3. Theory of Contact-Impact

3.1. Introduction

This chapter reports the mechanics of contact and the theory of the finite element method. Both the classical and the computational theories of contact mechanics are reviewed first, followed by the solution methods for FEM. The techniques of FE modelling of contact impact are also presented.

3.2. Brief Review of Mechanics of Contact

3.2.1. Classical theories

Contact is one of the common research topics because of its wide applications in the engineering field. The earliest theory of contact mechanics is due to the pioneering researcher Heinrich Hertz who published a classical paper on contact in 1882 in the German language. Subsequently several researchers improved the Hertz contact theory by relaxing the limitations and extending its application to more practical situations.

(a) Normal contact of elastic solids – Hertzian contact theory

Hertz contact theory (HCT) is established based on some basic assumptions: elastic contact bodies, frictionless contact surfaces, continuous and non-conforming surfaces, small strains and small contact area relative to the potential area of contacting surfaces (Johnson, 1985) .

Fig. 3.1 shows two non-conforming solids (Body 1 and Body 2) which contact at an area that is finite and small compared to their dimensions. Assuming that the profile of each surface is topographically smooth in both micro and macro scales, the profiles of the contacting bodies are expressed in Eq. (3.1) and (3.2).

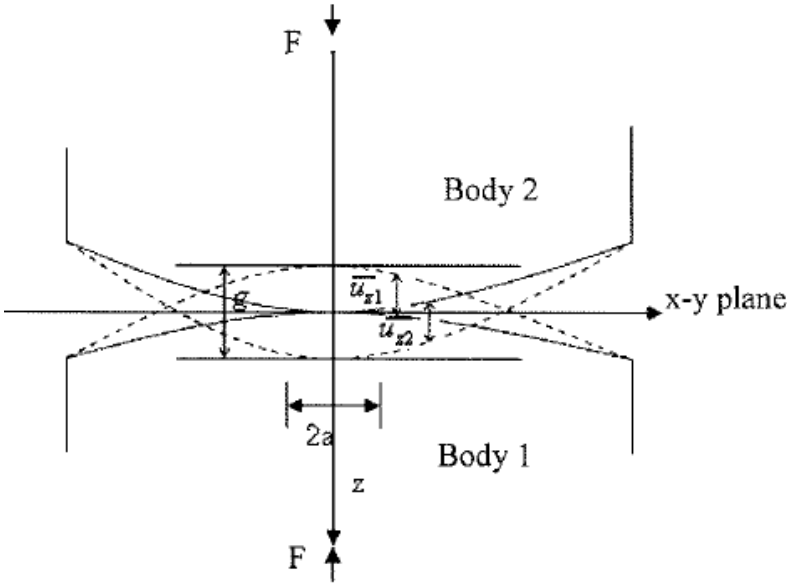

Figure 3.1 Contact between non-conforming solids

$$Z_1 = \frac{1}{2R_1^{'}} x^{2} + \frac{1}{2R_1^{''}} y^{2} \tag{3.1}$$

$$Z_2 = -(\frac{1}{2R_2^{'}} x^{2} + \frac{1}{2R_2^{''}} y^{2}) \tag{3.2}$$

The separation between the two surfaces is then calculated as follows:

$$g = Z_1 - Z_2 = \frac{1}{2R^{'}} x^2 + \frac{1}{2R^{''}} y^2 \tag{3.3}$$

where $\begin{cases} \frac{1}{R_1^{'}} + \frac{1}{R_2^{'}} = \frac{1}{R^{'}}, \\ \frac{1}{R_1^{''}} + \frac{1}{R_2^{''}} = \frac{1}{R^{''}} \end{cases}$.

Defining the $\overline{u_{z1}}$ and $\overline{u_{z2}}$ as the displacements of points on each surface and g as the compression displacement of two bodies, when points are in the contact area, the following expression can be written:

$$\overline{u_{z1}} + \overline{u_{z2}} = g - \frac{1}{2R'}x^2 - \frac{1}{2R''}y^2 \tag{3.4}$$

If Eq. (3.4) is not satisfied (as in Eq. (3.5)), the bodies are said to be separated.

$$\overline{u_{z1}} + \overline{u_{z2}} < g - \frac{1}{2R'}x^2 - \frac{1}{2R''}y^2 \tag{3.5}$$

In Eq. (3.4) and (3.5), $\overline{u_{z1}}$ and $\overline{u_{z2}}$ are obtained implementing the elasticity theory with the contact pressure P that is yet to be determined:

$$\begin{aligned}\overline{u_{z1}} &= \frac{1-\nu^2}{\pi E}\iint\frac{p(x,y)}{R}dxdy,\\ \overline{u_{z2}} &= \frac{1-\nu'^2}{\pi E'}\iint\frac{p(x,y)}{R}dxdy\end{aligned} \tag{3.6}$$

Inserting Eq. (3.6) into Eq. (3.4), an integral equation is obtained employing potential theory. The resulting pressure distribution is then worked out as:

$$p(x,y) = \frac{3F}{2\pi ab}\sqrt{1-\frac{x^2}{a^2}-\frac{y^2}{b^2}} \tag{3.7}$$

where a and b represent the major and minor axes respectively of the elliptical contact zone and can be determined by resolving the following set of integral equations once the curvatures of contact surfaces R' and R'' are determined (Eq. 3.8):

$$\begin{aligned}\frac{1}{2R'} &= \frac{3F}{4\pi}\left(\frac{1-\nu^2}{E}+\frac{1-\nu'^2}{E'}\right)\int_0^\infty\frac{d\zeta}{(a^2+\zeta)\sqrt{(a^2+\zeta)(b^2+\xi)\zeta}},\\ \frac{1}{2R''} &= \frac{3F}{4\pi}\left(\frac{1-\nu^2}{E}+\frac{1-\nu'^2}{E'}\right)\int_0^\infty\frac{d\zeta}{(b^2+\zeta)\sqrt{(a^2+\zeta)(b^2+\zeta)\zeta}}\end{aligned} \tag{3.8}$$

The analytical solution of contact dimensions and pressure distributions between two smooth elastic bodies is obtained through the above process. This problem is strictly nonlinear because the displacement at any point of contact depends on the distribution of contact pressure throughout the whole contact zone. This leads to a significant complexity to solve the integral equations of contact pressure for each step in the dynamic contact condition. As a simplification, the 'Hertz contact spring' is developed. Assuming a simple Winkler elastic foundation rather than elastic half space, the model is illustrated in Fig. 3.2 which shows an elastic foundation resting on a rigid base and contacted with a rigid indenter.

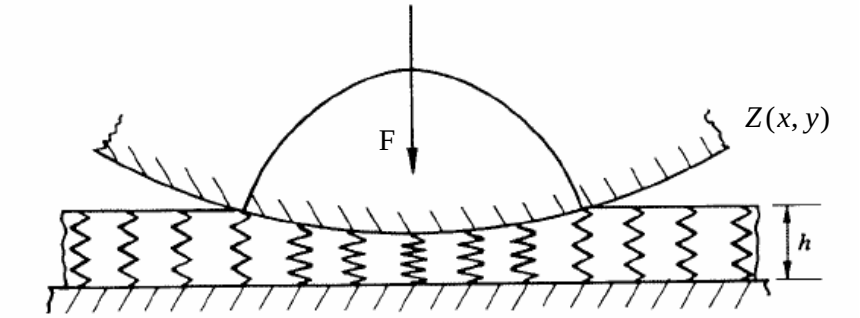

Fig 3.2 Hertz contact foundation model (Johnson, 1985)

Using the profile of the indenter $Z(x,y) = \frac{1}{2R'}x^2 + \frac{1}{2R''}y^2$ and the original compressed displacement g, the displacement profile of the contact surfaces is written as:

$$\overline{u_z}(x,y) = \begin{Bmatrix} g - Z(x,y), g > Z \\ 0, g < Z \end{Bmatrix} \tag{3.9}$$

The contact pressure at any point is assumed to be dependent only on the displacement at that point as in Eq. (3.10).

$$p(x,y) = (K/h)\overline{u_z}(x,y) \tag{3.10}$$

Inserting Eq. (3.9) into Eq. (3.10), the pressure distribution is expressed as:

$$p(x,y) = (K/h)(g - \frac{1}{2R^{'}}x^2 + \frac{1}{2R^{''}}y^2) \quad (3.11)$$

By integrationof the pressure distribution the total contact force is obtained as:

$$F = K\pi g^2\sqrt{R^{'}R^{''}}/h \quad (3.12)$$

where h is the depth of elastic foundation. The relationship of contact force and contact indention is thus generated.

(b) Non- Hertz normal contact of elastic bodies

HCT application to practical problems is limited due to its assumption of strict smooth elastic half space. To solve practical problems, non-Hertz normal contact solutions are, therefore, developed. For the wheel/rail contact at IRJs, the Hertzian assumptions are violated because of edge effect, discontinuous surface profile and interface frictions; Hertz solutions are therefore not strictly applicable for contact problems at IRJs.

(i.) Edge effect

The HCT half space assumption is violated for problems encountering contact at non-continuous profiles such as the edge of bodies. Many researchers have examined the edge effect in recent decades (Dundurs & Lee (1972), Gdoutos & Theocaris (1975), Comninou (1976), Bogy (1971), Khadem & O'Connor (1969)). Unfortunately analytical solutions are not possible, with the problems requiring idealisations or gross simplifications.

A rigid punch with a square corner was considered as a case of non-Hertzian contact theory as the edge of the punch was not continuous. These tilted punch problems were solved by Muskhelishvili (1949). The pressure distribution close to a corner ($s = a - x << a$) can be expressed as:

$$p(s) = \frac{2(1-\nu)}{\pi(3-4\nu)}(2as)^{-1/2}\cos\{[(1/2\pi)\ln(3-4\nu)]In(2a/s)\} \qquad (3.13)$$

where s is the distance from the contact edge corner and a is the contact patch dimension.

Furthermore general edge problems that contain angles at corners other than 90° were considered by Dundurs & Lee (1972) for frictionless contact and by Gdoutos & Theocaris (1975) and Comninou (1976) for frictional situations and by Bogy (1971) for no slip.

(ii.) Discontinuous surface profiles

When there is curvature change within the the contact area, the Hertz continuous surface assumption is violated. The geometries of edge effect problems are idealised as a wedge or cone to formulate analytical solutions. The pressure distribution was given in Johnson's (1985) book as:

$$p(x) = \frac{E^*\cot\alpha}{2\pi}\ln\{\frac{a+(a^2-x^2)^{1/2}}{2}\} = \frac{E^*\cot\alpha}{\pi}\cosh^{-1}(a/x) \qquad (3.14)$$

where $\frac{1}{E^*} = \frac{1-\nu^2}{E} + \frac{1-\nu'^2}{E'}$ and α denotes the semi-angle of the wedge or cone.

Love (1939) used the indentation of a flat surface by a blunt cone and gave similar results. Similar work has also been done by Sneddon (1948) and Spence (1968). However, the analytical solution for problems defined with generalised contact profiles is not yet found in the literature.

(iii.) Interface friction

The interface friction is inevitable in practical situations. In the normal direction, the material elastic deformation in the tangential plane causes traction even without any relative tangential movements. However, this is only applicable to the cases that deal with contacting bodies made of different materials. Johnson (1985) has maintained that the relationship for the normal pressure and traction ($q = \mu p$) still is valid for the slip case. For stick situations, Mossakovski (1954,1963) and Goodman (1962) studied this using a 2D problem firstly, and Spence (1968) improved their findings to show that under appropriate conditions the stress field is self-similar at all stages of loading. The traction distribution $q(x)$ is given as:

$$q(x) = \frac{\beta p_0}{\pi a}[(a^2 - x^2)^{1/2} \ln\left|\frac{a+x}{a-x}\right| + x \ln\{\frac{a + (a^2 - x^2)^{1/2}}{a - (a^2 - x^2)^{1/2}}\}] \tag{3.15}$$

where β represents the measure of difference between the elastic materials of the two elastic bodies and can be calculated as in Eq. (3.16):

$$\beta = [\frac{\{1 - 2\nu / G\} - \{1 - 2\nu'/ G'\}}{\{1 - \nu / G\} + \{1 - \nu'/ G'\}}] \tag{3.16}$$

where G is the shear modulus.

In summary, although the theory of classical contact mechanics is widely used in the study of wheel/rail contact, the limitation imposed by the basic assumptions and the difficulty to obtain the analytical solution introduce significant challenges to the specific problem of contact impact at IRJs. This is because classical contact mechanics, especially Hertz contact theory, does not account for the edge effect and material plasticity. Although several non-Hertz contact solutions are proposed in the literature, analytical solutions for more general cases are not yet available and hence, their application to railway engineering still remains far from being realised.

3.2.2. Computational theories

Computational contact mechanics is developed on the basics of non-linear continuum mechanics by employing numerical methods such as the finite element method. The contact is considered as a boundary condition. In this section, the basis of the finite element method is reviewed prior to presenting the computational contact theory.

(a) Basics of finite element method

Zienkiewicz (1971) has provided a displacement approach to solve the generalised elastic continuum problems numerically as described below:

i. *The continuum is separated by imaginary lines or surfaces into a number of 'finite elements'.*
ii. *The elements are assumed to be interconnected at a discrete number of nodal points located on their boundaries. The displacements of these nodal points are the basic unknown parameters of the problem.*

iii. A set of functions are chosen to define uniquely the state of displacement within each 'finite element' in terms of its nodal displacements.

iv. The displacement functions define uniquely the state of strain within an element in terms of nodal displacement. These strains, together with any initial strains and constitutive properties of material will define the state of stress throughout the element and, hence, also on its boundaries.

The finite element method introduces some approximations to the solution. The first is the displacement function which only approximately represents the displacement profile of the elements. The second relates to equilibrium conditions that are satisfied to within a prescribed level of tolerance.

The process of solving the equilibrium condition is equivalent to the minimisation of total potential energy of the system in terms of the prescribed displacement field. Therefore, finite element method applications can be extended to almost all problems where a variational formulation is possible.

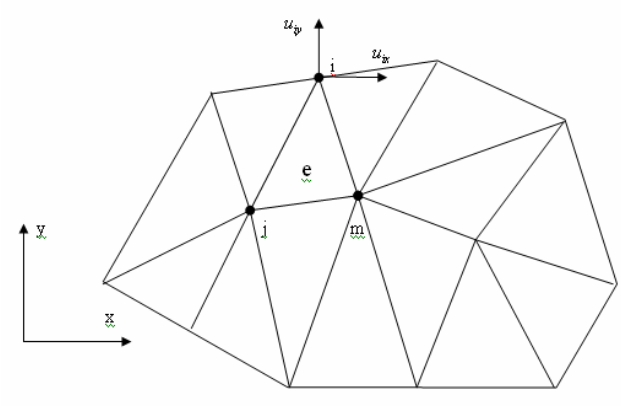

Figure 3.3 2D discrete plane with elements

For simplicity a two dimensional plane stress analysis formulation is provided here. In Fig. 3.3, a typical finite element, e, is defined by nodes, i, j, m and straight line boundaries. The displacement field within this element at any point can be represented as:

$$u = Nu_e \tag{3.17}$$

where N is the shape function and u_e represents the nodal displacement for an element. The strain-displacement relations are then expressed as:

$$\varepsilon = Bu_e \tag{3.18}$$

Matrix B is strain-displacement transformation matrix. Stresses are determined from:

$$\sigma = D\varepsilon \tag{3.19}$$

where D is the elastic matrix.

By imposing a virtual nodal displacement du_e, equilibrium with the external and internal work is achieved. Eqs. (3.17) and (3.18) are then rewritten as:

$$du = Ndu_e, d\varepsilon = Bdu_e \tag{3.20}$$

The work done by the nodal forces is the sum of the products of the individual force components and the corresponding displacement,

$$\Pi_{ext} = (du_e)F_e \tag{3.21}$$

where F_e is the nodal force.

In the same way, the internal work per unit volume done by stresses and body forces is

worked out as:

$$\Pi_{\text{int}} = (d\varepsilon)\sigma - (du)f \tag{3.22}$$

or

$$\Pi_{\text{int}} = (du_e)(B\sigma - Nf) \tag{3.23}$$

in which f is the body force.

Employing the virtual work principle that equates the external work to the total internal work, Eq. (3.24) is obtained:

$$(du_e)F_e = (du_e)(\int B\sigma dxdy - \int Nfdxdy) \tag{3.24}$$

When the material elasticity is valid, substituting Eqs. (3.18) and (3.19) into Eq.(3.24), the following equation can be obtained:

$$F_e = \int (B^T DBdxdy)u_e - \int Nfdxdy \tag{3.25}$$

In Eq. (3.25), $k^e = \int B^T DBdxdy$ is the matrix of element stiffness. F_e is a set of unknown parameters. In order to determine the displacement field u_e, boundary conditions must be employed to resolve these equations at the overall system level.

The stiffness of the whole system is obtained by assembling the stiffness matrices of all elements together.

$$K = \sum k^e \tag{3.26}$$

The principle of virtual displacement used above ensures the equilibrium of the system

for the displacement pattern that minimises the potential energy. The equilibrium would be complete only if the virtual work equality for all arbitrary variations of displacement were ensured.

Balancing the internal energy with the external work, Eq. (3.27) is obtained:

$$\int (d\varepsilon)\sigma dV - [\int (du)\, fdV + \int (du)\, f_b dS] = 0 \tag{3.27}$$

The first term of the above equation will be recognized as the variation of the strain energy, Π_{int} of the structure, and the second term that is in the brackets is the variation of the potential energy of external loads, Π_{ext}.

Rewriting Eq. (3.27), we obtain:

$$d(\Pi_{int} + \Pi_{ext}) = d(\Pi_p) = 0 \tag{3.28}$$

where Π_p is the total potential energy. This means the finite element method seeks a displacement field that keeps the total potential energy stationary and minimised. In that case, finite element method can be used in any problem in which function Π_p could be specified or in the following minimum condition:

$$\frac{\partial \Pi_p}{\partial u} = \begin{Bmatrix} \dfrac{\partial \Pi_p}{\partial u_1} \\ \dfrac{\partial \Pi_p}{\partial u_2} \\ \vdots \end{Bmatrix} = 0 \tag{3.29}$$

In practical application, the equilibrium equations can be obtained by descretising the

virtual work equation and expressed as:

$$F(u) = 0 \tag{3.30}$$

The displacement field can be obtained by solving Eq. (3.30), and other terms such as the strain and the force are derived from the obtained displacement.

So far the finite element process to the linear elastic problem is introduced. However, in this thesis, because of the material plasticity and contact boundary condition, the non-linearity is involved. Thus the approach is generalized to accommodate the non-linear problems. Galerkin Treatment is commonly used as a weighted residual method to the general finite element process. On top of that, the weak form of the differential governing equations is introduced first. The governing equations are written in the general form as:

$$H(u) = \begin{Bmatrix} H_1(u) \\ H_2(u) \\ \vdots \end{Bmatrix} = 0 \tag{3.31}$$

In a domain Ω, with the boundary conditions

$$J(u) = \begin{Bmatrix} J_1(u) \\ J_2(u) \\ \vdots \end{Bmatrix} = 0 \tag{3.32}$$

The equivalent weak-form is expressed as

$$\int wH(u)d\Omega + \int \overline{w}J(u)d\Gamma = 0 \tag{3.33}$$

Where w and $\overline{w}$ are arbitrary parameters called weighted coefficient. Eq.(3.33) is called the weakform of Eq.(3.31) and Eq.(3.32) with lower requirement of connectivity

for displacement function.

The solution in approximation form is written as following:

$$u \approx \sum N_i d_i = Nd \tag{3.34}$$

Where d is the nodal displacement field. The approximation to the Eq. () is written as:

$$\int wH(Nd)d\Omega + \int \overline{w}J(Nd)d\Gamma = 0 \tag{3.35}$$

The $H(Nd)$ and $J(Nd)$ represent the residual obtained by substitution of the approximation into the differential governing equations. Eq.(3.35) is a weighted integral of such residuals. The approximation thus is called the method of weighted residuals. To the weighted residual method, there are a few treatments; among which, the Galerkin method is most commonly used. The Galerkin method chooses the shape function as the weighted coefficient and written as:

$$w_j = N_j \tag{3.36}$$

As a result, in the Galerkin method, Eq.(3.37) is derived:

$$\int NH(Nd)d\Omega + \int NJ(Nd)d\Gamma = 0 \tag{3.37}$$

(b) Computational contact theory

For contact problems, the contact between two bodies is treated as a boundary condition for each body. The contact pressure and traction represented by term f_b (Eq. 3.27) are considered as boundary constraints. The Lagrange Multiplier method and the Penalty method of contact constraint enforcement are employed to solve the

equilibrium equations.

Contact is a complex boundary condition because of its nonlinearity. Before employing the contact constraint enforcement to solve the equilibrium equations, the relation between contact pressure/traction and displacement needs to be set up. As the state of contact affects the relationship between the contact pressure/traction and the displacement, first the computational approach should establish the occurrence of contact. The following conditions are required to be assessed in each computational step.

$$\begin{cases} non-contact \\ contact \begin{cases} stick \\ slip \end{cases} \end{cases}$$

A potential algorithm is presented as a simple illustration. Consider Fig. 3.4 showing two elastic bodies $B^i, i = 1,2$. x_i denotes coordinates of the original configuration. In the normal direction of contact, non-penetration condition is defined as gap function g_N given by:

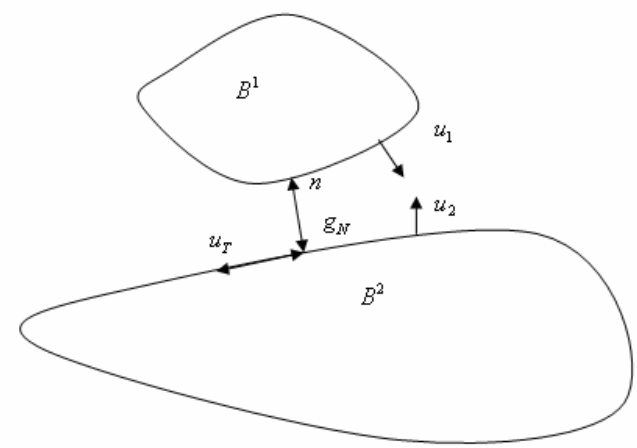

Figure 3.4 Two bodies in contact

$$g_N = g_{N0} - (u_1 - u_2) \cdot n \begin{cases} > 0(non-contact) \\ = 0(contact) \\ < 0(penetrate) \end{cases} \quad (3.38)$$

Eq. (3.38) is used to judge the state of contact/non-contact, in which n is the normal vector to the contact surface, g_{N0} is the original gap, expressed as Eq. (3.39):

$$g_{N0} = (x_2 - x_1) \cdot n \quad (3.39)$$

In Eq. (3.38), in the condition $g_N < 0$, the contacting bodies penetrate into each other and the penetration is defined as g_N'.

The tangential motions of contact state are associated with stick and slip. Stick refers to no relative motion between the two contact bodies while slip refers to existence of relative tangential motion. The motion can be defined using a function u_T in the tangential direction.

For stick condition:

$$u_T = [I - n \times n](u_1 - u_2) = 0 \quad (3.40)$$

while in slip conditions:

$$u_T = [I - n \times n](u_1 - u_2) \neq 0 \quad (3.41)$$

where I is the unit matrix. Through Eq. (3.38) to Eq. (3.41), the contact states are determined.

The compressive contact pressure p within the contact patch can be expressed as:

$$p = n \cdot \bar{\sigma} \cdot n \tag{3.42}$$

where $\bar{\sigma}$ is the boundary value of stress on the contact surface. For the slip zone, the frictional tangential traction employs Coulomb friction law and is defined as:

$$q = \mu p \tag{3.43}$$

For the stick zone, the frictional traction is expressed as:

$$q = \bar{\sigma} \cdot n - pn \tag{3.44}$$

The stress $\bar{\sigma}$ is converted to displacement based on the elastic or elasto-plastic material model. Thus, the relation between contact pressure/traction and displacement is developed.

(i.) Contact constraint enforcement

To solve the equilibrium equations, the contribution of total potential energy from the contact boundary is extracted and Eq. (3.29) is rewritten as:

$$\delta\Pi_p = \delta(\Pi_{ext,\text{int}} + \Pi_c) = 0 \tag{3.45}$$

where $\Pi_{ext,\text{int}}$ is the sum of internal and external energies except from the boundary of contact, and Π_c is the energy contribution from contact. The $\Pi_{ext,\text{int}}$ term in Eq. (3.45) is further extended as:

$$\Pi_{ext,\text{int}} = \int \{\varepsilon\}^T \{\sigma\} dV + \int \{\ddot{u}\}^T \{m\} \{u\} dV + \int \{f\} \{u\} dV + \int \{u\}^T \{f_b\} dS] \tag{3.46}$$

The term Π_c is expressed in different forms depending on the type of contact constraint method used. In this research, two common methods, the Lagrange

multiplier method and the Penalty method are employed in the static and dynamic analysis respectively.

1) Lagrange Multiplier method

In this method, the contact potential energy Π_c is written as:

$$\Pi_c = \int (\lambda_N g_N + \lambda_T u_T) dS \tag{3.47}$$

To get the solution of the multipliers λ_N, λ_T, variation principle is employed as per Eq. (3.45). In that process, multipliers λ_N, λ_T are treated as the unknown variables. The variation of the total potential energy generates a set of equations from which multipliers is determined using Newton iteration algorithm. The overall process of solving the contact boundary problem with Lagrange Multiplier method is illustrated in Fig.3.5. The multipliers (λ_N and λ_T) correspond to the normal and tangential pressures (p and q) respectively.

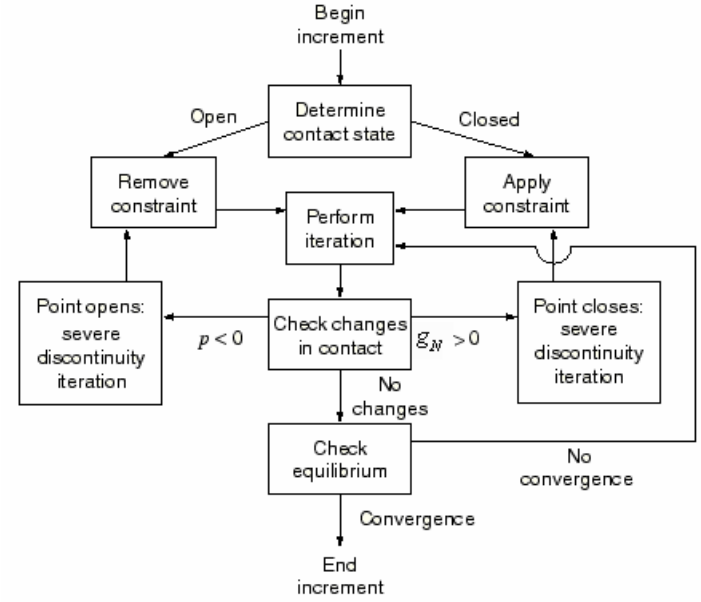

Figure 3.5 Process of solving the contact boundary problem using Lagrange Multiplier method (ABAQUS, 2003)

2) Penalty method

Relative to the Lagrange method, the Penalty method has the advantage that in the variational form the contact pressure and traction p and q are explicitly removed. Similar to Eq. (3.47), the contact potential energy can be expressed as:

$$\Pi_c = \frac{1}{2}\int (\chi_N (g_N^{'})^2 + \chi_T u_T \cdot u_T) dS \qquad (3.48)$$

where χ_N, χ_T are penalty parameters, and $g_N^{'}$ is the penetration function. The values of penalty parameters χ_N, χ_T are properly set to avoid the ill-conditioned numerical problem.

For ABAQUS/Explicit, which is employed for dynamic analysis of wheel/rail contact, the process of solving the contact constraint using the Penalty method can be described as follows:

1) Surfaces of the two contacting bodies are firstly defined as a 'master-slave' pair.
2) The Penalty method searches for slave node penetration g_N in the current configuration.
3) Contact forces as a function of the penetration distance g_N are applied to the 'slave' nodes to oppose the penetrations, while equal and opposite pressures p are applied on the master nodes as equivalent forces. The penalty stiffness is used to calculate contact forces.
4) The equilibrium equations with the contact forces are then solved

Another constraint enforcement method named Kinematic method is also available in the ABAQUS/Explicit exclusively for the explicit time-integration method. The steps

of this method are listed as follows:

1) The kinematic state of the model is advanced into a predicted configuration without considering the contact conditions.
2) The depth and the associated mass of the penetrated ‘slave’ nodes are then determined.
3) The resisting force required to oppose the penetration by using the penetration depth g_N, mass M and the time increment Δt is then calculated.
4) The resisting forces are then applied to the ‘master’ and the ‘slave’ surfaces to adjust the contact body from penetrating to contacting.
5) The equilibrium equations containing the contact forces are then solved.

(c) ALE Formulation

For contact problems, Lagrangian formulation employed in this thesis, is well understood and frequently used to solve the practical engineering problems. However, this formulation requires considerable computational cost especially when the contact model is large in size and the contact area requires refined mesh. For that reason, another efficient formulation namely, Arbitrary Lagrangian Eulerian (ALE), is recognized and developed in the recent years by many researchers such as Nackenhorst (2004),Ponthot and Belytschko (1997), Brinkmeier etc (2007). The major ALE advantages for rolling contact problems can be briefly concluded as:

1) A spatially fixed discretisation is introduced, which enables local refinement in the contact zone for more accurate analysis
2) Error control and adaptive mesh refinement can be performed with respect to

the spatial discretisation only

3) Superimposed transient dynamics is immediately described in space domain, which is required for example for rolling noise analysis

4) Within a purely Lagrangian description the whole circumference of the wheel has to be discretised as fine as needed for a detailed contact analysis. The number of unknows is drastically reduced when the rolling process is observed in a spatial observer framework

5) For the treatment of the explicit time dependency time discretisation schemes have to be involved. A stationary operating point has to be computed starting from the resting state

However, due to its rare application in the commercial code, which is important for practical modelling, in this research the Lagrangian formulation is employed. The basics of ALE formulation is briefly reviewed in this section for possible further model development in the future.

For rolling contact problems, the general idea of ALE formulation is the decomposition of motion into a pure rigid motion (φ) and the superimposed deformation (ϕ). The material deformation gradient is

$$O = \hat{O} \cdot Q \tag{3.49}$$

Where the Q is the pure rigid body motion and the $\hat{O}$ is a measure for the deformation of rolling body.

The elementary balance laws of solid mechanics in the ALE formulation contain two section: balance of mass and balance of momentum. The balance of mass is represented as Eq. (3.50)

$$M = \int_{\hat{\phi}} \rho dV = \int_{\varphi} \hat{\rho} d\hat{V} = \int \rho_0 dV = const. \tag{3.50}$$

Where the M is the mass, ρ is the mass density and the V is the mass volumn. On the other hand the balance of momentum is written as following with respect to the reference configuration,

$$Div\hat{P} + \hat{\rho} f = \hat{\rho} \frac{dv}{dt} \tag{3.51}$$

The $\hat{P}$ denotes the First Piola-Kirchhoff stress tensor, f is the body force density and the v the velocity of the material particals. The boundary condition can be described as:

$$\begin{aligned} &\phi = \bar{\phi} \\ &\hat{P} \cdot \hat{N} = \bar{T} \end{aligned} \tag{3.52}$$

In addition the contact conditions should be satisfied.

For approximate solutions using the finite element method the balance law is re-written in a weak form as Eq. (3.53)

$$\int (Div\hat{P} + \hat{\rho} f - \rho \frac{dv}{dt}) \cdot \eta d\hat{V} \tag{3.53}$$

This equation can be further developed to the incremental finite element representation of the equations of motion,

$$M\ddot{d} + G\dot{d} + [K - W]\Delta d = \mathrm{f}_{\mathrm{ext}} + \mathrm{f}_{\mathrm{inertia}} - \mathrm{f}_{\mathrm{int}} \tag{3.54}$$

To be solved for the evolution of the displacement field

$$d^{t+\Delta t} = d^t + \Delta d \tag{3.55}$$

The K is the tangential stiffness matrix, M is the standard mass matrix

$$G = \int \hat{\rho}(N^T A - A^T N) d\hat{V} \tag{3.56}$$

is the gyroscopic matrix and

$$W = \int \hat{\rho} A^T A d\hat{V} \tag{3.57}$$

is the ALE inertia matrix obtained from the linearization of the centrifugal forces.

For the contact boundary condition, the normal and tangential contact can be treated locally decoupled. For the normal contact, the enforcement of the Signorini condition is written as

$$g_N \leq 0,\ p \geq 0,\ p g_N = 0 \tag{3.58}$$

Well established algorithm for contact computation can be applied directly to enforce the normal contact constraints. The penalty method for example leads to the contact force contribution

$$f_{contact} = -\int \bar{N}^T \chi_n^t g_N da \tag{3.59}$$

Contribution to the tangent matrix:

$$K_{contact} = \int \chi_n \bar{N}^T \alpha_n \alpha_n{}^T \bar{N} da \tag{3.60}$$

However, the well established techniques developed within a pure Lagrangian framework can not be applied directly to enforce the tangential contact constraints within the ALE picture. This leads to the additional treatment from the Lagrangian to ALE formulation and can refer to Ziefle's (2007) work.

3.3. Review of Solution Methods for Finite Element Method

The solution methods for non linear problems can be classified into two types:

- Time independent
- Time dependent

The time independent algorithm is explored for static problems without considering the inertial effect, while the time dependent algorithm is suitable for the dynamic problems in which the inertial effect is not negligible. Both of the two methods are employed in the finite element model used in this research.

3.3.1 Algorithm for time-independent problems

For static non-linear problems, iteration methods such as the Newton's method are widely used in the finite element analysis to solve the system of equilibrium equations. The entire procedure of solving the non-linear equations is divided into several increments and each increment is subdivided into iterations.

Eq.(3.30) can be written as follows with the superscript n representing the increment n:

$$F^n(u) = 0 \tag{3.61}$$

The u is the exact solution of displacement. To obtain that solution, assume that an approximation u_i is obtained after the iteration i. The Δu_i is the difference between u and u_i, so:

$$F^n(u_i + \Delta u_i) = 0 \tag{3.62}$$

Expanding the left-side of this equation in a Taylor series gives:

$$F^n(u_i) + \frac{\partial F^n}{\partial u_i}(u_i)\Delta u_i + \frac{\partial^2 F^n}{\partial u_i^2}(u_i)\Delta u_i^2 + ... = 0 \tag{3.63}$$

Since u_i is a close approximation to the solution, Δu_i should be small. As a result, the second and higher order terms of Δu_i can be neglected. Eq. (3.63) is simplified as:

$$K_i^n \Delta u_i = -F_i^n \tag{3.64}$$

Where $F_i^n = F^n(u_i)$ and K_i^n is the Jacobian matrix which is solved as:

$$K_i^n = \frac{\partial F^n}{\partial u_i}(u_i) \tag{3.65}$$

Δu_i can then be obtained from Eq. (3.64) and the next approximation is expressed as:

$$u_{i+1} = u_i + \Delta u_i \tag{3.66}$$

The iteration continues until the Δu_i is small enough that the solution is considered convergent.

3.3.2 Algorithm for time-dependent problems

For dynamic problems, two algorithms have been widely used in the finite element method: explicit time integration method and implicit time integration method. Wriggers (2002) gives basic instructions about these two methods:

- *Explicit time integration methods are easy to implement, since the solution at time t_{n+1} depends only upon known variables at t_n. These methods are*

extremely efficient when the mass matrix is approximated by a lumped mass matrix which is diagonal. Explicit methods are conditionally stable, which means that the time step size is governed by the Courant criterion (a condition on numerical method calculations requiring that the time interval employed be no greater than that required for a stress wave to cross the characteristic length of elements).

- *Implicit time integration method schemes approximate time derivatives by quantities which also depend upon the last time step* t_n *and upon the still unknown values at time* $t_{n+\alpha}$. *These methods require a solution of a nonlinear equation at each time step. They are much more expensive, since they have to be combined with, for example, the Newton procedure. However, implicit schemes can be constructed so that they are unconditionally stable, and hence can be applied with a far bigger time step than the explicit schemes.*

The time step size for both these two methods depends on the nature of the problem. For high frequency response problems, such as impact, a small step size is necessary which should be lower than the time period of the sound wave travelling through the characteristic length of element.

For dynamic problems, the inertial force is not negligible and the system is in dynamic equilibrium which is expressed as:

$$M\ddot{u} + C_d\dot{u} + Ku = F \tag{3.67}$$

(a.) Explicit time integration

In the finite element method, a central difference scheme is widely applied where velocities and accelerations at time t_n are approximated by:

$$\dot{u}_n = \frac{u_{n+1} - u_{n-1}}{2\Delta t}, \qquad \ddot{u}_n = \frac{u_{n+1} - 2u_n + u_{n-1}}{(\Delta t)^2}. \tag{3.68}$$

Inserting the above functions into Eq. (3.63), Eq. (3.69) can be obtained:

$$(M + \frac{\Delta t}{2} C_d) u_{n+1} = (\Delta t)^2 [F_n - K u_n] + \frac{\Delta t}{2} C_d u_{n-1} + M(2u_n - u_{n-1})\mathbb{F} \tag{3.69}$$

To solve, initial conditions u_0 and $\dot{u}_0$ are required. Note the term u_{n-1} exists, which means at the first step u_{-1} needs to be determined first. By using a Taylor series expansion at time t_{-1}, we obtain:

$$u_{-1} = u_0 - \Delta t \dot{u}_0 + \frac{(\Delta t)^2}{2} \ddot{u}_0 \tag{3.70}$$

where $\ddot{u}_0$ is obtained from Eq. (3.67) as follows:

$$\ddot{u}_0 = M^{-1}[-C_d \dot{u}_0 - K u_0 + F_0] \tag{3.71}$$

The process introduced above is the concept of a classical approach of solving the equations explicitly. Different finite element codes adopt different algorithms. In ABAQUS/Explicit, the equations of motion for the body are integrated using the explicit central difference integration rule:

$$\dot{u}_{i+\frac{1}{2}} = \dot{u}_{i-\frac{1}{2}} + \frac{\Delta t_{i+1} + \Delta t_i}{2} \ddot{u} \tag{3.72}$$

$$u_{i+1} = u_i + \Delta t_{i+1} \dot{u}_{i+\frac{1}{2}} \tag{3.73}$$

where $\dot{u}$ is velocity and $\ddot{u}$ is acceleration. The subscript i refers to the increment number and $i-\frac{1}{2}$ and $i+\frac{1}{2}$ refer to mid-increment values. The central difference integration operator is explicit in that the kinematic state can be advanced using known values of $\dot{u}_{i-\frac{1}{2}}$ and $\ddot{u}_i$ from the previous increment:

$$\ddot{u}_i = M^{-1} \cdot (F_{ext\,i} - F_{\mathrm{int}\,i}) \tag{3.74}$$

where M is the nodal mass matrix, F_{ext} is the applied external load, and F_{int} is the internal force.

Special treatment of the mean velocities $\dot{u}_{i+\frac{1}{2}}$, $\dot{u}_{i-\frac{1}{2}}$ etc. is required for initial conditions, certain constraints, and presentation of results. For presentation of results, the state velocities are stored as a linear interpolation of the mean velocities:

$$\dot{u}_{i+1} = \dot{u}_{i+\frac{1}{2}} + \frac{1}{2}\Delta t_{i+1} \ddot{u}_{i+1} \tag{3.75}$$

The central difference operator is not self-starting because the value of the mean velocity $\dot{u}_{-\frac{1}{2}}$ needs to be defined:

$$\dot{u}_{\frac{1}{2}} = \dot{u}_0 + \frac{\Delta t_1}{2} \ddot{u}_0 \tag{3.76}$$

Substituting this expression into the updated expression for $\dot{u}_{i+\frac{1}{2}}$ yields the following definition of $\dot{u}_{-\frac{1}{2}}$:

$$\dot{u}_{-\frac{1}{2}} = \dot{u}_0 - \frac{\Delta t_0}{2}\ddot{u}_0 \tag{3.77}$$

The explicit procedure requires no iterations and no tangent stiffness matrix (See Eq. (3.74)), thus explicit integration dynamic analysis requires less computation cost for each time increment. However, as the central difference operator is conditionally stable, the increment should be significantly small. The stability limit for the operator is given in terms of the highest Eigenvalue in the system as:

$$\Delta t \le \frac{2}{\varpi_{\max}}(\sqrt{1+\xi^2}) - \xi) \tag{3.78}$$

where ξ is the fraction of critical damping associated with the highest mode. Another conservative estimate of the stable time increment can be given by the minimum taken over all the elements:

$$\Delta t = \min(L_e / C_d) \tag{3.79}$$

where L_e is the characteristic element dimension and C_d is the current effective dilational wave speed of the material which is related with density, elastic modulus, and Poison ratio of the material:

$$C_d = \sqrt{\frac{E(1-\nu)}{\rho(1+\upsilon)(1-2\upsilon)}} \tag{3.80}$$

ABAQUS/EXPLICIT uses the explicit integration algorithm for solving equilibrium equations. Simulations using this method generally take of the order of 10,000 to 1,000,000 increments, but the computational cost per increment is relatively cheap.

(b.) Implicit time integration

One of the most widely applied implicit methods is the Newmark (1959) method. The approximations of displacement and velocity at time t_{n+1} are based on the following two functions:

$$\begin{aligned} u_{n+1} &= u_n + \Delta t \dot{u}_n + \frac{(\Delta t)^2}{2}[(1-2\vartheta)\ddot{u}_n + 2\vartheta\ddot{u}_{n+1}], \\ \dot{u}_{n+1} &= \dot{u}_n + \Delta t[(1-\upsilon)\ddot{u}_n + \upsilon\ddot{u}_{n+1}]. \end{aligned} \tag{3.81}$$

where the constant parameters ϑ and υ can be chosen freely and the order and accuracy of the method is determined. By inserting Eq. (3.81) into Eq. (3.67), we can get the equilibrium equation which can now be solved by using some iteration method such as the previously introduced Newton method. By obtaining the solution of acceleration $\ddot{u}_{n+1}$, other variables like displacement and velocity can be worked out using Eq. (3.81).

In summary, for the solution of wheel/rail dynamic contact at IRJs, both implicit and explicit methods may be used. However, there are some significant differences between them. The implicit method calculates the overall dynamic response of the structure in each iteration while the explicit method employs the wave propagation solutions associated with relatively local response in continua. The implicit method is unconditionally stable because of the iteration process. In contrast, the conditionally stable explicit method is only stable when the increment is small enough relative to the stress wave propagation.

The nature of impact problems determines that the time increment should be small and hence the number of increments would be numerous. By using the implicit method,

the computational cost would be unacceptablely expensive as every increment would involve a number of iterations. By contrast, the explicit method would provide a much cheaper solution by computing local response in each increment; a reasonably accurate result can be guaranteed if the increment step is kept small.

3.4. Discussion of Contact Impact

The impact condition emerges as the rate of loading is high and the dynamic effects are important. In other words, in wheel/rail rolling or sliding contact, the material inertia flows through the deforming region and influences the stress field. This leads to the stress propagation wave in the contact bodies and material plasticity may be caused under the high rate of loading. Referring to Johnson (1985), the stress wave amplitude is expressed as:

$$\sigma = \rho c_0 v \tag{3.82}$$

where σ is the stress, ρ is the contact body density, c_0 is the stress wave propagation velocity and v is the deformation velocity of the contact body. If the stress value exceeds the yield stress Y, the material yields. To keep the material in elastic condition, the deformation velocity must be less than the certain value:

$$v < Y / \rho c_0 \tag{3.83}$$

For steel material employed in this research, the yield stress is 780MPa, the density is 7800Kg/ m^3 and therefore the stress propagation speed is 5900m/s. As a result, the maximum impact velocity in the deformation direction for elastic deformation is 16.95m/s. Deformation rates above this magnitude causes material yield.

3.5. Summary

In this chapter contact mechanics was first briefly reviewed. For classical theory, the Hertz contact theory has provided the analytical contact solution with the elastic half space assumption. Non-Hertz theory has also been discussed and it was shown that it better represents some special contact situations. However, it has also been shown that both Hertz and non-Hertz theory did not provide a practical solution for wheel/rail contact at IRJs. For computational contact mechanics, the contact boundary conditions have been introduced through constraint enforcement. The Lagrange Multiplier method and the Penalty method appear advantageous for the contact solutions.