in the current load increment. Therefore, using a small load increment is the most common remedy when convergence cannot be obtained. Many commercial programs have the capability to automatically control the load increment. When a given load increment does not converge, then the current increment is reduced by half or a quarter and convergence iteration is retried. This bisection process is repeated until the convergence can be achieved or the program stops when the maximum allowed bisections are consumed or the minimal size of load increment is not converged.

If the Newton–Raphson iteration failed to converge with the smallest load increment, the problem resides in fundamental issues. The basic assumption in Newton–Raphson method is that the nonlinear function is smooth with respect to input parameters. In the context of contact analysis, this can be interpreted as the contact force varies smoothly throughout deformation. Unfortunately, this is a strong assumption in finite element analysis because of discretization. As shown in Fig. 5.26, the slope of finite elements is discontinuous across the element boundary, especially when the contact boundary is curved. As illustrated in the figure, this discontinuity can make the contact force oscillate between two master elements and discontinuously change the direction of contact force. In order to minimize such a situation, it is necessary to use more elements to represent the curve boundary. As a rule of thumb, it is recommended to generate about 10 contact elements along the 90° corner fillet or use higher-order elements.

A nonsmooth contact boundary can also affect the accuracy of contact analysis. As an example, Fig. 5.27 shows contact between a shaft and a hole. In Fig. 5.27a, both the shaft and hole are discretized by 15 linear elements along the circumference. When the mesh locations of both parts are different, the inaccuracy of representing circular geometry significantly affects contact results. Some nodes

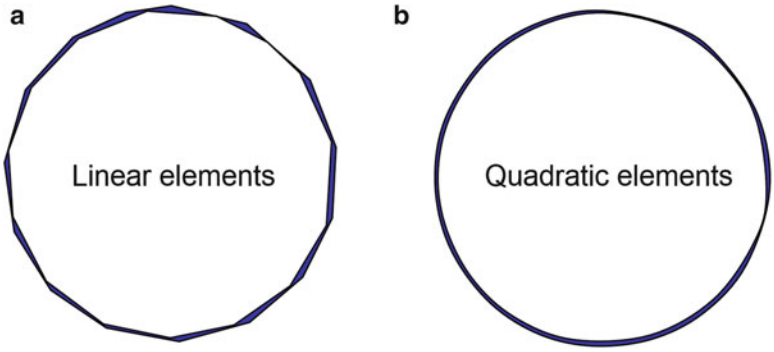

Fig. 5.27 Discretization of circular shaft and hole using (**a**) linear and (**b**) quadratic elements

are out of contact, while others are under excessive contact force due to over-penetration. Therefore, the contact stress contour does not show a smooth variation of contact stress. Rather, a localized random and discrete contact stress distribution may be observed. On the other hand, if higher-order elements are used as in Fig. 5.27b, the two contact boundaries become much more conforming and smooth contact stress distribution can be obtained.

5.7 Exercises

P5.1 For the beam contact problem in Sect. 5.2.1, determine the contact force and tip deflection using the Lagrange multiplier method. Choose the gap g as a Lagrange multiplier.

P5.2 For the beam contact problem in Sect. 5.2.1, determine the contact force and tip deflection using the Lagrange multiplier method. Model the beam using a two-node Euler beam element. Compare the results with the results in Sect. 5.2.1, and explain the reason for different results.

P5.3 For the frictional contact problem in Sect. 5.2.2, determine the frictional force and slip displacement using the Lagrange multiplier method. Choose the slip u_{tip} as a Lagrange multiplier.

P5.4 During a Newton–Raphson iteration, a rectangular plane element is in contact with a rigid surface as shown in the figure. Due to the penalty method, the penetration of $g = -1 \times 10^{-4}$ m is observed with penalty parameter $\omega_{\text{n}} = 10^6$. In the two-dimensional problem, the element has eight degrees of freedom $\{u_{1x}, u_{1y}, u_{2x}, u_{2y}, u_{3x}, u_{3y}, u_{4x}, u_{4y}\}^{\text{T}}$. Calculate the contact force and contact stiffness matrix in terms of 8×1 vector and 8×8 matrix, respectively.

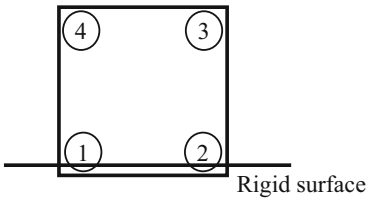

Fig. P5.4 Contact of a rectangular block

P5.5 A sphere of radius $r = 8$ mm is pressed against a rigid flat plane. Using a commercial program, determine the contact radius, a, for a given load $F = (30 \times 2\pi)$ N. Assume a linear elastic material with Young's modulus $E = 1{,}000$ N/mm^2 and Poisson's ratio $\nu = 0.3$. Use an axisymmetric model. Compare the finite element result with the analytical contact radius of $a = 1.010$ mm.

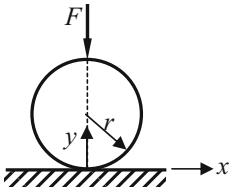

Fig. P5.5 Contact of a sphere

P5.6 A long rubber cylinder with radius $r = 200$ mm is pressed between two rigid plates using a maximum imposed displacement of $\delta_{\max} = 200$ mm. Determine the force–deflection response. Use Mooney-Rivlin material with $A_{10} = 0.293$ MPa and $A_{01} = 0.177$ MPa. Assume a plane strain condition and symmetry. Compare the results with the target results of $F = 250$ N at $\delta = 100$ mm and $F = 1{,}400$ N at $\delta = 200$ mm.

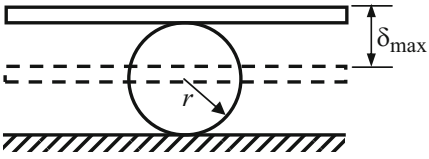

Fig. P5.6 Rubber cylinder contact problem

P5.7 Two long cylinders of radii $R_1 = 10$ mm and $R_2 = 13$ mm, in frictionless contact with their axes parallel to each other, are pressed together with a force per unit length, $F = 3{,}200$ N/mm. Determine the semi-contact length b and the approach distance d. Both materials are linear elastic with $E_1 = 30{,}000$ N/mm^2 and $v_1 = 0.25$ for Cylinder 1 and $E_2 = 29{,}120$ N/mm^2 and $v_2 = 0.3$ for Cylinder 2. Assume a plane stress condition with a unit

thickness and symmetry. Compare the results with the target results of $d = -0.4181$ mm and $b = 1.20$ mm.

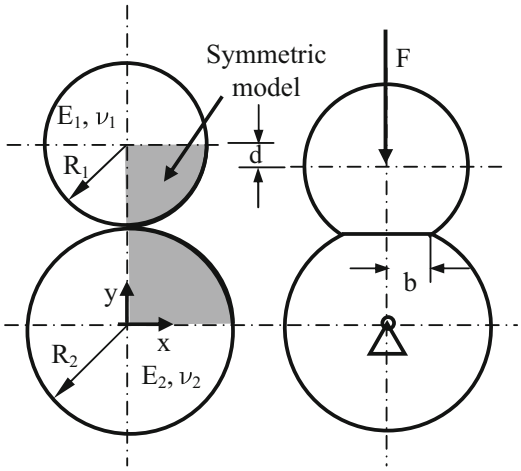

Fig. P5.7 Hertzian contact problem

P5.8 Deep drawing is a manufacturing process that can create a complex shape out of a simply shaped plate (blank). The deep-drawing configuration is shown in the figure, which is composed of a blank, punch, die, and blank holder. The thickness of the initial blank is 0.78 mm. The die is fixed throughout the entire process, while the punch moves down by 30 mm to shape the blank. The holder controls the slip of the blank by applying friction force. The fillet radii of both punch and die are 5 mm. After the maximum downstroke of the punch, both the punch and holder are removed. Then, the blank will experience elastic springback. The objective of this project is to simulate the final geometry of the blank after springback.

Model the process using an axisymmetric problem. You many use CAX4R elements. The whole simulation is divided by three steps. (1) The blank holder is pushed (displacement control) to provide about 100 kN of holding force. (2) While the blank holder is fixed at the location of step (1), the punch is moved down by 30 mm. (3) Punch, die, and blank holder are removed so that the blank is elastically deformed by springback. It is possible to change processes.

The following results need to be submitted: (1) deformed shape plots of five different steps, (2) graph of radial position vs. radial strain, and (3) graph of radial position vs. thickness change, (4) graph of punch displacement vs. punch force, and (5) comparison of deformed shapes at the maximum stroke and after springback.

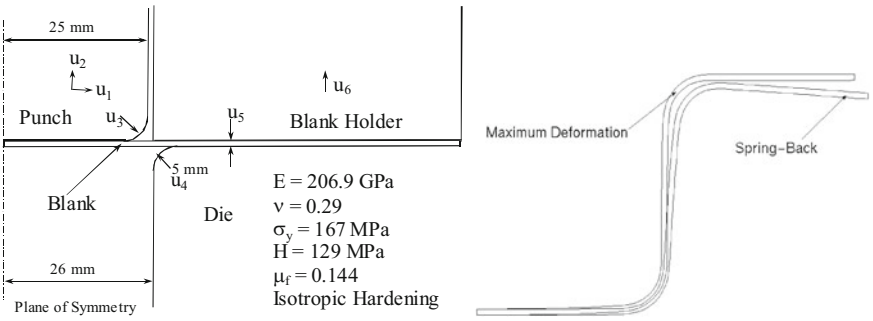

Fig. P5.8 Deep-drawing problem

References

1. Duvaut G, Lions JL. Inequalities in mechanics and physics. Berlin: Spring-Verlag; 1976.
2. Kikuchi N, Oden JT. Contact problems in elasticity: a study of variational inequalities and finite element method. Philadelphia: SIAM; 1988.
3. Ciarlet PG. The finite element method for elliptic problems. New York: North-Holland; 1978.
4. Klarbring A. A mathematical programming approach to three-dimensional contact problems with friction. Comput Methods Appl Mech Eng. 1986;58:175–200.
5. Kwak BM. Complementarity-problem formulation of 3-dimensional frictional contact. J Appl Mech Trans ASME. 1991;58(1):134–40.
6. Luenberger DG. Linear and nonlinear programming. Boston: Addison-Wesley; 1984.
7. Barthold FJ, Bischoff D. Generalization of Newton type methods to contact problems with friction. J Mec Theor Appl. 1988;7 Suppl 1:97–110.
8. Arora JS. Introduction to optimum design. 2nd ed. San Diego: Elsevier; 2004.
9. Wriggers P, Van TV, Stein E. Finite element formulation of large deformation impact-contact problems with friction. Comput Struct. 1990;37:319–31.
10. Kim NH, Park YH, Choi KK. An optimization of a hyperelastic structure with multibody contact using continuum-based shape design sensitivity analysis. Struct Multidiscip Optim. 2001;21(3):196–208.
11. DeSalvo GJ, Swanson JA. ANSYS engineering analysis system, user's manual, vols. I and II. Houston: Swanson Analysis Systems Inc.; 1989.
12. Laursen TA, Simo JC. A continuum-based finite element formulation for the implicit solution of multibody large deformation frictional contact problems. Int J Numer Methods Eng. 1993;36:2451–3485.
13. Kim NH, Yi KY, Choi KK. A material derivative approach in design sensitivity analysis of 3-D contact problems. Int J Solids Struct. 2002;39(8):2087–108.